Original May 2013 version has been annotated to include information from Wiley contract newly revealed by Wegman/Said lawsuits.
WIREs;CS became visible July 2009. Here, Jan 2006-June 2009 is called "Pre Launch" and highlighted.

FOIA Facts: Ed Wegman, Yasmin Said, George Mason University

John R. Mashey 05/16/13

Executive Summary

George Mason University (GMU) mishandled easy academic misconduct complaints, following a process at best incompetent. New FOIA replies reveal GMU violations of Federal reporting rules, but also expose multi-year misuse of Federal research grants by GMU's **Edward Wegman** and **Yasmin Said**. They claimed credit for much clearly-unrelated work, *false claims* alleged here that may rise to *grant fraud*, while $492K in grants produced almost nothing relevant in peer-reviewed research journals.

Background

GMU Professor Wegman was recruited to study the 1999 climate "**hockey stick**" of Michael Mann, Raymond Bradley and Malcom Hughes.
He led the 2006 "**Wegman Report**" (**WR**), promoted to Congress as "independent, impartial, expert" work by a team of "eminent statisticians," working "pro bono," mostly *false statements* to Congress. Much work was actually done by grad students and alcoholism-modeling new PhD Said.
From the latest FOIA replies, **Wegman and Said claimed their substantial WR efforts for credit from grants, not pro bono.**

In December 2009 Canadian blogger "Deep Climate" started documenting plagiarism in the Wegman Report. Based on those findings, in March and May 2010 Bradley lodged formal plagiarism complaints to GMU for:

- 2.5 pages of the WR's text, taken from Bradley's **own** book
- 5.5-page Social Networks Analysis (SNA) in the Wegman Report
- 1.5-page subset of that SNA text re-used in an Elsevier-published paper **that acknowledged 3 Federal grants**, as Bradley pointed out to GMU

GMU took nearly two years to rule that the WR had no plagiarism, while contradictorily admitting the paper did. GMU ruled misconduct on Wegman's part, placed a reprimand in his file and *oddly* told him to retract the paper that Elsevier had retracted 10 months' earlier. People had lodged more complaints of plagiarism and falsification, but GMU explicitly declined to investigate. In February 2012 **Provost Peter Stearns** described the investigation to faculty in a letter with factual untruths, as revealed by mid-2012 FOIAs. **The latest FOIAs show GMU failed to properly report results to the Federal agencies as required.**

Brief introduction to US Federal research grants via universities

- Via their schools, people propose specific research to Federal agencies, with funding for their "direct" costs, plus "indirect" costs for the school.
- If accepted, the **school signs a contract** and grant money starts flowing through the school to the researcher(s), often via periodic payments.
- *Works* (papers, talks) are expected to include acknowledgements *(acks)* of funding. People later *claim* works in summary reports to funders.
- People are supported to do relevant work. Productive researchers might list some less-relevant efforts, for little credit, but without raising concerns. Major diversion of effort (in this case, more than half) from relevant *(fit)* into clearly-irrelevant *(unfit)* work may rise to *grant fraud*.
- People may have multiple grants, but funds must not be *commingled*.
- Schools should oversee funds use and handle legitimate misconduct complaints properly, not do their utmost to ignore them.

Take the money

Wegman had successfully won many Federal research grants, the most recent pair administered by the **Army Research Office** (**ARO**):

1. W911NF-04-1-0447 11/01/04-4/30/08 $217K "Analytical and Graphical Methods for Streaming Data with Applications to Netcentric Warfare"
2. W911NF-07-1-0059 12/15/06-12/15/07 $100K "Adaptive Multi-modal Data Mining and Fusion For Autonomous Intelligence Discovery"

Said got money from both, plus a postdoctoral fellowship grant from the **National Institute on Alcohol Abuse and Alcoholism** (**NIAAA**):

3. F32AA015876 05/26/06-05/29/09 $164K
"A Social Network Model of Ecological Alcohol System"

She also got a $30K supplemental Commonwealth of VA grant via GMU.

Papers and talks were compared against the grant proposals, whose abstracts are included on p.4 to help explain each work's classification.
A work might have been plausibly claimable either to **ARO** (0447 or 0059) or to **NIAAA** (5876), but never both, and often neither.
Many works, such as attempts to discredit the hockey-stick and climate science, fit none of the grants, but were claimed for credit anyway.

Classification	1. **0447(W, S)**	2. **0059(W, S)**	3. **5876(S)**
ARO-relevant **papers**, talks	**fit**	**fit**	unfit ③
NIAAA-relevant -alcoholism	unfit ①	unfit ②	**fit**
Hockey-stick, miscellaneous	unfit ①	unfit ②	unfit ③

"CONFIDENTIAL DO NOT DISTRIBUTE WITHOUT PERMISSION" legends should be considered removed throughout.

Break the rules

Specific *allegations* are made against Wegman, Said and GMU, for Inspectors General of the **Department of Defense** (**DoD**) for ARO and **Department of Health and Human Services** (**DHHS**) for NIAAA. Many are easily-verifiable *false claims* of irrelevant work or *false statements* to the government. Other *concerns* may not rise to the level of these allegations, but are documented to help possible investigations. Each allegation stands or falls on its own merits, but a pattern of interrelated activities and funds commingling encourages explanation as a group.

Allegations are summarized here and explained further in the 12-page main discussion, backed by detailed appendices. They fall into 5 groups:

1	0447 (DoD)	Grant fraud by Wegman
2	0059 (DoD)	Grant fraud by Wegman, Said
3	5876 (DHHS)	Grant fraud by Said
4	(DoD)	Attempted Wegman grant fraud, Feb 2009 proposal
5	(DoD, DHHS)	Violation of misconduct handling rules by GMU

P*nnn* (T*nnn*) represent papers (talks) as numbered in Wegman's 2010 resume, with Said's claims and some others added for relevant context.

1 0447 Wegman (DoD)

1.1 False claims. The Final Report claimed credit for 75 works, of which 2 were never published, leaving 73. By examination of works or at least titles/abstracts, 30 of 73 were 0447-unfit, hence 30 false claims for ①.

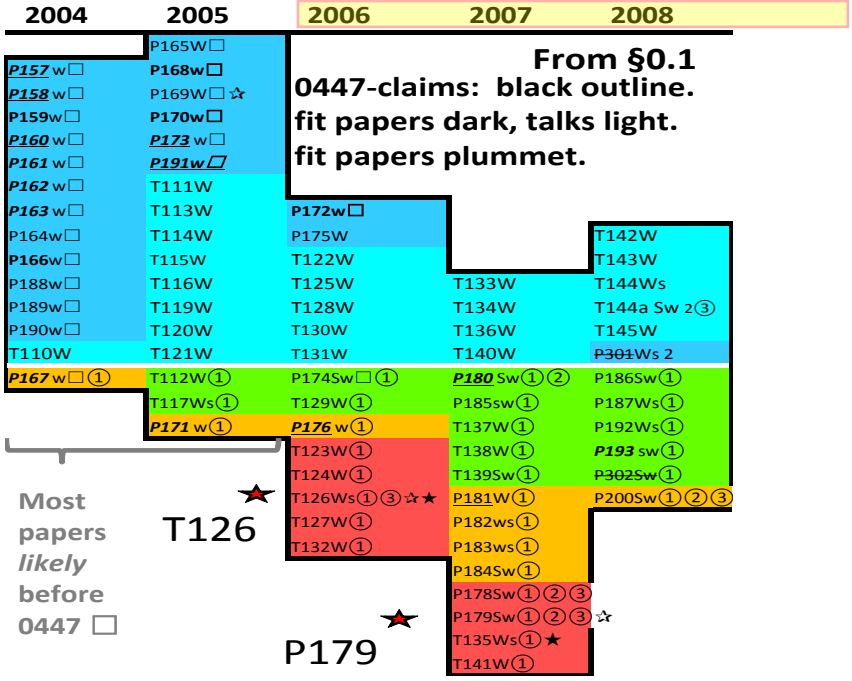
Classification	P claim	P publ	Talks	P+T	Totals publ
ARO-relevant **papers**, talks	21	20	23	43	43 **fit**
NIAAA-alcoholism ①	8	7	6	13	
Miscellaneous ①	8	8	0	8	30 unfit①
Hockey-stick①	2	2	7	9	
Totals published	39	37	36	73	

In chart at right, of 20 published possibly-fit papers (dark blue), 12 were done in 2003-2004 and 7 were *likely* or mostly done before 0447 (□), leaving just one fit paper *clearly* done 2006-2008. That yields 12-19 false claims by date. Later, 4 more unfit works ack'd 0447, giving 30+4 = 34 false claims by topic ①, for a grand total of **46-53 false claims** among 73 works, i.e., 60-70% by count alone. To calibrate Wegman's emphasis, ignore **dark blue** and compare cyan efforts to those below the line.

1.2 False statements. Wegman claimed T126, the 91-page WR and presentations to Congress, an 11-month effort in unfamiliar fields. He and Rep. Barton often told Congress the WR was "pro bono," effectively false statements. Detailed evidence is given elsewhere, not just of pervasive plagiarism in the WR, but that it and Wegman's testimony contained many misleading false statements Congress. Wegman refused a 2010 FOIA request for related communications, saying that "work was done offsite," despite ARO funding. He claimed follow-up paper P179 and 7 related works 2006-2007. His claims, visible actions and plunge in fit work consistently imply that he spent more than 60% of his 0447-funded effort on irrelevant works. Plagiarism or falsification are alleged for 5 works.

1.3 False statements. Wegman stated that 39 papers ack'd 0447, but at least 25 did not, thus **25 false statements.**

1.4 False statements. Wegman promised WR code to Rep. Waxman in 2006, but said he needed Navy approval. An earlier FOIA showed this to be a false statement, in fact for work now known to be claimed to ARO.



2 0059 Wegman, Said (DoD)

2.1 False claims. Wegman and Said claimed 5 works, of which 2 were 0559-unfit papers (social network attacks on climate scientists), including the paper whose retraction Elsevier forced. Another unfit paper, against which plagiarism is alleged, ack'd 0059, totaling at least **3 false claims** by topic ②. A 4th paper was never published, leaving one fit work.

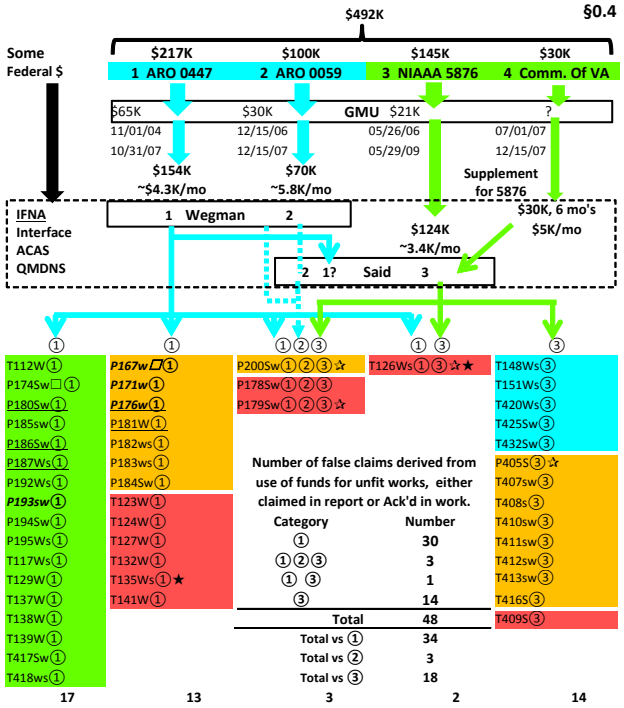
3 (DHHS) 5876: Said (DHHS)

3.1 False claims. Said claimed / ack'd 5876 for 18 unfit works, including T126 and P179, false claims by topic ③. She claimed 3 works done 9-12 months earlier (☐), thus totaling **21 false claims.**

3.2 Plagiarism. To get grant, Said cited works with plagiarism, P401, P402. She also ack'd 5876 or claimed 4 more, for a total of 6.

GMU added $30K of Commonwealth of VA funds as a supplemental grant, for which GMU could provide no oversight information.

Funds Commingling
This displays funds commingling of the grants. This is not an issue of university release time or weak relevance, but blatant funds misuse. **At least 48 works were unfit for acks or claims made.** Wegman claimed alcoholism work for ARO. Said claimed ARO work for NIAAA. Both claimed miscellaneous unfit works and hockey-stick attacks. T126 was clearly the largest single 0447.



4 Wegman February 2009 proposal (DoD)

4.1 Attempted false claim. Wegman sent a grant proposal to the ARO, but ~50% of the text was copy-paste-edited without attribution, not only from students' dissertations, but from other unrelated sources.

5 GMU misconduct handling, failure to inform agencies (DoD, DHHS)

5.1 Failure to inform. By all evidence, GMU failed to inform the funding agencies properly at any required steps in P179 misconduct proceedings.

5.2 GMU broke many rules of handling misconduct, described in other reports. Plagiarism/falsification complaints were sent to GMU 2010-2011, against Wegman, Said and others, including 4 grant-related works:

Work	0447 ARO	0059 ARO	5876 NIAAA	GMU Action
P169	Wegman	-	-	ignored
P179	Wegman①	Wegman + Said ②	Said③	misconduct
P200	Wegman①	Wegman + Said ②	Said③	ignored
T126	Wegman①	-	Said③	Bradley's rejected, 25+ pages ignored,

More problems have been found since in Federally-funded works. GMU declined to investigate well-documented complaints, despite obvious grant problems in P179, known to GMU by May 2010. A few minutes' perusal of 0447 was enough to see many false claims, starting with T126. GMU ignored not just academic norms, but Federally-mandated rules. In February 2013, GMU's new misconduct policy increased its opacity.

6 Summary

Wegman and Said seemed to pay little notice to Federal rules:

- At least 48 unfit works were claimed in reports or had wrong acks.
- At least 60% Wegman's of 0447-claimed works were unfit or done previously. Given the WR's size, yet more of his 0447-effort was unfit.
- Commingled ARO and NIAAA funds flowed to unfit works and for activities that might raise more concerns of funds-misuse.
- They produced almost nothing fit in peer-reviewed research journals.

With Federal funding now known, the WR is now open to more scrutiny. When did Scott get asked? Who was the "unknown 4th person" who quit? Who wrote what? What emails went among Wegman, Said, others? Etc,etc.

In Fall 2012, Wegman was appointed to the GMU College of Science Tenure and Promotions Committee for a 3-year term.

Abstracts of the grant proposals, in case of any doubt

Blue/ cyan shading was given to all that could even marginally fit 0447/0059 for the ARO, but readers might compare works against the proposal and assess actual progress. Research on Internet data streams cannot possibly overlap with an alcoholism postdoc work for the NIAAA.

0447, from **§Q.1** for ARO blue papers and cyan talks fit, all others unfit

Analytical and Graphical Methods for Streaming Data with Applications to Netcentric Warfare' *submitted by GMU on behalf of Edward Wegman.*

Abstract (p.C-1)

'This proposal focuses on a **new data structure**, namely **massive streaming data**. We propose to develop recursive algorithms and evolutionary graphics for handling massive streaming data. Streaming data is essentially a new data acquisition paradigm, in which data becomes constantly available. Older data has less value and therefore must be discounted. Strategies for discounting are proposed as well as strategies for multi-scale resolution of data streams. **The particular example we have in mind is streaming Internet packet headers, although theoretical and practical results will not be limited to this form of data**. However, network traffic data are especially important to military and the U. S. Army in particular as Netcentric Warfare and joint operations between services and with allies become increasingly important. **The ultimate goal of these techniques is to detect intrusion and fraud in streaming data systems.'**

pp.D-1 to D-13 described the work in detail, ending with:

'• I propose what I like to call *evolutionary graphics*, i.e. graphics, which evolve as a function of new data being added. The combination of recursive algorithms and evolutionary graphics will provided *(sic)* a fundamental approach for analyzing streaming data.
• I propose to develop scalable graphics devices (pan and zoom) for discrete data that has many more points than screen resolution would allow.
• I propose to investigate the theoretical properties and implementation of the quantizing algorithm for truly massive streaming data .
• I propose to investigate recursive kernel density estimators in the critical multidimensional case.
• I propose to investigate adaptive mixtures density estimation algorithms using orthonormal bases such as wavelets to reduce or eliminate spurious terms.
• I propose to investigate classical exponential smoother with adaptive time scaling and their potential use as multiscale data analysis.
• I propose to develop evolutionary graphics tools including waterfall diagrams, transient geographic mapping, and related multivariate pan and zoom methodology.'

0059, from **§R.1 for ARO** blue papers and cyan talks fit, all others unfit

p.3 'Adaptive Multi-modal Data Mining and Fusion For Autonomous Intelligence Discovery

My proposal addresses the challenges of **autonomous discovery and triage of contextually relevant information in massive, complex, dynamic text and imagery streams.** I will develop a prototype system to mine, filter and fuse multi-modal data streams and dynamically interact with the analysts to improve their efficiency through feedbacks and autonomous adaptation of the algorithms. I plan to implement four core capabilities:
• Text and image mining for feature extraction
• Multi-modal data fusion
• Agent-based adaptive information filtering
• Cognitively friendly information visualization.

5876, from **§S.2 for NIAAA** green works fit, all others unfit

This type of award is described by NIAAA and the proposal summarized:

'To provide postdoctoral research training to individuals to broaden their scientific background and **extend their potential for research in specified health-related areas**.'

'A Social Network Model of Ecological Alcohol Systems

We recognize that alcohol abuse leads to acute outcomes for both society and individuals. Among these, we identify DWI crashes with fatalities, assault and battery, suicide, murder, sexual assault, domestic violence, and child abuse. Alcohol abusers are embedded in a social network that involves the user, family and friends, producers and distributors of alcohol products, law enforcement, the judiciary, remediation, education, and intervention facilities such as Alcohol Safety Action Programs, and detox and treatment facilities, which are, of course, coupled to insurance and managed-care programs. This complex network is reminiscent of more traditional biologic ecology systems, hence the name. The basic idea is to formulate a model of this network with the goal of exploring short- and long-term interventions that reduce the overall probability of acute outcomes. The unique feature of this proposed work is that we are attempting to explore the simultaneous reduction of all acute outcomes. Historically, modeling attempts have focused on specific acute outcomes. The framework we are pursuing is a dynamic agent-based simulation.'

Red-shaded attempts to discredit the hockey-stick consumed much of Wegman and Said's grant-funded time and such cannot fit any grant here. Wegman claimed greens for 0447, Said claimed Blue/ cyan for 5876, and both claimed red. Readers might disagree with some orange works, but even if all were ignored, 32 false claims by topic remain.

Jan 2006-June 2009 is called "Pre Launch" and highlighted in various sections, also highlighted here.

Table of Contents

Front Matter

Advice on reading this very dense, cross-referenced report

Given names and titles are usually omitted for brevity, no discourtesy intended to any. Opinions and speculations[1] are *Italicized*, **Emboldening** in quotes is generally mine. Reading locality introduces some redundancy.

Wiki's are considered useful starting collections of reference sources, not authoritative. *Online readers are advised to open several windows on the PDF and use Full Search in lieu of an Index. Some discussion is regretfully complex, with many cross-connections, inherent in dissecting opaque, confusing situations of which people spoke untruths and their contradictions exposed only by integrating multiple information sources.*

WebCitation URLs are often given for archived files, especially as important files sometimes get lost. Some GMU files have disappeared after being publicly identified [MAS2010a §A.11]. In one case, a crucial Said lecture file not only disappeared, but someone edited its existence out of the GMU seminar record, necessarily purposeful.

Acknowledgements.

Infinite thanks to Canadian blogger "Deep Climate" (DC) for unearthing the first, long-overlooked plagiarism cases, plus many more. **None of this could have been done without his discoveries and tenacious work.** Many thanks to Dan Vergano for FOIAs and good journalism. Thanks to Brian Schmidt and "Eli X. Rabett" for careful reading and good suggestions and likewise to many anonymous reviewers. Thanks also to a community effort by others who helped find more cases, submitted complaints or offered advice. Thanks to FOIA folks at ARL, ARO, NSWC, NIAA, all of whom were courteous, responsive and helpful.

Dr. Mashey is an easy-to-Google semi-retired Bell Labs (1973-1983) / Silicon Valley (1983-) computer scientist / corporate executive, profiled in *Science* for efforts against climate anti-science: www.desmogblog.com/science-article-recognizes-john-mashey
His professional experience applies directly to some of the technical topics here.
He is a member of AAAS, AGU, APS, ACM, and IEEE CS.
JohnMashey (at) yahoo DOT com PLEASE REPORT ERRORS, updates do occur.

[1] Sometimes "connect the dots" alternatives must be enumerated to alert readers of relevant possibilities or guide further investigation for those with subpoena power.

Glossary

~	approximately, as in mathematics
0047	W911NF-04-1-0447 ARO-managed research grant, Wegman
0059	W911NF-07-1-0059 ARO-managed research grant, Wegman
5876	F32AA015876 NIAAA alcoholism postdoc fellowship, Said
nnnn-fit	might possibly fit grant *nnnn* (0447, 0059 or 5876), generous
nnnn-unfit	obviously did not fit *nnnn*(0447, 0059 or 5876)
ACAS	Army Conference on Applied Statistics, managed by IFNA
ack	acknowledgement in paper/talk, of people or especially funding
ARO	Army Research Office, manages external research
ARL	Army Research Laboratory, does internal research
Bipartite graph	standard term in graph theory **§K.6** T138,
CEI	Competitive Enterprise Institute
claim	claim of paper, talk or other work in grant report to agency
CSDA	*Computational Statistics and Data Analysis,* Elsevier journal
DARPA	Defense Advanced Projects Agency
DC	Canadian blogger "Deep Climate" (person*), Deep Climate* (blog)
DHHS	US Department of Health and Human Services
DoD	US Department of Defense, includes ARO and ARL
FCA	False Claims Act
FOIA	Freedom of Information Act
GMI	George C. Marshall Institute (think tank)
GMU	George Mason University, Fairfax, VA
IFNA	Interface Foundation of North America, Wegman treasurer
IG	Inspector General (of DoD or DHHS, for example)
JHU	Johns Hopkins University, Baltimore, MD strong public health
NIAAA	National Institute on Alcohol Abuse and Alcoholism (in DHHS)
NIH	National Institutes of Health, of which NIAAA is one
NSWC	Naval Surface Warfare Center
ORI	Office of Research Integrity, research watchdog of DHHS[2]
SNA	Social Network Analysis, mis-applied in WR and [SAI2008]
SoPP	School of Public Policy at GMU
WR	Wegman Report (2006),[3] [WEG2006], most of T126,

[2] ori.hhs.gov main ORI page
ori.hhs.gov/research-misconduct-0 research misconduct, especially plagiarism
ori.hhs.gov/case_summary 2011: Jagannathan, Lushington, Visvanathan, Weber

[3] archives.republicans.energycommerce.house.gov/108/home/07142006_Wegman_Report.pdf

Frequently-cited references

BAR2006 07/14/06 Report Raises New Questions About Climate Change Assessments, House Energy and Commerce Committee.[4]
BAR2006a 07/27/06 Complete Transcript, of Wegman Report[5]
DEE2009 12/17/09 Contrarian scholarship: Revisiting the Wegman Report[6]
DEE2010p 09/15/10 Wegman report update, part 2: GMU dissertation review[7]
DEE2010r 11/16/10 Replication and due diligence, Wegman style[8]
DEN2005 2005 Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj, *Exploratory Social Network Analysis with Pajek (used to be online, not found)*
MAS2010 03/15/10 Crescendo to Climategate Cacophony[9]
MAS2010a 09/26/10 Strange Scholarship in the Wegman Report[10]
MAS2011 01/04/11 Strange Inquiries at George Mason University[11]
MAS2011a 05/26/11 Strange Tales and Emails – Said, Wegman, et al[12]
MAS2011b 05/27/11 Strange Falsifications in the Wegman Report[13]
MAS2011d 10/30/11 Curious coincidences at George Mason University …[14]
MAS2012 02/13/12 Fake Science, Fakexperts, Funny Finances, Free of tax[15]
MAS2012b 07/13/12 Ed Wegman Promised Data to Rep. Henry Waxman…[16]
MAS2012c 08/20/12 See No Evil, Speak Little Truth, Break Rules, …[17]
MAS2012d 10/25/12 Fakery 2: More Funny Finances, Free of Tax[18]
REZ2009 Spring'09 Dissertation *Enhancement of Network Robustness and Efficiency through Evolutionary Computing, Statistical Computation and Social Network Analysis*[19]
SHA2008 10/31/08 *Multi-Mode and Evolutionary Networks*[20]
SAI2007 09/07/07 *Experiences with Congressional Testimony: Statistics and The Hockey Stick, GMU Data and Statistical Sciences Colloquium*[21] T414
SAI2008 2008 Social networks of author–coauthor relationships,[22] P179
SAI2010 Yasmin H. Said, Edward J. Wegman, and Walid K. Sharabati, "Author–Coauthor Social Network and Emerging Scientific Subfields," F. Palumbo et al. (eds.), *Data Analysis and Classification*, Studies in Classification, Data Analysis, and Knowledge Organization, DOI 10.1007/978-3-642-03739-9_30, ©Springer-Verlag 2010, pp.257-268.[23] P200.
VER2010 10/08/10 University investigating prominent climate science critic[24]
VER2010a 11/22/10 Experts claim 2006 climate report plagiarized[25]
VER2010c 11/23/10 Wegman report round-up[26]
VER2011 05/15/11 Climate study gets pulled after charges of plagiarism[27]

[4] archives.republicans.energycommerce.house.gov/108/home/07142006_Wegman_fact_sheet.pdf This was the announcement.
[5] Joe Barton, et al, 07/19/26, 07/27/06. DC provides page-numbered PDF version: deepclimate.files.wordpress.com/2010/04/hockey-stick-hearings-2006-ec-committee.pdf The original, with unnumbered pages is at: frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=109_house_hearings&docid=f:31362.wais
[6] deepclimate.org/2009/12/17/wegman-report-revisited DC's sleuth work uncovered first plagiarisms in WR.
[7] deepclimate.org/2010/09/15/wegman-report-update-part-2-gmu-dissertation-review
[8] deepclimate.org/2010/11/16/replication-and-due-diligence-wegman-style
[9] www.desmogblog.com/crescendo-climategate-cacophony 185p
[10] www.desmogblog.com/sites/beta.desmogblog.com/files/STRANGE.SCHOLARSHIP.V1.02.pdf 250p originally hosted at Deep Climate, with many thanks. Now consolidated here.
[11] www.desmogblog.com/gmu-still-paralyzed-wegman-and-rapp-still-paranoid 45p
[12] www.desmogblog.com/mashey-report-reveals-wegman-manipulations 17p
[13] www.desmogblog.com/wegman-report-not-just-plagiarism-misrepresentation 12p
[14] www.desmogblog.com/curious-coincidences-george-mason-university-ed-wegman-milton-johns-and-ken-cuccinelli
[15] www.desmogblog.com/fake-science-fakexperts-funny-finances-free-tax 213p
[16] www.desmogblog.com/ed-wegman-promised-data-rep-henry-waxman-six-years-ago-where-it
[17] www.desmogblog.com/see-no-evil-speak-little-truth-break-rules-blame-others
[18] www.desmogblog.com/2012/10/23/fakery-2-more-funny-finances-free-tax
[19] Hadi Rezazad, PhD Dissertation gradworks.umi.com/33/64/3364566.html
[20] Walid Sharabati, Phd Dissertation digilib.gmu.edu:8080/dspace/bitstream/1920/3384/1/Sharabati_Walid.pdf [DEE2010p] and [MAS2012c §W.5.7] discuss the plagiarism problems.
[21] www.galaxy.gmu.edu/stats/colloquia/AbstractsFall2007/TalkSept7.pdf *Infinite thanks to DC for this, which revealed much hidden information.* This key file disappeared August 2010, but an annotated copy is included in [MAS2010a §A.11.2]. DC saved one also: www.webcitation.org/6E35F5rZr deepclimate.files.wordpress.com/2010/09/said-talksept7.pdf
[22] Yasmin H. Said, Edward J. Wegman, Walid K. Sharabati, John T. Rigsby, "Social networks of author–coauthor relationships," *Computational Statistics & Data Analysis* 52 (2008) 2177 – 2184. Recvd 8 July 2007; accepted 14 July 2007. The (2007) vs (2008) difference has caused citation confusion. **Retracted**.
[23] link.springer.com/chapter/10.1007%2F978-3-642-03739-9_30?LI=true
[24] content.usatoday.com/communities/sciencefair/post/2010/10/wegman-plagiarism-investigation-/1 UPDATE 05/26/11 on Walsch comments
[25] www.usatoday.com/weather/climate/globalwarming/2010-11-21-climate-report-questioned_N.htm
[26] content.usatoday.com/communities/sciencefair/post/2010/11/wegman-report-round-up/1

VER2011a 05/16/11 Retracted climate critics' study panned by expert[28]
VER2012 02/22/12 Univ. reprimands climate science critic for plagiarism[29]
WAS1994 1994 Stanley Wasserman and Katherine Faust, *Social Network Analysis – Methods and Applications*, Cambridge
WEG2006 07/14/06 Wegman Report (**WR**) - Ad Hoc Committee Report On The 'Hockey Stick' Global Climate Reconstruction[30] T126 includes the WR and several days' testimony to Congress, including a follow-up report.
WEG2006b 07/27/06 Testimony of 07/27/06[31] This is also part of T126.
WEG2009 02/16/09 Proposal written for ARO, not accepted[32]
WEG2010 06/18/10 Resume of Edward Wegman, archived[33]
WEG2010a Nov 2010 Wegman FOIA "Relevant emails" to Dan Vergano[34]
Caveat: seems a partial collection, very relevant, but one cannot know what might have been missing. Oddly, Wegman provided this, not GMU. IT had much irrelevant material, including some using his gmail.com address.
WSJ2006 07/14/06 Hockey Stick Hokum[35] Wall Street Journal editorial, appeared same day as Barton announcement and contained a falsified image.[36]

Further reading

Raymond S. Bradley, *Global Warming and Political Intimidation*, 2011.
Michel E. Mann, *The Hockey Stick and the Climate Wars…*, 2012.

[27] www.usatoday.com/weather/climate/globalwarming/2011-05-15-climate-study-plagiarism-Wegman_n.htm
[28] content.usatoday.com/communities/sciencefair/post/2011/05/retracted-climate-critics-study-panned-by-expert-/1
[29] content.usatoday.com/communities/sciencefair/post/2012/02/george-mason-university-reprimands-edward-wegmand-/1
[30] archives.republicans.energycommerce.house.gov/108/home/07142006_Wegman_Report.pdf Edward J. Wegman, David W. Scott, Yasmin H. Said
[31] archives.republicans.energycommerce.house.gov/108/Hearings/07272006hearing2001/Wegman.pdf part of [BAR2006]
[32] www.documentcloud.org/documents/524550-descriptf-network-science-proposal.html
[33] web.archive.org/web/20100609135746/www.galaxy.gmu.edu/stats/faculty/wegman.resume2.pdf
www.galaxy.gmu.edu/stats/faculty/wegman.resume2.pdf originally, but it and other key files vanished [MAS2010, §A.11.1], ~between 08/16/10 and 08/23/10.
[34] www.documentcloud.org/documents/527437-relevant-emails-redact-addresses.html
[35] archives.republicans.energycommerce.house.gov/108/News/07142006_1990.htm
[36] sg.wsj.net/public/resources/images/ED-AE505_1hocke_20060713182815.gif
The false citation's odd history is explained:
scienceblogs.com/stoat/2012/10/08/more-use-and-abuse-of-ipcc-1990-fig-7-1c

Key people

Most are discussed in [MAS2010a].
Stanley Azen, USC, past Editor-in-Chief of *CSDA* [MAS2011a]
Joseph Barton (R-TX), **Ed Whitfield** (R-KY), US Representatives who got Wegman recruited, and promoted the WR
Jerry Coffey, associate of Jim Tozzi (Data Quality Act), consultant for House Republicans, old associate of Wegman's, contacted him for WR
Myron Ebell, CEI, Cooler Heads Coalition, with Fred Singer, one of two key recruiters of McKitrick and McIntyre, [MAS2010a]
Milton Johns, lawyer for Wegman and Said, [MAS2011d].
Steven McIntyre, retired mining consultant, Ontario, Canada. With McKitrick, created talk that acted as WR blueprint [MAS2011a, p.17].
Ross McKitrick, Prof. Economics, U of Guelph, Ontario, Canada
Pat Michaels, was at U VA, now CATO. In 2010, was a GMU Distinguished Senior Fellow, taught course for School of Public Policy.
Fred Singer, SEPP [MAS2012] long affiliated with GMU's Institute for Humane Studies in the 1990s, worked closely with GMI.
Peter Spencer, Barton staffer, met Wegman after Coffey, sent materials

Contributors to WR and related work (Wegman, associates)
Edward J. Wegman, GMU
David W. Scott, Rice University
Yasmin H. Said, PhD 2005 (Wegman), Johns Hopkins University (2005-2006), then back at GMU by date of WR release.
An unknown 4th person, who later dropped out[37]

WR ack'd 2 Wegman students for help, vaguely:
John T. Rigsby III, Naval Surface Warfare Center, MS 2005
Denise M. Reeves, MITRE, PhD 2009, wrote the oft-copied SNA text, , clarified later by Wegman [MAS2011a]

Walid Sharabati, PhD, 2008. Unmentioned in the WR, he contributed much of response to Rep. Stupak in 2006, [WEG2006b], part of T126.

[37] *It might be quite worthwhile to locate this person and ask them to comment on how they got asked, what they did and why/when they dropped out.*

GMU administration – January 2011 and July 2012[38]

Wegman is shown twice as per *Connect2Mason*:[39]

> *'Wegman holds a 30 percent appointment in the Department of Statistics, but his 70 percent, majority appointment is in the School of Physics, Astronomy, and Computational Science in the College of Science.'*

Those marked (➔) were certainly involved, plus at least one of the Deans, one of whom seems responsible for a 5-month delay. Since no name has appeared, this report just uses "the Dean" in any discussion of process.

January 2011, July 2012, with organizational changes noted:

➔President **Alan G. Merten,**[40] now **Angel Cabrera**
J. Thomas Hennessy, Jr. SoPP, Chief of Staff, Office of President
➔Provost **Peter Stearns**[41]
➔VP for Research and Economic Development **Roger R. Stough**[42]
Also a 20-year Professor of Public Policy (SoPP)[43]
➔ Assistant to VP Stough **Donna Sherrard**[44]
? Dean, College of Science, **Vikas Chandhoke**[45] *(likeliest Dean)*
(2011) Dept. Hd, Computational and Data Sci, D. Papaconstantopoulos [46]
(2012) Director, School of Physics, Astronomy, and Computational Sciences (SPACS) Michael Summers[47]
➔Professor Edward **Wegman**[48] (70%)
? Dean, Volgenau School of Info. Tech. and Engr, Lloyd J. Griffiths[49]
Department Chair, Statistics, William F. Rosenberg[50]
➔Professor Edward **Wegman**[51] (30%)

[38] I have not rechecked all these since July 2012, but this covers the people in the most relevant period.
[39] www.connect2mason.com/content/university-committee-finds-professor-guilty-research-misconduct-document
[40] www.gmu.edu/resources/visitors/bio.html. retired➔ Ángel Cabrera July 2012.
[41] provost.gmu.edu/stearns; provost.gmu.edu/integrity/index.html (by Stearns)
[42] policy.gmu.edu/tabid/86/default.aspx?uid=80 (URL changed)
[43] www.webcitation.org/69bhBF8iw archived
[44] provost.gmu.edu/seniorstaff
[45] cos.gmu.edu/about/administration OK
[46] cds.gmu.edu/node/15; as of 03/05/11 cds.gmu.edu/node/36
[47] spacs.gmu.edu/content/about-spacs
[48] cds.gmu.edu/node/40
[49] volgenau.gmu.edu/about_ite/dean.php ➔ Kenneth Ball, August 2012.
[50] statistics.gmu.edu/pages/people.html

The Board of Visitors[52] was led by Rector Ernst Volgenau, recently replaced by C. Daniel Clemente. The July 2012 Board and others were:

'C. Daniel Clemente Rector
Nancy Mitchell Pfotenhauer Vice Rector53
Edward J. Newberry Secretary

Board Members
Karen Alcalde
B.G. Beck
Reginald J. Brown
Kimberly O. Dennis
Kathleen M. deLaski
Anthony R. Jimenez
Carol Kirby
Mark F. McGettrick
Stuart Mendelsohn
Steven P. Mullins
Robert F. Pence
M. Siddique Sheikh'

Faculty Representative
June Tangney

Student Representatives
Steven Scott
Alexander Williams

Faculty Committee Representatives
David S. Anderson
Gerald Hanweck
Michael Nickens
Martin Perlin'

[51] statistics.gmu.edu/people_pages/wegman.html
[52] bov.gmu.edu [MAS2012c, §A.5] discusses some other key members of the Board of Visitors, such as Vice Rector **Nancy Mitchell Pfotenhauer** (Independent Women's Forum, Koch Industries), Members **Kimberly Dennis** (Searle Freedom Trust, DONORS TRUST) and **Mark F. McGettrick** (EVP/CFO of Dominion Resources). Some of these entities have long been involved in funding climate anti-science activities. Koch Industries and Dominion have been major contributors to VA AG Ken Cuccinelli.
[53] Recently named Vice-Rector, she was on BoV in mid-2011.

Disclaimers, criticisms not implied

This report makes specific allegations against specific people, such as Wegman, Said and some members of the GMU administration.
Others are noted for necessary context, and in case their testimony is ever needed. No guilt-by-association is ever implied, unlike Wegman works that misused SNA to assert bad behavior in paleoclimate research, without evidence. No complaint is made about other GMU faculty or against Wegman's students or other coauthors. *In fact, some students may have been misused or mis-trained, perhaps to their detriment.*

Misleading the agencies or not?

If ARO and NIAAA searched for copies of the works and spent the large amount of time to assess them in detail, and if they still thought all the work was fit, starting with T126, then one might complain against the agencies themselves for misuse of funds and inadequate supervision.
It seems far more likely that volume and complexity of the reports obscured the reality, given reasonable limits on time people can or should spend to check everything.

Thus, no criticism of ARO is implied.

Wegman had a long history of productive research grants with ARO and other DoD branches. *The proposals for 0447 and 0059 seemed mostly plausible, from a senior researcher.* Much effort was needed to sort through the voluminous 0447-claims and confusing 0059 claims, find copies of the works and unearth obscured issues:

- Almost all the fit, claimed papers had been done earlier(☐), but often that was discovered only by finding copies of the papers themselves and checking publication histories.
- T126 was the largest single 0447 work, but buried among ~40 talks.
- While some works might have been obviously unfit by title/abstract, many required domain familiarity to know. For example, Wegman and Said claimed various SNA works that had nothing to do with potential SNA applications for 0447, 0059 or 5876.
- Peer-reviewed journal papers clearly derived from 0447/0059 were minimal, but publication lags make that more obvious in retrospect.

ARO then rejected Wegman's 2009 proposal. Even with no awareness of plagiarism issues, the referee reports were *negative*, **§Z.2**-05/07/09**.**
Wegman seems to have gotten no new grants from ARO since 2006.

Likewise, no criticism of NIAAA is implied.

This proposal *(reasonably)* required several iterations, but they were instructive. Reviewers raised legitimate concerns and the proposal changed to add (unkept) promises. NIAAA thought Said's statistics background was good, but worried about academic background and real connection with alcoholism research. NIAAA plausibly relied on:

- superlative recommendations from Wegman
- claims of P401 (dissertation) and P402 (genetic algorithms), both of which had plagiarism, but discovered only in 2010
- promises of coursework, **§S.1,** by year 50%, 25%, 0%**.** Then, she just audited a few courses, pleaded lack of money to pay for formal courses.
- promises to work closely with Dr. Wieczorek, but with few results
- claimed intent to devote her career to alcoholism research, quickly contradicted by many unrelated activities, visible mostly in retrospect.

Said's impressive-at-first-glance reports to NIAAA had numerous claims, some unfit or redundant, but lacked peer-reviewed journal papers. Works were mostly done for statistics, not health audiences. She listed SNA work, but as in 0447, scientific coauthorship SNA was unfit. As did Wegman in 0447she tried to justify everything she had done as relevant to 5876, including ARO work, text mining, T126 and foreign travel, **§S.3.1**.

*NIAAA obviously tried to help a young postdoc, but the pattern seemed to abuse NIAAA. Following is consistent with all evidence here, **§S**:*

- Wegman was personally motivated towards alcoholism studies, but had never gotten any grants or published any peer-reviewed work in it.
- *Said's academic background lacked relevant domain background for alcoholism research. She then did a dissertation with unusual rapidity.*
- *At least some of her claims seemed "manufactured" to get the grant or pad her resume. Wegman promoted her into many positions he controlled or influenced, far above her actual accomplishments.*
- *The proposal seemed designed to keep her at GMU, far less appropriate than nearby Johns Hopkins U, letting them do some alcoholism work, although never publishing a peer-reviewed research journal paper on it. She engaged in numerous distractions, including travel rather unusual for a postdoc. When the grant ended, so did alcoholism work.*
- *There was no evidence of dedication to health research or real ability to do it, but this funding might have gone to another who had them.*

0 Introduction

This report builds on evidence documented in earlier reports over the last few years. In 2009 Canadian blogger Deep Climate (DC) unearthed a few pages of plagiarism in the WR [54] and found more later. Each discovery generated more leads, connections, and problems, even when limited to public data. Some loose ends hint at evidence that would need subpoenas.

Reports so far have summarized contributions by many people:

- The people, organization and machinery of climate anti-science, following a coherent political strategy over many years [MAS2010].
- As a step in that strategy, the recruitment of Edward Wegman and creation of the WR, filled with errors, plagiarisms and untruths, but very well-connected with the machinery above and well-marketed by US Representatives and the *Wall Street Journal*. [MAS2010a]
- Plagiarism was easy for almost all to see, once found by DC or others. WR falsifications were harder to explain, but some were [MAS2011b].
- GMU seriously and even falsely mishandled a relatively simple academic misconduct complaint, and odd connections emerged. [MAS2011, MAS2011a, MAS2011d, MAS2012c]

The main discussion is **§0-§6**, *likely* more than enough for most readers, but backed by voluminous data and analysis for any who must dig deeper, as a few may, given serious allegations of actions *that may rise to felonies.*
§0 outlines the background common to the allegations.
§1-§5 describe the major allegation groups, plus any related concerns
§6 concludes the main discussion

Appendices **§A-$Z** record the details needed to support the allegations, including links to copies of FOIA replies. Readers unfamiliar with Federal grants might read **§B** first. *Grants cannot be treated as slush funds.*
§M-§P briefly describe earlier Wegman grants that *seemed far more productive and better-handled as expected from an senior researcher.*

§Q-§T dissect the 4 grants, their histories and results.

§K details the 2001-2010 papers and talks graphically-consolidated next in **§0.1**, which centers on the best-documented grant, 0447. Its Final Report was relatively complete and precise. Anything in the central black-outlined section was 0447- claimed, but those below the line were *unfit*. Fit papers are highlighted blue, but those tagged (☐)were clearly or *likely* done before 0447 started[55] and thus false claims. Fit papers essentially disappeared 2006-2008, leaving only talks, some *seeming mostly to present students' work*, often repetitively. Any instance of symbols ①②③☆★ is a clear problem and most ☐ symbols *very likely* are problems as well.

In **§0.1,** at left are shown earlier works, from years that *seemed* productive with many papers and coauthors. At right, more resume works are listed to allow for fit papers that might have been delayed, but very few appeared. Above and below are shown other works involving Wegman or Said. Whether claimed or not, many were distractions. Gray-highlighted *WIREs:CS* papers at lower right offer further context.

People might disagree on some classifications[56] without really changing the results, given the numerous problems, starting with T126 and P179.
Wegman and Said *likely* spent at least 60% of their time on *blatantly-unfit* efforts and claimed credit for them.[57] *This is not marginal misuse of funds.*

§0.1 displays a strong Wegman shift into 0447-unfit work. Said did mostly 5876-unfit work, **§0.2.** Both misused 0059. It also shows a real dearth of peer-reviewed research journal work from 2006-onward.

[54] deepclimate.org/2009/12/17/wegman-report-revisited; deepclimate.org/2009/12/22/wegman-and-rapp-on-tree-rings-a-divergence-problem-part-1 ; deepclimate.files.wordpress.com/2010/01/wegman-bradley-ice-cores-corals1.pdf ; deepclimate.org/2010/04/22/wegman-and-saids-social-network-sources-more-dubious-scholarship among the earliest of dozens

[55] It is rather problematical to claim work completed before a grant, but papers published afterwards might fairly ack it, given publication delays.

[56] For example, readers might choose to ignore orange-shaded works, and that changes the numbers, but not the overall conclusions on funds misuse.

[57] Academic friends say that acceptable boundaries differ, *but this seems far over. Few would care if Wegman had claimed some marginally-relevant efforts among substantial fit work. For instance, as a part-time computing historian I enjoyed* ***P181****, which recorded useful history of statistics, even if it had nothing to do with 0447. Numerous unfit claims obscure the paucity of relevant peer-reviewed research papers. The overall pattern raises seriosu concerns on use of funds.*

§0.3 shows a chronology for Wegman and Said's grants and activities, with pointers to supporting details and distractions, **§H.** *The chronology raises concerns.* Senior people often engage in many concurrent activities, but Said was also involved in many efforts *seemingly far removed from the purpose of her postdoc.* Her proposal **§S.1** claimed career dedication to alcoholism research and promised actions that were quickly displaced by distractions. Wegman had clearly initiated the alcoholism research, and for years authored or coauthored almost every alcoholism work with Said.

Wegman had a long history of productive success in various Federal research grants, especially from branches of the Department of Defense. Those supported papers and talks mostly about statistics and computing, credible and influential for decades through 2004.

Wegman 0447-claimed almost every 2004-2008 paper or talk in his resume. **In effect, it seemed that he treated 0447 as funding to be used for any work he wanted to do.** In 2006-2008, ARO-fit peer-reviewed journal papers disappeared almost entirely in favor of unfit work.
Talks seemed to be repetitive, with little evidence of much progress, **§K.2**.

- He got asked to do the hockey-stick work and certainly threw himself, Said and grad students into it with vigor, as P178, P179, T126, and 6 more talks were claimed 2006-2007, plus 4 later talks not 0447-claimed (P407,T409, T414, T415). During 09/01/05-07/27/06, Wegman and Said had to spend much effort reading 90+ papers and assembling the 91-page WR on paleoclimate hockey-stick and Social Networks Analysis, unfamiliar topics. More time was spent on testimony, the rest of T126. Wegman spoke often about the hockey-stick in 2006-2007 and Said spoke occasionally. Wegman's report of spending time on them was strongly corroborated by the striking plummet in fit work.[58] During 2007, Wegman and Said seemed to make many visits and spend much time in Saudi Arabia, and their patent proposal showed they were seeking funding there. During late 2007-early 2008, Said spent ~2 months in Saudi Arabia[59] and 3 months in the UK.
- Both Wegman and Said several times expressed financial worries, but 2007-2008 saw an upsurge in foreign travel. *That raises concerns* regarding possible distraction and funds use, for which the latter information is not publicly available. It is a normal part of academic life to attend conferences, but one still has to spend time actually doing research. People can now better keep in touch with students via the Internet, but extensive foreign travel can be time-consuming. The last ARO grant was 0059, for one year, starting in late 2006. *Concerns may be raised* on the heavy travel during 0059 and the quality of grad student supervision that year. *Various other details raise concerns*, especially with regard to possible mishandling of grad students listed in blue in center. *Concerns may be raised* about handling of intellectual property.
- Hockey-stick attacks almost stopped 2008 (just T147), but "Climategate" stirred a resumption in 2009-2010. Despite criticisms starting in 2006, even from his reviewers, he and Said kept repeating the same messages, mostly for non-expert audiences: T147, T423, T424, T427, T428, T429.
- Alcoholism work seemed to stop when Said's grant did, mid-2009.
- By 2009 their main efforts *seemed* to go into *WIREs:CS*, edited by Wegman, Scott and Said, a sample shown in gray cells in lower right corner, **§K.4.2**. In their own journal, Wegman and Said wrote 2 "peer-reviewed" papers that had serious plagiarism issues. After multiple complaints to Wiley over a year they (quietly) departed in June 2012 . *WIREs:CS* published many papers by Wegman students, against whom no complaint is made. However, c*oncerns about peer review quality must be raised about WIREs:CS, ironic in light of Wegman's repeated, baseless claims of bad peer review among climate scientists.*
- *All this seems a sad change from decades of productive work[60] through 2004,that evolved into increasingly strange behavior by 2006. By 2007 ARO-fit work was very weak.*

Given the major effort required for T126, Wegman and Said must have spent more than 60% of their time on work unfit for the grants.

Finally, **§0.4** displays the flows of funds to unfit works, *whose pattern seems strong evidence for commingling of funds.* ***In effect, Wegman and Said treated the grants as unrestricted support for any work they chose.***

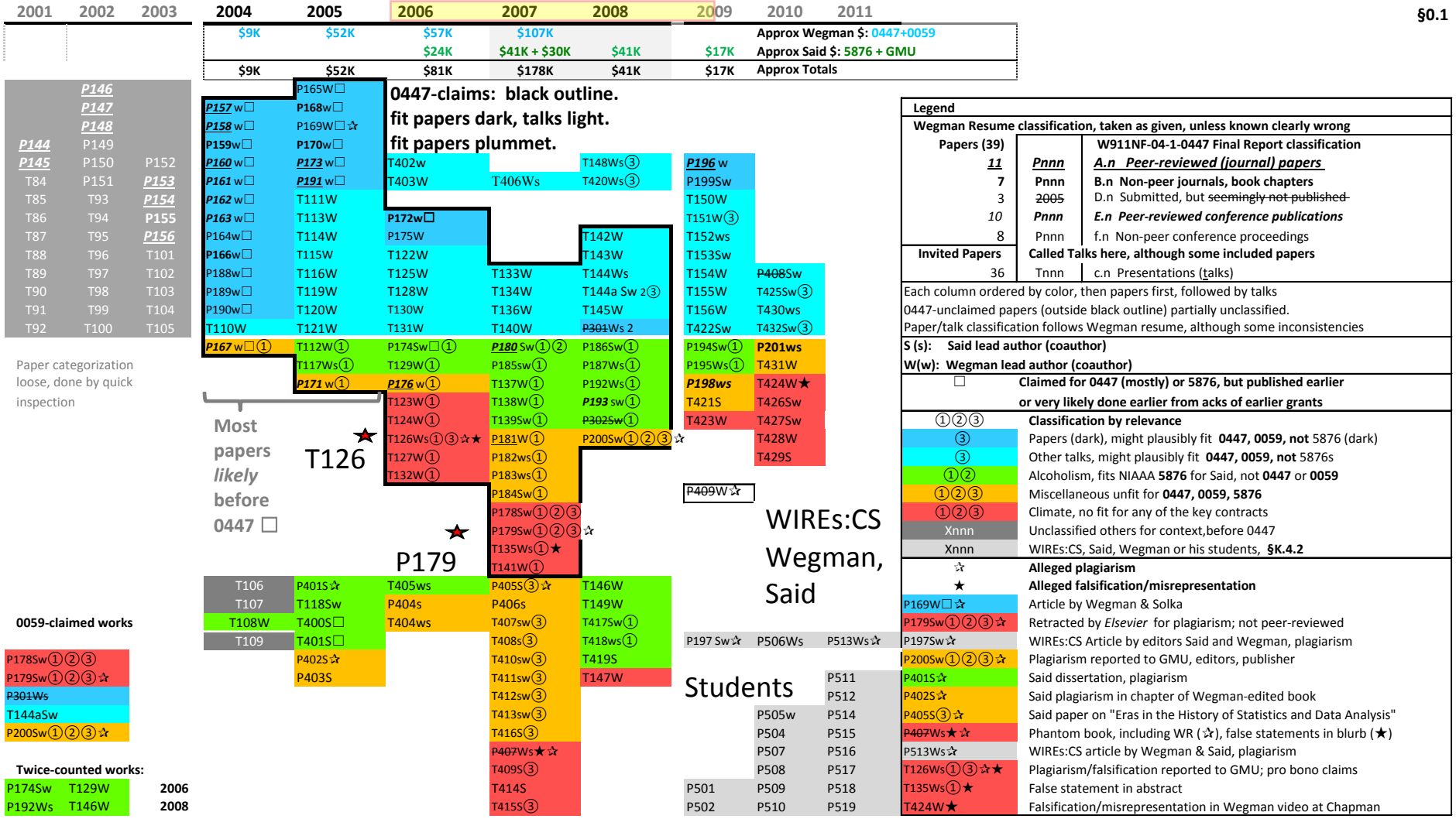
[58] Had Wegman not claimed these unfit works, funds misuse would still be alleged, given the other evidence. He correctly claimed he worked on these.
[59] *Saudi Arabia seems among the most unlikely conceivable locations for alcoholism research, especially by a female postdoc. No works resulted.*

[60] I admire Wegman's earlier work on computational statistics, he used SGI computers and SGI even co-sponsored one of the Interface meetings.

0.1 Papers & talks summary chart - all

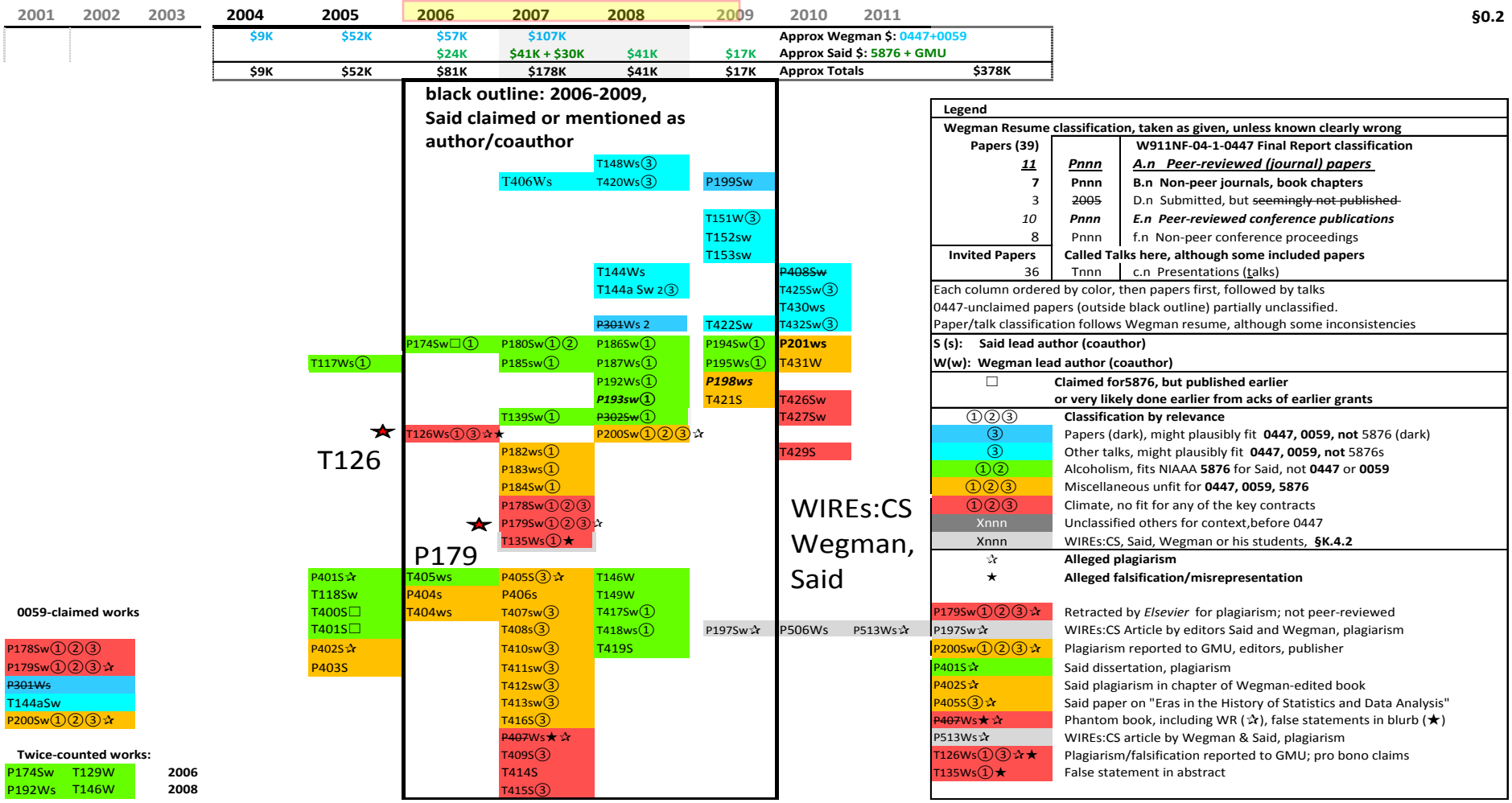
Wegman shift from 0447/0059-fit papers, talks to unfit hockey-stick attacks, alcoholism, others, peer-reviewed journal work ➔ zero, §K.
All fit claimed papers except **P175** were clearly or likely (☐) done before 0447, and then fit papers essentially disappeared, while unclaimed distractions rose. The actual spreadsheets are attached at www.desmogblog.com/foia-facts-2-no-pro-bono-federal-funds-mis-used
Inside the black outline, the 30 works below white line are problematical.

§0.1



0.2 Papers/talks summary chart - Said subset

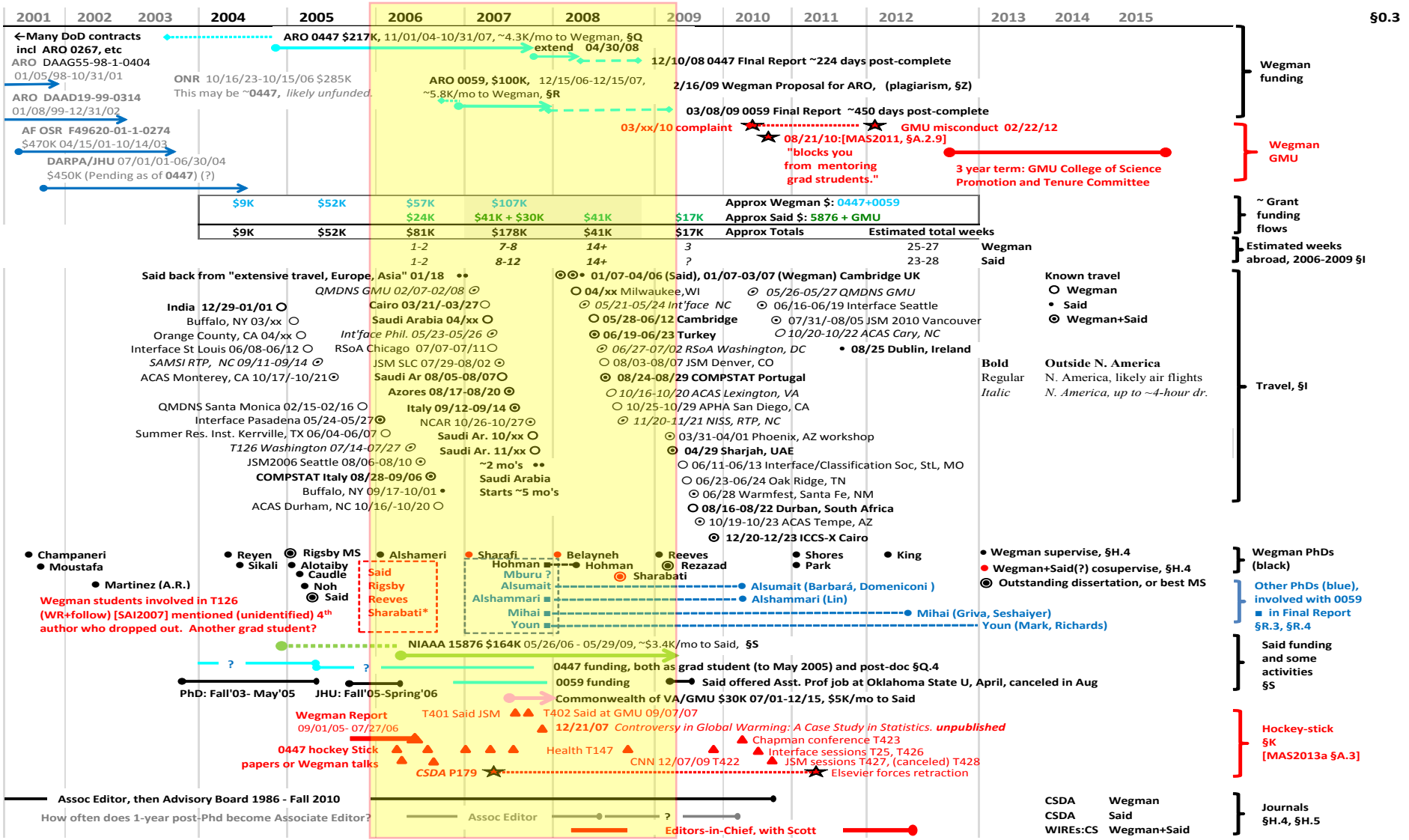
Said: 5876-fit alcoholism, unfit hockey-stick attacks and unfit others. This extracts cells from **§0.1** with Said as author, coauthor or ack'd by Wegman for joint work. The 0447/0059-fit papers, talks were distractions from 5876, but might have fit any ARO funding she got. Year 2007 was dominated by 5876-unfit work, despite the highest grant income. Alcoholism work disappeared by mid-2009 and papers in peer-reviewed research journals totaled zero. Plagiarism is alleged against 8 works authored/coauthored by her, including 6 related to 5876. *Does this chronology and small fraction of **alcoholism** works show intent to dedicate her career to alcoholism study?* From this and next page, diversions started immediately and continued, as 2006 and especially 2007 were dominated by unfit work. Almost nothing was done without Wegman, and as shown in **§K.2**, some unusual repetitiveness can be found *Given the redactions in reports, she may well have claimed more.*



0.3 Chronology of grants and other activities

Correlation is not causation, but funding and foreign travel *seem* to correlate well, especially in 2007, when 0059 was active and Said got $30K from GMU, followed by more foreign travel in 2008.

This level of travel may be perfectly fine for a senior person, *but it seems unusual for a postdoc in that crucial early career stage. The actual spreadsheet may be easier to read for anyone who wants to check details: www.desmogblog.com/foia-facts-2-no-pro-bono-federal-funds-mis-used.*



0.4 False claims and alleged funds commingling

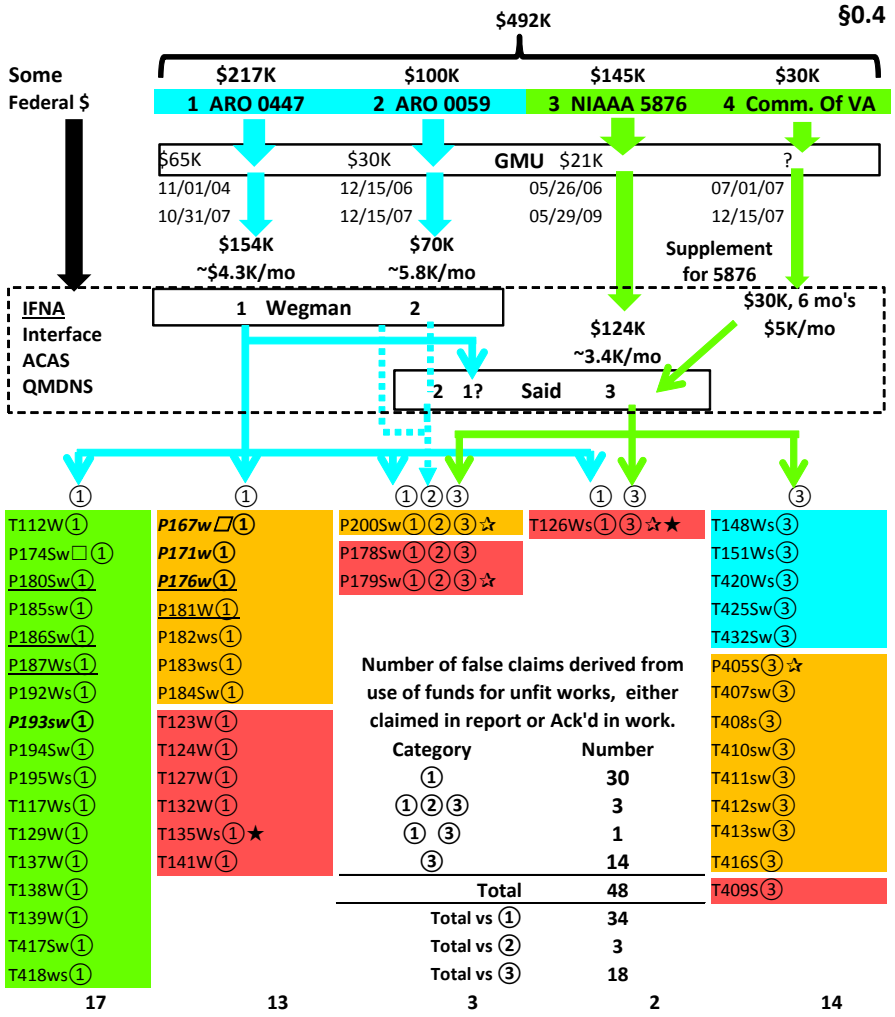
This chart extracts from **§0.1** those marked ① ② or ③ because a report claimed a unfit work or because a work itself ack'd a grant for which it was unfit. *Given redactions in reports, Said may have claimed more, adding more* ③. Copies of many works, especially talks, could not be found, but more acks might add ① ② or ③ to some of these:
P182 ws①, P183 ws①, P184 Sw①, P198 ws, P201 ws, ~~P301~~ Ws, T118 Sw, T146 W, T404 Ws, T405 Ws, T407 sw③, T410 sw③, T411 sw③, T412 sw③, T413 sw③, T422 sw.

This chart illustrates flows of grant funds through GMU to Wegman or Said, for any of the works marked ① ② or ③. Wegman and Said took the money, then claimed work for grants or ack'd them with little concern for their fitness. Alcoholism works often got claimed for 0447 and non-alcoholism works for 5876. Orange and red highlighted works fit none of the grants. **A total of 48 works had specific erroneous acks or claims.**[61]

In effect, all 3 key grants were used for many works that could not possibly fit, *as though funds had simply been commingled and used for anything they wanted. This seems strange for someone with Wegman's long experience with research funding and careful acks in earlier papers.*
The Commonwealth. of VA funding is mentioned in **§S.3.2**, but with no information about the purpose of the grant, progress reports or results.[62]
The FOIA reply from GMU showed only that Said got $2.5K twice/month, 07/01/07-12/15/07, signed by Dimitrios Papaconstantopoulos, Chair of the Department of Computational and Data Science, **§T.**

Concern. The IFNA box represents no allegation of wrongdoing, just a reminder of funding flows and frequent IFNA activities by Wegman and Said, both in organizing conferences and giving talks at them.[63]
IFNA plausibly paid some US travel expenses and it certainly *seemed* reasonable to have gotten some Federal funding. In 2007, Interface was in Philadelphia, QMDNS at GMU and ACAS in Houston. The 2007 $6,682 travel cost *seemed* a bit high for mostly local conferences, **§H.1.**

IFNA also got $18,100 of Federal funding that year and other entities often cosponsored these conferences. IFNA is a 501(c)(6) nonprofit, but in some ways it *seems* to act as a business that runs conferences. *Again, this is no allegation of IFNA wrongdoing, just a concern raised by likely commingling of funds elsewhere.*
§F offers a checklist of plagiarism, falsification and relevant authorities.



[61] This does not include the12-19 false claims tagged□ for having (likely) preceded 0447 and 2 such for 5876.
[62] *VA AG Ken Cuccinelli is reputed to be vigilant on the use of VA funds.*
[63] All financials seem to have been done by the Treasurer, Wegman, **§H.1.**

1 0447 Wegman (DoD) [64]

1.1 False claims. The Final Report claimed credit for 75 works, of which 2 were never published, leaving 73. By examination of works or at least their titles/abstracts, 30 of 73 were 0447-unfit, thus **30 false claims** for ①.

Classification	P claim	P publ	Talks	P+T	Totals publ
ARO-relevant **papers**, talks[65]	21	20	23	43	43 **fit**
NIAAA-alcoholism ①[66]	8	7	6	13	
Miscellaneous ①[67]	8	8	0	8	Unfit ①
Hockey-stick①	2	2	7	9	30
Totals published	39	37	36	73	

After 2005, 0447-fit work plummeted, especially peer-reviewed journal papers (***Pnnn***). Of 20 published fit papers, 12 were done in 2003-2004 before 0447 and 7 were *likely* or mostly done earlier (☐), leaving only one fit paper *clearly* done during 2006-2008, while many unfit papers were seen. After 0447, 4 more unfit works ack'd it, giving 30+4 = **34 false claims** by topic ①, plus **12-19 more false claims** for pre-0447 work, for a grand total of **46-53 false claims** among 73 works. All available evidence shows that **Wegman spent >60% of his ARO-funded time on unfit work.** At least 5 works are alleged to have plagiarism and/or falsification.

1.2 False statements. [68] Wegman claimed 91-page WR and presentations to Congress, T126, an 11-month effort in unfamiliar fields. He claimed follow-up paper P179 and 7 related works 2006-2007. He and Barton often told Congress the WR was "pro bono," effectively **false statements**, **§J.1.** [MAS2010a] alleges that the WR had many false statements and *testimonies to mislead Congress and affect public policy.*
Wegman refused a 2010 FOIA request for T126-related information:

> 'Separately in that response, Wegman said his "email was downloaded to my notebook computer and **was erased from the GMU mail server**," and he would not disclose any report communications or materials because the "work was done offsite," aside from one meeting with Spencer.' [VER2010a]

1.3 False statements. Wegman claimed acks in 39 papers, but at least 25 did **not** ack 0447, thus **25 false statements .** Oddly all 5 discoverable papers that ack'd 0447 were actually unfit: P179, ***P180***, ***P192***, ***P193***, P200, Unclaimed, unfit later works also ack'd 0447: P194, P195, T417, T418.

1.4 False statements. Wegman promised WR code to Rep. Henry Waxman in 2006, but asserted a need for US Navy approval as reason for delay. An earlier FOIA [MAS2012b] showed that was a **false statement** to Waxman, for work claimed to ARO. As of this writing, the code has not been published. *This might be viewed as a cover-up, or if code has been deleted, perhaps obstruction of justice. It likely had a 100:1 cherry-pick.*

***Concern 1.5 Quality*, §F**
Allegations of plagiarism (☆) and/or falsification (★) are made against **P169**, P179, P200, T126 and T135. *This raises concerns about other work.*

Concern 1.6 Misuse of grad students for WR?
Wegman knew many fine statisticians, but instead, much of T126 was done by Said and grad students Rigsby, Reeves, Sharabati and *possibly the "unknown 4th participant who dropped out"* [MAS2010a §A.11.2, slide 5]**.** Apart from falsely stating the WR as work by experts, *was this ethical use of grad students?* Concern 2.3 is another of the same sort.

Concern 1.7 Said use of 0447 funds? §Q.4
Wegman claimed 0.25-time support for Said as grad student and postdoc, concurrent with T126 and P179, but she coauthored none of the fit papers he claimed, and just 2 fit talks (T144, T144a). She never ack'd 0447 in any works found. *Did 0447 funds support her for work on T126?*

Concern 1.8 GMU oversight responsibility for alleged grant fraud?
Bradley raised the P179 Federal funding in May 2010 [MAS2011 §A.2.4]:

> 'The reported plagiarism also points to a number of Federal grants that may be implicated in this matter, which raises a whole set of additional oversight concerns of which you should be aware.'

GMU got 30% of the funds and of course had copies of the grant reports. When the ARO sent the 0447 report, a brief scan found T126 within a few minutes. *One might think GMU would have done at least minimal due diligence to fulfill its oversight duties.* See Concern 2.5.

[64] Backing detail for this section is **§0.1, §0.3, §F , §K, §Q.**
[65] P191 was published in 2008, but submitted in 2005, **§K.4**. As in {P172, P175}, Wegman coauthored with a past student on a paper derived from their PhD topic.
[66] Refereed abstracts P186 and P187 were claimed as refereed journal papers.
[67] P176 (consumer price indexes, Wegman 4th author), P181 (interview)
[68] www.law.cornell.edu/uscode/text/18/1001 18USC§1001

2 0059 Wegman, Said (DoD)[69]

2.1 False claims. Wegman and Said claimed 2 0559-unfit papers (social networks attacks on climate), P178 and P179.
Unfit P200 ack'd 0059, yielding at least **3 false claims** by topic ②.
That left only T144a and ~~P301~~, the latter never published.
P179 was retracted for plagiarism, also alleged against P200, **§F.**

ARO reporting rules, **§B.1**, require works to have "brief, but complete information" and be classified properly. Unlike 0447, the 0559 Final Report provided references, some vague, but just saying, **§R.4**:

> **p.10** 'Several papers were developed based on research carried out under this project. These papers exploited aspects of the developments here. They are Said et al. (2007, 2008), Said and Wegman (2009), and Wegman and Said (2007). Presentations were given in a number of forums that credited this contract.'

Their report was also filed 449 days after completion, about a year late.
§R.4 summarizes the resulting analysis as:

P178 Sw①②③	0059-unfit, *likely* similar to P179, would add ☆
P179 Sw①②③☆	0059-unfit, retracted for plagiarism
~~P301~~ Ws	0059-fit (maybe), *unpublished so unclear*
T144a Sw	0059-fit
P200 Sw①②③☆	0059-unfit, not claimed, but ack'd 0059

At least 5 later talks ack'd 0059 and *seemed* fit, but also ack'd 5876:
T148 Ws③, T151 Ws③, T420 Sw③, T425 Sw③, T432 Sw ③.

Concern 2.2 Potential mismatch of proposal, implementation, results
The proposal described an *ambitious* software system, *perhaps derived* from an earlier, larger proposal to another agency, not funded, **§R.1**.
Although the 11/27/06 proposal was accepted by 12/15/06, **§R.2**, some implemented modules sometimes *seemed* not to match. The proposal mentioned no Arabic, but by the 05/05/07 progress report, **§R.3**, the Arabic module seemed well along and the student doing it "eager to graduate." It is impossible to know without detailed investigation, *but this proposal seems cobbled together quickly, and coherent results are unclear,* [70] *although individual parts seemed plausible efforts by the students.*

***Concern 2.3 Possible misuse of grad students?* §R.3, §R.4**
Concerns may be raised by the staffing history of this short project. The 05/07/07 progress report described the work of 6 students. The final report was very late, and mentioned the work of only 4.[71]
Unlike many Wegman students, they wrote no/few papers with him.

Name[72]	Final	PhD	Advisor(s)
① Alshammari, Eiman Tameh	Y	2010 (IT)	Lin (CS)
② Mihai, Felix	Y	2012 (?)	Griva (CDS), Seshaiyer(Math)
③ Youn, In-ja	Y	not yet (CS)	Mark (CS), Richards (EE)
④ Mburu, Peter	N	not found	Claimed as Said student(?)
⑤ Hohman, Elizabeth Leeds	Y	2008 (CS&I)	Wegman (CDS)
⑥ Alsumait, Loulwah	N	2010 (CS)	Barbará,Domeniconi (CS)

Except for Hohman, who was certainly Wegman's student, it is unclear which, if any, of the others was then actually working on a PhD with Wegman. *At one time or another, Said was claimed to be "supervising" Alshammari, Mburu, and 3 other grad students, slightly odd role for an alcoholism postdoc busy with many other activities, and who spent months away from GMU in 2007, **§0.3**, **§H.4**, **§I**.*

Concern 2.4 The invention?
§R.5 discusses a proposal by Wegman and Said for a patent that *seemed* mostly based on the students' work, but mentioned none of them.
This may be perfectly innocent, but raises concerns about Wegman and Said's approach to intellectual property.[73] *An IG might investigate further*, as most of the students and their advisors were easily findable at GMU or elsewhere, **§R.3, §R.4.**

Concern2.5 GMU responsibility for alleged grant fraud?
Does GMU bear any financial responsibility for 0059? This is the same issue as in Concern 1.8.

---

[69] Backing detail for this is **§0.1, §0.3, §F , §K, §R.**
[70] As elsewhere, grad student staffing of development work can be chancy.
[71] Interdisciplinary projects may reasonably include students from different groups. Professor might "borrow" students or offer them a chance to work on a short-term project for experience.
[72] en.wikipedia.org/wiki/Al- Al-names are often used inconsistently.
[73] Although the available files *raise concerns*, other unseen emails and files might ease them, perhaps by naming students as co-inventors on proposals. Intellectual property ownership can be a murky issue, although plagiarism is usually clear.

3 5876 Said (DHHS)[74]

3.1 Said claimed / ack'd 5876 for unfit works: **18 false claims** by topic ③, with plagiarism (☆) in 4:

P178 Sw①②③ "Implications of co-author networks on peer review," Said, Wegman, Sharabati and Rigsby (2007) (if found, ☆ likely)

P179 Sw①②③☆ "Social networks of author–coauthor relationship," Said, Wegman, Sharabati and Rigsby (2007/2008) RETRACTED

P200 Sw①②③☆ "Author-coauthor social networks and emerging scientific subfields," Said, Wegman, Sharabati, 2008 (really 2010).

P405 S③☆ "On the Eras in the History of Statistics and Data Analysis" J. Washington Academy of Sciences 93(1) Spring 2007, 17-35

T126 Ws①③☆★ Presentation of the WR to Congress 07/19/06, etc.

T148 Ws③ "Text Mining and Social Networks: Some Unexpected Connections," ACAS 2008

T151 Ws③ "Document Clustering and Social Networks," Classification Society Annual Meeting, St. Louis, MO, June, 2009' (really, Interface).

T407 sw③ "Relationship Between Competitiveness of Colleges and Levels of Tobacco Use" JSM2007

T408 s③"Access Control Model for E-Learning System" *JSM2007*

T409 S③ "Pro Bono Statistics and Public Policy JSM2007

T410 sw③ A Statistical Social Network Approach to Computer Network Optimization" *JSM2007 (July29-August 2, 2007, Salt Lake City)*

T411 sw③ "Style of Author-Coauthorship Social Networks: Statisticians of Prominent U.S. Universities" *JSM2007 (July29-August 2, 2007, SLC)*

T412 sw③ "Effect of Intravenous Insulin Infusion on Coronary Artery Disease" JSM2007 *(July29-August 2, 2007, Salt Lake City)*

T413 sw③ "Update on Cardiometabolic Risk Reduction: Role of the Endocannabinoid System Disease" JSM2007

T415 S③ ASA/NCAR climate workshop in Boulder, CO

T420 Ws③ "Mixture Models for Document Clustering" U MD, 2008.

T425 Sw③ "Preserving semantic content in text mining using multigrams." given by Said for QMNDS 2010

T432 Sw③ "High Dimensional Vector Space Methods for Characterizing Semantic Content," (2011)

She claimed P174, T400, T401, done 9-12 months early(□), 18+3 = **21 false claims.** T126, P179 falsified "pro bono" assertions in T409, T415.

[74] Backing detail for this is **§0.1, §0.3, §F, §H, §I, §K, §S.**

3.2 Plagiarism. As part of her grant proposal, she cited 2 works with plagiarism: P401 (her dissertation) and P402 (genetic algorithms). P179 and P200 have already been reported, as they ack'd 5876. Claims of P405 and T126 yield a total of 6 related to 5876.

Concern 3.4 Wegman had a strong interest in alcoholism research,[75] but it is unclear that Said really did. Promises made in **§S.1** in response to concerns *seemed unfulfilled* amid many distractions. Postdoc work was coauthored with Wegman, but only one conference paper and talk with Wieczorek. During the grant's 36 months, she spent 6-7 months abroad..

When 5876 ended, so did alcoholism works. Given timing and similarity of titles, it is unclear whether much research progress actually happened:

2006.08 P174 Sw① or T129W① "Geospatial distribution of alcohol-related violence in Northern Virginia" **COMPSTAT 2006**

2007.07 P185 sw "Temporal statistics for consequences of alcohol use" (***JSM2007***, *Salt Lake City, August)*

2007.10 ***P180*** Sw①② "Quantitative assessments of alcohol-related outcomes" *Chance.* **§R.6**-08/28/07 completed.

2007.11- Said in Saudi Arabia and UK ~5 months, Wegman at least ~3 months.

2008.06 P186 Sw① "Using administrative data to estimate cyclic effects of alcohol usage (refereed abstract)," *Alcoholism: Clinical and Experimental Research*" June 2008. *Said: a poster*, **§S.3.2.** P187 Ws① "Modeling spatiotemporal effects for acute outcomes in an alcohol system (refereed abstract)," *Alcoholism: Clinical and Experimental Research*" June 2008. *Said: a poster*, **§S.3.2.**.

The 5876 grant started 05/26/06, but she inquired about part-time work, **§S.5**-08/30/06, and soon was helping get 0059, **§R.6**-11/27/06. She pleaded insufficient funding to take formal courses, saying she "sat in" on courses, not a very concrete claim, and in any case, seemed not to broaden her exposure to the alcoholism research field. In **§Z.2**-12/15/08, Wegman was seeking ARO funding for her. **§S.6** consolidates the results of 5876 and compares them with journals where Wieczorek published, as well as other NIAAA awardees. Wegman and Said mostly spoke at statistics meetings and simply did not publish in those relevant journals.

[75] *It seems puzzling that Wegman did not just seek NIAAA funding himself, if he really wanted to spend so much effort on alcoholism research.*

4 Wegman February 2009 proposal (DoD)[76]

4.1 Attempted false claim. Wegman submitted a grant proposal to the ARO, but the text was ~50% copied without attribution, **§Z**.
It included the 5th known re-use of SNA text from [WAS1994, DEN2005] that first surfaced in T126 and P179, and was re-used in 2 dissertations.
According to the NSF, in some cases, this can rise to felony, **§B.2** (p.66), so perhaps Wegman was fortunate the proposal was rejected.
A few quotes may be useful for context:

§Z.2-10/01/08 Wegman to ARL

'I am still hoping we can get into a new contract arrangement with ARL. This past summer **I had no research support, which is very hard for me**.'

§Z.2-12/15/08 Wegman to ARO

'I hope this is of interest to ARO. Last summer, I went without research support and wound up having to **take a second mortgage on my home in order to make ends meet**. …
Would it be possible to include **some support for Dr. Yasmin Said** who has been working closely with me over the last 2 ½ years?'

§Z.2-05/07/09 ARO to Wegman, quoting one of the reviewers

'"The proposal outlines some potential new capabilities that can arise from the proposed study. **Some of those are very interesting**, like task **2** focusing on conversion of multimode non-binary adjacency tensors and matrices into lower degree networks or evolutionary algorithms for optimizing network assessment metrics. **Other tasks,** like **3** (evolving social networks), **4** (missing edges), **7** (connection between text mining and social networks), or **8** (limiting behavior of agent-based systems) **have been already studied and lack of references to the relevant work makes this reviewer doubtful of the likelihood that the proposed research will lead to new capabilities in this** *(sic)* **tasks."**'

Even without knowing of the plagiarism, that reviewer recognized serious problems with this proposal. Wegman had a long history of proposing and executing plausible research, **§0.3,** *but this was very weak. He and Said seemed dependent on a flow of research grants that had dried up.* [77]

Combined with other evidence from 2004 onward, this seems a sad pattern for a distinguished statistician who had written well-cited papers for years, ***§K.4.1.*** *T115 in April 2005 seemed to lament a downturn in funding for statistics, but did identify areas of opportunity:*

'Conclusions
Best of times and the worst of times …
Basic research support in statistics by federal government (and industry?) severely eroded,
but statistics … data … is everywhere.
Where is some action?
Social network modeling – homeland security
Streaming data
Text and Image data mining'

Interdisciplinary work can be quite productive. Application specialists may find interesting intersections or may apply existing techniques to new problems, with help from {statisticians, mathematicians, computer scientists, for example}. The latter must engage with application experts, as described by statistician Jim Berger at the 2007 NCAR meeting attended by Wegman [MAS2010a §A.4]. Good results lead to joint papers.

Wegman did not seem to do that, but seemed rather to send PhD students off to look quickly at unfamiliar topics. Then they would write papers or give talks about that, but for other statisticians, rather than in venues that would expose them to application domain experts. That makes it hard to evaluate the caliber of the work, especially as Wegman shifted to authoring almost entirely with Said and students.

The sorts of comments by the reviewer at left might also apply to:

- *SNA, graph theory [MAS2010a §W.5]* or ***§J.3***.
  They thought coauthorship analysis was new, seemed to be reinventing old ideas with new names, without clear knowledge of the literature.
- *Climate [MAS2010a], where they made numerous mistakes*
- *Perhaps text processing, perhaps, as seen in papers here*
- *Alcoholism, especially with H.I.V. connections*

From ***§K.4.1****, it seems application experts did not notice or pay much attention to Wegman and Said's work, despite all the talks.*

[76] Backing detail for this is **§B, §Z.**
[77] Google: climate science gravy train Whether in climate or statistics, *professors and especially postdocs rarely get rich on research grants*. Still, Wegman had an external business (dMining Technology) and statisticians may have reasonable consulting opportunities, so this *seems strange.*

5 GMU misconduct handling, failure to inform (DoD, DHHS)

5.1 Failure to inform. By all evidence, **GMU did not inform the funding agencies** at any of the 3 Federally-required steps in the misconduct proceedings for P179. October 2012 FOIA replies in **§C** offer consistent evidence that GMU did not properly inform funders.[78] **Neither ARO nor GMU could find any documentation or email on plagiarism**.

5.2 GMU broke many rules of handling misconduct, shown in earlier reports. Most complaints were ignored, despite possibility of grant fraud administered by GMU. Federal funding issues had been repeatedly mentioned to GMU, starting in May 2010. The then-applicable GMU Policy 4007, excerpted in **§A.1,** noted that funding agencies must be given the reports at the following steps, with dates from [MAS2012c 1.2.2b]:

- ~03/07/11 When GMU decides investigation warranted 2013
- ~10/11/11 When an investigation is completed
- ~02/22/12 When adjudication is completed

[VER2012] quoted GMU Provost **Peter Stearns:**

> '"We took these charges very seriously," Stearns said, in a telephone interview, **adding that the university will forward the investigation reports to federal authorities.** The National Institutes of Health and the Department of the Army supported the 2008 study.'

Plagiarism/falsification complaints were lodged 2010-2011, against Wegman, Said and others, including 4 grant-related works:

Work	MAS2012c	0447 ARO	0059 ARO	5876 NIAAA	GMU Action
P169	§4.3	Wegman	-	-	ignored
P179	§2.3	Wegman①	Wegman+Said ②	Said③	misconduct
P200	§4.6	Wegman①	Wegman+Said ②	Said③	ignored
T126	§2, §3	Wegman①	-	Said③	Bradley's rejected, 25+ pages ignored,
		DoD IG	DoD IG	DHHS IG, ORI	

GMU declined to investigate these well-documented complaints, despite possibility of grant fraud in P179. The **old** GMU Policy 4007, **§A.1** said:

> '**Pursues diligently** all significant issues and leads discovered that are determined relevant to the investigation, including any evidence of additional instances of possible research misconduct, and continues the investigation to completion.'

[78] *Perhaps GMU informed them since the October 2012 FOIAs, a year and a half late, but that does not nullify the seeming deliberate rules violations.*

The larger problem became rapidly clear from 0447 final report, as so many works were clearly unfit, including T126. Some may have been concerned at the time of the WR, [79] ***but then GMU administrators ignored Federal rules***[80] ***in many ways.*** GMU gets Federal funds, **§A.5.3.**[81]

[MAS2012c] chronicled George Mason University's mishandling of simple academic misconduct complaints. FOIA replies had exposed more details of an academic misconduct process that was at best incompetent. Even with repeated coverage by *USA Today,* a critical editorial by *Nature*, comments in *Science* and other publications, GMU ignored almost all 80+ pages of well-documented plagiarism, ratifying only plagiarism in the paper retracted in May 2011, P179. In February 2012, this was ruled academic misconduct, albeit with minimal[82] penalties.

Here, **§5.1** and **§5.2** are specifically alleged against GMU past President Alan G. Merten, Provost Peter Stearns, VP Research Roger R. Stough and perhaps the Investigation committee.[83] New GMU President Angel Cabrera was sent email and certified letter, in August/September 2012, **§D. In Fall 2012, Wegman was named to 3-year term on the Promotion and Tenure Committee for College of Science, SPACS.**[84] Whether this was an honor or not, Wegman's actions were totally acceptable at GMU.

In February 2013, AVP Elizabeth Brock and Provost Stearns signed a new policy, **§A.1**. They kept most words from the old one, but GMU now guarantees that it will never "reveal the nature or substance of the evidence or reasoning employed throughout the proceedings." The procedure was already more opaque than those of most schools. Now it is more so.

[79] [MAS2010 §A.11.2**,** slide 23 of [SAI2007]] "Invitations: **Bad Ones**: We were invited by the Dean of the College of Science, and the Vice President of Research at GMU to explain our testimony." *That meeting might be worth investigating.*
[80] www.gpo.gov/fdsys/pkg/FR-2000-12-06/html/00-30852.htm Institutional Notification of the Agency.
[81] $129M total, 85% from Federal sources, but Koch and allies' gave as well.
[82] Wegman had a reprimand placed in his file, had to retract the already-retracted paper and had to apologize to the journal, i.e., his long-time associate Azen.
[83] Physics, the School of Public Policy, and the Provost Office people, unnamed.
[84] cos.gmu.edu/sites/default/files/COS_Committees_Elected_2012_0.pdf www.webcitation.org/6CS6Eh9Rc

6 Conclusion (DoD, DHHS)

Wegman and Said *paid little obvious attention* to Federal rules:

- Reports claimed works that did not actually ack the grants.
- At least 48 unfit works were claimed or contained improper acks.
- Wegman claimed 12-19 works completed before 0447.
  Said claimed 3 works done 9-12 months before 5876.
- Reports were often late, sometimes as much as a year after deadline.
- ARO and NIAAA funds flowed to unfit works, clearly commingled.
  Some activities need investigation of possible funds mis- use.
- Wegman and his lawyer have asserted that neither he nor Said have plagiarized, in the face of consistent evidence to the contrary.
- Whether ack'd or not, many talks and papers simply had zero relevance to Wegman's 0447/0059**,** for which **§B** should be reviewed.
  Similar comments apply to Said and her 5876 award.

All this seems inexplicable for a well-known, experienced researcher, but it is hard not to conclude that something went badly wrong ~2005.

GMU's response to a simple complaint seems absurd. *Other academics may want to offer opinions as GMU behavior could tarnish the "brand" of academe as a whole, although it should not.* No over-generalization should be made about the GMU faculty,[85] but top GMU administrators and some faculty were involved throughout, gave Wegman minimal wrist-slaps and then named him to a tenure and promotion committee.

Why has GMU behaved so far outside academic norms and Federal rules? Possible hypotheses proposed for this behavior include:

- Serious incompetence is always possible, as per Napoleon's dictum.
- T126 was an effort to mislead Congress [MAS2010a §A.10]**.** Perhaps someone thought admission of WR misconduct would open the door to worse. GMU and Wegman knew of earlier felony concerns [MAS2010a §4, #29-30.] Perhaps the follow-on process may some day rise to cases of misprision (18USC§4) or obstruction of justice (18USC**§**1519).
- GMU may be so enmeshed in funding and influence by the Kochs and allies and so involved with thinktanks and politicians pursuing climate anti-science, that it simply could not allow any criticism or retraction of the WR. That is speculation, but see **§A.5** or [MAS2012c §A.5, §A.6.]

*Wegman and Said broke the rules and GMU has badly broken the bargain American universities make with their funders and the rest of academe. Taxpayers may be unhappy funding **any** research at a school whose administration has shown that it could not:*

- handle even a simple misconduct complaint and tell the truth about it
- follow its own policies on intervals, but take ~2 years to assess 9.5 pages, then ignore all plagiarism except that already retracted
- follow its own policy on diligent investigation and pursue complaints
- follow Federal rules requiring notification and reports at steps in the misconduct process, recognized by GMU's own policies
- recognize the legal implication of possible serious funding misuse or grant fraud, for funds of which GMU took 30%
- treat a distinguished academic (Ray Bradley) with normal courtesy, but tell him almost nothing, mislead him on status, break promises, then blame him falsely, with acts perhaps rising to retaliation/defamation.

What now?

It has repeatedly proved useless to file formal complaints with GMU and its latest Policy 4007 makes it even more fruitless. Its opacity *seems designed* to shield GMU from misconduct complaints.

It is time for various agencies to pursue appropriate legal actions against Wegman, Said and GMU, using subpoena powers as needed and perhaps interviewing people mentioned here. Given the large number of clear allegations plus other concerns, more problems may well surface, beyond those found by merely integrating publicly-available information.

Perhaps it is time for an institutional debarment until GMU has shown consistent ability to act within academic norms and Federal rules.[86]

[85] I have met or communicated with credible GMU academics and have heard of many others. Much of GMU seems like a typical university. *The inquiry committee seemed to act within academic norms, but not the investigation committee, although nothing is really public about them but Stearns' letter, which contained other falsehoods. The report seems not to have left GMU.*

[86] I know this would be a hardship, but the many honest academics at GMU might take existing grants elsewhere for their and their students' sake. *GMU has proved itself totally unable or unwilling to manage normal academic misconduct proceeding. After criticism the process was changed to make it more opaque.*

A.1 GMU University Policy 4007, the old and the new

This short excerpt from [MAS2012 §A.1] highlighted Federal reporting rules from GMU's policy, with a few additions. This policy applied through 2012, but a new one was approved 02/22/13 by AVP Elizabeth Brock and Provost Stearns.

The old text[87] includes many "as soon as possible" notes and possibilities of challenges. For something as simple as a few pages of obvious copy-paste-trivial change plagiarism, one would expect most steps to go faster. The rest should be assumed as quoted except for annotations. Bold is mine. The agencies need to be notified at steps ❶ ❷ ❸, comparable to **§B.3**.

> **Notifying Federal agencies as required**[88]
> For proceedings that involve Federal support and research misconduct as defined by the funding agency, **the university meets the reporting requirements of the funding agency relating to the decision whether an investigation is warranted.** ❶ The university may be required to provide the research records and evidence reviewed during the inquiry, transcripts or recordings of any interviews, and copies of all relevant documents, among other materials.
> **The university also meets the reporting requirements of the funding agency pertaining to –**
> (a) Any plans to close a case at the inquiry, investigation, or appeal stage on the basis that the respondent has admitted guilt, a settlement with the respondent has been reached, or for any other reason than a finding that an investigation is not warranted or that no misconduct occurred; and
> (b) **The outcome of the investigation ❷ and any administrative actions against the respondent. ❸ …**'

> **Cooperation with Federal agencies**
> For proceedings that involve Federal support and research misconduct as defined by the funding agency, the university cooperates fully and on a continuing basis with Federal agencies during any oversight reviews of the university and its research misconduct proceedings and during the process under which the respondent may contest the agency's findings of research misconduct and proposed administrative actions. The university cooperates with and assists the appropriate Federal agency, as needed, to carry out any administrative actions it may impose as a result of a final finding of research misconduct by that agency.
> In conducting the investigation, the committee –
> (a) Uses diligent efforts to ensure that the investigation is thorough and sufficiently documented and includes examination of all research records and evidence relevant to reaching a decision on the merits of the allegations;
> (b) Interviews each respondent, complainant, and any other available person who has been reasonably identified as having information regarding any relevant aspects of the investigation, including witnesses identified by the respondent; and
> **(c) Pursues diligently all significant issues and leads discovered that are determined relevant to the investigation, including any evidence of additional instances of possible research misconduct, and continues the investigation to completion.'**[89]

A cursory comparison of old and new versions found some new text:

> 'The University reserves the right to communicate general information regarding the results of any proceedings, where necessary in the judgment of the administration to preserve or restore the reputation of the University, the repondent, *(sic)* or the complainant. **At no time**, however, shall the University reveal detailed personal information regarding the complainant, the respondent or witnesses, **nor shall the University reveal the nature or substance of the evidence or reasoning employed throughout the proceedings**.'

GMU's previous process allowed complainants to be told nothing and it never told Bradley or any other related complainant any result. The new version explicitly **assures** that none would ever be told any substance.

[87] universitypolicy.gmu.edu/policies/misconduct-in-research-and-scholarship www.webcitation.org/6GJ5ylUbz archived 05/01/13.
The procedure is now in a file separate from the policy: universitypolicy.gmu.edu/wp-content/uploads/2013/02/University-Procedure-for-Policy-4007.docx most of the new procedure is similar, but GMU now has: oria.gmu.edu/contact-us AVP Research Integrity and Assurance
They reorganized the website, once universitypolicy.gmu.edu/4007res.html www.webcitation.org/65v5bf0CL Archived for [MAS2012c] www.webcitation.org/6C6Yeg67L A quick scan in 2012 of the quoted text showed no obvious changes since January 2011 when this text was excerpted.

[88] www.gpo.gov/fdsys/pkg/FR-2000-12-06/html/00-30852.htm

[89] *One might think this should include investigation of issues regarding possible misuse of government funds. Such issues were brought to GMU's attention by Bradley May 2010, and various times thereafter. GMU seems to get 30% of contract money for overhead, so it might have some responsibility. If it turns out there was grant fraud that rose to the level of felony, hiding it might be one too.*

A.5 GMU funding, adapted from [MAS2012c]

A.5.1 GMU funding from C.G.Koch and allies

Like many universities, GMU gets most of its research funding from the Federal Government, directly or indirectly, shown on later pages.

GMU appeared often in the tobacco archives, legacy.library.ucsf.edu, but a major funding source is for health research, *an odd juxtaposition*.

GMU has a strong political Libertarian/small-government theme, [90] especially clear in the Mercatus Center, Institute for Humane Studies (IHS), Economics, Law and Economics Center, Public Policy, and the Law School, whose most famous graduate may be VA AG Ken Cuccinelli, well-known for his repeated U VA lawsuits [MAS2012c §A.6].

GMU is strongly connected with thinktanks and others that do climate and environmental anti-science and help tobacco companies. More analysis is needed, but frequent funders of climate anti-science [MAS2010, MAS2012d] *seem very influential.*

DONORS CAPITAL+DONORS TRUST form a conduit for money from some of the same donors [MAS2012d, §I], but anonymizes it.
Charles Koch can give:

- Personally, directly to any of these (not public)
- Through C.G.Koch Foundation (public)
- Through C.L.Lambe Foundation (public)
- Through the Knowledge and Progress Fund (K&P F) [91] (public)
- Through DONORS, whose donations are public, but not the original donors. Money could arrive through any of the previous routes. DONORS TRUST and CAPITAL have transferred funds in both directions.

In addition, donations can:

- Go directly to entities with their own EINs, such as Mercatus and IHS.
- Go to the GMU Foundation, but tagged for Mercatus, IHS.
- Go to GMU Foundation, but tagged for Law and Economics Center [92]or other entities inside GMU that have no EIN, such as Economics.
- Be tagged as general operating expenses or completely untagged.[93]

It is very difficult to track the money through this maze.
On next page, drop in Koch funding of GMU and Mercatus seem synchronous with a large jump in DONORS funding there. *DONORS is designed to anonymize funding, so nothing can be proved, but it hints at possible movement to a less-visible path in last few years*. Koch has given about the same amount of money through K&P F to DONORS, as DONORS has given to GMU, but the actual sources are hidden. *One would guess that GMU knows which actual sources are giving via DONORS.*

On next page are summaries of some key donors' gifts to the GMU Foundation, Mercatus Center and IHS.[94] Many other foundations give as well, especially to the GMU Foundation, so this is a small fraction of overall giving, which is spread over many activities.[95] It is easy to miss donations, so these may be considered lower bounds.

The core foundation group's donations to MC+IHS were 25-33% of total contributions from 990s, easily enough for strong influence, even without Board membership or Chairman role.

---

[90] That's fine. The oddity is the combination of anti-science efforts (tobacco, climate) and small-government policy efforts in some parts of GMU, while other parts seek and get substantial science and large-government funding.

[91] www.eri-nonprofit-salaries.com/index.cfm?FuseAction=NPO.Summary&EIN=541899251
This is funded by Koch, seems to give only to DONORS. Of course, funds need not be given from DONORS until later years, so this is very hard to follow.

[92] LEC is specifically mentioned, given appearances in the Legacy Tobacco Documents Library (UC San Francisco) (LTDL), such as: legacy.library.ucsf.edu/tid/off46a00/pdf, in which GMU sought money from RJ Reynolds for the LEC, noting past support.

[93] Foundations and their major donors talk, and may well have a shared agreement on the uses of funds, whether written in Form 990s or not
When C.G.Koch gives nearly $5M for "General/operating support; Program Development, (2009), is it possible he has any idea or influence on its usage?

[94] As seen in , [MAS2012c §A.6**],** Koch chairs IHS and is on the Board of Mercatus, cofounded by lieutenant Richard Fink, who is also on the IHS Board.

[95] Some donations go to obvious non-advocacy purposes, such as Visual and Performing Arts, Zotero (www.zotero.org/about, *nice work*), etc. However, much of the money is so vaguely described that it is hard to know its real intent.

~$ in 1,000s to DONORS	2005	2006	2007	2008	2009	2010	Total
K&P F (Koch) ➔ DONORS	390	3210	1,125	1,240		2,000	7,965

~$ in 1,000s to GMU	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**Total**
C. G Koch F + C. Lambe F	3,450	350	408	2,873	5,119	3,667	**15,767**
DONORS CAPITAL+TRUST	164	370	1,020	800	495	1,020	3,869
Searle Freedom Trust		145	174	169	443	1150	2,080
S. Scaife F + Carthage F	350	350	200	200	250	250	1,600
Earhart F	299	203	48	192	161	135	1,037
L&H Bradley F	85	95	85	90	90	65	510
TOTALS	**4,348**	**1,513**	**1,934**	**4,324**	**6,457**	**6,287**	**24,863**
TOTAL Contribs (990)	**26,980**	**21,134**	**23,001**	**22,720**	**46,067**		

~$ in 1,000s to Mercatus	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**Total**
C. G Koch F	1,015	3,900	2,683	1,050	600	0	**9,248**
DONORS CAPITAL+TRUST	1	104	276	156	107	1,157	1,800
Searle Freedom Trust		100	150	230	250	250	980
S. Scaife F	80	80	80	80	50	50	420
Earhart F	83		20	45	60	30	238
L&H Bradley	10	10	10				30
TOTALS	**1,189**	**4,194**	**3,218**	**1,561**	**1,067**	**1,487**	**12,716**
Total Contributions (990)	**6,562**	**4,846**	**6,561**	**9,631**	**8,042**	**11,603**	**47,245**

~$ in 1,000s to IHS	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**Total**
C. G Koch F		1,082	886	1,169	2,461	2,160	**7,757**
DONORS CAPITAL+TRUST	14	168	1,088	757	27	364	2,428
Searle Freedom Trust	315	245	73	250	300		1,183
S. Scaife F	50	50	50	0	35	35	220
Earhart F	30		137				167
L&H Bradley F	15	35	40	30	25	5	150
TOTALS	**424**	**1,580**	**2,273**	**2,216**	**2,848**	**2,564**	**11,905**
Total Contributions (990)	**2,935**	**3,049**	**5,764**	**6,059**	**6,324**	**9,163**	**33,295**

GMU+MC+IHS TOTALS	**5,961**	**7,287**	**8,101**	**9,004**	**10,372**	**10,337**	**49,483**
C.G.Koch totals	**4,465**	**5,332**	**3,976**	**5,092**	**8,079**	**5,827**	**32,772**
C.G.Koch %	**75%**	**73%**	**54%**	**63%**	**78%**	**56%**	**66%**
Foundation Sponsored Research[96]			**5,453**	**6,718**	**6,637**		

Even given calendar years vs fiscal years, and different categorizations, donations from just these foundations to GMU, Mercatus and IHS are nearly 2X larger than the total foundation **research** funding for GMU.

[96] irr.gmu.edu/factbooks/1011/Factbook1011_Sponsored.pdf and next page

A.5.2 Sponsored research expenditures

GMU gets most of its research funding from government and a very small fraction of its funding from private industry, *which seems curious, given the strong free-market/private enterprise/small government views for which GMU is known.*

The most recent Factbook gives some detail on sponsored research:[97]

TABLE 8.1 Expenditures by Sponsor Type, FY 2009-2011 (Dollars)

	FY 2009	FY 2010	FY 2011
Federal Sources			
Federal	67,336,712	69,949,715	60,164,604
Federal funds passed through state	6,014,699	5,265,661	6,109,124
Federal funds passed through industry, non-profit and universities	10,097,690	10,211,552	11,641,786
Federal funds passed through local governments/schools	163,562	292,488	271,962
Subtotal	83,612,664	85,719,417	78,187,475
Other Sources			
Associations/Non-Profit	2,373,951	1,826,373	1,296,928
Foreign Entities	3,105,348	2,118,304	1,549,565
Foundations	6,718,537	6,637,356	6,787,948
Industry	1,309,144	759,540	733,132
Local Government/Schools	705,633	776,799	1,301,268
Revenue/Other Governments	405,429	339,892	163,050
State (including SCHEV)	1,490,083	1,619,730	712,700
Universities	443,809	489,165	824,917
Subtotal	***16,551,932***	***14,567,159***	***13,369,507***
Total	$100,164,596	$100,286,575	$91,556,982

Note: Numbers may not add up to subtotals/totals due to rounding.
** Pooled expenditures not included.*

From previous page:[98]

~$ in 1,000s GMU,MC,IHS	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**Total**
GMU+MC+IHS TOTALS	**5,961**	**7,287**	**8,101**	**9,004**	**10,372**	**10,337**	**49,483**
C.G.Koch totals	**4,465**	**5,332**	**3,976**	**5,092**	**8,079**	**5,827**	**32,772**

C.G.Koch and close allies, who jointly fund many thinktanks active in climate anti-science,[99] form a major funding source for GMU, *but very likely not for research, since their numbers are larger than those claimed for foundation research funding.*

[97] irr.gmu.edu/factbooks/1112/Factbook1112_Sponsored.pdf pp.2-6
[98] Foundations use calendar years, GMU's fiscal year seems to be July-June.
[99] Many other familiar foundations appear as donors to GMU, MC or IHS, but this seemed enough to make the point. None of this is a surprise of course, but the numbers seemed worth recording.

A.5.3 Sponsored research expenditures

The 4 largest sources are marked. The funds from Koch and close allies are definitely not negligible on this scale, would rank roughly #3 in 2010.

TABLE 8.2 Expenditures by Federal Agencies, FY 2009 - FY 2011 (Dollars)

	FY 2009	FY 2010	FY 2011
Agency for International Development	25,654	0	0
Department of Agriculture	376,524	307,434	278,057
Department of Commerce	3,521,937	4,353,119	4,173,611
Department of Defense	13,200,642	17,106,725	15,131,050
Department of Education	3,036,451	2,998,257	4,051,162
Department of Energy	2,182,675	2,168,787	2,637,016
Department of Health and Human Services *	20,483,668	18,093,719	7,234,686
Department of Homeland Security	849,247	1,302,938	1,427,609
Department of Interior	630,765	525,796	294,251
Department of Justice	1,266,177	1,405,123	1,265,399
Department of State/USIA	130,979	613,249	1,594,998
Department of Transportation	1,558,786	1,403,629	2,095,467
Environmental Protection Agency	10,032	0	0
FDIC	17,638	30,688	206,154
General Services Administration	758,590	689,396	707,625
NASA	9,116,130	8,234,626	7,025,298
National Endowment of Humanities/Arts	575,085	532,961	592,195
National Science Foundation	5,888,252	6,564,194	7,649,883
Small Business Administration	2,320,792	2,639,130	2,932,939
Other Federal Agencies**	1,386,689	979,944	867,205
Subtotal Direct Federal Funds	***$67,336,712***	***$69,949,715***	***$60,164,604***
Federal Funds Passed Through State	***$6,014,699***	***$5,265,661***	***$6,109,124***
Federal Funds Passed Through			
1) Industry	7,266,862	6,777,265	6,757,489
2) Associations/Non-profits	1,002,570	805,251	1,163,797
3) Local Governments and Schools	163,562	292,488	271,962
4) Universities	1,828,258	2,629,036	3,720,500
Subtotal Federal Funds Passed Through 1,2,3 and 4	**$10,261,252**	**$10,504,040**	**$11,913,748**
Total Federal Agencies Expenditure	**$83,612,664**	**$85,719,417**	**$78,187,475**

** Includes construction grant funds from Department of Health and Human Services (DHHS) for Regional Biocontainment Lab (RBL): FY09 $9,740,200, FY10 $11,841,683, FY11 $416,817.*

*** FY09: U.S. Institute for Peace, Library of Congress, National Libraries, Securities and Exchange Commission, IMLS, NARA, Veterans Affairs, Congress, and National Gallery of Art; FY10: Smithsonian, Securities and Exchange Commission, Central Intelligence Agency, USPS, USIP, Federal Trade Commission, Library of Congress, NHPRC, IMLS; FY11: Smithsonian, Securities and Exchange Commission, USPS, USIP, Federal Trade Commission, Library of Congress, NHPRC, IMLS, NARA, NEA.*

~$ in 1,000s GMU,MC,IHS	2005	2006	2007	2008	2009	2010	Total
GMU+MC+IHS TOTALS	5,961	7,287	8,101	9,004	10,372	10,337	49,483
C.G.Koch totals	4,465	5,332	3,976	5,092	8,079	5,827	32,772

A.5.4 Expenditures by Mason Organizational Unit

Again, figures from the Factbook are:

TABLE 8.4 Expenditures by Mason Organizational Unit, FY 2009 - FY 2011 (Dollars)

Academic Units	FY 2009	FY 2010	FY 2011
COLLEGE OF EDUCATION AND HUMAN DEVELOPMENT (CEHD)			
Center for the Advancement of Public Health	493,195	0	0
Center for Restructuring Education	255,165	0	0
College of Education and Human Development	2,369,648	7,763,278	9,003,549
Kellar Institute for Human Disabilities	5,598,170	0	0
School of Recreation, Health and Tourism	77,515	502,470	553,053
CEHD Subtotal	***$8,793,694***	***$8,265,748***	***$9,556,602***
COLLEGE OF HEALTH AND HUMAN SERVICES (CHHS)			
Center for Chronic Illness and Disease	280,237	226,530	371,034
Center for Discovery Science and Health Informatics	0	0	0
Center for Health Policy Research and Ethics	0	6,629	359,933
Department of Global and Community Health	0	13,285	64,518
Department of Health Administration and Policy	838,668	698,904	517,468
Office of the Dean	230,973	15,240	110,619
School of Nursing	270,863	369,282	607,543
Social Work	45,833	50,595	67,438
CHHS Subtotal	***$1,666,575***	***$1,380,465***	***$2,098,553***
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES (CHSS)			
Communication	378,090	572,734	781,977
Center for Biodefense[4]	16,617,675	13,424,183	1,847,349
Center for Social Science Research	174,755	205,845	220,837
Criminology, Law and Society Program (formerly Admin of Justice	4,779,416	4,774,700	3,458,390
Economics	2,205,658	1,687,112	1,498,965
English (Association of Writers and Writing Programs)	1,239,885	1,211,731	1,144,221
History and Art History	3,751,838	3,223,180	2,871,415
Modern and Classical Languages	41,661	54,247	646
New Century College	270,056	597,562	633,616
Office of the Dean	23,719	341,022	131,630
Philosophy	0	0	52,997
Psychology	4,101,717	4,498,074	5,314,331
Public and International Affairs	678,817	696,997	666,214
Religious Studies	14,775	41,283	67,960
Sociology/Anthropology	59,230	41,941	40,017
Women's Studies	0	39,040	0
CHSS Subtotal	***$17,719,616***	***$17,985,467***	***$16,883,214***
COLLEGE OF SCIENCE (COS)			
Atmospheric Oceanic and Earth Sciences	1,286,391	1,568,490	1,263,384
Bioinformatics and Computational Biology[1]	733,973	824,761	*
Center for Applied Proteomics	1,749,348	1,886,120	1,815,428
Center for Clean Water and Sustainable Technologies	0	39,659	0
Center for Computationals Fluid Dynamics[2]	1,606,007	1,207,827	*
Center for Computational Material Science[2]	331,665	325,643	*

Note, for fields containing asterisks, please refer to the following numbers accompanying Academic Units *(Continues on next page)*

[1] See SSB

[2] See SPACS

[4] Center for Biodefense includes construction grant funds from Department of Health and Human Services (DHHS) for Regional Biocontainment Lab (RBL): FY09 $9,740,200, FY10 $11,841,683, FY11 $416,817.

TABLE 8.4 Expenditures by Mason Organizational Unit, FY 2009 - FY 2011 (Dollars) (Continued)

Academic Units	FY 2009	FY 2010	FY 2011
COLLEGE OF SCIENCE (COS) *(Continued)*			
Center for Earth Observing and Space Research	6,063,883	4,791,374	3,891,814
Center for Intelligent Spatial Computation	18,149	55,007	0
Center for Quantum Studies[2]	97,742	391,941	*
Center for Spatial Information Science and Systems	1,586,230	1,448,228	1,688,683
Chemistry and Biochemistry	218,906	84,592	61,378
Computational and Data Sciences[2]	1,123,705	1,457,020	*
Computational Statistics Center	0	33,651	0
Environmental Sciences and Public Policy	440,288	425,538	1,384,898
Geography and Geoinformation Science	825,430	912,752	2,027,967
Life Sciences	284,767	298,533	1,046,866
Mathematical Sciences	487,795	522,044	922,388
Microbiomic Analysis Center[1]	0	407,256	*
Molecular and Microbiology[1]	285,431	457,760	*
Office of the Dean	722,276	446,051	158,770
Physics and Astronomy[2]	1,919,458	2,429,286	*
School of Systems Biology (SSB)	0	0	1,294,368
School of Physics/Astronomy and Computational Sciences (SPACS)	0	0	6,666,356
COS Subtotal	***$36,399,118***	***$33,437,716***	***$24,069,649***
COLLEGE OF VISUAL AND PERFORMING ARTS (CVPA)			
Office of the Dean	0	108,924	62,650
Music	25,586	N/A	0
CVPA Subtotal	***$25,586***	***$108,924***	***$62,650***
SCHOOL FOR CONFLICT ANALYSIS AND RESOLUTION (SCAR)			
SCAR Subtotal	***$535,838***	***$419,929***	***$1,108,949***
THE VOLGENAU SCHOOL OF ENGINEERING (VSE)			
Air Transportation Systems Research Center	1,673,746	1,559,201	1,467,407
Applied Information Technology	986,428	750,685	454,023
C4I Center	3,776,841	5,058,644	3,154,493
Civil, Environmental, and Infrastructure Engineering	67,461	177,807	655,402
Computational Statistics Center	34,536	0	0
Computer Science	1,328,008	1,300,466	1,860,727
E-Center for E-Business	93,484	246,276	55,822
Electrical and Computer Engineering	3,023,329	3,374,350	4,388,551
Intelligence and Security Research Center	0	33,690	0
Learning Agents Center	315,268	581,854	659,204
Office of the Dean	525,183	402,343	93,841
Secure Information Systems Center	3,112,101	3,244,450	4,012,637
Statistics	228,993	179,978	248,416
Systems Engineering and Operations Research	1,147,310	1,315,837	1,495,644
VSE Subtotal	***$16,312,689***	***$18,225,583***	***$18,546,168***

Note, for fields containing asterisks, please refer to the following numbers accompanying Academic Units

[1] See SSB

[2] See SPACS

[+] Center for Biodefense includes construction grant funds from Department of Health and Human Services (DHHS) for Regional Biocontainment Lab (RBL): FY09 $9,740,200, FY10 $11,841,683, FY11 $416,817.

(Continues on next page)

TABLE 8.4 Expenditures by Mason Organizational Unit, FY 2009 - FY 2011 (Dollars) (Continued)

Academic Units	FY 2009	FY 2010	FY 2011
SCHOOL OF LAW (LAW)			
LAW Subtotal	***$3,664,601***	***$4,637,048***	***$3,534,383***
SCHOOL OF MANAGEMENT (SOM)			
SOM Subtotal	***$346,607***	***$482,158***	***$382,706***
SCHOOL OF PUBLIC POLICY (SPP)			
SPP Subtotal	***$9,623,978***	***$7,966,811***	***$8,152,498***
KRASNOW INSTITUTE (KRAS)			
KRAS Subtotal	***$3,154,946***	***$4,990,955***	***$5,137,340***
Administrative Units			
ACADEMIC AFFAIRS			
Academic Affairs	18,865	39,296	58,535
Office of the Provost	287,051	1,388,019	914,290
Professional Training and Development	1,018,847	407,694	359,617
Robinson Professors	15,390	0	0
Academic Affairs Subtotal	***$1,340,153***	***$1,835,008***	***$1,332,441***
ADMINISTRATION			
Administration/Facilities	0	808	0
Environmental Safety and Risk Management	10,424	1,370	132,223
Sustainability	0	0	11,785
University Police	73,703	0	8,923
Administration Subtotal	***$84,127***	***$2,178***	***$152,930***
INFORMATION TECHNOLOGY			
Library	318,130	312,666	315,150
Information Technology Subtotal	***318,130***	***312,666***	***315,150***
UNIVERSITY LIFE			
Early Identification Program	150,783	210,152	184,548
Student Services/Sexual Assault Services	28,155	25,566	39,201
University Life	0	200	0
University Life Subtotal	***$178,938***	***$235,918***	***$223,749***
University Total	**$100,164,596**	**$100,286,575**	**$91,556,982**

Note, for fields containing asterisks, please refer to the following numbers accompanying Academic Units

[1] See SSB

[2] See SPACS

[+] Center for Biodefense includes construction grant funds from Department of Health and Human Services (DHHS) for Regional Biocontainment Lab (RBL): FY09 $9,740,200, FY10 $11,841,683, FY11 $416,817.

According to a May 2012 letter, GMU got $129M in research awards in 2011, 85% in Federal grants.[100] This was an 18.5% increase over 2010, somewhat ironic in light of the Kochs' long drive to downsize Federal government.

[100] http://masonresearch.gmu.edu/2012/05/letter-from-roger-r-stough/

B. Grant rules, grant fraud and contracts

Almost any research can be made unfairly to sound silly by the Proxmire "Golden Fleece" approach. Real research often finds surprises or fails to achieve the proposed objectives. ***However, Federally-funded researchers are supposed to focus on work in their proposals, not spend most of their time on clearly-unrelated works for which they claim credit.***[101] Some sources about Federal-funded research are collected here as examples of policies more-or-less typical of Federal funders. A good overview is "University Research: The role of Federal Funding."[102] The "grant" and "contract" terminology can be confusing. Grants come with contracts, usually signed by institutions, but some contracts have no associated grants. For simplicity, 0447, 0059, 5876 are all called "grants" here, although all had contract numbers. No contract number was found for the Commonwealth of VA grant via GMU to Said. Grant policies of GMU and Ohio State U are given as examples.[103] Universities bear major responsibilities for both research and financial compliance.[104]

B.1 ARO

A recent Terms and Conditions document is available.[105]
That document notes:

> '**24. Publication and Acknowledgment.** Publication, acknowledgement and disclosure of federal funding under this grant shall be as described in the DODGARs §32.36.
> **25. Technical Reporting Requirements**. For detailed technical reporting requirements and instructions, see ARO Form 18, "Reporting Instructions," found on ARO's website at http ://www.aro.army.mil …
> **27. After-the-Award Requirements.** Closeout, subsequent adjustments, continuing responsibilities, and collection of amounts due are subject to requirements found in DODGARs §32.71through 73.'

[101] People, especially senior researchers, may well produce unrelated works, possibly distractions, but possibly quite reasonable. Claiming credit for much totally-unrelated work is troublesome. Taxpayers deserve to have funds spent on the research proposed or at least closely related, not on efforts far removed.
[102] www.aau.edu/WorkArea/DownloadAsset.aspx?id=11588 They distinguish between "grant" and "contract." Common usage sometimes uses either one.
[103] oria.gmu.edu; controller.osu.edu/acc/fedgrant-home.shtm
[104] www.gpo.gov/fdsys/pkg/FR-2000-12-06/html/00-30852.htm
[105] www.arl.army.mil/www/pages/386/termcond05.pdf
I do not know how much that has changed since 2004.

The most recent (May 2011) ARO Reporting Instructions[106] are instructive:
p.4 'A Final Progress Report (FPR) is due within 90 days following the expiration of agreement.'[107]

p.5 describes some of the required documentation, which seems to match the categories in the FPR for 0447, but which 0059 failed to provide, **§R.4.**

> 'b. "The Report Documentation Page (SF298) Continuation Sheet (Enclosure 2)" or a plain piece of paper **must include brief, but complete, information for each of the following categories:**
> (1) Submissions or publications under ARO sponsorship **during this reporting period**. List the title of each and give the total number for each of the following categories:
> **(a) Papers published in peer-reviewed journals**[108]
> (b) Papers published in non-peer-reviewed journals
> (c) Presentations
> i. Presentations at meetings, but not published in Conference Proceedings
> ii. Non-Peer-Reviewed Conference Proceeding publications (other than abstracts)
> iii. Peer-Reviewed Conference Proceeding publications (other than abstracts)
> (d) Manuscripts
> (e) Books
> (f) Honor and Awards
> (g) Title of Patents Disclosed during the reporting period
> (h) Patents Awarded during the reporting period'

It can be nontrivial for anyone not an expert in a field to assess the credibility of work claimed. *What can be done?*

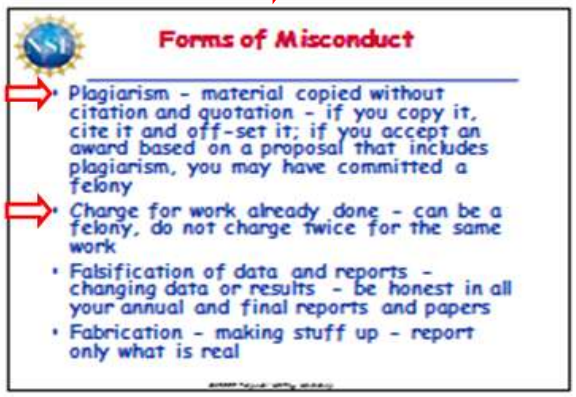
- Peer-reviewed papers in credible research journals (a) really count, *as the ARO obviously understands from the rule above*.
- Citation analysis can help assess the post-publication impact, but that does not help the ARO much, as it takes years.
- One can find a few experts and ask them, typically done when proposals for new funding are being considered.

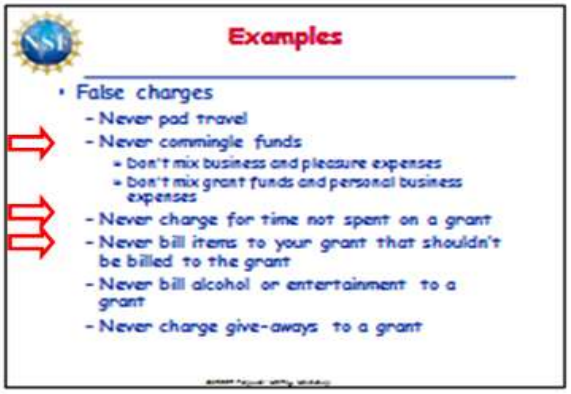
[106] www.arl.army.mil/www/pages/218/form18.pdf, May 2011.
[107] If that rule was in force earlier, then at least 6 grants violated it, ranging from 109 days (not too bad) to 449 days, **§K.1.**
[108] Peer-review in credible journals does not prove correctness, *but inability to publish there should be a concern.* Most of the peer-reviewed journal papers 0447-claimed actually were done (or likely were done) before 0447 started. The work of 2005-2009 generated almost nothing in peer-reviewed research journals.

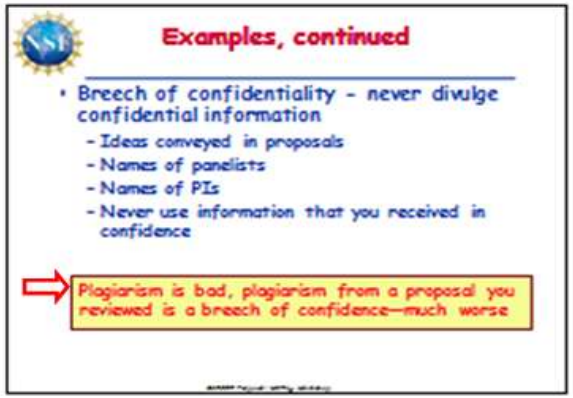
B.2 NSF CAREER Proposal Writing Workshop 2009 - GMU[109]

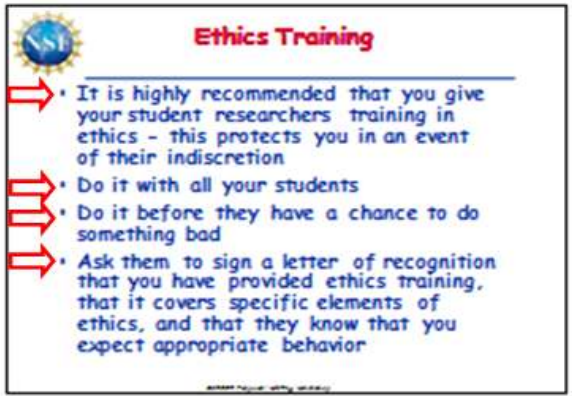
Although no NSF grants were involved here, NSF rules seem typical and are especially well-documented in accessible forms. NSF and other agencies take plagiarism seriously in grant proposals and grant-ack'd and claimed work.[110] Wegman had received various NSF grants in years before 2001, **§K.1**.
This and next page show some helpful slides from an NSF workshop at GMU, which seems well-organized to help researchers get grants.

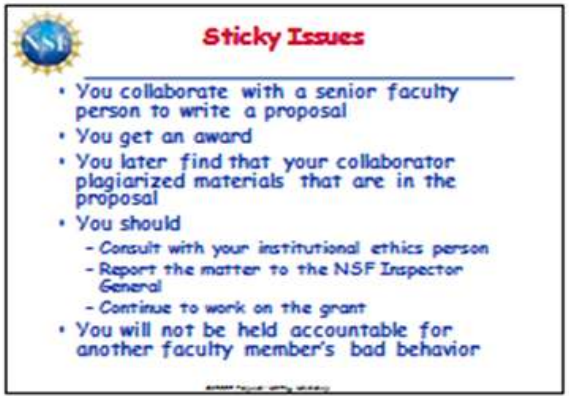
CAREER Writing Workshop: pp. 65, 66, 68; 69, 71, 72; annotated with ⇨







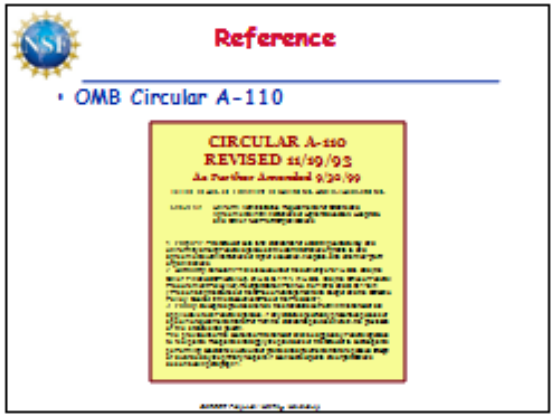


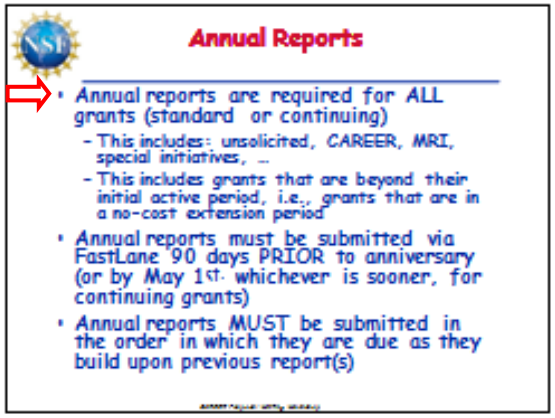


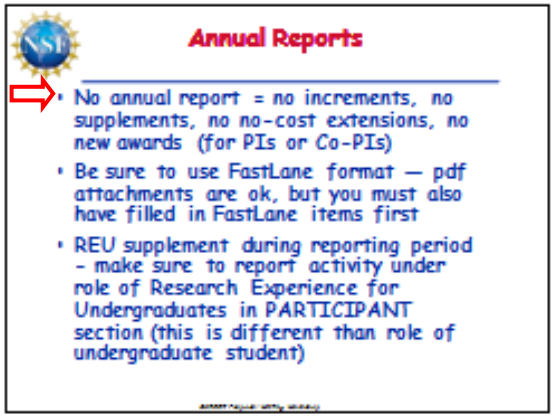
[109] 2009 NSF CAREER Proposal Writing Workshop CAREER Program Development, March 2009, copy found at
grants.soe.ucsc.edu/sites/default/files/2%20**George**.ppt
carlbatt.com/Carl_Batt/Welcome.html
www.k-state.edu/career/2009/09documnets/General%20information.pdf
'This web site is established for access by George Mason faculty and staff. It has been specifically developed for NSF CAREER workshop program.'
research.gmu.edu/ResearchDev/research_development_resources.html

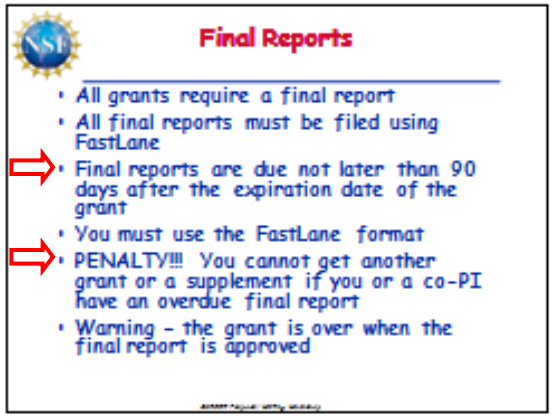
[110] www.nsf.gov/oig/index.jsp www.nsf.gov/pubs/2012/oig12002/oig12002.pdf for example. Search: fraud.

CAREER Writing Workshop continued: pp. 73, 74, 77; 79, 80, 81; annotated





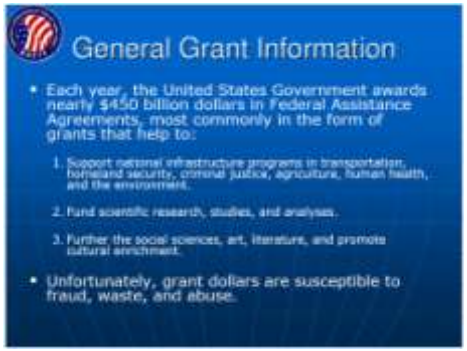




The OMB Circular A-110 is: www.whitehouse.gov/omb/circulars_a110
See also discussion by Scott Moore and John Dahlberg, new.livestream.com/coloradostateu/events/2088755, at this 05/08/13 workshop: ricro.colostate.edu/Plagiarism2013/PlagiarismEvent.html Moore gives example of P.I. copying from student dissertations.

B.3 Grant fraud, US[111] Notification

pp. 1-4



2:

'Each year, the United States Government awards nearly $450 billion dollars in Federal Assistance Agreements, most commonly in the form of grants that help to: …
2. Fund scientific research, studies, and analyses.

3:

'1. Grant recipients are stewards of federal funds.
2. Grant dollars must be used for their intended purpose.
3. Where applicable, grant recipients must account for costs and justify expenditures.
Using federal grant dollars for unjust enrichment, personal gain, or other than their intended use is a form of theft, subject to criminal and civil prosecution under the laws of the United States'

4:

'**Federal grant dollars are susceptible to several forms of financial theft, most commonly in the form of specific federal violations, including: …**
3. False Statements[112]
4. False Claims[113]
5. Mail Fraud and Wire Fraud
Each of these violations of law are subject to criminal prosecution, fines, restitution, and civil penalties.

[111] www.grants.gov/assets/GrantFraud.pdf
www.ignet.gov

[112] www.law.cornell.edu/uscode/text/18/1001 18USC§1001, discussed along with 18USC§371, 18USC§4, and 18USC§1519 in [MAS2010].
See also codes.lp.findlaw.com/uscode/18/I/73/1505 18USC§1505: "Obstruction of proceedings before departments, agencies, and committees" and codes.lp.findlaw.com/uscode/18/I/73/1519 18USC§1519, "Destruction, alteration, or falsification of records in Federal investigations and bankruptcy"

[113] en.wikipedia.org/wiki/False_Claims_Act
www.falseclaimsact.com/common_frauds_research.php
www.law.cornell.edu/uscode/text/31/3729 31USC§3729-3733

pp. 5-8

6:

6: 'Grant fraud occurs in many ways, but some of the most common fraud scenarios include:
1. Charging personal expenses as business expenses against the grant.
2. Charging for costs which have not been incurred or are not attributable to the grant. [114]

7:

'**As a grant recipient**,[115] you can protect your organization and the source of your federal funds by detecting and preventing fraud:
1. Establish an adequate and effective system of accounting, internal controls, records control, and records retention.
2. Implement an internal compliance and ethics program that encourages the recognition and reporting of fraud, waste, or abuse. …'

Notification

In 2000, the US Office of Science and Technology Policy set ground rules for handling academic misconduct.[116] Specifically relevant to §5 is this subsection:

'**Institutional Notification of the Agency. Research** institutions will notify the funding agency (or agencies in some cases) of an allegation of research misconduct if (1) the allegation involves Federally funded research (or an application for Federal funding) and meets the Federal definition of research misconduct given above, and (2) **if the institution's inquiry into the allegation determines there is sufficient evidence to proceed to an investigation.❶ When an investigation is complete**, ❷ the research institution will forward to the agency a copy of the evidentiary record, the investigative report, recommendations made to the institution's adjudicating official, and the subject's written response to the recommendations (if any). **When a research institution completes the adjudication phase**,❸ it will forward the adjudicating official's decision and notify the agency of any corrective actions taken or planned.'

The agency or agencies must be informed at stages ❶❷❸. All universities that get Federal funds are required to follow this and misconduct policies generally incorporate it explicitly or indirectly, as at GMU, **§A.1**. *It is almost unthinkable for large research school to be unaware of this policy or be willing to ignore it.*

[114] Generally, FOIA replies do not provide the specific charges, as for travel, but the Federal agencies get them. Both DoD and DHHS have Inspector Generals:
www.dodig.mil/hotline/hotlinecomplaint.html
https://oig.hhs.gov/fraud/report-fraud/index.asp
https://oig.hhs.gov/hotline/report-fraud-form.aspx is Internet complaint page
Inspector Generals can seek subpoenas: II. B. 'Administrative Subpoena Authority Held By Inspectors General of the Various Agencies' in:
www.justice.gov/archive/olp/rpt_to_congress.htm

[115] Federal grants to a university researcher are generally handled by contract through the university, whose indirect charges vary among schools and specific activities, perhaps from roughly 30% (for off-campus) to 60%.
Presumably, this allows cost recovery and includes some monitoring of grantees.

[116] www.gpo.gov/fdsys/pkg/FR-2000-12-06/html/00-30852.htm
ori.hhs.gov/federal-research-misconduct-policy copy at ORI

B.4 Grant fraud examples, a few with jail time

None of these may be an exact analog of this case, but can be summarized:

- Grant fraud clearly **can** rise to felony, but usually *seems* not to, depending on the circumstances.
- It may help (but not always) to admit guilt, clearly and early.
- Universities have responsibility and may have to pay financial penalties if they do not handle and monitor funds well.

11/30/12 'Former Penn State Professor Charged in $3 Million Federal Research **Grant Fraud'**[117]

'The United States Attorney's Office for the Middle District of Pennsylvania announced that a **felony Information** has been filed in United States District Court in Harrisburg against Craig Grimes, age 55, of Raleigh, North Carolina, charging him with wire fraud, false statements, and money laundering.'

'Ex-Penn State professor sentenced on research grant fraud counts'[118]

A former Penn State engineering professor will spend more than three years in prison for defrauding federal agencies of $3 million in research grant money. Craig Grimes, 56, formerly of Boalsburg and now of Raleigh, N.C., was sentenced in Harrisburg on Friday by federal Judge Yvette Kane. Grimes was ordered by pay $640,660.37 in restitution on top of the sentence of 41 months behind bars.

Grimes pleaded guilty in February to **defrauding** the National Institutes of Health, making **false statements** to the U.S. Department of Energy and laundering money he received from the NIH.

Prosecutors said Grimes got a $1.2 million grant from the NIH for medical research through his company. Between June 2006 and February 2011, Grimes was to study gases in patients' blood that could help detect the presence of a disease in infants.

Prosecutors allege Grimes promised to use about $509,000 for research at Penn State's Hershey Medical Center, but that never happened. Instead, prosecutors said Grimes used the money on himself.'

09/17/12 'Cornell University Guilty of Misusing HIV Grant Funds'[119]

p.3 'Professor Settles Civil complaint … a **false claims** case …'

06/01/12 'When Good Money Goes Bad," Examples from the EPA'[120]

p.3 'Professor Settles Civil complaint … a **false claims** case …'

p.4 'University Agrees to Pay $2.5 Million to Settle **False Claims Allegations.**[121] … In addition, the university must certify that it has in place an adequate compliance program for preventing fraud and false billings to federal grants. This investigation was conducted jointly with the Defense Criminal Investigative Service, the Defense Contract Audit Agency, and the U. S. Army Criminal Investigation Command.'[122]

12/05/11 'Number of research grant fraud cases on the rise, as government works to stop trend' [123]

'Reports of fraud in federal research awards have risen in recent years, and universities and the federal government are acting to halt the trend. …

For example, between October 2010 and March 2011, the NSF saw four convictions and guilty pleas for criminal cases involving research funds, according to the agency's semiannual report to Congress. …

Among various oversight offices, policies and educational programs, UCLA has a large infrastructure in place to ensure research funds are used ethically, said Marcia Smith, associate vice chancellor of the UCLA Office of Research Administration, whose office employs about 160 staff members.'

06/28/11 Luk Van Parijs, prompt confession avoids jail

'Former MIT professor **spared jail for grant fraud**'[124]

'A **prompt confession**, and letters from senior scientists begging for clemency, helped a former MIT professor Luk Van Parijs avoid jail for grant fraud. …

See also earlier articles on this case:

Former MIT biologist penalized for falsifying data *Nature* (February 2009) [125]

Data irregularities in the work of MIT professor *New Scientist* (Oct …2005)'[126]

[117] www.justice.gov/usao/pam/news/2012/Grimes_01_31_2012.htm

[118] www.centredaily.com/2012/11/30/3420398/ex-penn-state-professor-sentenced.html

[119] www.aidsresponseeffort.org/2012/09/17/cornell-university-guilty-of-misusing-hiv-grant-funds

[120] www.epa.gov/oig/reports/ARRA/EPA_OIG_Grant_Fraud_Brochure.pdf

[121] The university was directly involved.

[122] Agencies cooperate.

[123] dailybruin.com/2011/12/05/number_of_research_grant_fraud_cases_on_the_rise_as_government_works_to_stop_trend/

[124] eugenier.wordpress.com/2011/06/28/former-mit-professor-spared-jail-for-grant-fraud

[125] www.nature.com/news/2009/090203/full/news.2009.74.html

[126] www.newscientist.com/article/dn8230

11/15/2006 NACUA Fall Workshop, November 15-17, 2006. 'LEGAL ISSUES IN HIGHER EDUCATION SPONSORED RESEARCH, COMPLIANCE AND TECHNOLOGY TRANSFER, 07B. Enforcement of Federal Grant Accounting: The Legal Perspective'[127] This has many examples and seems a good summary of the issues.

08/31/05 'United States: Research Misconduct: New Enforcement Actions and Developments'[128]

> 'ii. Mayo Foundation Pays $6.5 Million To Settle **Grant Fraud** Charges … Mayo Clinic, paid the United States $6.5 million to "resolve allegations that it charged the government under federal grants for research costs unrelated to the research projects sponsored by those grants." …
> iii. Cornell University's Weill Medical College Settles **Fraud Charges** For $4.3 Million …
> These developments concern three distinct but related areas of research misconduct – Butkovitz was accused of outright fabrication, the Mayo Clinic and Cornell University **settled cases alleging improper accounting of research funds**, and the Nature study inquired about **areas of ethical doubt that may or may not reach the level of a federal crime but are nevertheless alarming** given their prevalence. Occurring within weeks of each other and within months of the University of Alabama settling a research fraud case, the University of Vermont settling a research fraud case, and the Department of Health and Human Services publishing new research misconduct rules (all discussed in our earlier memorandum), the emerging trend is clear. The government, whistleblowers, and the media alike are taking a new and robust interest in research fraud. Rules will be interpreted broadly, and investigations will be conducted aggressively.'

12/10/94 'Uconn professor, associate guilty of science **grant fraud**' [129]

> 'Instead of using the house as a lab, the defendants used graduate and undergraduate students and Connecticut labs to conduct the research, … guilty of two counts of false statements, one count of filing a false claim.'

06/10/92 'UCLA Professor Gets **2-Year Prison Term** in **Research Grant Fraud**'[130]

> 'In what federal prosecutors said was the stiffest punishment ever for **research grant fraud** nationwide, UCLA electrical engineering professor Cavour W. Yeh was sentenced Tuesday to two years in prison and ordered to pay $1.75 million in fines and restitution. Yeh, a UCLA professor since 1967, also resigned his tenured post.
> Two of Yeh's sisters and a brother, all of whom were secretly put on the UCLA payroll by the professor, were each sentenced to three years' probation. They had pleaded guilty to covering up the professor's misconduct.'

[127] Bob Kenney, Director, Federal Research Practice, Hogan & Hartson LLP Washington, DC www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=19&ved=0CGoQFjAIOAo&url=http%3A%2F%2Fwww.higheredcompliance.org%2Fwp-content%2Fuploads%2F2012%2F02%2Fxxv-06-11-9.doc&ei=pqVDUcWaD-ifyQHV24G4Dg&usg=AFQjCNGgl6lldalZBbNdJToo8dOrlXaemA&bvm=bv.43828540,d.aWc

[128] www.mondaq.com/unitedstates/x/34644/Government+Statutory+Law/Research+Misconduct+New+Enforcement+Actions+and+Developments

[129] news.google.com/newspapers?nid=1915&dat=19941210&id=2wIhAAAAIBAJ&sjid=b3QFAAAAIBAJ&pg=2805,2346812

[130] articles.latimes.com/1992-06-10/local/me-282_1_research-grant

C. FOIAs

C.1 ARO results for 0447, 0059, 0267, 0314

I had sent FOIA requests that were consolidated at ARO, such as:

> '**I request access to and copies of information and documentary materials, including electronic mail and other communication, made by Dr. Edward J. Wegman or George Mason University (GMU) in connection with / or related to the following grant: W911NF-04-1-0447**
> He wrote two papers that included the following:
> 'The work of Dr. Wegman is supported in part by the Army Research Office under contract W911NF-04-1-0447.'
> Both papers had obvious plagiarism, one has already been forced to be retracted by the publisher Elsevier.
> I am trying to find and understand:
> A) The research contracts and any reports to ARO that involve the papers, as there may be misuse of funds, besides the plagiarism problems.
> For context, see the second following file, which shows the problems, i.e. anything highlighted in cyan is copy-pasted.
> www.desmogblog.com/see-no-evil-speak-little-truth-break-rules-blame-others introduction
> www.desmogblog.com/sites/beta.desmogblog.com/files/see.no_.evil_.speak_.little.truth_.pdf
> This details numerous examples of alleged plagiarism involving Wegman and Said, but the two relevant to ARO are as follows (and are included for convenience):
> p.19: Usually labeled [SAI2008], Elsevier forced retraction of this paper May 2011. Yasmin H. Said, Edward J.Wegman,Walid K. Sharabati and John T. Rigsby; Social networks
> of author–coauthor relationships; Computational Statistics and Data Analysis; 52(2008) 2177 – 2184.
> p.29-30: labeled [SAI2010] GMU has ignored this, having known of it since June 2011. Yasmin H. Said, Edward J. Wegman, and Walid K. Sharabati, "Author–Coauthor Social Network and Emerging Scientific Subfields," F. Palumbo et al. (eds.), Data Analysis and Classification, Studies in Classification, Data Analysis, and Knowledge Organization, DOI 10.1007/978-3-642-03739-9_30, ©Springer-Verlag Berlin Heidelberg
> 2010, pp.257-268.
> B) GMU has stated that it is not investigating anything else. Since GMU has had an inquiry and investigation of the first and agreed with the retraction, I trying to find out if they followed Federal law and reported this appropriately to ARO. As shown in the report above, GMU's misconduct process was extremely unresponsive.'

ARO supplied the Proposals, Awards and Final Technical Reports described in **§M, §N, §O, $P, $Q**, and **$R,** plus the following, excerpted from the attached original for 0447 and 0267:[131]

> 'Reference your requests for information on the two agreements listed above. For agreements #W911NF-07-01-0447 and #DAAH04-94-G-0267, clearly releasable information includes the final technical reports, the basic contracts, and the original proposals, which are attached.
>
> For agreement #WW911 NF -07-01-044 7, not released were cost and budgetary documents which we believe are exempt under FOIA Exemption 4. Electronic mail between Dr. Wegman, the Principal Investigator, or George Mason University, was not found for either agreement. Regular mail and other documentation in our files address only the submitting of the proposals in accordance with our operating procedures.
>
> **There is nothing in the files of either agreement pertaining to George Mason University reporting any accounts of plagiarism to ARO.'**

They also covered 0559:[132]

> Reference your FOIA request on the above grant agreement (W911NF-07-01-0059) between George Mason University and the Army Research Office, received at ARO on August 27, 2012.
> We apologize for the response delay which occurred during the end of the Government fiscal year. Attached are clearly releasable documents. They include a copy of the award and supporting documentation, a copy of the proposal and the final technical report from DTIC. Not released are budget and cost documents which our office believes are exempt under FOIA Exemption 4. **A search for related emails did not turn up anything other than what is attached.** We believe that is most likely due to a system migration a number of years ago and the length of time involved between the grant effort and your request.'

Since GMU did not investigate the plagiarism alleged against a 0267-funded paper**,** there was no communication about that. But GMU should have sent 3 separate reports to ARO about P179. Anything on 0059 would have been sent from 2010 onward, after the system migration. **In any case, GMU effectively confirmed the above, on next page.**

---

[131] www.desmogblog.com/sites/beta.desmogblog.com/files/ARO.Response.0447.0267_0.pdf ARO FOIA response

[132] www.desmogblog.com/sites/beta.desmogblog.com/files/ARO.Response.0059.pdf ARO FOIA response

C.2 GMU Results

Dan Vergano made a FOIA request to GMU and got back several replies from Phil Hunt, excerpted from the cited originals:

FOIA reply 10/18/12[133]

'Subject: Re: USA TODAY: FOIA request 10/12/2012

Hi Dan -

In a preliminary search, I do not see that we have anything that would be responsive to this request but I will confirm that shortly.

I hope all is well.

Phil

On 10/12/2012 5:52 PM, Vergano, Dan wrote:
> Phil,
>
> Thanks again for the previous FOIA responses. A number of observers have asked me a few questions about the February interview with the provost on the Prof. Wegman matter, and I'm afraid I would like to make another request for information related to the case. In the interview (link below), the provost said that the university was alerting funding bodies to the results of the GMU investigation.
>
> Questions are as follows. Please consider this request limited to records from 2010 and afterwards.
>
> 1) **I would like to find the dates and copies of any reports made to the US Army regarding the inquiry, investigation and final resolution related to academic misconduct by Edward Wegman** for the article in Computational Statistics and Data Analysis, for which funding was acknowledged for:
> W911NF-04-1-0447 (Wegman)
> W911NF-07-1-0059 (Wegman, Yasmin Said)
>
> 2) Likewise, I would like to request the dates and copies of any reports to the Army for plagiarism related to the same two contracts, but for Yasmin H. Said, Edward J. Wegman, and Walid K. Sharabati, "Author-Coauthor Social Network and Emerging Scientific Subfields," F. Palumbo et al. (eds.), Data Analysis and Classification, Studies in Classification, Data Analysis, and Knowledge Organization, DOI 10.1007/978-3-642-03739-9_30, Springer-Verlag Berlin Heidelberg 2010, pp.257-268.
>
> 3) With regard to Army Research Office DAAH04-94-G-0267, I ask for the dates and copies of any reports to them regarding plagiarism in this report from 1996:
> www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=A
> DA313999
>
> 4) **I make the same requests on items 1) and 2), but with regard to reports sent to the Office of Research Integrity, for:**
> F32AA015876 from the National Institute on Alcohol Abuse and Alcoholism, for Yasmin Said.
>
> Thanks again for your time and attention to this matter.
>
> Dan Vergano
> USA TODAY
> 703 854 3400
>
> Link:
> content.usatoday.com/communities/sciencefair/post/2012/02/georg
> e-mason-university-reprimands-edward-'

FOIA reply 10/26/12[134]

'Dan,
Per my last email, we do not have any documentation that would be responsive to your resent *(sic)* request.'

Thus, as of October 2012, neither ARO nor GMU had records indicating ARO had been informed in a timely fashion or at all.

GMU had no responsive records that they had ever informed ORI.[135]

[133] www.documentcloud.org/documents/524432-re-usa-today-foia-request-10122012.html

[134] www.documentcloud.org/documents/524433-foia-request.html

[135] ORI (properly) does not discuss progress of any open cases, so I did not ask.

D. Attempted communication to President Cabrera

Sent 08/26/12 10PM via GMU People Finder form[136]

To
Cabrera, Ángel
Office of the President, President
Why can't I see this email address?

Your Name
John Mashey

Your Email
johnmashey@yahoo.com

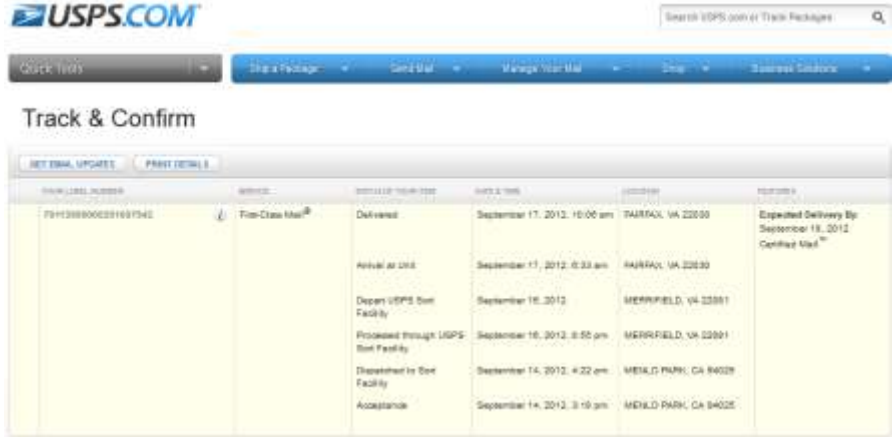
Subject
Academic misconduct problems I hope you deal with, openly and expeditiously

Message

```
Dear President Cabrera:
I am sorry to have to bother a busy person in a new job, but I must
alert you to a serious problem you have inherited:

http://www.desmogblog.com/see-no-evil-speak-little-truth-break-rules-
blame-others
http://www.desmogblog.com/sites/beta.desmogblog.com/files
/see.no_.evil_.speak_.little.truth_.pdf

p.41 of the 69-page report gives a list of 15 formal academic
misconduct complaints I now lodge: plagiarism and/or falsification
by Professor Edward Wegman and/or some of his students, items
1-12,14-16.  Item 13 is needed to fix a Wegman falsehood.  The
report links to many pages of detailed analyses to be included in
the complaints.  p.42 suggests other actions, including apologies.

Sincerely,
John R. Mashey, PhD, Portola Valley, CA
I WOULD APPRECIATE AN ACKNOWLEDGMENT OF RECEIPT OF THIS EMAIL.
```

Automatic response:

09/14/12 Sent (same letter) via Certified Mail, delivered 09/17/12



GMU neither replied to these nor even acknowledged receipt within a month, at which point the additional FOIAs in **§C.1** seemed worth trying, primarily to discover if GMU had informed the agencies at the appropriate times, with requests for the proposals and reports included almost as afterthoughts. Wegman's 2010 resume was already familiar from study in [MAS2010a §A.6] and a truncated and somewhat confusing 0059 report had been found [MAS2010a §A.7]. It was assumed that:

- GMU surely had reported to the agencies.
- Given Wegman's long experience with Federal grants, surely the sloppiness on 0059 was just an error there.

It was surprising to find no evidence that ARO had ever been informed and even more stunning to see, within minutes of receiving the 0447 report, that Wegman had claimed the WR and so many other obviously-unfit works. *That GMU behavior under President Merten would change under President Cabrera was the last hope, but it was forlorn.*

[136] peoplefinder.gmu.edu/contact.php?id=1050&type=faculty&search=angel%20cabrera

F. Plagiarism, falsification and relevant authorities

This table is a consolidated checklist of plagiarism/falsification issues, separate from funds misuse. Most were reported to GMU years ago, which ignored or rejected everything except the already-retracted P179.

The plagiarism/falsification problems in most were known, but they lacked ack's. FOIAs exposed the claims on Federal grants, discoverable in a few minutes' perusal of the reports that GMU transmitted to the agencies from Wegman and Said. A few were recently discovered, labeled "new."

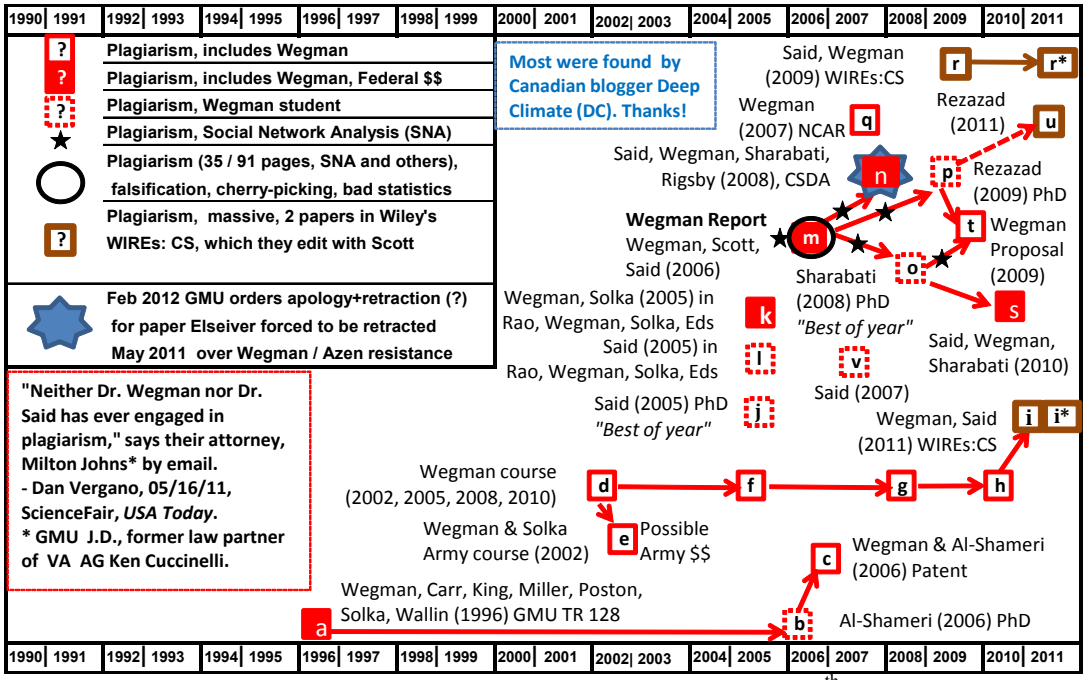
Works involving Wegman and/or Said with plagiarism and/or falsification. Does not include the much larger set of works that are just grant-unfit.
Does not include other students' PhD dissertations or work before 2005, see (§G).
S (s): Said lead author (coauthor)
W(w): Wegman lead author (coauthor)
☆ Alleged plagiarism
★ Alleged falsification/misrepresentation
Problem: (W or S) shows which person, ①②③: funds mis-use against grants 0447, 0059 or 5876, ☆★ is problem
W①☆★ Means: Wegman mis-use of 0447, plagiarism/falsification,for DoD IG

				Relevant Funding Authorities							Reference,section in MAS2012c, or (**section here**)	
		☆	★	DoD IG		DHHS IG	ORI	Publisher	Reported to GMU	GMU Action	MAS2012c, "See No Evil, Speak Little Truth,…" DeSmogBlog, 08/20/12 § (§)	Description
	Authors			0447	0059	5876	5876					
												Cited by Said in support of gaining grant
P401	S	☆				S③	S☆		2011	ignore	☆§3.4	Said dissertation, plagiarism
P402	S	☆				S③	S☆	Elsevier	2011	ignore	☆§4.4	Said plagiarism in chapter of Wegman-edited book
												Acknowledged grants or were claimed for credit
P169	W	☆		W☆					2011	ignore	☆§4.3	Article by Wegman & Solka, claimed for 0447, but fit
P179	Sw	☆		w①☆	Sw②☆	S③	S☆	Elsevier-Retracted	2010	misconduct	☆§2.3	Retracted by *Elsevier* for plagiarism; not peer-reviewed
P200	Sw	☆		w①☆	Sw②☆	S③	S☆	Springer-Verlag	2011	ignore	☆§4.6	Plagiarism reported to GMU, editors, publisher
P405	S	☆				S③	S☆	Washington Acad.Sci	New	-	☆(**§X**)	Said paper on "Eras in the History of Statistics and Data Analysis"
T126	**Ws**	**☆**	**★**	**W①☆★**		**S③**	**S☆★**	**US House of Rep.**	**2010**	**Reject/Ignore**	**☆§2,§3 ★**	**Plagiarism/falsification reported to GMU [MAS2010a, MAS2011b]**
				s①☆★ ?? Said may have also used 0447 funds, (§Q.4)								Of 35 pages' plagiarism documented, GMU accepted 0
T135	Ws		★	W①★				-	New	-	★(**§K.6**)	False statement in abstract
												Plagiarsm in grant proposal, also cited (retracted) P179
~~P409~~	W	☆		W☆ (for DoD, not for 0447/0059)				-	New	-	☆(**§Z**)	Wegman proposal to Army, 02/16/09, rejected
												No known Federal funding
P197	Sw	☆						Wiley	2011	ignore	☆§4.8	WIREs:CS Article by editors Said and Wegman, plagiarism
P513	Ws	☆						Wiley	2011	ignore	☆§4.7	WIREs:CS article by Wegman & Said, plagiarism
~~P407~~	Ws	☆	★					Wiley	2010	ignore	☆★(**§H.7**)	Phantom book, including WR (☆), false statements in blurb (★)
T424	W		★					-	New	-	★(**§K.6**)	Falsification/misrepresentation in Wegman video at Chapman
Totals		11	4	7	2	6	6					

G. Plagiarism chains of Wegman and/or his students

This update from [MAS2012c] alleges a long history of plagiarism by Wegman and some of his students, now totaling 100+ pages with the addition of [t, u, v] at right.

Wegman and some of his students often "borrowed" text with poor or often no attribution, then made trivial changes. Sometimes, unattributed Wikipedia text was copied, while adding some of its citations as disconnected references. People often introduced errors, some ludicrous, especially when departing from the original text. Some parts of the WR displayed falsification, not by faking data, but by plagiarizing expert text, then weakening or inverting its conclusions without basis. Likewise, the WR used statistics code that relied on a 1:100 cherry-pick, also falsification.



Relevant summary sources: generally link back to sources of first identification. Formal complaints have made to GMU on all but [q] (minor) and [t, u] (new), but were almost all ignored.

a, b, c	deepclimate.org/2010/12/02/wegman-et-al-miscellany	Earliest, 4th PhD
a	www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA313999&Location=U2&doc=GetTRDoc.pdf	ARMY ARO
d, e, f, g, h, i	deepclimate.org/2011/03/26/wegman-and-said-2011-dubious-scholarship-in-full-colour	Successive iterations
	deepclimate.org/2011/05/15/wegman-and-said-2011-part-2	Eventually *WIREs:CS* color
j, o, p	deepclimate.org/2010/09/15/wegman-report-update-part-2-gmu-dissertation-review	Said, Sharabati, Rezazad PhDs
m,n	deepclimate.org/2010/07/29/wegman-report-update-part-1-more-dubious-scholarship-in-full-colour	WR total: 10 pages
m, n, o, p, q	[MAS2010a §W.11]	WR: +25 pages, total 35
m	deepclimate.org/2010/11/16/replication-and-due-diligence-wegman-style, [MAS2010a §W.4.1]	Bad statistics, cherry-picking
q	[MAS2010a §A.4]	slides (minor)
m	www.desmogblog.com/wegman-report-not-just-plagiarism-misrepresentation	Alleged falsification
k,l	deepclimate.org/2011/06/07/mining-new-depths-in-scholarship-part-1	2 articles in famous series
r	deepclimate.org/2011/10/04/said-and-wegman-2009-suboptimal-scholarship	WIREs: CS from Wikipedia
s	www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.W.5.7.pdf	Sharabati PhD, new antecedents
*l, *r	deepclimate.org/2012/03/16/wiley-coverup-complete-wegman-and-said-redo-hides-plagiarism-and-errors	Cover-up, [MAS2012c, §5.3]
t	**§Z** this document, regarding WEG2009	*~half plagiarism in grant proposal*
u	www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.W.5.10.pdf.	Self-plagiarism of PhD *WIREs:CS*
v	**§X** Said(2007) article on statistics, in Washington Academy of Sciences (P403 here)	Plagiarism

H. Possible distractions

University professors rarely punch time cards, but often write books, edit journals, run conferences, consult, start businesses and may thus have complex finances.[137] In light of grant funding that expected some focused effort, the following possible concurrent distractions seem worth noting.

H.1 Interface Foundation of North America

[MAS2010a §A.6.2-§A.6.4] summarizes Wegman and Said's work on IFNA, a 501(c)(6) that manages several conferences: the Symposium on the Interface, the Army Conference on Applied Statistics (ACAS), and Quantitative Methods in Defense and National Security (QMDNS).[138]

Unsurprisingly, Wegman was involved with Interface and ACAS almost every year. Said was involved many years, including co-editing both ACAS and Interface Proceedings 2003 and 2004[139], while doing coursework for her PhD. Other Wegman students have spoken often. The following are summarized from [MAS2010a §A.6] and IFNA 990s.[140]

Conferences	2004	2005	2006	2007	2008	2009	2010	2011
Wegman - Interface	X	X	X,N	X	X	X	PC	PC
Wegman - ACAS	S1,C2,O	C3,C4	I1,O1		I2	C6,O3,O5	O4	
Wegman - QMDNS			I	PC			PC, O	
Said - Interface	E	X	PC	X	X		PC,X	PC
Said - ACAS	E	C7	I1			S4,O3, O5		
Said -QMDNS							I	

Wegman signed the 990s that showed roles and hours/week. Wegman's claim of 10 hours/week on IFNA *might be reasonable,* as organizing conferences can be time-consuming. *One would expect the revenues to cover Wegman's and Said's expenses, trips to these conferences.*

IFNA 990 Forms:	2004	2005	2006	2007	2008	2009	2010	2011
Roles, hours/week								
Said, Yasmin				Ex-Of BoD Editor, 10	BoD, Editor,5	- Editor,5	Pres, 2	Pres, 2
Wegman, Edward	Treas,10	Treas,10	Treas,10	Treas,10	Treas,10	Treas,10	Treas,10	Treas,10
Pgm Service Rev I.2	$74,683	$1,800	$71,805	$109,751	$53,428	$19,205	$80,661	$40,425
Total Revenue I.9	$97,998	$98,930	$73,447	$115,316	$56,646	$19,205	$80,725	$54,068
Prof. Fees I.13	$99,946	$76,055	$88,270		$59,969	$21,801	$61,399	$34,372
Total Expenses I.17	$113,640	$82,458	$94,738	$121,294	$74,044	$27,237	$79,200	$38,122
Net assets I.21	$126,010	$142,468	$12,178	$115,199	$97,801	$89,769	$91,294	$107,239
Conferences	I,A	I,A	I,A	I,A,Q	I,A	A	Q,I,A	I,A

Wegman thus claimed that Said was also spending 5-10 hours/week on IFNA during most of her 5876 fellowship.[141] Most years used the 990 EZ (short form), but in 2007, the longer form provided travel costs, which were $6,682, with Interface in Philadelphia, PA, ACAS in Houston, TX, and QMDNS at GMU.[142] IFNA also got $18,100 in fees from government.

H.2 2009–? Quran Research Institute

From its IRS Form 990s, this was a 501(c)(3) public charity,[143] located at 23044 Winged Elm Drive, Clarksburg, MD 20871-4433.[144]

'Ruling Date: 06/2009

NTEE: X05 - RESEARCH INSTITUTES & PUBLIC POLICY ANALYSIS'

No revenues or expenses have been reported. All say (IX-A, 1):

'The Foundation had not begun to function during the year.'

The 10 hours/week numbers here seem dubious, as the reality of this seems unclear. Faxon is also at GMU.

Quran Research Institute, Ruling Date: 06/xx/09		2009	2010	2011
Said, Yasmin	President, hours/week	10	2	2
Wegman, Edward	Treasurer	10	2	2
Faxon, Don	Director		0	0
Aljammaz, Abdulla	Director		0	0

[137] All this is fine, as long as done within university agreements, and if care is taken to avoid commingling of grant funds with others. Senior professors often seem like entrepreneurial small business owners and in some sense, IFNA seems a business for running conferences. Many such activities *may be good ideas*. Generalizations are difficult, because the details matter. Intellectual property ownership can be especially tricky. Good ethics avoids misuse of grad students. **Use of a new PhD and 3-4 grad students to work on the WR seemed *strange*.**

[138] www.qmdns.org At GMU 2007, 2010, 2012.

[139] www.interfacesymposia.org/I04/master.pdf Said, Marchette, Solka Eds.

[140] www.eri-nonprofit-salaries.com/index.cfm?FuseAction=NPO.Summary&EIN=541522706

[141] That may be fine or it might be a distraction from 5876.

[142] As seen in **§I** and **§0.3**, 2007 saw a burst of foreign travel with unknown funding sources. No allegation is made of wrongdoing, but concern is raised.

[143] 990finder.foundationcenter.org/990results.aspx?fn=quran+research+institute&action=Find OR www.eri-nonprofit-salaries.com/index.cfm?FuseAction=NPO.Summary&EIN=262976607

[144] Said's home address, as seen in [WEG2010a, p.10].

H.3 2006– dMining Technology[145]

Wegman, Ali S. Hadi and Rida E. Moustafa are the principals of this software consulting and product company, founded in 2006.

H.4 Supervision of doctoral students

Wegman's Mathematics Genealogy entry has a long student list, **§0.3,**[146] but Said *seemed or claimed* to be co-supervising 5 students during 2006-2008, none of whom seemed to be doing PhDs in alcoholism research:
Alshammari, Sharafi, Belayneh, Sharabati, Mburu
That period included 6-7 known foreign trips for her, including ~5 months in Saudi Arabia and UK, **§I,** and involvement with IFNA, **§H.1**.

The Mathematics Genealogy entry for Said lists [147] 3 students, each with Wegman as Advisor 1 and Said as Advisor 2:[148]

2007 Homayoun Sharafi
Barriers to Teaching Computing from a Distance at Community Colleges
2008 Sirak Belayneh
Identity of Zeros of Higher and Lower Dimensional Filter Banks
2008 Walid Khaled Sharabati
Multimode and Evolutionary Networks

§R.3 and **§R.4** describe student participation in 0059. **§R.4** includes:

> 'The work reported here was implemented by **Eiman Alshammari**, a Ph.D. student who is a native Arabic speaker, **under the supervision of Dr. Edward J. Wegman and Dr. Yasmin H. Said**,[149] who is also a native Arabic speaker and a U.S. citizen.'

In T419 *(***§K.5**), Said wrote:

> 'I would like to acknowledge the contributions of Dr. Rida Moustafa and **my students, Peter Mburu and Walid Sharabati**, who assisted me with the development of various figures and computations in this paper.'

Using postdocs to co-supervise PhD students may or may not make sense at all universities. How much work did they do on T419?

[145] www.dmining-technology.com/bios.html
[146] genealogy.math.ndsu.nodak.edu/id.php?id=41964&fChrono=1
[147] genealogy.math.ndsu.nodak.edu/id.php?id=90582&fChrono=1
[148] Sharabati listed Said as co-supervisor, Belaynek did not. Sharafi was unfound.
[149] On 12/15/06, Said was 1.5 years post-PhD, working on an alcoholism postdoc. Although an Arabic speaker, *there was little else in her background to that point that suggested she would be an appropriate co-supervisor of work on this topic.*

H.5 1986-2010 *CSDA*

Wegman and then Said's close involvement with *CSDA* were detailed in [MAS2010a §W.5.6.1, MAS2011a]. From Wegman's resume and *CSDA* Editorial page archives,[150] he was an Associate Editor from 1986, but the earliest archive, 03/27/04, listed him as member of the Advisory Board. So did 07/31/10, but not 12/04/10, *likely due to [VER2010] and related publicity or perhaps complaints filed with publisher Elsevier on P179*.

Said was named an Associate Editor sometime after 02/18/06 and before 10/08/06, was listed through 07/31/08, but was gone by 07/31/10.[151]

Ted Kirkpatrick noted some oddities about her role:[152]

> 'Now look more closely at the editorial board for the journal. Edward Wegman is on the Advisory Board and Yasmin Said is an Associate Editor. (**Wegman remains listed on the Board for the August 2010 issue** but Said is no longer listed as an Editor.) Reviewing their lists of interests, Wegman's list is appropriately diverse for a scholar of his stature and seniority.
> Said's list, however, is an absolute grab-bag: **"Biostatistics, epidemiology, public health, statistical modeling and graphics, adaptive design, social network theory, data mining, time series analysis, computer intrusion detection, climatology, metadata.'** … **More importantly, this is far too diverse a range of topics for a new scholar (two years out from her Ph.D.)**. Bear in mind that topic lists for Associate Editors are given to indicate areas where they have sufficient expertise to select referees, evaluate their reports, and make a final publication decision. This is a higher level of expertise than required for listing as an interest on one's professional Web page. **It defies likelihood that any young scholar could already have sufficient expertise to act as Associate Editor for such a diverse range of topics**.[153] **Said is the only Associate Editor at the time who listed social networks—or "social" anything—in her interest list.** Presumably, her article was given to some other editor to handle for review,[154] but it's not clear that any of the possible editors knew the topic area well."

[150] wayback.archive.org/web/*/http:/www.elsevier.com/wps/find/journaleditorialboard.cws_home/505539/editorialboard
[151] For a few months, she and many others were erroneously shown on Advisory Board, *likely* via editing errors not actual role changes.
[152] deepclimate.org/2010/04/22/wegman-and-saids-social-network-sources-more-dubious-scholarship/#comment-3389 OR partially [MAS2010a §W.5.6.1]
[153] Many more than most Associate. Editors. *Oddly,* her affiliation was always given as 206 Crabb Avenue, Rockville, MD, not GMU.
[154] Nobody then guessed that E-i-C Azen had done it all himself [MAS2011a].

In 2013, it seemed that WIREs:CS had been long in planning and execution, i.e., somewhat leisurely.
The Wiley contract's schedules, heavy commitments, failures to deliver and long-delayed payments were big surprises.

H.6 2006-2012 Wiley *WIREs:CS*, edited by Wegman, Said, Scott

The 03/05/08 Guide for Authors, p.2 showed:[155]

> 'Dr. Yasmin H. Said
> Co-Editor in Chief and Managing Editor' p.8
> p.8 'The manuscript you submit should be the final version that you wish to be sent for **peer review**.'
> p.12 Articles **commissioned** for *WIREs: Computational Statistics* will be submitted and **peer-reviewed**… In order to facilitate the **peer review** process, at the time you submit your manuscript, please be sure to enter the names and email addresses of 4–5 potential reviewers who are familiar with the topic.'

In 2010, Wegman, Said and Scott described the origin of the journal:[156]

> 'WIREs is a WINNER
> A group of us met with the editorial management of John Wiley and Sons in Hoboken, NJ in **December of 2005.** The original idea was to create an *Encyclopedia of Computational Statistics* analogous to Wiley's highly successful *Encyclopedia of Statistical Sciences*. After we signed a contract to develop the *Encyclopedia*, we met Wiley-Blackwell Vice President and Publisher Sean Pidgeon. Sean floated the idea of *Wiley Interdisciplinary Reviews* (*WIREs*) to us. We agreed that the *WIREs* concept was intriguing and after a number of months of soliciting manuscripts our new journal was **launched officially in July-August 2009** titled as ***Wiley Interdisciplinary Reviews: Computational Statistics***. As earlier readers will know, *WIREs Computational Statistics* is a hybrid review publication that is **by invitation only**, is **refereed**, and publishes in color as needed. Over time, its content will build to provide an encyclopedic coverage of the field. …
> We are proud and elated that our efforts with *WIREs Computational Statistics* have played a significant part in earning these prestigious awards for the WIREs program. We now have more than 120 submissions and **are looking forward to a bright future for our journal**.'

It may not have worked out that way.[157] *This is too bad, as the idea seemed good*. The other WIREs journals *seem credible*, run as normal journals by senior editors, who not only solicited articles, but also accepted proposals and had them clearly peer-reviewed, thanking reviewers.[158]

WIREs:CS has run fine articles,[159] but *WIREs:CS* also had many papers by Wegman students, Editors or their associates and those might be fine, *or might raise concerns about peer review*, **§K.4.2.** After multiple complaints about Said/Wegman plagiarism in 2 papers written for their own "peer reviewed journal" Wiley allowed them quiet rewrites. Further complaints were made to Wiley executives and Board.[160] No problem was admitted, but in June 2012 Wegman and Said quietly disappeared from the *WIREs:CS* masthead, leaving only Scott.[161] Some time in early 2013, several well-known and productive statisticians joined as Editors-in-Chief.[162]

2009-2011- Said and (not) Oklahoma State University

[MAS2010a, §6.5] raised an issue of false rank and affiliation on the *WIREs:CS* masthead, although the full story was then unknown:[163]

> 'Yasmin H. Said, **Professor, Oklahoma State University**, Ruth L. Kirschstein National Fellow,[164] George Mason University'

[MAS2012 §A.3] showed the 04/24/11 email in which Wiley was told that Said had **never** been employed by OSU. In Spring 2009, she accepted a tenure-track Statistics Assistant Professorship offer, for August 2009. She later requested a lighter teaching load, granted, and then ~August asked to start in January. The offer was withdrawn. For a few months, she and OSU believed she would be coming, but both knew otherwise by September 2009. By 09/16/11, this had been partially fixed:[165]

> 'Yasmin H. Said, **Professor**, George Mason University'

Later, it was finally and properly corrected to:[166]

> 'Yasmin H. Said, **Research Assistant Professor**, George Mason University'

---

[155] media.wiley.com/assets/2205/94/WIREs_comp_stats_author_guide.pdf www.webcitation.org/5xyt55RyU

[156] 'WIREs is a WINNER' onlinelibrary.wiley.com/doi/10.1002/wics.85/full

[157] [MAS2012c, §4.7, §4.8, §5.2, §5.3, §A.3.]

[158] wires.wiley.com/WileyCDA/WiresJournal/wisId-WCC.html
For example, where I know some of the people and *think them quite credible*.

[159] For example, "Statistical methods for computer performance evaluation," by David Lilja and student Shruti Patil: Lilja is a leading expert. I've used his book, and the area is familiar, from past work, i.e., SPEC benchmarking group. onlinelibrary.wiley.com/doi/10.1002/wics.192/abstract

[160]

[161] deepclimate.org/2012/07/13/wegman-and-said-leave-wiley-journal-and-said-disappears-from-gmu

[162] web.archive.org/web/20130201230218/wires.wiley.com/WileyCDA/Section/id-398002.html?al=eb James Gentle (GMU) and Karen Kafadar (Indiana U)

[163] www.webcitation.org/5zTFlnebI That was still true 06/15/11.

[164] *The Fellowship was completed 05/29/09, so that was no longer quite true.*

[165] www.webcitation.org/61kfdnQPc

[166] www.webcitation.org/66lLftW7N

H.7 2007 Attempt(s) to publish WR

The following (~~P407~~ Ws☆★)was not 0447-claimed, but gives more evidence of possible distraction from grant work, updated from [MAS2010a §A.3]:

'DC found the strangest connection, an Amazon entry for:
Edward Wegman, *Controversy in Global Warming: A Case Study in Statistics*, December 21, 2007.www.amazon.com/Controversy-Global-Warming-Study-Statistics/dp/0470147849/ref=sr_1_14?ie=UTF8&s=books&qid=1279606110&sr=8-14
"Product Description
Commissioned by the House Committee on Energy and Commerce and the House Subcommittee on Oversight and Investigation, the authors of the book detail time lines, findings, and **interpretations that helped shape the misconception behind the effects of global warming as we know it today. Fraught with the potential for damaging political innuendo and inappropriate social networking overtones,** the authors steer clear of passing personal judgments in favor of outlining the accepted controversies surrounding the topic, this for historical and reconstructive purposes. **The authors' report to Congress is included in its entirety as an appendix at the rear of the book.**[167]
Product Details Paperback: 288 pages; Publisher: Wiley-Interscience (December 21, 2007);
Language: English ISBN-10: 0470147849 ISBN-13: 978-0470147849"'

The reference to "authors" might be a typo, or perhaps this was the book mentioned by Yasmin Said in [SAI2007, p.24]:

'Book • By Wiley – The Heated Debate – under contract..'

Or as of 03/05/08, in *WIREs:CS* biography: [168]

'She is currently writing a book, *Controversies in Global Warming, …*'

Amazon has

'Sign up to be notified when this item becomes available.'

Various bookselling search engines have offered sets of bookseller websites that have called it (concurrently, to amusement):

"In stock – ship in 2-3 days" or "out of stock" or "not yet printed."

Wiley does not seem to list this book, [169] but Google Books said it would appear February 28, 2013,[170] as did overstock.com.[171]
Both showed a (*Wegman-authored?*) blurb:

'**This book recounts the story and facts – in layman's terms -- behind what many believe is a major statistical flaw in recent paleoclimate conclusions regarding global warming, i.e. that the Earth is becoming hotter due to human emissions of carbon dioxide (CO2).**'
(followed by "Commissioned by the House…" text earlier)

Others have shown the WR's own statistics to be seriously incompetent, at best.[172] *The blurb is also seriously incompetent, or a false statement (★)t, but perhaps reveals of the true purpose of the WR.*
The technical criticism of the hockey stick was of the "shaft," not the "blade" [MAS2010a §1.4], whereas *an underlying purpose of the WR and many Wegman talks matches doubt-creation on the credibility of the "blade,"* i.e., the strong human-caused rise in global temperature. For example, the abstract for T135 Ws①★ made a similar **false statement.**

Wegman and (maybe) Said took the effort to propose a book for Wiley, to the point of getting an ISBN and publication date. Presumably the work would have been done during August 2006-December 2007, a period overlapped by all 3 key grants.

[167] Given [MAS2010a], this would have generated both plagiarism ☆ and copyright complaints, had it ever been published.
[168] www.webcitation.org/5xyt55RyU p.6

[169] www.wiley.com/WileyCDA/Section/id-WILEY2_SEARCH_RESULT.html?query=0470147849
[170] books.google.com/books?id=IRIGPQAACAAJ&dq=ISBN:0470147849,ISBN:9780470147849&hl=&source=gbs_api 11/23/12
[171] www.overstock.com/Books-Movies-Music-Games/Controversy-in-Global-Warming/2436732/product.html; www.webcitation.org/6COkehbZ8 11/23/12
[172] deepclimate.org/2010/11/16/replication-and-due-diligence-wegman-style
moyhu.blogspot.com/2011/06/effect-of-selection-in-wegman-report.html
deepclimate.org/2010/10/25/the-wegman-report-sees-red-noise
At worst, this includes falsification, since the code used by McIntyre and the WR contained an obvious 1:100 cherry-pick atop statistical parameters unlike those of the real world. Some problems were shown by David Ritson and others in 2006.

In 2013, this travel schedule seemed unusual, especially for (postdoc) Said. The 2008 WIley contract mentioned the unanticipated foreign travel that interfered with difficult commitments ... having zero relationship to the NIAAA fellowship or ARO grants.

I. Known travel

Academics normally attend conferences, often combine trips and may add talks to vacations, so a heavy travel schedule alone is absolutely not evidence of grant fraud. However, it seemed *especially odd* for postdoc Said to have traveled so much in 2007-2008. Both Wegman and Said expressed financial worries, so some travel *may be a concern* as to funding sources, especially given the frequency of talks tagged red, orange or ①, ②, ③ (in any color). Following assumes attendance at all conference days, but ignores travel days. They may or may not have returned home between some trips. Actual presenters are not always certain. Some Washington-local talks are omitted for brevity.
§0.3 shows these as part of overall graphical timeline.

Bold	**Outside N. America**
Regular	N. America, likely air flights
Italic	*N. America, up to ~4-hour drive*[173]
○ ○⊙	Code, as in **§0.3,** ○Wegman, ○Said, ⊙Wegman+Said
	⊙ used when adjacent trips made that more likely than ○

Dates	C	Place/conference
1999		
… 12/xx	○	**Lausanne, Switzerland, 2 lectures** T73
06/09-06/12	○	Interface Schaumberg, IL
2000		
04/05-05/08	○	Interface New Orleans, LA
11/xx	○	**Dortmund, Germany** T81
2001		
05/xx	○	**U of Aalborg, Denmark, 5 lectures** T86
06/13—6/16	○	Interface Costa Mesa, Orange County, CA
2002		
04/17-04/20	○	Interface Montreal
05/xx	○	**Lahti, Finland** T94
10/xx	○	**Rain am Lech, Germany** T97
2003		
03/12-03/15	○	Interface Salt Lake City
05/xx	○	**Mullingar, Ireland** T102
10/29-10/31	○	ACAS Davis, CA

[173] These were not major/expensive trips, but illustrate time spent on conferences.

Detailed annotations begin

2004

05/26-05/29	○	*Interface Baltimore, MD*
06/xx	○	RSoA Vancouver, BC T108
08/23-08/27	○	**COMPSTAT Prague?** P161, P162, P163[174]
10/22-10/22	○	ACAS Atlanta, GA
11/01		**0447 started**

2005

12/29-01/01	○	**Hyderabad, India** T111 W, T112 W①
….03/xx	○	Buffalo, NY T113 W
….04/xx	○	Orange County, CA T114 W
04/xx	○	*Murray Hill, NJ* T115 W[175]
06/08-06/12	⊙	Interface, St L, MO, T116 W, T117 Ws①, T118 Sw①
09/01		Jerry **Coffey contacted Wegman to do WR**
09/11-09/14	⊙	*SAMSI, Research Triangle Park, NC* T120 W
10/17-10/21	⊙	ACAS Monterey, CA T121 W, T401 S□

2006

02/15-02/16	○	QMDNS Santa Monica, CA T122 W
05/24-05/27	⊙	Interface Pasadena, CA T123 W①, T124 W①
05/26		**5876 started**
06/4-06/07	○	Summer Research Conf, Kerrville, TX T125 W
07/14-07/27	⊙	*Washington, DC* T126 Ws①③☆★
08/06-08/10	⊙	JSM2006 Seattle, WA
		T127 W①, T402 w, T403 W, T404 Ws, T405 Ws
08/28-09/01	⊙	**COMPSTAT Rome**
		P174 Sw①, P175 W, T128 W, T129 W①
09/04-09/06	⊙	**Anacapri, Capri, IT** T130 W[176]
09/xx	⊙	**Naples, IT** T131 W
09/17-10/01	•	Buffalo, NY, visit Wieczorek (fit trip for Said)
10/16-10/20	○	*ACAS Durham, NC* T132 W①[177]
12/15		**0059 started**

[174] ?: Wegman was coauthor on 3 papers, so he might have given 0-3.
[175] This is ~4 hours' drive, but trains also work. I used to live a mile away.
[176] www.stat.unipg.it/iasc/Proceedings/2006/COMPSTAT_Satellites/KNEMO/
[177] www.armyconference.org/ACAS06/default.htm Proceedings not online.

2007

?-01/18	•	**"extensive travel, Europe, Asia" §R.5**
02/07-02/08	⊙	*QMDNS GMU (no travel)*
03/21-03/26	○	**Cairo, Egypt**[178] T133 W, T134 W, T135 Ws①★, T136 W
04/xx	○	**Saudi Arabia (sponsor briefing) §R5**
05/23-05/26	⊙	*Interface Philadelphia, PA* T137 Sw①
07/07-07/11	○	RSoA, Chicago, IL T138 W①
07/29-08/02	◉	JSM2007 Salt Lake City[179] P182 ws①, P183 ws①, P184 Sw①, P185 sw①, T139 Sw①, T407 sw③, T408 s③, T409 S③, T410 sw③, T411 sw③, T412 sw③, T413 sw③ — *Seems strange*
08/05-08/07	○	**Saudi Arabia §R.6**-08/05/07
08/17-08/20	◉	**Azores §R.6**-08/28/07 "we" T140 W
09/12-09/14	◉	**Macerata, IT**, **P178** Sw①②③
10/26-10/27	◉	ASA/NCAR, Boulder, CO T141 W①, T415 S③
10/xx	◉	**Saudi Arabia (sponsor briefing) §R5**
11/xx	◉	**Saudi Arabia §R.5**
11/10-11/13	•	**Riyadh, Saudi Arabia**, §S.3.2, T416 S③
11/xx-12/xx	•	**Saudi Arabia §R.5, §S.3.2 "two months"**
12/15		**0059 ended**

2008

01/07-03/17	◉	**Visiting Fellow at Newton Institute, UK**[180] T142 W
03/18-04/06	•	**Visiting Fellow at Newton Institute, UK**
… 04/xx	○	U Wisconsin, Milwaukee T143 W
04/30		**0447 ended**
05/21-05/24	⊙	*Interface Durham, NC*
05/28-06/12	○	**Visiting Fellow at Newton Institute, UK**
06/19-06/23	◉	**Kayseri, Turkey** T144 Ws, T144a Sw③
06/27-07/02	◉	RSoA, Washington, DC P186 Sw①, P186 Sw①
08/03-08/07	○	JSM, Denver, CO T145 W
08/24-08/29	◉	**COMPSTAT Porto, Portugal** ***P192*** Ws①, ***P193*** sw①, T146 W
10/22-10/24	○	*ACAS, Lexington, VA* T148 Ws③
10/25-10/29	○	American Public Health Assoc, San Diego, T147 W
11/20-11/21	⊙	*NISS, RTP, NC* T149 W

2009

02/16		**Wegman proposal to ARO**
03/31-04/01	◉	Workshop, Phoenix, AZ P194 Sw①, P195 Ws①
04/25	◉	**Sharjah, UAE** T150 W
05/29		**5876 ended**
06/11-06/13	○	Interface/Classification Society, StL, MO T151 Ws③
06/23-06/24	○	Oak Ridge, TN P198 ws, T152 ws
06/28/09	◉	Swarmfest, Santa Fe, NM T153 Sw
08/16-08/22	○	**Durban, South Africa** T155 W
10/19-10/23	◉	ACAS Tempe, AZ T422 sw
12/20-12/23	◉	ICCS-X Cairo, Egypt P199 Sw, T156 W

2010

05/26-05/27	⊙	*QMDNS, GMU (no travel)* T425 Sw③
06/16-06/19	◉	Interface Seattle, WA T426 Sw, T427 Sw
07/31-08/05	◉	JSM2010, Vancouver, BC T428 W, T429 S
08/31	?	ICWIIA, Toronto T430 ws (presenter unknown)
10/20-10/22	○	*ACAS Cary, NC* T431 W

2011

06/01-06/03	○	*Interface, Cary, NC*
08/25/11	•	Dublin, Ireland, T432 Sw③ *likely* ◉ ACAS: neither Wegman nor Said spoke

2012

QMDNS (GMU), Interface, ACAS: neither spoke

2013

04/04-04/06 ○ Interface Chapman University, CA

Est. weeks abroad	*2006*	***2007***	***2008***	*2009*	*Total*
Wegman	*1-2*	***7-8***	***14+***	*3*	*25-27*
Said	*1-2*	***8-12***	***14+***	*?*	*23-28*

Travel is broadening, but from personal experience, *this sort of schedule is not easy*. Said seemed to supervise students, and often ack'd them for help on her papers, *creating some ambiguity of credit*. As seen in **§0.3**, 6 Wegman PhD students finished during 2007-early 2009.[181] 5 other students were reported working on 0059. [182]

[178] www.aucegypt.edu/sse/math/NewsEvents/Pages/20072008Events.aspx
[179] This seems very odd. Almost every talk had nothing to do with their grants.
[180] www.newton.ac.uk/programmes/SCH/AllPart.html

[181] Sharafi, Belayneh, Hohman, Sharabati, Reeves, Rezazad.
[182] Mburu, Alsumait, Alshammari, Mihai, Youn

J. Pro bono, plagiarism, peer review and publish-before-study

Wegman seems to have adopted unusual ideas of pro bono, plagiarism and peer review and publishing that may help explain recent strange behavior.

J.1 Pro bono by experts or not … in this case, NOT

Pro bono funding of the WR was discussed [MAS2010a §A.7].
Indeed, it is *unlikely* anyone on the committee got checks specified for work on the WR,[183] **but pro bono claims have now been self-refuted.**
Wegman claimed the WR presentations and testimony, T126, accurately, as it and other unfit work replaced fit work in 2006-2007.
§0.1 corroborates his claim quite well.
Wegman and Barton repeated a clear message, and these were thus **false statements** by Wegman and Barton to Congress, echoed later by Said:
WR p.1:

> 'This Ad Hoc Committee has worked **pro bono**, has received no compensation, and has no financial interest in the outcome of the report.'

The announcement [BAR2006]:

> **'About the Wegman committee:** Dr. Wegman assembled a committee of statisticians, including Dr. David Scott of Rice University and Dr. Yasmin Said of The Johns Hopkins University. Also contributing were Denise Reeves of MITRE Corp. and John T. Rigsby of the Naval Surface Warfare Center. **All worked independent of the committee, pro bono, at the direction of Wegman.** In the course of Wegman's work, he also discussed and presented to other statisticians on aspects of his analysis, including the Board of the American Statistical Association.'

Joe Barton prepared statement [BAR2006a, p.8]:

> 'I would especially like to thank Dr. Edward Wegman who, **on his own time and his own expense, assembled a pro bono committee of statisticians** to provide us with independent and **expert** guidance concerning the hockey stick studies and the process for vetting this work.'

Wegman testimony [BAR2006a, p.18]:

> 'Our panel is composed of myself, Edward Wegman at George Mason University, David W. Scott at Rice University, and as mentioned, Yasmin Said at the Johns Hopkins University. This ad hoc panel has worked on a **pro bono** basis. **We have received no compensation**, not even taxi fare, and **no financial interest** and we have no financial interest in this.'

Wegman prepared statement [BAR2006a, pp.21-22:

> 'Our panel is composed of Edward J. Wegman (George Mason University), David W. Scott (Rice University), and Yasmin H. Said (The Johns Hopkins University). **This Ad Hoc Panel has worked pro bono, has received no compensation, and has no financial interest in the outcome of the report**.'

Wegman testimony [BAR2006a, p.131]:

> 'Our panel was composed of myself from George Mason University, Dr. David W. Scott from Rice University, and Yasmin Said, Dr. Said, from the Johns Hopkins. **This ad hoc panel has worked pro bono, has received no compensation, and has no financial interest in the outcome.**'
> The debate over Dr. Mann's principal

Wegman testimony [BAR2006a, p.134]:

> 'We certainly agree that modern global warming is real. We have never disputed this point.[184] We think it is time to put the 'hockey stick' controversy behind us and move on.
> I would like to make it clear that our role as statisticians in the hockey stick game is not as players in the hockey game, but as referees. What we have seen and continue to see is that, not withstanding the efforts by Dr. Nychka and others at NCAR, there is relatively little interaction[185] between the statistical community and the climate science/meteorology communities although the latter frequently use statistical techniques. **Statisticians in general have to pay their mortgages just like everyone else and in general cannot afford to do pro bono work such as we have been doing.'**[186]

Said presentation 09/27/07 [SAI2007, p.7]:

> '**We agreed to serve Pro Bono**.
> – To avoid the perception that we were "bought" by the Republican Congress
> – To preserve our independence of either side of the debate.
> – To avoid being coerced into a schedule that would be inconsistent with our other duties.'

She claimed T126, **§S.3.1**, **so her work was not pro bono either**.
Wegman wrote of her 0447-funding, as grad student and postdoc, **§Q.4,** *so perhaps that also applied***.** Students Reese and Rigsby had other jobs.

---

[183] *Given financial pressures, why did he do this? Were promises made or implied for financial or nonfinancial rewards? Why did he keep on with this in 2006-2007 and then again in 2009-2010, having "taken 2nd mortgage to make ends meet"?*

[184] But they did, later: T135: "When the PCA methodology is correctly used, the hockey stick essentially disappears (i.e. the rapid rise from 1850 disappears." and in P407, **§H.7:** 'behind what many believe is a major statistical flaw in recent paleoclimate conclusions regarding global warming, i.e. that the Earth is becoming hotter due to human emissions of carbon dioxide (CO2).' *Those were not for experts and both were false statements or misrepresentations* (★).

[185] This was not a well-informed claim.

[186] Wegman was getting regular payments from 0447. Money is fungible.

The remaining sections of **§J** are offered in hopes of offering context of Wegman's recent actions, which *seem very strange* in the light of his credible work and experience for decades before. None of the allegations depend on these sections, but they may add some context.

J.2 Plagiarism

[MAS2012c] displayed the alleged plagiarisms by Wegman and/or his students, to which are added those in **§X** and **§Z.** Even the earliest cases found were instantly obvious to plagiarism experts [VER2010a], but Wegman has consistently asserted that there has been no plagiarism. **Except for the single case where a paper had already been retracted for plagiarism, GMU agreed completely with Wegman, and refused to investigate the additional cases reported.** Following are a few quotes:

Wegman, 09/06/10 email to Joseph Kunc of USC [MAS2011 p.11]:

> 'The web blog deepclimate.org is, in my opinion, a totally unsavory operation. They have developed conspiracy theories and have consistently made charges of plagiarism not only against Dr. Rapp, but against me and my colleagues in our report to Congress. They have never spoken with me and **have jumped to wild conclusions that have nothing to do with reality.'**

Wegman, 09/13/10 email to Donald Rapp [MAS2011 p.11]:

> 'Thank you for your recent email. It is at least some comfort to know that the zealots aren't targeting me alone. **…** However, **the official definition of plagiarism** involves copying the ideas or words of someone else and presenting them as your own. Of course, in the so-called Wegman report, we make it clear that we were not trying to represent ourselves as the inventors of paleoclimate reconstruction via tree rings as Bradley implies. Indeed, we explicitly say that these materials **were included so as to give the Congressional audience a balanced picture**[187] of the area. The deepclimate website is full of crackpot conspiracy theories. I avoid reading it in order to keep a semblance of normalcy. …I do agree that this is a **shabby attempt at a smear campaign that attempts to deflect scrutiny from the real misconduct revealed by the climategate emails**.'[188]

Vergano, 11/22/10, article [VER2010a]:

> In an earlier e-mail Wegman sent to Joseph Kunc of the University of Southern California, however, he called the **plagiarism charges "wild conclusions that have nothing to do with reality**." **The plagiarism experts queried by USA TODAY disagree** after viewing the Wegman report: …'

Vergano, 11/23/10, article [VER2010c]:

> 'He *(Wegman)* added, "we have never intended that our Congressional testimony was intended to take intellectual credit for any aspect of paleoclimate reconstruction science or for any original research aspect of social network analysis."
>
> "Unattributed near-verbatim copying is plagiarism, period, " Mashey says, by email. "In fact, **much plagiarism exists not to claim intellectual credit for discovering something, but to claim expertise as a base for making strong statements.**" … Some critics of climate science have suggested any plagiarism in the Wegman report is a side show. However, without commenting directly on the case, the federal Office of Research Integrity's John Dahlberg, says, "No, that's not right. We take plagiarism very seriously."'

Wegman, 03/16/11, email to Elsevier [MAS2011a pp.6-10] annotated:

> '**Let me say at the outset that we would never knowingly publish plagierized *(sic)* material**. … When Denise *(Reeves)* returned from her short course at Carnegie-Mellon**, I took her to be the most knowledgeable among us on social network analysis**, and I asked her to write up a short description we could include in our summary. She provided that within a few days, **which I of course took to be her original work. … I gave it as reading material to Walid as background material. Walid included it as background material in his dissertation with only minor amendments.**'

Milton Johns (Wegman's lawyer), 03/16/11 email to Elsevier [VER2011]:

> '"**Neither Dr. Wegman nor Dr. Said has ever engaged in plagiarism …**"'

Elsevier disagreed, even for one of the shorter plagiarized texts.

Summary

Even if Wegman really thought the SNA text was Reeves' own work, it was used at least 5 times without proper credit, typically to support (dubious) claims of expertise. Wegman himself passed it to a student and used some of it himself in **§Z.** Said was involved in at least 10 different works with plagiarism, **§0.2,** including 7 authored/coauthored with Wegman, **§F. Despite this and the misconduct charge upheld (minimally), Wegman was named to a Tenure and Promotions Committee 2012-2015, so all this seemed quite acceptable to the GMU administration**.[189] *This report likely never would have been written, if Wegman had simply admitted the obvious problems and retracted the WR.*

[187] Their view was both incompetent and false.

[188] *This seems a truly strange comment. Does that make the slightest sense?*

[189] But not all GMU professors, as seen in notes by Professor James Gentle: www.webcitation.org/6EgoEZWAo He communicates firm rules in section: "Make sure that work that is supposed to be yours is indeed your own."

J.3 Peer review

[MAS2010a §A.1] discussed the strange interpretation of peer review used by Wegman in his Congressional testimony. Peer review usually means serious review by reasonably-objective experts, usually anonymous at the beginning of the process[190] and chosen by an editor or other third-party, not just by the authors.

- The external review seemed to have been a gathering of quick comments from a few associates, not reviews in the usual sense, certainly not peer-reviews and not anonymous.
- Wegman sent the draft report to a few long-time statistics associates, saying they were not part of his social network because he had not coauthored any papers with them. He had coauthored at least paper with one of them, had written book chapters for some, had recruited at least one for Said's PhD committee, etc.
- While most of the WR was about paleoclimate and SNA, Wegman never sent the draft to paleoclimate or SNA experts, but to statisticians (including fine ones), but sometimes with little relevant expertise.

He 0447-claimed 11 works as papers published in peer-reviewed journals. Of these, 3 were **false (or at best misleading) statements**:[191]

P179 Sw①②③☆

This SNA paper was only given a quick review by long-time associate Editor-in-Chief Stanley Azen[192] and Wegman knew this [MAS2011a]. The paper was published in a peer-reviewed journal [MAS2010a §W.5.6] but **this** paper was not peer-reviewed.

P186 Sw① and P187 Ws①

These alcoholism presentations were later called "peer reviewed abstracts" and Said called them "posters."

Wegman's production of papers for peer-reviewed research journals almost entirely halted by ~2006-2007.

By 2009, they had started *WIREs:CS*, promoted as peer-reviewed:[193]

'As a WIREs author:
Your review will be published alongside other world-class contributions from leading researchers in the field. All WIREs article topics and authors are selected by an internationally renowned Editorial Board, and **all content is rigorously peer reviewed by experts.**'

P197 Sw☆ and P513Ws could not have been peer-reviewed, especially the former, filled with ludicrous errors and stitched together from Wikipedia.

P509 was a lightly-edited extract from Rezazad's dissertation, one of many papers written by Wegman students, P501-P519. It is clear that *WIREs:CS* has published many good articles, *but the quality/consistency of WIREs:CS peer review under Wegman, Said and Scott is in serious doubt.*[194]

As in T115 W, Wegman's traditional support had "eroded," and he was looking for other ideas. Interdisciplinary efforts can be tricky, so it is a good idea to partner with a real expert from another area, start publishing in relevant journals[195] and co-supervise students with them, to make sure the students know the current research. But a different pattern emerged:

- Students seemed to be sent to look at unfamiliar fields, many of which were long-established with vast literatures of their own.
- Works were published, new terminology coined, sometimes without clear evidence of field familiarity. Sometimes basic definitions were plagiarized, with errors, rather than just adopting and citing them.
- Works were concentrated in familiar statistics journals and conferences, not usually where domain experts might be as involved.
- Such domains included climate science, alcoholism, operations research, some computer networking and SNA, the last detailed next in **§J.4.**

In many ways, this is very sad[196]*, both for Wegman and some of his students. Perhaps this was a scramble for fundable work?*

[190] Sometimes conference program committees do stringent peer review, where everyone looks at every abstract or paper. An author knows their identities, but not how anyone voted. They may just get accept or reject notices.
[191] However, they may not have been deliberately false, as it is possible that Wegman simply had a different interpretation of "peer review" than others.
[192] Azen had no obvious publications on this topic. He is no longer an Editor.
[193] wires.wiley.com/WileyCDA/Section/id-398153.html
[194] Expert peer reviewers' comments are valuable. *Lack of them is not a favor.*
[195] Statisticians can bring invaluable expertise, but need to understand an application area well enough to be useful.
[196] As before, I am sympathetic to computational statistics and Wegman's role, but that does not excuse the alleged academic and grant frauds documented so far.

J.4 Publish first, study later - Social Network Analysis

T126 tried to discredit peer review of paleoclimate researchers.[197] SNA consumed at least 13.5 pages of the WR, and much of the testimony in [WEG2006b]. The use of SNA methods was heavily criticized, so P179 was written to create a "peer-reviewed" justification, but with obvious lack of expertise [MAS2010a §W.5]. Wegman's group used SNA without much exposure to the field, while using its tools to generate impressive, but irrelevant graphs. The recent FOIAs expose clarify the chronology, leading to harsher conclusions. *Publishing seemed precede study:*

2004 Wegman MS student Rigsby applied SNA (human) terminology *somewhat oddly* to computer networks.

2005.06 PhD student Reeves took SNA short course from experts.[198] Wegman wrote in 2011 [MAS2011a p.7]

'Her *(Reeves)* company sent her to take a short course on social network analysis from Kathleen Carley, a professor at Carnegie-Mellon University. Dr. Carley is an internationally recognized expert on social network analysis.[199] When Denise returned from her short course at Carnegie-Mellon, **I took her to be the most knowledgeable among us on social network analysis**, and I asked her to **write up a short description** we could include in our summary.'

2005.08 Said applied to NIAAA, with zero visible SNA coursework, also proposing to take "CSS 692 – Social Network Analysis," **§S.1**:

'To this end I am **seeking to develop skills in social networks**, geographic information systems, neuroscience, biology, and alcohol studies in general, to complement my background in mathematics, statistics, and computer science.'

2005.09 Effort began on the WR, whose SNA text was created sometime before July 2006. At that time,[200] Wegman deemed Reeves as most SNA-expert, via a short course. She *seems* to have just given him the summary text, **leaving even less-expert others to do the analysis**.

2006.05 NIAAA awarded 5876 grant to Said.

2006.07 T126 published: WR plus testimonies, including [WEG2006b]. **In 2007, Said claimed credit from NIAAA for the SNA work.**

2006.09 Earliest Said could have "sat in on" an SNA course at GMU.

2007.07 Said sent an interim report to NIAAA, **§S.3.1:**

'I have had to take coursework informally. I have **sat in on** CSS 692, and CSS 645 as planned. **I have written two papers on social network analysis**.'[201]

'I am developing the theory behind bipartite social network analysis for alcohol studies. I have been exercising this model on co-author networks (one set of actors being authors, the other set being papers). **The social network research was reported in the Congressional testimony during July 19 and July 27, 2007*(sic)* and in the report that was submitted the U.S. House Subcommittee on Oversight and Investigations.'**[202] That is T126.

2007.10 Said took a 1.5-day short course on SNA, 3 months after P179 was online, 15 months after T126, as she reported later, **§S.3.2**:

'I also studied social network analysis under Dr. Stanley Wasserman whose book is widely regarded as the authoritative treatment of social network analysis. Dr. Wasserman gave the short course at the Army Conference on Applied Statistics in October, 2007.'

The WR was claimed to be "expert," including by Joe Barton, **§J.1**., but SNA experts were unimpressed by its SNA, a critical element. Wegman attempted to use SNA to discredit paleoclimate research, based on work of grad students with minimal SNA experience, and especially Said, a postdoc who seemed not to have taken an SNA course for credit, or even sat in on one, before T126. Most people acquire expertise first, before writing reports to Congress. **It is here alleged that this was deliberate deception of the public and the US government, 18USC§1001, both in doing this and falsely claiming expertise.**

[197] Missions #1 and #2: discredit hockey-stick, paleoclimate research in general.
[198] www.casos.cs.cmu.edu/events/summer_institute/2005 June 25-30, 2005
[199] Carley **is** a real expert and in 2011 she "panned" P179 [VER2011a].
[200] Someone with subpoena power might ask Reeves more about her involvement.
[201] These were presumably P178 and P179.
[202] She thus claimed the incompetent SNA work in the WR, got its year wrong and then *wrote nonsense*: SNA alcoholism models are not SNA coauthorship models. Neither WR nor testimony had **bipartite** coauthorship models. Bipartite graph theory is old, found in my 1960s textbooks, and also in SNA textbooks no later than 1994 [WAS1994]. **Rigsby or Sharabati may have done the work and then she claimed credit.** Most of the later testimony's SNA work was by Sharabati.

K. Grants, papers and talks tabulated

K.1 Introduction

[MAS2010a] used a version of Edward Wegman's Resume [WEG2010], downloaded February 2010, of which earlier versions were subsets. Following are a few relevant excerpts:

'CONTRACTS' *(pp.1-2)*[203]
Research Associate with Professor George Nicholson on both AFOSR and NSF Grants, 1969, 1970
Research Associate with Professor R. C. Bose on NSF Grant GP-23520, 1971, 1972
Research Associate with Professors W. L. Smith and M. R. Leadbetter on ONR contract, 1973, 1974
Principal Investigator on AFOSR contract, 1975-1977, 1987-1989, 2001-2005
Senior Faculty Fellow Award from NSF, 1976
Principal Investigator on ARO contract, 1987-2008(1)
Principal Investigator on ARL contract, 2006-2007(2)
Principal Investigator on NSF grant, 1987-1993, 1996-2001
Principal Investigator on a Virginia CIT contract, 1987-1990
Principal Investigator on an IBM equipment grant, 1987
Principal Investigator on an Apple Computer, Inc. equipment grant, 1987
Principal Investigator on DoD Instrumentation grant, 1988, 1989, 1992, 1993
Principal Investigator on ONR contract, 1989-1995, 1999-2002
Principal Investigator on NSF SCREMS grant, 1988, 1990, 1993, 1999
Principal Investigator on NIH/NIAAA grant, 2006-2009(3)
Co-Investigator on DARPA contract, 1999-2005(7)(8)
Navy-ASEE Distinguished Faculty Fellow Award from ONR,1996,1997,2008
NSF/ASA/BLS Senior Faculty Fellow, 2000'

(1) W911NF-04-1-0447 and various predecessor contracts.
(2) W911NF-07-1-0059 This was actually handled via ARO.
(3) Yasmin Said is listed as the P.I. / Project Leader for NIAAA F32AA015876-03,[204] 3-year Kirschstein Fellowship.
A search for Wegman in the same database yields zero hits.[205]
(7)(8) DARPA *This seems to be* 8905-48174 (via Johns Hopkins, AF OSR) or 8105-48267 (via Johns Hopkins)

Other sections of the *impressive* resume say:

EMPLOYMENT
- Assistant Professor, University of North Carolina, Department of Statistics, 1968-1973
- Associate Professor, University of North Carolina, Department of Statistics, 1973-1978
- Visiting Professor, University of Manchester (England), Department of Mathematics (on leave from the University of North Carolina), 1976-1977
- Director, Statistics and Probability Program, Office of Naval Research, 1978-1983
- Head, Mathematical Sciences Division, Office of Naval Research, 1982-1986
- Professor and Director, Center for Computational Statistics, George Mason University, 1986- 2006
- Founding Chairman, Department of Applied and Engineering Statistics, George Mason University, 1992-1999
- Chairman, Data Sciences Program, School of Computational Sciences, George Mason University, 2004-2006
- Professor and Director, Center for Computational Data Sciences, George Mason University, 2008-

RESEARCH INTERESTS
- Statistical Graphics and Scientific Visualization
- Computational Statistics
- Time Series Analysis
- Function and Curve Estimation including Splines
- Inference under Order Restrictions
- Parallel Computing
- Massive Data Sets, Streaming Data, and Data Mining
- **Modeling Alcohol Use Behavior**
- **Text Mining**
- **Social Networks**
- Statistical Methods for Computer Intrusion Detection
- **Paleoclimate Reconstruction'** *this is a very strange claim*

Unlike many of the other topics, he has published few or no peer-reviewed research journal papers on the **bold** topics.
Wegman was clearly DoD-experienced and had been successful in proposing, winning and executing DoD research contracts.

ARO had reason to expect good work from 0447 and 0059.

[203] *This seems an impressive record and some of Wegman's papers are well-cited.*
[204] projectreporter.nih.gov/project_info_details.cfm?aid=7491624&icde=0
[205] search2.google.cit.nih.gov/search?q=wegman&site=REPORT&client=REPORT_frontend&proxystylesheet=REPORT_frontend

Appendices **§M, §N, §O, $P** briefly review earlier Wegman grants for context and comparison with 0447 and 0059. *These grants seemed to produce plausible, substantial work qualitatively quite different from most of the fit results of 0447, even ignoring the unfit work there. These seem examples of the kinds of reports a funder might expect from an experienced senior researcher, unlike the later ones.*

For brevity, the full grant numbers are usually shortened to the last 4 digits, and in dense tables, the 1-2 digit codes preceding the grant are used. Especially important are ①②③, whose appearance alleges grant fraud of a specific paper/talk against grants 0447, 0059 or 5876, respectively, either because a report claimed one, or the work itself ack'd a grant improperly.

Researchers are generally supposed to provide Final Reports within 90 days after the end of the grant, **§B.1**. *Wegman seemed to be a bit lax in that,* but was clearly improving through 2004, then

§	Funder	#	Grant	Start	End	Final	Days
§M	ARO	a	DAAH04-94-G-0267	07/15/94	10/01/97	08/07/98	310
§N	ARO	b	DAAG55-98-1-0404	01/05/98	10/30/01	08/05/02	279
§O	ARO	c	DAAD19-99-0314	01/08/99	12/31/02	07/21/03	202
§P	AF OSR	d	F49620-01-1-0274	04/15/01	10/14/03	01/31/04	109
Key grants:							
§Q	ARO	1	W911NF-04-1-0447	11/01/04	04/30/08	12/10/08	224
§R	ARO/ARL	2	W911NF-07-1-0059	12/15/06	12/15/07	03/08/09	**449**
§S	NIAAA	3	F32AA015876	05/26/06	05/29/09	??[206]	>125

The ARO grants specified GMU as the contractor, saying (as in 0059):

'This grant award supports the research identified in the Recipient's proposal'

The following, from resume pp.6-12 are coded T*nnn* (talks) here:

'**INVITED PAPERS**

110. "Visual Data Mining of Streaming Data," … December 2004 …

145. "Mixture Models for Document Clustering," … August 2008

Wegman transferred[207] Talks T110-T117, T119-T145 to pp.21-22 of the Final Report and added a lecture by Said (T144A), all under heading:

'Invited Addresses by Edward J. Wegman Acknowledging Support of ARO Contract W911NF-04-1-0447'

Then, Résumé pp.14-23 has a section for papers, coded P*nnn* here:

'**PAPERS**

157. "Data mining strategies for detection of chemical warfare agents" … 2004 …

193. "Spatial and computational models of alcohol use and problems" … 2008

200."Author-coauthor social networks and emerging scientific subfields"2010'

In 2008, this had been submitted, but not yet published.

Wegman transferred Papers P157-P193 and P200 to pp.23-25 of the 0447 Final Report, plus others that *seem* never to have been published, (labeled here P300, P301) all under the heading:

'**Papers Published by Edward J. Wegman Acknowledging ARO Support under Contract W911NF-04-1-0447 (or its predecessor contract DAAG55-98-1-0404, but not report in the predecessor final report)**'[208]

Wegman then allocated the Papers and Talks to ARO categories, listed on pp.5-10 of the Final Technical Report.

Finally, pp.25-26 of the Final Technical Report shows:

'**Dissertation and Theses Directed by Edward J. Wegman while under Contract W911NF-04-1-0447**

Yasmin H. Said, *Agent Based Simulation of Ecological Alcohol Systems* (PhD) **– Yasmin was supported as a postdoc partly by Contract W911NF-04-1-0447. She has worked on contract W911NF-07-1-0059** with the Army Research Laboratory in our technology transfer effort. She currently holds an F32 award from NIH.'

Grant 0059 proposed a set of activities that effectively included 0447, so any 0447-fit work was considered 0059-fit, to avoid yet more colors.

[206] **§S.5 10/01/09** shows email request from NIAAA to Said asking for termination notice.

[207] The transfer included editing to recover standard reference format.

[208] This seems *strange* - 0404 ended 3 years before 0447 began and the claimed papers were published in late 2003-2004, but none found ack'd 0404.

Wegman's 0447 Final Report 12/10/08 (224 days after completion) claimed almost every 2004-2008 paper and talk. **§K.3** and **§K.5** use Wegman's Feb 2010 Resume, **(# Resume** as an identifier) also coded according to the 0047 Final Report (**0447 Code**).

- Any on hockey-stick or climate issues were filled **red**.
- Any on alcoholism were filled green, fit for Said's 5876, no others.
- Any others unfit for 0447 or 5876 were orange. Some SNA topics could be relevant, but not science coauthorship, for example.
- All other papers (talks) were filled blue (cyan) to give benefit of the doubt for relevance to 0447 (or 0059, in some sense a superset). It seemed enough to identify misuses obvious even from titles.

The identifier includes authors, also shown under **Auth,** (W, S) for lead, (w, s) for coauthor, for Wegman and Said. Other tags may appear:

☐ 0447/5876-claimed, but published or mostly done before, **§B.2**
①②③ 0447-unfit, 0059-unfit 5876-unfit, **either** claimed or ack'd
☆ Plagiarism alleged
★ Falsification / misrepresentation alleged

Auth shows author letters. See **§K.2** on meaning of **Series.**
Ack Lack(■) shows a 0447-claimed work that did not actually ack 0447.
Ack shows grants actually ack'd in inspected works.
Claim shows which final reports claimed support for a paper.
Any circled number ① under Claim or Ack alleges that work to be 0447-unfit (30 of 75, or 41% of 0447-claimed). Given the large effort for the WR, it is *very likely* that more than half the work was 0447-unfit.
Some are also alleged to be unfit for 0059② or 5876 ③.
MAS § gives section number showing plagiarism in [MAS2012c].
§ shows Appendix for the key grants and a few more for context.

F/f indicates **F**oreign trips or **f**lying within US. Sometimes conferences provide travel grants or people add talks to already-sponsored trips or vacations. For conference presentations with multiple authors, actual attendee(s) may be unclear. Only ARO and GMU know if Wegman billed any of the red or green-labeled flights to 0447 or 0559.
Cites gives Google Scholar citations, Total and Self, for some works.

Full references are given in **§K.4** and **§K.5,** but most of the reference text appears here, truncated as needed. The full spreadsheet is attached at: www.desmogblog.com/foia-facts-2-no-pro-bono-federal-funds-mis-used

	Agency	0447 Code	Ack lack	Ack	Claim	MAS §	MAS §	Cites Tot	Cites Oth	Papers (P###) and Talks (T###) From wegman.resume2.pdf, February 2010), plus a few later additions People	Contract #s, all via GMU, unless otherwise noted	§K.1
Key contracts	**ARO**			1①	1①		§Q			**Wegman**	**W911NF-04-1-0447** Almost all work in Wegman's resume in relevant period were claimed to have Ack'd 0447. Many did not.	
	ARL(ARO)			2②	2②		§R			**Wegman,Said**	**W911NF-07-1-0059** This claimed only a few papers, proposal topic/period overlapped with 0447	
	NIAAA			3③	3③		§S			**Said**	**F32AA015876,** Kirschstein Fellowship (health, alcoholism) - no circle version since no actual claims known	
	Virginia/GMU				4					**Said**	**Commonwealth of Virginia, Supplemenntal Grant through GMU**	
In 0447, but lacked ack, so false statement		■				①②③				Unfit talk/paper claimed in Reports (1, 2, or 3) OR ack'd in paper/talk (1, 2, 3),hence any circle is alleged false claim.		
Other contracts	ARO			a	a		§M			Wegman	DAAH04-94-G-0267	
	ARO			b	b		§N			Wegman	DAAG55-98-1-0404 (mentioned in **0447**)	
	ARO/ONR			c	c		§O			Wegman	DAAD19-99-0314	
	AF OSR			d	d		§P			Wegman	49620-01-1-0274	
	DARPA			e	e		-			Wegman	8905-48174 (via Johns Hopkins, AF OSR)	
	DARPA			f	f		-			Wegman	8105-48267 (via Johns Hopkins)	
	DARPA			g	g		-			Wegman	F49620-01-0395 (for P168)	
	ARO			h	h		-			Wegman	DAAD19-01-1-0464 (for P172)	

Color	Type	0059 (W or S)	0447 (W or S)	5876 (S)
Papers	ARO-	**fit**	**fit**	unfit ③
Talks	relevant	**fit**	**fit**	unfit ③
Papers or Talks	alcoholism	unfit ①	unfit ①	**fit**
Papers or Talks	miscellaneous	unfit ①	unfit ①	unfit ③
Papers or Talks	Hockey-stick	unfit ①	unfit ①	unfit ③

0447 categories, not all consistent with resume :P175, T128
Before 0447 ☐
Plagiarism☆
Falsification★

A. **(a) Papers published in peer-reviewed journals (N/A for none)**, p.5. **This is the most important category.** ***Sometimes claims of peer-reviewed journals were*** **dubious.**
B. **(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none) These are basically invited book chapters.** pp.5-6.
c. **(c) Presentations** pp.6-8
D. **(d) Manuscripts** p.10
E. **Peer-Reviewed Conference Proceeding publications (other than abstracts):** p.9
f. **Non Peer-Reviewed Conference Proceeding publications (other than abstracts):** pp.8-9

Bold Italic Underlined = *peer-reviewed jounral*
Bold Italic = *peer-reviewed,*
UPPER = papers, book chapters
lower = presentations, non-peer-reviewed conference proc.

Year.mo	Auth	Series	# Resume	0447 Code	Ack lack	Ack	Claim	MAS §	MAS (§)	F/f	Cites Tot	Cites Oth	Papers from Wegman resume, using its number plus others found later, P301- and P401- (Foreign trip or likely flight in USA)	§K.3(a)

K.2 Chronological series of related works

ARO specifies clear categorization, **§B.1**, but actual progress can be also be calibrated by checking series of related works, briefly:

A **Talks on ideas or early results at research conference**, perhaps to get knowledgeable feedback. This might just be a set of slides or a "poster session."[209] Review, if any, might be limited to abstracts. Sometimes, anyone can present such just by asking, but usually a program committee at least does minimal checks. Posters are usually intended to give wide access, even to early results. They also offer time for longer discussions than a Q&A session following a formal talk.

B **Talks about substantive results, likely accompanied by formal paper published in proceedings**. Review may range from minimal check of an abstract to review of paper by program committee or full peer review by anonymous referees. Acceptance rates vary widely.

C **Paper published in appropriate journal**, ideally with good reputation and strong peer review by experts. Of course, anyone can have papers rejected, which is no shame. This is just the first hurdle, no guarantee of correctness or importance. Valuable work not only survives scrutiny but accumulates positive citations from others. *Persistent absence of such papers and citations for academic papers might raise concerns.*

D **Talks about the above** might be given for years, usually not at research conferences, except perhaps as keynotes, but often at other meetings. *Well-known experts usually have several talks they could give at short notice when invited to speak.*

It is plausible for similar talks to appear several times, but:

- Talks/papers at research conferences (A or B) would be expected to show progress over time.
- Sooner or later, relevant peer-reviewed journal papers (C) should appear.
- So, it is fine for evolving work to appear as A, B, C … then D.
- It is important to speak at relevant research conferences, but that must be balanced against the time to do actual research. *Presenting more-or-less similar talks at minimally-relevant venues does not signify progress.*

[209] en.wikipedia.org/wiki/Poster_session
For example, every Fall, 20,000 people attend American Geophysical Union (AGU) in San Francisco. The exhibit hall gets filled twice a day with thousands of posters, which require hours to even walk past without stopping to talk. Some posters attract many people, others not.

Sometimes, sections of students' dissertation sections were turned into papers with little change but with Wegman and/or Said as authors, sometimes earlier in the author list than the students, as in T153, T406. *Such might be fine, or might raise concern over credit.*[210]

Chronological series emerged[211] from study of all accessible works. *Some seem rather repetitive, with talks at minimally-reviewed conferences, chapters in books edited by the authors, but almost nothing in peer-reviewed research journals. Readers may judge the dividing line between reasonable series and recycling of old content. Series analysis concentrates on the two groups that might have grant-fit work, shown on next page. A few series emerged among the generally-unfit works, shown next, but not tagged as the others.*

{*all*} **2006-2010** are treated implicitly as a series, since the same themes were repeated and since none of this ever generated any peer-reviewed research papers, although such were promised in support of T126. In any case, none of this work could possibly fit any of the grants, so was irrelevant to assessing progress on possibly fit work.

{P167, P171} **2004-2005** Student Alotaiby papers
{P183, P200} **2008** SNA offshoots from Sharabati dissertation
{P403, P405} **2005** Said articles on statistics

In the following, long series are tagged with black circles (❶), short series are tagged with gray circles (❷).[212]

[210] Dissertations must be original work by their authors, but of course, good PhD advisors may spend much time with the student, so papers that expand on dissertation sections might be reasonably coauthored by dissertation advisors. This is a gray area among academics. *It might raise concern if a new paper is composed almost entirely of unchanged sections of a dissertation, without the student as lead author.*

[211] This property emerged from the study of individual articles, as similarities of title and/or material were increasingly noticeable.

[212] No wrongdoing is alleged for any of this and I make no claim to be a domain expert in these topics. This analysis is just a rough attempt to start with a long list of works and try to understand the degree to which new, relevant research was actually happening, or not, for context.

❶ 2004-2007 Visual streaming data
{T110, T114, T116, T119, T131, T133, T134}
These seemed to fit stage D above as Wegman had been talking on this often since 1999 [WEG2010]: T67, T71, T74, T78, T81, T82, T85, T86, T87, T95, T98, T100, T103, T104, T105, T106, T107.

❷ 2005-2006 Text analysis with minimal spanning trees,
{P170, T125}
Solka, Bryant ,Wegman, reasonable evolution from 2004's P190.

❸2005-2007 Automated metadata
{T120, T121, T122, T130, T136, T403, T406}
These seem related to Alshameri(2006) dissertation,[213] who Wegman listed as coauthor in {T130, T403}and may have ack'd him elsewhere. *Strangely,* T406 authors were Wegman, **Said**, Alshameri. *(Why Said?)*

❹ 2007-2010. Text mining.
{T140, T143, T144, T144a, T148, T150, T151, T155, T422, P199, T425, T432}
These seem closely related with much overlap. They might fairly be split into more groups, but impeded by many missing. P163 and P189 might be included, but seemed part of earlier Martinez work.[214] T421 might even fit here, if mostly the same. No peer-reviewed journal papers were found.

❺ 2008 Mixture Models for Document Clustering
{T145, T420}

❻ 2009 (Agent-based models for) Multi-Mode Social Networks
{T152, T153}
T153 seemed mostly direct from Sharabati's dissertation, although the authorship was listed as Said, Sharabati, Wegman.

People might compare these against the 0447 proposal, p.4 here.

[213] deepclimate.org/2010/12/02/wegman-et-al-miscellany This was embedded in a patent application, www.freepatentsonline.com/20100223276.pdf
[214] Angel R. Martinez and Wegman, " A Text Stream Transformation for Semantic-Based Clustering" (2002) was a paper related to his 2002 dissertation: armyconference.org/ACAS00-02/ACAS02/MartinezAngel/MartinezAngel.pdf Martinez also wrote "Data Mining for Text Files" for the 2005 *Handbook of Statistics*, Rao, Wegman, Solka, Eds.

❿ 2004-2006 Said dissertation or close
{T112, P401, T117, T118, T400, T401, T405, P174=T129}
Most of these *seemed* either directly from Said dissertation, or close.

2007 *Chance* article, *likely* same material as above, but not found
{***P180***}

⓫ **2007** HIV/AIDS attempt, see Testa, **§S.3.2**-05/08/07.[215]
{T137, T139}

⓬ **2007-2008** bipartite graphs[216]
{T138, T149}

⓭ 2007-2009
Add temporal data to the geographic/spacial data in first large group
{P185, T417, P186, P187, ***P192***=T146, T419, P194, P195}

⓮**2008** analysis of Wieczorek data, but **not** using Said model
{T418, ***P193***}

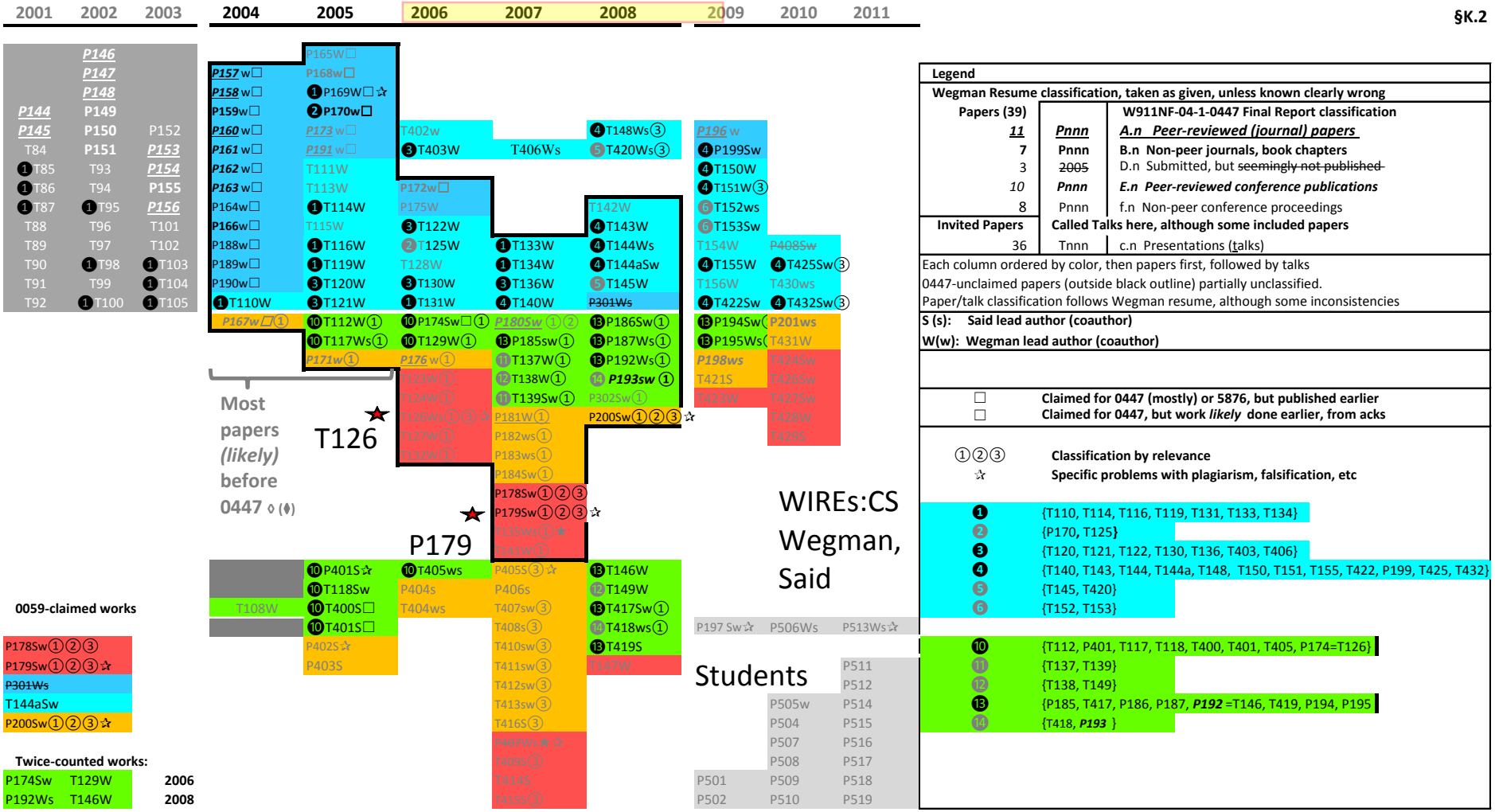
The next page displays these series, in same form as §0.1.

- Many fit 0447-claimed works *seemed* repetitions of efforts for earlier grants, or from dissertations, but again, failed to generate peer-reviewed research papers.
- Said's dissertation was recycled into talks often in 2004-2006. In 2007-2009, they added temporal data to the existing model, and then repeated that often. Wegman authored/coauthored almost every work. All this generated zero papers in peer-reviewed research journals.
- The dearth of work arising from the grants in peer-reviewed research journals and low citation counts *raise concerns*. Only domain experts could properly critique the quality of these works, but that is mostly irrelevant to the allegations.[217]

[215] It seemed they were trying to extend the Said model into HIV/AIDS, but a domain expert explained why this was unlikely to be useful, **§S.5**-05/08/07.
[216] See discussion of bipartite graphs under T138.
[217] If quality was low, that is not grant fraud. These series emerged from seeing repetitions and trying to understand their history.

Graphical version of the series, just blue (0447, 0059) and green (5876)
Many of the talks *seemed* repetitions of a few, those tagged with black numbers. Three sequences (❶❸❹) account for 16/23 of the fit talks Wegman claimed for 0447 and (❿⓭) for a majority of alcoholism works.

This may be fine, or hint that little real research progress occurred. It is hard to know without finding copies of more of the talks. This is not part of any allegation, just a concern, but does tentatively conclude that a few talks comprised the bulk of blue and green works and did not lead to peer-reviewed research journal papers.



K.3 Paper summary table

Year.mo	Auth	Series	# Resume	0447 Code	Ack lack	Ack	Claim	MAS §	MAS (§)	F/f	Cites Tot	Cites Oth	Papers from Wegman resume, using its number plus others found later, P301- and P401- (Foreign trip or likely flight in USA) §K.3(a)
2003.07	w		***P157*** w□	***A.1***	■	d,f	1,c,d				2	1	Solka, Jeffrey L., **Wegman, Edward J.**, and Marchette, David J. (2004) "Data mining strategies for detection of chemical warfare agents," *Statistical Data Mining and Kr*
2004.01	w		***P158*** w□	***A.2***	■	NO	1,c,d				6	1	Marchette, David J. and **Wegman, Edward J.** (2004) "Statistical analysis of network data for cybersecurity," *Chance* , 17(1), 8-18
2004.09	w		**P159**w□	**B.1**	■	ARO	1,c,d				1	0	(error) **Wegman, Edward J.** and Chow, Winston (sic) (2004) "Modeling continuous time series driven by fractional Gaussian noise," in Time Series Analysis and Applica
2004.09	w		***P160*** w□	***A.3***	■	NO	1,d				4	3	Johannsen, D.A., **Wegman, E.J.**, Solka, J.L. and Priebe, C.E. (2004) "Simultaneous selection of features and metric for optimal nearest neighbor classification," *Commur*
2004.09	w		***P161*** w□	***E.1***	■	d	1				3	0	Kafadar, Karen and **Wegman, Edward J.** (2004) "Graphical displays of Internet traffic data," COMPSTAT 2004, (Antoch, J., ed.), Berlin: Physica-Verlag, 287-302. Confer
2004.09	w		***P162*** w□	***E.2***	■	NO	1,d				15	5	Priebe, C.E., Marchette, D.J., Park, Y., **Wegman, E.J.**, Solka, J.L., Socolinsky, D.A., Karakos, D., Church, K.W., Guglielmi, R., Coifman, R.R., Lin, D., Healey, D.M., Jacob, M.
2004.09	w		***P163*** w□	***E.3***	■	NO	1				1	0	Martinez, A.R., **Wegman, E.J.** and Martinez, W.L. (2004) "Using weights with a text proximity matrix," COMPSTAT 2004, (Antoch, J., ed.), Berlin: Physica-Verlag, 327-33
2004.08	w		P164w□	f.1	■	d	1				3	3	Faxon, Don, King, R. Duane, Rigsby, John T., Bernard, Steve, and **Wegman, Edward J.** (2004) "Data cleansing and preparation at the gates: A data-streaming perspectiv
2005.12	W		P165W□	**B.3**	■	c,d,f,h	1,c,d				1	0	**Wegman, Edward J.** (2005) "On some statistical methods for parallel computation," Handbook of Parallel Computing and Statistics (Erricos John, Ed.) 285-307 (2005
2004.12	w		**P166**w□	**B.2**	■	d,f	1,d				2	1	Solka, J.L., Adams, M.L., and **Wegman, E.J.** (2004) "Man vs. machine - A study of the ability of statistical methodologies to discern human generated ssh traffic from m
2004.10	w		***P167*** w□①	***E.4***	■	NO	①				1	1	Alotaiby, Fahad T., Chen, Jim X., **Wegman, Edward J.**, Wechsler, Harry, and Sprague, Debra (2004) "Teacher-driven: web-based learning system," in Proceedings of the
2005.05	w		**P168**w□	**B.4**	■	g	1,d				2	2	Marchette, David J., **Wegman, Edward J.** and Priebe, Carey E. (2005) "Fast algorithms for classification using class cover digraphs," Handbook of Statistics: Data Mining
2005.05	W		P169W□☆	**B.5**	■	e	1	☆§4.3			5	4	**Wegman, Edward J.** and Solka, Jeffrey L.(2005) "Statistical data mining," Handbook of Statistics: Data Mining and Data Visualization (Rao, C. R., Wegman, E. J. and Solk
2005.05	w	❷	**P170**w□	**B.6**	■	e	1				8	3	Solka, Jeffrey L., Bryant, Avory C., and **Wegman, Edward J.** (2005) "Text data mining with minimal spanning trees," Handbook of Statistics: Data Mining and Data Visua
2005.04	w		***P171*** w①	***E.5***	■	NO	①				5	5	Alotaiby, Fahad T., Chen, Jim X., Wechsler, Harry, **Wegman, Edward J.**, and Sprague, Debra (2005) "Adaptive web-based learning system," in the Proceedings of the12
2006.07	w		**P172**w□	**B.7**	■	c,d,f,h	1				?	?	Moustafa, Rida E. A. and **Wegman, Edward J.** (2006) "Multivariate continuous data, generalizations of parallel coordinates," Graphics of Large Datasets: Visualizing a M
2005.08	w		***P173*** w□	***A.4***	■	d	1				4	3	Kafadar, Karen and **Wegman, Edward J.** (2006) "Visualizing 'typical' and 'exotic' Internet traffic data," *Computational Statistics and Data Analysis* , 50(12), 3721-3743
2006.08	Sw		*P174* Sw□①	***E.6***	■	NO	① 3			F	1	0	**Said, Yasmin H** and **Wegman, Edward J.** (2006) "Geospatial distribution of alcohol-related violence in Northern Virginia," in COMPSTAT 2006, (Alfredo Rizzi and Mauri
2006.08	W		P175W	c.18	■	NO	1			F	1	1	**Wegman, Edward J.** and Caudle, Kyle A. (2006) "Density Estimation from Streaming Data Using Wavelets," COMPSTAT 2006, Rome, Italy, August, 2006
2006.12	w		***P176*** w①	***A.5***	■	NO	①,d				8	6	Dorfman, Alan H., Lent, Janice, Leaver, Sylvia G. and **Wegman, Edward J.** (2006) "On sample survey designs for consumer price indexes," *Survey Methodology* , 32(2),
2007.09	Sw		P178Sw①②③	***E.8***		?	①②③			F	2	1	**Said, Yasmin H., Wegman, Edward J.**, Sharabati, Walid K. and Rigsby, John T. (2007) "Implications of co-author networks on peer review," in Classification and Data An
2007.08	Sw		P179Sw①②③☆	***A.6***		①②③	①②③	☆§2.3			40	37	**Said, Yasmin H.**, **Wegman, Edward J.**, Sharabati, Walid K. and Rigsby, John T. (2008) "Style of author-coauthor social networks," *Computational Statistics and Data An*
2007.10	Sw		***P180*** Sw①②	***A.7***		①② 3	①				5	0	**Said, Yasmin H.** and **Wegman, Edward J.** (2007) "Quantitative assessments of alcohol-related outcomes," *Chance* , 20(3), 17-25. Acks from [GMU2010, pp.12-13]
2007.05	W		P181W①	***A.8***	■	NO	①				1	1	**Wegman, Edward J.** and Martinez, Wendy L. (2007) "A conversation with Dorothy Gilford," *Statistical Science* , 22(2), 291-300
2007.07	ws		P182ws①	f.2		?	①				?	?	Alnoshan, Abdullah, Rotenstreich, Shmuel, **Wegman, Edward**, **Said, Yasmin** and Rajput, Adil (2007) "Microeconomic approach to resource allocation in P2P grids," Prc
2007.07	ws		P183ws①	f.3		?	①				?	?	Sharabati, Walid K., **Wegman, Edward J.** and **Said, Yasmin H.** (2007) "A model of preferential attachments for emerging scientific subfields," Proceedings of the Joint S
2007.07	Sw		P184Sw①	f.4		?	①				?	?	(error) **Said, Yasmin H.** and **Wegman, Edward J.** (2007) "Restrictions of trans fatty acids: Health benefits and economic impact in the Washington, DC Metro Area," Pr
2007.07	sw		P185sw①	f.5		?	①				1	0	Mburu, Peter K., **Said, Yasmin H.** and **Wegman, Edward J.** (2007) "Temporal statistics for consequences of alcohol use," Proceedings of the Joint Statistical Meetings,
2008.06	Sw		P186Sw①	***A.9***		?	①				0	0	**Said, Yasmin H.** and **Wegman, Edward J.** (2008) "Using administrative data to estimate cyclic effects of alcohol usage (refereed abstract)," *Alcoholism: Clinical and Exp*
2008.06	Ws		P187Ws①	***A.10***		?	①				0	0	**Wegman, Edward J.** and **Said, Yasmin H.** (2008) "Modeling spatiotemporal effects for acute outcomes in an alcohol system (refereed abstract)," *Alcoholism: Clinical a*
2004.05	w		P188w□	f.6	■	NO	1				0	0	Lin, Chien-Chih, Noh, Eun Young, Yan, Younggping, and **Wegman, Edward J.** (2008) "User profiling in window title and process table," Computing Science and Statistics
2004.05	w		P189w□	f.7	■	DARPA	1				0	0	Martinez, Wendy L., Martinez, Angel R. and **Wegman, Edward J.** (2008) "Classification and clustering using weighted text proximity matrices," Computing Science and
2004.05	w		P190w□	f.8	■	DARPA	1				?	?	Solka, Jeffrey L., Bryant, Avory C. and **Wegman, Edward J.** (2008) "Identifying cross corpora document associations via minimal spanning trees," Computing Science ar
2005.11	w		P191w□	***A.11***	■	e	1				5	5	Reyen, Salem S., Miller, John J. and **Wegman, Edward J.** (2008) "Separating a mixture of two normals with proportional covariances," *Metrika* , doi:10.1007/s00184-0
2008.08	Ws		P192Ws①	***E.9***		① 3	①				2	0	**Wegman, Edward J. and Said, Yasmin H.** (2008) "A directed graph model of ecological alcohol systems incorporating spatiotemporal effects," COMPSTAT 2008, (Paula
2008.08	sw		***P193*** sw①	***E.10***		① 3	① 3			F	0	0	Wieczorek, William F., **Said, Yasmin H. and Wegman, Edward J.** (2008) "Spatial and computational models of alcohol use and problems," COMPSTAT 2008, (Paula Brit
2009.03	Sw		P194Sw①			① 3					0	0	**Said, Yasmin H.** and **Wegman, Edward J.** (2009) "Estimating cyclic and geospatial effects of alcohol usage in a social network directed graph model," **in** *Social Computi*
2009.03	Ws		P195Ws①			①, 3					1	0	**Wegman, Edward J. and Said, Yasmin H.** (2009) "A social network model of alcohol behaviors," in *Social Computing and Behavioral Modeling* , (H. Liu, J. Salerno, M. Y
2009.10	Ws		***P196*** w			NO					7	6	**Caudle, Kyle A.** and **Wegman, Edward J. (2009)** "Nonparametric density estimation of streaming data using orthogonal series," *Computational Statistics and Data An*
2009.07	Sw		P197Sw☆			NO		☆§4.8			1	1	**Said, Yasmin H.** and **Wegman, Edward J. (2009)** "Roadmap for optimization," *Wiley Interdisciplinary Reviews: Computational Statistics* , 1(1): 3-11, DOI: 10.1002/wics
2009.06	ws		P198ws			?				f	0	0	Sharabati, W.K., **Wegman, E.J.** and **Said, Y.H.** (2009) "Multi-mode social networks" *Proceedings of the Human Behavior-Computational Modeling and Interoperability C*
2009.12	Sw	❹	P199Sw			NO				F	0	0	**Said, Yasmin H.** and **Wegman, Edward J. (2009)** "Preserving semantic content in text mining using multigrams," *Proceedings of the Tenth Islamic Countries Conferenc*
2008.12	Sw		P200Sw①②③☆	D.1		①②③	①	☆§4.6		F	0	0	**Said, Yasmin H., Wegman, Edward J.** and Sharabati, Walid K. (2008) "Author-coauthor social networks and emerging scientific subfields," to appear *Data Analysis and*
2010	ws		**P201**ws			?					0	0	Sharabati, Walid K., **Wegman, Edward J.**, **Said, Yasmin H.** (2010) "A ten-mode social network of Afghan society" to appear *International Journal of Computational Inte*
2008	Ws		~~P301~~Ws 2	~~D.2~~		?	1,2						2. **Wegman, Edward J. and Said, Yasmin H.** (2008) "Text mining with application to fraud discovery," submitted to *Applied Stochastic Models in Business and Industry*
2008	Sw		~~P302Sw~~①	~~D.3~~		?	①						3. **Said, Yasmin H. and Wegman, Edward J.** (2008) "Agent-based simulation of the alcohol ecological system," submitted *Journal of the American Statistical Associatio*

This part includes all 0447-claimed papers, and all from that part of Wegman's resume, next page has others.

Year.mo	Auth	Series	Resume	Code	Ack lack	Ack	Claim	§	F/f	Tot	Oth	§K.3(b) (Foreign trip or likely flight in USA)
2005.05	S		P401S☆					☆§3.4		0	0	**Said, Yasmin H**. (2005) *Agent-Based Simulation of Ecological Alcohol Systems* , PhD Dissertation
2005.07	S		P402S☆					☆§4.4		0	0	**Said, Yasmin H**. (2005) "On Genetic Algorithms and Their Application" in *Handbook of Statistics, Volume 24: Data Mining and Data Visualization* by C.R. Rao, E. J. W
2005.08	S		P403S							0	0	**Said, Yasmin H**. (2005) "Statistics" in Encyclopedia of the Modern World, Peter N. Stearns, Ed, Volume 7, pp.135-137.
2006.05	s		P404s			NO				4	4	Banks, D., and **Said, Y. H.** (2006). "Data Mining in Electronic Commerce," *Statistical Science* , 21, 234-246.
2007	S		P405S③☆			NO		③ ☆(§X)		0	0	**Said, Yasmin H**. (2007) On the Eras in the History of Statistics and Data Analysis, Journal of Washington Academy of Sciences, 93(1), 17-35.
2007	s		P406s			NO				0	0	Banks, D., and **Said, Y**. (2007). "New Issues in Human Rights Statistics," in *Statistical Methods for Human Rights* , ed. by J. Asher, D. Banks, and F. Scheuren, Springer: N
2007.12	Ws		~~P407Ws~~★☆					☆(§H.7)		0	0	Wegman and Said, Controversies in Global Warming: A Case Study in Statistics , December 2007, but unpublished.
2009	Ws		~~P408Ws~~									Edward Wegman **(Editor), David W. Scott (Editor),** Yasmin H. Said **(Editor), Jeffrey L. Solka (Editor)** *Encyclopedia of Computational Statistics* , New York: John Wiley ar
2009.02	W		~~P409~~W☆					☆(§Z)				Edward Wegman, grant proposal to ARO 02/16/09, 'Mathematical and Statistical Foundations of Networks,' rejected
			Wiley WIREs:CS									
2009.08			P501									Rida E. Moustafa and Ali S. Hadi, "Grand tour and the Andrews plot" *Wiley Interdisciplinary Reviews: Computational Statistics* , 1(2) 245-250 Sep/Oct 2009 DOI: 10.10
2009.12			P502									Winston C. Chow, "Brownian bridge" *Wiley Interdisciplinary Reviews: Computational Statistics* , 1(3) 325-332 Nov/Dec 2009 DOI: 10.1002/wics.38 online 12/01/09.'
2010.03			P503									David Marchette, "Class cover catch digraphs" *Wiley Interdisciplinary Reviews: Computational Statistics* , 2(2) 171-177 March/April 2010 DOI: 10.1002/wics.70 onlin
2010.03			P504									Angel R. Martinez, "Bigram proximity matrix" *Wiley Interdisciplinary Reviews: Computational Statistics* , 2(2) 251-25, March/April 2010 DOI: 10.1002/wics.67 online
2010.08	w		P505w									**Shores, Roger W.** and **Wegman, Edward J. (2010)** "Bounds on Delaunay tessellations" *Wiley Interdisciplinary Reviews: Computational Statistics* , 2(5) 571-580 Sep/Oc
2010.11	Ws		P506Ws			NO						**Wegman, Edward J**. and **Said, Yasmin H.** (2010) "Natural homogeneous coordinates" *Wiley Interdisciplinary Reviews: Computational Statistics* , 2010 2(6) 678–685 DC
2010.10			P507									Rida E. Moustafa, "Pseudograd tour" *Wiley Interdisciplinary Reviews: Computational Statistics* , 2(6) 711-718 Nov/Dec 2010 DOI: 10.1002/wics.133 online 10/13/10
2010.10			P508									Kyle A. Caudle, Gary O. Fowler, Leah R. Jager, David M. RuthWiley, "Discounting older data" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(1) 30-33 Jan/
2010.11			P509									Hadi Rezazad, "Computer network optimization" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(1) 34-36 Jan/Feb 2011 DOI: 10.1002/wics.135 online 1
2010.11			P510									Wendy L. Martinez, "Graphical user interfaces" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(2) 119-133 March/April 2011 DOI: 10.1002/wics.150 onl
2011.02			P511									Wendy L. Martinez, "Computational Statistics in MATLAB®" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(1) 69-74 Jan/Feb 2011 DOI: 10.1002/wics.1
2011.02			P512									Rida E. Moustafa, "Parallel coordinate and parallel coordinate density plots" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(2) 119-133 March/April 201
2011.02	Ws		P513Ws☆			NO		☆§4.7				**Wegman, Edward J**. and **Said, Yasmin H.** (2011) "Color theory and design" *Wiley Interdisciplinary Reviews: Computational Statistics* ,3(2) 104-117 March/April 2011, fi
2011.02			P514									Winston C. Chow, "Fractal (fractional) Brownian motion" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(2) 149-162 March/April 2011 DOI: 10.1002/wi
2011.03			P515									D. M. Reeves, G. M. Jacyna, "Support vector machine regularization" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(3) 204-215 May/June 2011 DOI: 1(
2011.03			P516									Rida E. Moustafa, "Andrews curves" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(3) 373-282 Jul/Aug 2011 DOI: 10.1002/wics.160online 03/23/11
2011.07			P517									David Marchette, "Implicit translation" *Wiley Interdisciplinary Reviews: Computational Statistics* , 4(1) 28-34 Jan/Feb 2012 DOI: 10.1002/wics.181 online 07/15/11
2011.09			P518									Angel R. Martinez, "Part-of-speech tagging" *Wiley Interdisciplinary Reviews: Computational Statistics* , 4(1) 107-113 Jan/Feb 2012 DOI: 10.1002/wics.195 online 0
2011.06			P519									David Marchette, "Scan statistics in graphs" *Wiley Interdisciplinary Reviews: Computational Statistics* , 3(5) 466-473 Sep/Oct 2011 DOI: 10.1002/wics.1217 online (

§K.4.2 discusses the *WIREs:CS* papers for context, not as basis for any allegation of false claims.
Reminder: actual spreadsheet is attached to www.desmogblog.com/foia-facts-2-no-pro-bono-federal-funds-mis-used

K.4 Papers annotated[218]

K.4.1 Papers related to allegations

***As elsewhere, any of these is alleged to be a clear problem*, §B.2:**[219]

□ 0447-claimed, but published before 0447 started, or likely done before, by evidence of the acknowledgements, **§B.2**. They either explicitly ack'd earlier grants or exhibited publishing history with long delays. Wegman's own numbering showed when he added to resume, sometimes forgetting and then adding them later. YYYY.mo is given for these.

①②③ 0447-unfit, 0059-unfit or 5876-unfit, claimed or ack'd

☆ Plagiarism alleged

★ Falsification / misrepresentation alleged

For some papers, *n/m* gives (number of citations from Google Scholar (GS)) / (number by other authors), i.e., subtracting self-references.[220]

Some works appeared in series that seemed to report the same research, with progress unclear, **§K.2.**

Wegman's resume generally just listed "with" other authors, but 0447 had the real citations, with authors in order, although with occasional errors. A GMU search page gave Said's publications from 2008-2010 [221] and a cross-check for Wegman.[222]

[218] Paper dates must be approximations, given publication delays and asynchrony of journal dates and actual release. Dates here use the earliest when the work was known to be done. If online and print versions were both available, the former was used. Sometimes long delays intervened between submission and publication.

[219] NSF CAREER Workshop @ GMU, p.66:
'Forms of misconduct …Charge for work already done – can be a felony, do not charge twice for the same work.'

[220] That is approximate, as GS can get confused, but still a useful hint. The 2nd number approximates use of the work for other authors, although it included Wegman students, which arguably raised counts. Most of the papers claimed for 0447 gathered few citations from nonrelated authors.
This is not a serious citation analysis, just a hint.

[221] https://par.cos.gmu.edu/biblio/author/Said

[222] https://par.cos.gmu.edu/biblio/author/wegman

Wegman had written in the 0447 Final Report:

'Papers Published by Edward J. Wegman Acknowledging ARO Support under Contract W911NF-04-1-0447 (or its predecessor contract DAAG55-98-1-0404, but not report in the predecessor final report)'[223]

For context, a few earlier samples from [WEG2010] are:

Most frequently cited:

P022 551/?

'"Hyperdimensional data analysis using parallel coordinates," Washington Operations Research and Management Science Council, Washington, DC, January, 1986'

P039 259/?

'"Splines in statistics," with I. W. Wright, *Journal of the American Statistical Association*, 78, 351-365, 1983'

P040 259/?

'"Optimal nonparametric function estimation," *Journal of Statistical Planning and Inference*, 9, 375-388, 1984

Quick sample from 2000-2003, for comparison with 2005- [224]

P139 11/10

'"Affordable environments for 3D collaborative data visualization," *Computation in Science and Engineering*, 2(6), 68-72, 74, 2000'

P146 10/9

'"On methods of computer graphics for visualizing densities," with Q. Luo, *Journal of Computational and Graphical Statistics*, 11(1), 137-162, 2002'

P148 28/24

'"On some mathematics for visualizing high dimensional data," with Jeffrey L. Solka, *Sanhkya (A)*, 64(2), 429-452, 2002'

P150 14/11

'"On some techniques for streaming data: A case study of Internet packet headers," with David J. Marchette, *Journal of Computational and Graphical Statistics*, 12(4), 893-914, 2003'

P156 30/23

'"On some techniques for streaming data: A case study of Internet packet headers," with David J. Marchette, *Journal of Computational and Graphical Statistics*, 12(4), 893-914, 2003'

[223] This may have been a fair comment, but see **§B.2** about work already done. Of course, those could not ack 0447, but none actually ack'd 0404 either, **§N**.

[224] This claims no pretense of being a random sample, just a few that looked interesting from years just before P157. Strong papers accumulate references over time, so one cannot over-interpret comparisonswith newer papers. Still, it is clear that other people noticed and cited these papers, not so clear of later efforts.

P157 w□[225] 2003.07 2/1

'Solka, Jeffrey L., **Wegman, Edward J**., and Marchette, David J. (2004) "Data mining strategies for detection of chemical warfare agents," *Statistical Data Mining and Knowledge Discovery*, 71-92.'

Wegman claimed this as peer-reviewed journal paper, but it really was as book chapter was edited by Hamparsum Bozdogan July 29, 2003, from a 2002 conference.

P158 w□[226] 2004.01 6/1

'Marchette, David J. and **Wegman, Edward J**. (2004) "Statistical analysis of network data for cybersecurity," *Chance*, 17(1), 8-18'

Chance[227] is peer-reviewed, but is not a research journal:

'CHANCE is not a technical magazine, but rather a cultural record of an evolving field, intended to entertain as well as inform.'

P159 w□[228] 2004.09 1/0

'**Wegman, Edward J.** and Chow, Winston *(sic)* (2004) "Modeling continuous time series driven by fractional Gaussian noise," in Time Series Analysis and Applications to Geophysical Systems, a book in the series : The IMA Volumes in Mathematics and its Applications , Vol. 139, New York: Springer-Verlag, (Brillinger, David R.; Robinson, Enders A.; Schoenberg, Frederic P., Eds.), 239-256'

This erroneously reversed order of authors from the actual paper.

P160 w□[229] 2004.09 4/3

'Johannsen, D.A., **Wegman, E.J.**, Solka, J.L. and Priebe, C.E. (2004) "Simultaneous selection of features and metric for optimal nearest neighbor classification," *Communications in Statistics: Theory and Methods*, 2137-2158'

P161 w□[230] 2004.09 3/0 This likely led to ***P173*** w□

'Kafadar, Karen and **Wegman, Edward J**. (2004) "Graphical displays of Internet traffic data," COMPSTAT 2004, (Antoch, J., ed.), Berlin: Physica-Verlag, 287-302.' Acknowledgement:

'Funding from Grant No. F49620-01-1-**0274** from the Air Force Office of Scientific Research, awarded to George Mason University, is gratefully acknowledged. Part of this research was conducted during the first author 's appointment as faculty visitor at National Institute of Standards and Technology.'

P162 w□[231] 2004.09 15/5(?)

'Priebe, C.E., Marchette, D.J., Park, Y., **Wegman, E.J.**, Solka, J.L., Socolinsky, D.A., Karakos, D., Church, K.W., Guglielmi, R., Coifman, R.R., Lin, D., Healey, D.M., Jacob, M.Q., and Tsao, A. (2004) "Iterative denoising for cross-corpus discovery," COMPSTAT 2004, (Antoch, J., ed.), Berlin: Physica-Verlag, 381-392'

P163 w□[232] 2004.09 1/0

'Martinez, A.R., **Wegman, E.J**. and Martinez, W.L. (2004) "Using weights with a text proximity matrix," COMPSTAT 2004, (Antoch, J., ed.), Berlin: Physica-Verlag, 327-338'

P164 w□[233] 2004.08 3/3

'Faxon, Don, King, R. Duane, Rigsby, John T., Bernard, Steve, and **Wegman, Edward J.** (2004) "Data cleansing and preparation at the gates: A data-streaming perspective." In 2004 *Proceedings of the American Statistical Association' That is really JSM2004.*

[225] binf.gmu.edu/jsolka/PAPERS/chapter04.pdf
wayback.archive.org/web/*/binf.gmu.edu/jsolka/PAPERS/chapter04.pdf
[226] amstat.tandfonline.com/doi/abs/10.1080/09332480.2004.10554881
dmarchette.com/Papers/chance.pdf
www.webcitation.org/6D5qZe5FQ
This was sometimes cited as Wegman and Marchette.
[227] chance.amstat.org/about
[228] citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.87.8501&version=0
[229] wayback.archive.org/web/*/binf.gmu.edu/jsolka/PAPERS/paper_4_auth.pdf

[230] rd.springer.com/chapter/10.1007/978-3-7908-2656-2_23
www.amazon.com/COMPSTAT-2004-Proceedings-Computational-ebook/dp/B000U5FXMU
[231] old-site.clsp.jhu.edu/~damianos/cs04reprint_final.pdf
www.webcitation.org/6D5s9lly3
[232] rd.springer.com/chapter/10.1007/978-3-7908-2656-2_26
[233] osf1.gmu.edu/~dfaxon/JSM2004-000742.pdf
www.webcitation.org/6D5sGHFWJ

P165 W□[234] 2005.12 1/0

'**Wegman, Edward J.** (2005) "On some statistical methods for parallel computation," *Handbook of Parallel Computing and Statistics* (Erricos John, Ed.) 285-307'

This Chapter 9 in a book published December 2005 was carefully ack'd:

'This work was completed under the sponsorship of the Office of Naval Research under contract DAAD19-99-1-**0314** administered by the Army Research Office, by the Air Force Office of Scientific Research under contract F49620-01-1-**0274** and contract DAAD19-01-1-**0464**, the latter also administered by the Army Research Office and finally by the Defense Advanced Research Projects Agency through cooperative agreement 8105-4**8267** with the Johns Hopkins University. The author would like to thank three anonymous referees and also Prof. Karen Kafadar for careful reading and helpful suggestions that substantially improved the presentation of this chapter.'

This was tagged (□), pre-0447:

- It was sequenced between works from 08/xx/04 and 12/xx/04.
- It ack'd an earlier contract, but not 0447.
- The 0314 final report, claimed this as "submitted but not yet published.," 07/21/03, **§O.4.** Thus, it seemed done long before 0447 started.

P166 w□[235] 2004.12 2/1

'Solka, J.L., Adams, M.L., and **Wegman, E.J.** (2004) "Man vs. machine - A study of the ability of statistical methodologies to discern human generated ssh traffic from machine generated scp traffic," in *Statistical Methods in Computer Security*, (W. Chen, ed.), Marcel-Dekker, New York, 169-181'

This was published in 12/xx/04 as a book chapter, adding Wegman as 3rd author to paper published earlier[236] by the others. The ack added:

'The work of the third author (EJW) was completed under the sponsorship of the Air Force Office of Scientific Research under the contract F49620-01-1-**0274** and the Defense Advanced Research Projects Agency through cooperative agreement 8105-4**8267** with Johns Hopkins University.'

P167 w□①[237] 2004.10 1/1 {P167, P171}

'Alotaiby, Fahad T., Chen, Jim X., **Wegman, Edward J**., Wechsler, Harry, and Sprague, Debra (2004) "Teacher-driven: web-based learning system," in Proceedings of the 5th Conference on Information Technology Education, ACM SIGITE, 284'

This one-page paper was on Alotaiby's work, no attendance by Wegman assumed.

P168 w□[238] 2005.05 2/2

'Marchette, David J., **Wegman, Edward J.** and Priebe, Carey E. (2005) "Fast algorithms for classification using class cover digraphs," *Handbook of Statistics: Data Mining and Data Visualization*, (Rao, C. R., Wegman, E. J. and Solka, J. L., eds.), 331-358'

This was slightly mis-titled. The correct title was

"Fast algorithms for classification using class cover **catch** digraphs."

This was tagged (□), pre-0447:

- The final report for 0314, 07/21/03, **§O.4,** claimed this as "submitted but not yet published," although it also showed it as submitted to the *Journal of Computational and Graphical Statistics. Either that was an error, or the paper was not accepted, and instead appeared in the book.*
- The book was published in May 2005, and usual publishing delays imply it was done before 0447 started.

[234] www.crcnetbase.com/doi/abs/10.1201/9781420028683.pt2

[235] dx.doi.org/10.1201/9781420030884.ch11 (paywall)

[236] www.amstat.org/sections/srms/proceedings/y2002/files/JSM2002-000805.pdf
Original was by Solka and Adams, the later book chapter added Wegman, but was almost identical, with minimal typographical editing. Figure 4's 'The scp sessions are plotted as blue circles while the ssh sessions are plotted as red plus signs.' was missed in conversion from color to black/white.

[237] dl.acm.org/citation.cfm?id=1029617&dl=ACM&coll=DL&CFID=242530769&CFTOKEN=37020151

[238] www.sciencedirect.com/science/article/pii/S0169716104240123
lnfm1.sai.msu.ru/~rastor/Books/Rao_et_al-Handbook_of_Statistics_Data_Mining_and_Data_Visualization.pdf

P169 W□☆[239] 2005.05 5/4

'**Wegman, Edward J.** and Solka, Jeffrey L.(2005) "Statistical data mining," *Handbook of Statistics: Data Mining and Data Visualization* (Rao, C. R., **Wegman, E. J.** and Solka, J. L., eds.), 1-46'

'The work of E.J.W. was supported by the Defense Advanced Research Projects Agency via Agreement 8905-4**8174** with The Johns Hopkins University. This contract was administered by the Air Force Office of Scientific Research. The work of JLS was supported by the Office of Naval Research under "In-House Laboratory Independent Research." Figures 21 through 24 were prepared by Professor Karen Kafadar who spent time visiting E.J.W. During her visit, she was support by a Critical Infrastructure Protection Fellows Program funded at George Mason University by the Air Force Office of Scientific Research. Much of this chapter summarizes work done with a vast array of collaborators of both of us and we gratefully acknowledge their contributions in the form of ideas and inspiration.'

This book chapter has plagiarism alleged [MAS2012c §4.3].
It was tagged (□), pre-0447:

- It ack'd a contract that ended 06/30/04, but did not ack 0447.
- The book was published in May 2005, and usual publishing delays imply it was done before 0447 started.

P170w□[240] 2005.05 8/3 {②}

'Solka, Jeffrey L., Bryant, Avory C., and **Wegman, Edward J**. (2005) "Text data mining with minimal spanning trees," *Handbook of Statistics: Data Mining and Data Visualization*, (Rao, C. R., Wegman, E. J. and Solka, J. L., eds.), 133-170'

This book chapter acknowledged:'

'This work was sponsored by the Defense Advanced Research Projects Agency under "Novel Mathematical and Computational Approaches to Exploitation of Massive, Non- physical Data." Some of the authors (J.L.S. and A.C.B.) also acknowledge the support of the Office of Naval Research (ONR) under "In-House Laboratory Independent Research." The DARPA support of one of the authors (E.J.W.) was obtained via Agreement 8905-4**8174** with The Johns Hopkins University. His DARPA contract was administered by the AFOSR.'

Although this was clearly an outgrowth of P190[241], and included some similar material, it was not grouped, because it clearly added new text.
It was tagged (□), pre-0447:

- It ack'd a contract that ended 06/30/04, but did not ack 0447.
- The book was published in May 2005, and usual publishing delays imply it was done before 0447 started.

P171 w①[242] 2005.04 5/5 {P167, P171}

'Alotaiby, Fahad T., Chen, Jim X., Wechsler, Harry, **Wegman, Edward J.**, and Sprague, Debra (2005) "Adaptive web-based learning system," in the Proceedings of the12th Annual IEEE International Conference and Workshop on the Engineering of Computer-Based Systems, 423-430' *April 2005 in MD'*

This fit Wegman student Alotaiby's research and dissertation.

[239] www.elsevier.com/books/handbook-of-statistics/rao/978-0-444-51141-6
lnfm1.sai.msu.ru/~rastor/Books/Rao_et_al-Handbook_of_Statistics_Data_Mining_and_Data_Visualization.pdf
www.webcitation.org/6D5v1mVpx

[240] www.sciencedirect.com/science/article/pii/S0169716104240056
www.webcitation.org/6D5v1mVpx

[241] P188-P190 were really from 2004, but erroneously claimed for 2008.

[242] ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1409944

P172 w□[243] 2006.07 ?/? (could not find)

'Moustafa, Rida E. A. and Wegman, Edward J. (2006) "Multivariate continuous data, generalizations of parallel coordinates," Graphics of Large Datasets: Visualizing a Million, (Antony Unwin, Martin Theus, Heike Hofmann, eds.) 143-156

It carefully acknowledged earlier grants:

'*Rida Moustafa* completed this work on a postdoctoral appointment at the Center for Computational Statistics, George Mason University, USA. His work was funded by the Air Force Office of Scientific Research under contract F49620-1-**0274**. *Ed Wegman* is Director, Center for Computational Statistics, George Mason University, USA. His work was funded by the Air Force Office of Scientific Research under contract DAAD19-99-1-**0314** administered by the Army Research Office, by the Air Force Office of Scientific Research under contract F49620-01-1-**0274** and contract DAAD19-01-1-**0464**, the latter also administered by the Army Research Office, and finally by the Defense Advanced Projects Agency though cooperative agreement 8105-4**8267** with the Johns Hopkins University.'

This is related to work reported to ARO for 0404,[244] which mentions Moustafa's dissertation, *Fast Conceptual Clustering Algorithm for Data Mining and Visualization* (2001).

P173 **w□**[245] 2005.08 4/3 A plausible earlier version was ***P161*** w□.

'Kafadar, Karen and Wegman, Edward J. (2006) "Visualizing 'typical' and 'exotic' Internet traffic data," *Computational Statistics and Data Analysis*, 50(12), 3721-3743

Kafadar was Corresponding Author and the history was:

'Received 28 July 2004; received in revised form 27 June 2005; accepted 29 June 2005 Available online 1 August 2005.'

Comparison with an earlier version[246] was instructive. The old ack was:

'Funding from Grant No. F49620-01-1-**0274** from the Air Force Office of Scientific Research, awarded to George Mason University, is gratefully acknowledged. Part of this research was conducted during the first author 's appointment as faculty visitor at National Institute of Standards and Technology.'

The published acknowledgement showed peer review, unlike P179, **§H.5.**

'Funding from Grant No. F49620-01-1-0274 from the Air Force Office of Scientific Research, awarded to George Mason University, is gratefully acknowledged. **We also thank** the Editor, Associate Editor, and an anonymous referee for **useful suggestions on an earlier version of this article**.'

The original had 7 references, of which 4 were by Wegman or students. Vardeman and Jobe is a textbook, and Tukey(1977) is an old classic.

'[1] Marchette D.J . (2001). Computer intrusion detection and network monitoring. Springer.

[2] Khumbah N.-A., Wegman , E.J . (2003). Data compression by geometric quantization. Recent Advances and Trends in Nonparametric Statistics, M. Akritas, D.N. Politis (eds) , North Holland Elsevier , Amsterdam.

[3] Silverman B.W. (1986). Density estimation. Chapman and Hall: London.

[4] Tukey J.W. (1977). Exploratory data analysis. Addison-Wesley, Reading, Massachusetts.

[5] Vardeman S.B., Jobe J.M. (1999). Statistical quality assurance methods for engineers. Wiley, New York.

[6] Wegman E.J ., Marchette D.J . (2003). On some techniques for streaming data: A case study of Internet packet headers. J. Comput. Graph. Stat . **12** (4) , 893-914.

[7] Wegman E.J .; Marchette D.J.[247] (2004). Statistical analysis of network data for cybersecurity. Chance, 9-19.

The final published version added ~26 additional references. *I speculate that* a knowledgeable referee wanted more context.[248] This ack'd only a grant that ended in 2003, and most work was clearly done pre-0447 (□).

[243] The book can be found on Amazon.com, and the ack is on p.27. www.amazon.com/s/ref=nb_sb_noss?url=search-alias%3Daps&field-keywords=%22Graphics+of+Large+Datasets%3A+Visualizing+a+Million%22 Google Scholar did not seem to be able to find it.

[244] www.dtic.mil/dtic/tr/fulltext/u2/a414241.pdf, which ended 10/31/01.

[245] www.sciencedirect.com/science/article/pii/S0167947305001489 Changes have not been compared in detail, but one kind was striking: 9 references (of which 5 were Wegman and/or students) got expanded to 33, including additions of Cleveland and Tufte classics references.

[246] The earlier version was "ARMY2010_IDVIS\kkew1.pdf" dated 07/27/04. www.documentcloud.org/documents/550097-kkew1.html

[247] This was P158, a slight mis-cite, as it was Marchette and Wegman.

[248] *This seemed a credible paper, and if it got more statisticians interested in the problems, it was useful.* Networking is a huge field with a long history and huge literature. Interdisciplinary papers may be published in a wide range of journals. *CSDA* was plausible, but *analysis of computer networks was not an obvious specialty there*, and even the expanded set of references had few from networking journals. Wegman and/or students published SNA papers, but not generally in SNA journals. *P173 showed much more expertise. One might wish it to have been published in a networking journal to assess its credibility and novelty than to include 2 pages of tutorial on IP addresses for a statistics journal.*

P174Sw□①[249]=T129W① 2006.08 1/0 {❿}

'**Said, Yasmin H** and **Wegman, Edward J.** (2006) "Geospatial distribution of alcohol-related violence in Northern Virginia," in COMPSTAT 2006, (Alfredo Rizzi and Maurizio Vichi, eds.), 197-208 Rome, IT' August

It was acknowledged (p.206):

'This paper benefitted greatly from the help and encouragement of Professor William Wieczorek of Buffalo State, part of the SUNY system and of Professor David Anderson of George Mason University.'

Said's affiliation is given as Johns Hopkins U. Most text seemed taken word-for-word from Said's dissertation, P401S, but with no reference. Wegman was added as coauthor, and "I" changed to "we," or "my" to "our." Of the 19 references, 14 were from the dissertation, (Ezz02) was cited with no reference, and 4 were new (Mor01, Poik77, Ross96, Smoo85). Corresponding pages were:

P174 Sw	P401S (PDF pages)	"We" ← "I" (or equivalent), notes
197	10, 11, 13	1 case
197 (1)	13	
198	13, 14, 15	2 cases
198 (2)	23	1 case
199	23, 24,25	2 cases
199 (2.1)	28	
200	28, 29, 30, 31	2 cases
201 N/A		
202 N/A		
203	40	
203 (3.2)	41	
203 (4)	49, 51	4 cases
204	51	2 cases
204 (5)	New text condensed descriptions from P401S	
205	83	Fig.2 ← Figure 13
205 (6)	New conclusion text	
206	94	Fig. 3 ← Figure 23
206	95	Fig. 4 ← Figure 24
207	105,106,107	References

If a new PhD can convert a part of their dissertation into a credible paper, *that is good*, but usually the paper references the dissertation, especially when much of the paper is extracted almost word-for-word.
Here, it seemed Wegman was taking coauthorship credit for Said's work, or that Said's dissertation should have credited Wegman for it.
Wegman claimed this (①), but the paper did not ack 0447. Said claimed it in **§S.3.1**, but the paper was mostly taken from her dissertation, with no evidence of any further development from its status as of May 2005, a year before 5876 started, so (□). *Was this worth a trip to Italy?*

P175 W[250]=T128 W 1/1

'**Wegman, Edward J**. and Caudle, Kyle A. (2006) "Density Estimation from Streaming Data Using Wavelets," COMPSTAT 2006, Rome, Italy, August, 2006'

This seemed to have been derived from Caudle's 2005 dissertation, and also seems to have been the only claimed, fit paper to have actually have been done and published during the 0447 grant period. It does actually fit the 0447 abstract ("density estimation") on p.4 here.

P176 **w①**[251] 8/6 {P167, P176}

'Dorfman, Alan H., Lent, Janice, Leaver, Sylvia G. and **Wegman, Edward** J. (2006) "On sample survey designs for consumer price indexes," *Survey Methodology*, 32(2), 197-216'

P177 Wegman's resume had no entry numbered 177.
It might have been omitted or possibly was submitted, but not accepted.

[249] link.springer.com/chapter/10.1007%2F978-3-7908-1709-6_15?LI=true
books.google.com/books?hl=en&lr=&id=kO9nNrrT1MoC&oi=fnd&pg=PA196&ots=CH1j4qUvOE&sig=Ot2UO_ckfGwIlvANl4XxEP1UukQ#v=onepage&q&f=false

[250] link.springer.com/chapter/10.1007%2F978-3-7908-1709-6_18

[251] www.statcan.gc.ca/ads-annonces/12-001-x/9554-eng.pdf
www.webcitation.org/6D5vtNgy0

P178 Sw①②③[252] 2/1

'**Said, Yasmin H., Wegman, Edward J**., Sharabati, Walid K. and Rigsby, John T. (2007) "Implications of co-author networks on peer review," in *Classification and Data Analysis*, Macerata, Italy: EUM-Edizoni Università di Macerata, 245-248 (Sept 12-14 2007)'

This conference paper in Italy was claimed in **§S.3.2**. P178 and P179 had identical authorship, and the latter was published online 08/09/07, but not yet in print. If P179 had similar SNA text, ☆ would be appropriate.

P179 Sw①②③☆ 40/37[253]

'Said, Yasmin H., Wegman, Edward J., Sharabati, Walid K. and Rigsby, John T. (2008) "Style of author-coauthor social networks," *Computational Statistics and Data Analysis*, 52, 2177-2184, 2008; doi:10.1016/j.csda.2007.07.021, 2007' *NOTE: Online Aug 2007, hence placed in 2007 on charts.*

This was [SAI2008] in earlier reports. Authors have sometimes cited it as:

"Style of author-coauthor social networks"[254] *but the published title was:*

"Social networks of author–coauthor relationships."

DC's key discovery[255] led to much discussion [MAS2010a §W5.6.1]. P179 tried to convert WR's SNA attack on climate science peer review into a peer-reviewed paper [MAS2010a §A.1.4]. **P178** and P179 are highlighted red, as part of the attacks on the hockey-stick and climate science.

It acknowledged funding from the 3 Federal grants, non relevant:

'The work of Dr. Yasmin Said was supported in part by the National Institutes on Alcohol Abuse and Alcoholism under grant 1 F32 AA01**5876**-01A1. The work of Dr. Edward Wegman was supported in part by the Army Research Office under contract W911NF-04-1-**0447**. The work of Dr. Said and Dr. Wegman was also supported in part by the Army Research Laboratory under contract W911NF-07-1-**0059**.'

The claim of peer review was also false. Wegman sent it to Editor-in-Chief[256] Stanley Azen, a 20-year Wegman associate lacking SNA publications, for a journal that did not generally publish SNA. He briefly reviewed it himself and accepted it in a few days [MAS2011a §5]. Elsevier forced retraction in May 2011 for plagiarism,[257] over Wegman's request to let them just add citations.

Bradley's May 2010 complaint to GMU [MAS2011 §A.2.4] mentioned the funding issue on this paper.[258] Provost Stearns recognized it [VER2012], but the new FOIAs give evidence that GMU broke the reporting rules, **§5.**

P180 Sw①② 5/0 {?}

'Said, Yasmin H. and Wegman, Edward J. (2007) "Quantitative assessments of alcohol-related outcomes," *Chance*, 20(3), 17-25'

Said asked for the 3 key grants to be ack'd by email, **§S.5**-08/31/07. *Chance*[259] is peer-reviewed, but is not a research journal:

'CHANCE is not a technical magazine, but rather a cultural record of an evolving field, intended to entertain as well as inform.'

P181 W①[260] 1/1

'**Wegman, Edward J**. and Martinez, Wendy L. (2007) "A conversation with Dorothy Gilford," *Statistical Science*, 22(2), 291-300'

This was a transcript of an interview.

[252] eum.unimc.it/catalogo/catalogo-2007/classification-and-data-analysis-2007 eum.unimc.it/catalogo/catalogo-2007/classification-and-data-analysis-2007 cladag2007.unimc.it/ClaDAG-2007-Programme.pdf
Springer, actual papers seem unavailable for online purchase. A 2010 book says that it has revised versions of papers, but the paper there is P200 [SAI2010], not P178, so is quite different. searchworks.stanford.edu/view/9113791

[253] *Surprisingly*, although it used SNA incorrectly and has been retracted, this paper got the most non-coauthor citations of any claimed papers examined.

[254] JSM2007 had a paper of that title, P183, by Sharabati, Said, Wegman, p.306. www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf

[255] deepclimate.org/2010/04/22/wegman-and-saids-social-network-sources-more-dubious-scholarship [called DEE2010f in MAS2010]

[256] Sometime between October 2012 and April 2013, he stepped down as Editor. www.webcitation.org/6FiPDmnOY

[257] www.sciencedirect.com/science/article/pii/S0167947307002861

[258] As did [MAS2010a §A.7] and [MAS2012a, MAS2012c].

[259] chance.amstat.org/about
Article itself was unfound, but ack was given to *Chance* by Said in **§S.5.**

[260] arxiv.org/pdf/0710.4768.pdf This history of her and ONR was interesting, and worth having. "Beaches in Rio" was amusing. It was worthwhile for the profession, but had nothing to do with 0447.

P182 ws① [261] ?/?

'Alnoshan, Abdullah, Rotenstreich, Shmuel, **Wegman, Edward, Said, Yasmin** and Rajput, Adil (2007) "Microeconomic approach to resource allocation in P2P grids," Proceedings of the Joint Statistical Meetings, 1975-1980' *(July29-August 2, 2007, Salt Lake City)*

Oddities were discussed in 2011.[262] The JSM2007 Abstract Book calls this "Grid Computing."

P183 ws① [263] ?/? {P183, P200}

Sharabati, Walid K., **Wegman, Edward J**. and **Said, Yasmin H**. (2007) "A model of preferential attachments for emerging scientific subfields," Proceedings of the Joint Statistical Meetings, 2048-2055
(July29-August 2, 2007, Salt Lake City)

This citation seems to be an earlier conference presentation of P200, and is shown as Wegman claimed it in 0447, but, the JSM book says:

Walid Sharabati, **Yasmin H. Said**, **Edward Wegman**,
"Style of Author-Coauthorship Social Networks: Statisticians of Prominent U.S. Universities"

'In the past 20 years social networks have been used to analyze relations and ties among individuals of the same network and similarities between different networks in an attempt to obtain a better understanding on how societies interact. One of the applications of social networks is the author coauthor networks also known as the citation networks. This branch of social networks tries to answer the question of "who-wrote-with-who" and with what frequency. It also investigates other important features such as cliques, structural equivalence, MDS and CONCOR clustering. Social networks can be treated as directed graphs in which actors (individuals) are represented by vertices (nodes) while interactions between actors are represented by edges (ties) which may have weights. In this paper, we study in depth coauthorship social networks of statisticians from prominent U.S. universities.'

People sometimes used this title for P179.

P184 Sw① [264] ?/?

'**Said, Yasmin H**. and **Wegman, Edward J**. (2007) "Restrictions of trans fatty acids: Health benefits and economic impact in the Washington, DC Metro Area," Proceedings of the Joint Statistical Meetings, 1523-1527'
(July29-August 2, 2007, Salt Lake City)

This was another JSM2007 presentation, but *seems a false citation,* as the JSM2007 Abstract Book showed:

'Restriction of Trans Fatty Acids in Washington, DC, Metropolitan Area: Health Benefits and Economic Impact **Joost Joost Bottenbley**, Johns Hopkins University; **Samer Ellahham**, Innovative Medical Institute/Paragon Cardiovascular Foundation; Yasmin H. **Said,** George Mason University; Edward **Wegman**, George Mason University; **Malda Bashi**, Paragon Cardiovascular Foundation'

If this was actually related to 5876, it might have been highlighted green.
See also T407, in same session.

P185 sw① [265] 1/0 {⓭}

'Mburu, Peter K**., Said, Yasmin H**. and **Wegman, Edward J.** (2007) "Temporal statistics for consequences of alcohol use," Proceedings of the Joint Statistical Meetings, 2005-*2009' (July29-August 2, 2007, Salt Lake City)*

This was yet another JSM2007 presentation. The abstract was[266]:

'Time-of-day, day-of-week, and month-of-year statistics are important for modeling the acute consequences of alcohol use and abuse. These consequences are not homogeneous through the year, although that assumption is often made when planning intervention strategies. In this paper we consider data from the Virginia Department of Motor Vehicles concerning alcohol-related fatal crashes, also Fairfax County, Virginia DWI arrests and finally Fairfax County, Virginia alcohol treatment admission data. We investigate the cyclic effects over the period 2000–2005.'

Mburu was mentioned in the 0059 progress report as working on that, **§R.3**, but did not appear in the Final Report, **§R.4**.
He was ack'd in P194, T417 and T419, and in the last, Said wrote:

'I would like to acknowledge the contributions of Dr. Rida Moustafa and **my students, Peter Mburu and Walid Sharabati**, who assisted me with the development of various figures and computations in this paper.'

Said officially co-supervised Sharabati, but her role with regard to Mburu was unclear.

[261] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.54
GS had no mentions.
[262] deepclimate.org/2010/12/23/george-mason-universitys-endless-inquiry/#comment-7137 and later comments by Ted Kirkpatrick and DC.
[263] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.304
[264] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.21
[265] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.116
[266] Automobile topic *seems* close, not identical to Sharabati's [SHA2008 §5.4].

P186 Sw①[267] 0/0 {⓭}

'Said, Yasmin H. and Wegman, Edward J. (2008) "Using administrative data to estimate cyclic effects of alcohol usage (refereed abstract)," *Alcoholism: Clinical and Experimental Research*, 32(6) Supplement, 139A.' Washington.

Abstract (#516, p.129):

'Cyclic and geospatial effects in acute consequences of alcohol use play a crucial role in determining intervention policy issues. Knowing when and where to place such interventions can substantially mitigate the negative consequences. In this paper we use administrative records in Virginia for fatal crashes, DWI arrests and alcohol-related hospital admissions as well as records related to distribution of distilled spirits sales to estimate the daily, weekly, monthly, annual and geospatial intensity of alcohol use and, by extension, the prevalence of acute outcomes. We develop an estimation procedure for a comprehensive modeling framework based on agent-based simulation, social networks, and directed graphs that captures some of the complexity of the alcohol ecological system. We provide a framework for temporal and geospatial effects and discuss proxies for approximating these effects. To completely specify the framework, considerable additional data collection efforts must be undertaken. Nonetheless, with opportunistically available data sources, we are able to reproduce with reasonable accuracy the actual experiences for a diversity of alcohol-related acute outcomes in Fairfax County, Virginia. The novelty of the framework that we have proposed lies in the ability to modify the conditional probabilities on the various paths through the directed graph and thus experimentally determine what social interventions are most likely to have a beneficial effect for reducing acute outcomes. Although this model is cast in terms of acute alcohol-related outcomes, there is an **obvious applicability to expanding this framework to other settings such as homeland security and criminal activities.'**

Wegman 0447-claimed P186 and P187 as refereed journal papers, but then said they were refereed abstracts at conference cosponsored by the Research Society on Alcoholism (RsoA). Said called them posters,[268] **§S.3.2.** These were the only papers presented at an actual alcoholism conference, although T138 was presented at a RSoA[269] meeting. Most Wegman and/or Said alcoholism papers were presented at JSM, Interface, COMPSTAT, etc, broad statistics-oriented conferences, generally with minimal peer review.[270] Reviewers for the 5876 proposal expressed concern about lack of connection with the alcoholism research field.[271] They were assured that there would be close work with Wieczorek, but only a single paper was coauthored with Wieczorek, P193, at COMPSTAT. Wegman and/or Said generally did not publish in the alcoholism journals that often published articles by Wieczorek and others, **§S.6**.

P187 Ws①[272] 0/0 {⓭}

'Wegman, Edward J. and Said, Yasmin H. (2008) "Modeling spatiotemporal effects for acute outcomes in an alcohol system (refereed abstract)," *Alcoholism: Clinical and Experimental Research*, 32(6) Supplement, 140A, 2008' Washington DC.

Abstract (#517, p.130):

'Alcohol studies have traditionally focused on specific acute outcomes and intervention strategies for mitigating those particular acute effects. Such approaches often have unintended consequences so that intervening to reduce one acute effect may increase other undesirable outcomes. Previously we have modeled the regional alcohol ecology system as a hybrid social network directed graph model. This model omitted key features including temporal and spatial characteristics as well as gender and age. While adequate for high level overview of the alcohol ecology system, that model omitted key spatiotemporal effects as well as effects related to underage drinking and characterization of sexually transmitted disease. In the present model we consider these additional factors and describe the inference procedures for assessing the probabilities required for the model. Geospatial and temporal effects have been considered separately, but are not traditionally integrated with intervention models. It is difficult to find outlets for modeling innovations within the traditional alcohol literature. In the present paper, we outline a comprehensive modeling framework based on agent-based simulation, social networks, and directed graphs that captures some of the complexity of the alcohol ecological system. We aggregate agents into relatively homogeneous clusters, but provide diversity by stochastic modeling of outcomes. We provide a framework for temporal and geospatial effects and discuss proxies for approximating these effects. We calibrate the model, again using approximations based on actual data.'

[267] onlinelibrary.wiley.com/doi/10.1111/acer.2008.32.issue-s1/issuetoc onlinelibrary.wiley.com/doi/10.1111/j.1530-0277.2008.00685.x-i1/abstract Conference June 27-July 2, 2008 in Washington.

[268] Big conferences have numerous posters, whose acceptance bar tends to be low, so P186 and P187 were reclassified as "f. non-peer conference proceedings."

[269] rsoa.org/about.html Wegman was an elected member of RSoA.

[270] For instance, any attendee could present at JSM, *although I understand that rules have been tightened a bit in the last few years.*

[271] **§S.5,** 12/22/05, p.32, pp. 33-36, pp.36-39 of critique

[272] onlinelibrary.wiley.com/doi/10.1111/j.1530-0277.2008.00685.x-i1/abstract

P188 w☐[273] 2004.05 0/0

'Lin, Chien-Chih, Noh, Eun Young, Yan, Younggping, and **Wegman, Edward J. (2008)** "User profiling in window title and process table," *Computing Science and Statistics*, 36, 530-546'

All but (2008) were correct, as this was presented at Interface 2004[274] May 26-29, 5 months before 0447 started.[275] The ack was:

'The authors would like to thank Dr. Ryszard Michalski, Dr. Kenneth Kaufman and Mr.Jarek Pietrzykowski of the Machine Learning and Inference Laboratory for the use of the data set and preprocessing applications in the system.

P189 w☐[276] 2004.05 0/0

'Martinez, Wendy L., Martinez, Angel R. and **Wegman, Edward J**. (2008) "Classification and clustering using weighted text proximity matrices," *Computing Science and Statistics*, 36, 600-611'

This was mis-cited in same way, (2008) instead of (2004). The ack was:

'The first author would like to acknowledge the support of the Office of Naval Research ROPO program. The **third author** acknowledges the support of DARPA through a contract to AlgoTek, Inc.'

[273] www.interfacesymposia.org/I04www.interfacesymposia.org/I04/I2004Proceedings/LinChien-Chih/LinChien-Chih.paper.pdf
Computing Science and Statistics is proceedings of Interface conferences, not a journal, and not peer-reviewed conference.

[274] www.interfacesymposia.org/I04

[275] web.archive.org/web/20081227205633/www.galaxy.gmu.edu/stats/faculty/wegman.resume2.htm
has older resumes that grew to "Prepared September 2007," whose list ended at P180. P181-P193 were added in a PDF "Prepared September 2008." web.archive.org/web/20081010125010/www.galaxy.gmu.edu/stats/faculty/wegman.resume2.pdf *It seems that he had forgotten about these, then added them in 2008, consistent with his numbering.* His resume labeled other authors as "with," "User profiling in window title and process table," with Chien-Chih Lin, Eun Young Noh, Younggping Yan, *Computing Science and Statistics*, 36, 530-546, 2008.' Hence, either the transfer to 0447 occurred from a different standard-format list or required Wegman to edit the resume list.

[276] www.interfacesymposia.org/I04

P190 w☐[277] 2004.05 ?/?

'Solka, Jeffrey L., Bryant, Avory C. and **Wegman, Edward J.** (2008) "Identifying cross corpora document associations via minimal spanning trees," *Computing Science and Statistics*, 36, 952-961'

This was also mis-cited as (2008) instead of (2004). The ack was:

'The authors would like to acknowledge the support of the Defense Advanced Projects Research Agency.'

Although this was a predecessor to **P170**, it was not grouped with {**P170**, T125}, because the latter were not just repetitions, but examples of normal progression of research.

P191 w☐[278] 5/5 0447-relevant, maybe

'Reyen, Salem S., Miller, John J. and **Wegman, Edward J.** (2008) "Separating a mixture of two normals with proportional covariances," *Metrika*, doi:10.1007/s00184-008-0193-4'

This seemed oddly sequenced, as Reyen got his PhD[279] in Fall 2004, with Miller as Director and Wegman as Co-Director. The paper properly cited the dissertation and *seemed* an extension of that work, which *seemed* vaguely connected with 0447.[280] This is shown in chronology here as 2005, not 2008, given:

'**Received: 14 November 2005** / Published online 22 August 2008'

The acks were:

'Salem S. Reyen was supported by the Defense Advanced Research Project Agency through cooperative agreement 8105-4**8267** with the John Hopkins University.' *(p.297)* and Authors gratefully thank Carey E. Priebe for providing us the hyperspectral data set and Mark Werner for sharing his FORTRAN code of RMD.' *(p.310)*

This was the most difficult paper to date sensibly. People might argue about 2005 versus 2008, and whether or not the ☐ tag makes sense.

[277] www.interfacesymposia.org/I04
binf.gmu.edu/jsolka/PAPERS/serendipity_int04_paper.pdf
The title was misspelled "Copora" but that may have been a draft.

[278] link.springer.com/article/10.1007%2Fs00184-008-0193-4?LI=true#page-1

[279] citeseerx.ist.psu.edu doi=10.1.1.194.4168.pdf

[280] I examined the paper and it might posisbly fall udner the "multidimensional" estimation, but did not try to study this carefully.

P192 Ws①[281]=T146 W 2/0 {⓭}

'**Wegman, Edward J**. and **Said, Yasmin H.** (2008) "A directed graph model of ecological alcohol systems incorporating spatiotemporal effects," COMPSTAT 2008, (Paula Brito, ed.), 179-190 (Porto, Portugal)'

'Users of alcohol are incorporated into a societal system, which for many purposes resembles an ecological system. We have previously modeled such systems using a directed graph with acute outcomes reflecting undesirable individual and societal outcomes. In this paper, we expand the model to a hybrid social network directed graph model for modeling the acute outcomes associated with alcohol use and abuse. We describe the approximate estimates of conditional probabilities based on available data. In the present model, we also approximate geospatial effects related to transportation as well as temporal effects. Acute outcomes include assault, murder, suicide, sexual assault, infection with STDs or HIV, domestic violence, child abuse, and DWI and alcohol related fatal crashes. The model is calibrated using demographic, crime, alcohol-related, and alcohol outlet data from Virginia and Fairfax County in Virginia. We propose proxy data and methods for capturing temporal and geospatial effects. The goal is to investigate methods for simultaneous suppression of multiple negative public health consequences. The model may be used as a public policy decision tool by adjusting conditional probabilities in order to investigate the effect of interventions.'

Acknowledgement:

'The work of Dr. Wegman is supported in part by the U.S. Army Research Office under contract W911NF-04-1-**0447**. The work of Dr. Said is supported in part by Grant Number F32AA01**5876** from the National Institute On Alcohol Abuse And Alcoholism.'

P193 sw①[282] 0/0[283] {⓮}

'Wieczorek, William F., **Said, Yasmin H.** and **Wegman, Edward J.** (2008) "Spatial and computational models of alcohol use and problems," COMPSTAT 2008, (Paula Brito, ed.), 191-202'

Abstract:[284]

'This paper focuses on multivariate and computational approaches that are being developed in the alcohol field. There is substantial monetary support for conducting alcohol research. Alcohol use and problems are complex behaviors by individuals, across their life spans, while embedded in a number of social and economic networks. This complexity, coupled with the research support primarily from the National Institutes of Health (NIH), has led to numerous data collection and research projects, many of which require sophisticated multivariate and spatial statistical approaches. Some of the methods used to model alcohol use and problems are latent growth curves, multilevel models, and latent class analysis. These techniques allow for the examination and modeling of both individual and group level factors. However, these types of models are not suitable for mining large data sets. In this paper, we exploit **regional data in Erie County, NY** to illustrate the use of multivariate and spatial analysis tools in alcohol studies.'

This paper modeled alcoholism in Erie County around Wieczorek's Buffalo, NY location. Wieczorek was already quite well-published on health, addiction and alcoholism **(§S.6)** and a member of Said's dissertation committee. Said's revised 5876 application promised close cooperation and mentoring with him, as earlier reviews had expressed concerns the need for such, **§S.5**-12/22/05. This was the 3rd paper in a session with P192 and T419 and Wegman gave a same-titled talk earlier, T418. Google Scholar lists only this single paper in which Said or Wegman coauthored with Wieczorek.[285] The ack (p.189) was:

'The work of Dr. Wegman is supported in part by the U.S. Army Research Office under contract W911NF-04-1-**0447**. The work of Dr. Said is supported in part by Grant Number F32AA01**5876** from the National Institute on Alcohol Abuse and Alcoholism.'

This analysis used parallel coordinate displays (on which Wegman had often written) and conditioned choropleth maps (CCmaps), **but not Said's software on which so many of the others works were based.**
This may be fine or it may be a negative indicator.[286]

[281] "Paper 192\COMPSTAT2008_paper_Wegman_Said_Final.pdf" www.documentcloud.org/documents/550099-compstat2008-paper-wegman-said-final.html

[282] link.springer.com/content/pdf/10.1007%2F978-3-7908-2084-3_16

[283] Wieczorek has published much in this field, but even he did not seem to cite this. Of course, GS can easily miss cites.

[284] www.zentralblatt-math.org/portal/en/zmath/search/?q=ai:wieczorek.william-f

[285] *Given the long association, this seems odd. Is unclear who presented this, as COMPSTAT was a common venue for Wegman and Said, but not Wieczorek.*

[286] It is easy to construct computing models that can generate results, but the models need to be validated against the real world, and tested to see if they are actually useful for decision-making. It seems that Said had a model of Northern VA, but it is unclear if it applied elsewhere or was ever actually used for anything.

Except for P200, ~~P301~~ **and** ~~P302~~**, remaining papers were not 0447-claimed, although some did ack 0447.**[287] No attempt has been made to classify these into 0447 categories. Wegman's resume listed papers as "with" other authors, but the correctly-ordered author lists are given here. Post-0447 papers were examined, in case there had been substantial fit papers that had long publishing delays and appeared later, but little was found.

P194 Sw①[288] 0/0 {⓭}

'**Said, Yasmin H.** and **Wegman, Edward J.** (2009) "Estimating cyclic and geospatial effects of alcohol usage in a social network directed graph model," in *Social Computing and Behavioral Modeling*, (H. Liu, J. Salerno, M. Young, eds.), 172-179, 2009'[289] Abstract:

'We would like to acknowledge the contributions of Dr. Rida Moustafa and **our students Peter Mburu and Walid Sharabati**, who assisted us with the development of various figures and computations in this paper. The work of Dr. Said is supported in part by Grant Number F32AA01**5876** from the National Institute on Alcohol Abuse and Alcoholism. The work of Dr. Wegman is supported in part by the U.S. Army Research Office under contract W911NF-04-1-**0447**.'

The 2nd International Workshop on Social Computing, Behavioral Modeling, and Prediction[290] was in Phoenix AZ 03/21/09-04/01/09.[291] This 2-pager was in a poster board session.[292]

P195 Ws①[293] 1/0 {⓭}

'**Wegman, Edward J.** and **Said, Yasmin H.** (2009) "A social network model of alcohol behaviors," in *Social Computing and Behavioral Modeling*, (H. Liu, J. Salerno, M. Young, eds.), 236-243, 2009'

'The work of Dr. Wegman is supported in part by the U.S. Army Research Office under contract W911NF-04-1-**0447**. The work of Dr. Said is supported in part by Grant Number F32AA01**5876** from the National Institute on Alcohol Abuse and Alcoholism.'

This workshop was in Phoenix AZ 03/21/09-04/01/09. The poster board was 2 slides,[294] of which the 2nd corresponds to ***P192*** Fig.1 and Fig.3.

P196 w[295] 7/6 0447-relevant

'Caudle, Kyle A. and **Wegman, Edward J**. (2009) "Nonparametric density estimation of streaming data using orthogonal series," *Computational Statistics and Data Analysis*, doi:10.1016/j.csda.2009.06.014 , 2009' pp.3980-3986.

This *CSDA* paper had the same title as Caudle's dissertation(2005), and contained no acks. **This was relevant to the 0447 proposal,** but was submitted in February 2009 (9 months after 0447 ended), revised June 2009 and published then.

[287] While it makes no sense to 0447-claim works done before 0447, given nature of research and lag times, 0447-acks of fit work after the contract seem reasonable.

[288] www.google.com/search?tbm=bks&hl=en&q=%22Estimating+cyclic+and+geospatial+effects+of+alcohol+usage+in+a+social+network+directed+graph+model%22&btnG= (Search for Wegman, p.242).

[289] www.springer.com/computer/database+management+%26+information+retrieval/book/978-1-4419-0055-5
link.springer.com/content/pdf/bfm%3A978-1-4419-0056-2%2F1

[290] sbp.asu.edu

[291] www.public.asu.edu/~huanliu/sbp09/program.html

[292] www.public.asu.edu/~huanliu/sbp09/Presentations/poster%20presentations/09-SaidEstimating%20Cyclic%20and%20Geospatial%20Effects%20of%20Alcohol%20Usage.pdf

[293] books.google.com/books?id=AR6lxjZmeUkC&pg=PA242&lpg=PA242&dq=%E2%80%9CA+social+network+model+of+alcohol+behaviors%22++social+computing+behavior+modeling&source=bl&ots=NnIHG54Lz7&sig=4KXTsEtPrUz2xJ8oiFBvo7Xp548&hl=en&sa=X&ei=TePCUNH6Ks6ujAKDlIDQAg&ved=0CDkQ6AEwAg#v=onepage&q=%E2%80%9CA%20social%20network%20model%20of%20alcohol%20behaviors%22%20%20social%20computing%20behavior%20modeling&f=false

[294] www.public.asu.edu/~huanliu/sbp09/Presentations/poster%20presentations/01-WegmanA%20Social%20Network%20Model%20of%20Alcohol%20Behaviors.pdf

[295] www.sciencedirect.com/science/article/pii/S0167947309002394
"ARMY2010_IDSVIS\PDFs\COMSTAT_4421.pdf" 07/05/09
www.documentcloud.org/documents/550098-comsta-4421.html

P197 Sw☆[296] 1/1

'Said, Yasmin H. and **Wegman, Edward J.** (2009) "Roadmap for optimization," *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1): 3-11, DOI: 10.1002/wics.16, 2009, online 07/13/09.'

This was mostly assembled by mosaic plagiarism[297] of Wikipedia pages, in a Wiley journal edited by Wegman, Said and Scott.

After formal complaints were filed with Wiley on it and P513, the original was quietly revised, not retracted [MAS2012c, §4.8]. A few months' later, they quietly disappeared from the masthead. In late 2012 or early 2013, several well-known statisticians replaced them as Editors-in-Chief.

P198 ws[298] 0/0

'Sharabati, W.K., **Wegman, E.J**. and **Said, Y.H.** (2009) "Multi-mode social networks" Proceedings of the Human Behavior-Computational Modeling and Interoperability Conference (CD), Oak Ridge, TN: ORNL.' *June 23-24.*[299]

The actual program had:[300]

'Multi-Modal Social Networks
Edward Wegman (presenter) Yasmin Said and Walid Sharabati'

P199 Sw[301] 0/0 {❹}

'**Said, Yasmin H**. and **Wegman, Edward J.** (2009) "Preserving semantic content in text mining using multigrams," Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences (CD), 2009, Cairo. ICCS-X, 12/23-25 2009 pp.963-976'

Acknowledgement:

'We would like to acknowledge the contributions of several former students: Dr. Walid Sharabati who provided the examples for the bigram-document and bigram-bigram documents; Dr. Faleh Alshameri who provided the document clustering material; Dr. Jeffrey Solka who provided the CLUTO and SPACETREE illustrations; and Dr. Angel Martinez who provided the data for all of these examples. We are in debt to the support provided by the Isaac Newton Institute for Mathematical Science at the University of Cambridge in Cambridge, England, which has created the opportunity to formulate this work. We have filed a patent disclosure on certain aspects of this work.'

[296] onlinelibrary.wiley.com/doi/10.1002/wics.v1:1/issuetoc
onlinelibrary.wiley.com/doi/10.1002/wics.16/abstract

[297] This article was filled with errors obvious to anyone with the slightest background, proving that WIREs:CS had zero semblance of real peer review. In addition, *it seemed a strange attempt* to subsume under computational statistics a large part of operations research, a field with a long, deep history of its own.

[298] www.csiir.ornl.gov/HBIOC
Kathleen Carley of Carnegie Mellon was a keynote speaker, [MAS2011a, p.7]

[299] The correct name was "HUMAN BEHAVIOR-COMPUTATIONAL INTELLIGENCE MODELING CONFERENCE 2009"

[300] csiir.ornl.gov/HBIOC/program.shtml

[301] www.scribd.com/doc/56160595/Volume-2-conference-ICCS-X
www.scribd.com/doc/56160595/18/PRESERVING-SEMANTIC-CONTENT-IN-TEXT-MINING-USING-MULTIGRAMS

P200 Sw①②③☆[302] 0/0 {P183, P200} [SAI2010]

'**Said, Yasmin H**., **Wegman, Edward J**. and Sharabati, Walid K. (2008) "Author-coauthor social networks and emerging scientific subfields," to appear Data Analysis and Classification: From the Exploratory to the Confirmatory Approach, (Carlo Lauro, Francesco Palumbo, Michael Greenacre eds.) Berlin: Springer-Verlag ' *This actually was published in 2010.*

The author order was "Said, Wegman and Sharabati" although almost all text originated with Sharabati's dissertation [SHA2008], including plagiarism of SNA and other text.[303]

'The work of Dr. Said is supported in part by Grant Number F32AA01**5876** from the National Institute on Alcohol Abuse and Alcoholism. The work of Dr. Wegman is supported in part by the Army Research Office under contract W911NF-04-1-**0447**. Both were also supported in part by the Army Research Laboratory under contract W911NF-07-1-**0059.**'

The text was almost entirely from Sharabati, so how did Said and Wegman come to be ahead of him in author list?

P201 ws[304] 0

'Sharabati, Walid K., Wegman, Edward J., Said, Yasmin H. (2010) "A ten-mode social network of Afghan society" to appear *International Journal of Computational Intelligence*, 2010'

India-based Serials Publications[305] publication categories cover a wide range including astrology, homeopathy and occult sciences, [306] but IJCS does not list SNA among its many topics.[307]

The following did not appear in the Wegman resume, either because they did not involve Wegman, followed the February 2010 date, or seemed never to have been published.

The first pair seemed never to have been published, hinting that they tried to publish in peer-reviewed journals.

~~P301~~ Ws 0/0 (necessarily)

This was 0447-claimed, but unpublished:

'**Wegman, Edward J.** and **Said, Yasmin H**. (2008) "Text mining with application to fraud discovery," submitted to *Applied Stochastic Models in Business and Industry*'

I could find no trace of it in that Wiley journal.[308]

It might or might not have fit 0447, but was given benefit of doubt.

They seemed to have trouble getting papers published in independent peer-reviewed research journals, as seen in the next as well. Of course, the only way of finding such is to see the notation "submitted."[309]

~~P302~~ Sw① 0/0 (since unpublished)

'**Said, Yasmin H.** and **Wegman, Edward J**. (2008) "Agent-based simulation of the alcohol ecological system," submitted *Journal of the American Statistical Association*'

This was 0447-claimed, but also unpublished: Said may have claimed it, but since NIAAA redacted "not yet published" papers, yet more papers may have been submitted, but rejected.

[302] link.springer.com/chapter/10.1007%2F978-3-642-03739-9_30?LI=true "Author–Coauthor Social Network and Emerging Scientific Subfields," F. Palumbo et al. (eds.), *Data Analysis and Classification*, Studies in Classification, Data Analysis, and Knowledge Organization, DOI 10.1007/978-3-642-03739-9_30, ©Springer-Verlag Berlin Heidelberg 2010, pp.257-268. searchworks.stanford.edu/view/9113791

[303] www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.W.5.7.pdf

[304] www.serialspublications.com/contentnormal.asp?jid=235&jtype=1 www.webcitation.org/6DD2pL2Ni

[305] www.serialspublications.com/aboutus.asp

[306] www.serialspublications.com/categories.asp www.webcitation.org/6DD6cgvqh

[307] www.serialspublications.com/journals1.asp?jid=235&dtype=1&jtype=1 www.webcitation.org/6DD3HPQ9E

[308] onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291526-4025

[309] No allegation of wrongdoing is implied – most researchers get rejections. However, the lack of peer-reviewed publications, and hints of rejected attempts raise concerns. They tried to publish, but did not do well.

The following 3 works were cited by Said in her grant proposal, §S.1.

P401 S☆[310] 0/0 {⓾}

"Agent-based simulation of ecological alcohol systems" Dissertation May'05

Plagiarism allegations are summarized in [MAS2012c, §3.4].

P402 S☆[311] 0/0

"On Genetic Algorithms and Their Application" in Handbook of Statistics, Volume 24: Data Mining and Data Visualization by C.R. Rao, E. J. Wegman and J. L. Solka' *(Jul 5, 2005)*

DC showed plagiarism,[312] in which Said stitched together 3 antecedents and listed this in support of the 5876 proposal, **§S.5**-06/01/05.[313]

P403 S[314] 0/0 (not in original proposal, added in revisions)

"Statistics "in *Oxford Encyclopedia of the Modern World 1750 to the Present,* (2008) Peter N. Stearns, Ed.

Said listed this 2-page article in support of her proposal for 5876, **§S.1.**

'2006 "Statistics "to appear *Encyclopedia of the Modern World*, (Stearns, Peter N., ed.), New York: Oxford University Press.

The 8-volume *Encyclopedia* was published in 2008, not 2006, edited by GMU Provost Stearns [MAS201c §6]. *Wegman was well-qualified to write this, but no Said. See detailed analysis in* **§Y.**

The remaining works were collected from a variety of sources.

P404 S[315] 4/4

'Banks, D.,[316] and **Said, Y. H.** (2006). "Data Mining in Electronic Commerce," *Statistical Science*, 21, 234-246.' *(May 2006 publication)*

P405 S③☆[317] 0/0

'**Yasmin H. Said**, "On the Eras in the History of Statistics and Data Analysis" *J. Washington Academy of Science* 93(1) Spring 2007, 17-35.

This was not only 5876-unfit, but **§X** alleges substantial plagiarism.

P406 s[318] 0/0

'Banks, D., and **Said, Y**. (2007). "New Issues in Human Rights Statistics," in *Statistical Methods for Human Rights*, ed. by J. Asher, D. Banks, and F. Scheuren, Springer: NewYork, pp. 227–240.'

~~P407~~ Ws☆★ 0/0 (unpublished) details in **§H.7**

This Wiley book was supposed to be published December 2007 and is still listed on Amazon. It was supposed to include the full WR text, which would raise both plagiarism and copyright issues. The description of the book has a false statement (★), similar to that in T135 Ws①★. Wegman repeatedly and falsely portrayed the 20th-century measured temperatures as an artifact of the reconstruction.

~~P408~~ Ws[319] (unpublished)

'Edward Wegman (Editor), David W. Scott (Editor), Yasmin H. Said (Editor), Jeffrey L. Solka (Editor), *Encyclopedia of Computational Statistics, 2009/2010*

This effort likely evolved into WIREs:CS, cited by Said in **§S.3.2** as:

'9. Wegman, Edward J., Said, Yasmin H .. Scott, David W. (2009) Encyclopedia of Computational Statistics, New York: John Wiley and Sons.'

[310] www.galaxy.gmu.edu/stats/syllabi/IT871/MasterCopyDissertation.pdf was the original location of a file that disappeared August 2010, *which might be related to "terry" 08/03/10 at 9:26am, who identified some problems:* deepclimate.org/2010/08/03/what-have-wegman-and-said-done-lately/#comment-4755 and August 4 2:10 am comment. A copy is: web.archive.org/web/20060905150733/http:/www.galaxy.gmu.edu/stats/syllabi/IT871/MasterCopyDissertation.pdf See [MAS2010a §A.11].

[311] Rao, Wegman, Solka, Eds, ***Handbook of Statistics:*** *Data Mining and Data Visualization,* 2005. books.google.ca/books?id=fEgUjUPCtsEC "Key Features: Distinguished contributors who are international experts in aspects of data mining" Given the date, Said wrote this before she finished PhD defense.

[312] deepclimate.org/2011/06/07/mining-new-depths-in-scholarship-part-1 deepclimate.files.wordpress.com/2011/06/said-genetic-algorithms-v10.pdf

[313] She took a course in 2003 on this topic, but *further qualification is not obvious.*

[314] www.amazon.com/Oxford-Encyclopedia-Modern-World-Present/dp/B007MXUUSU

[315] projecteuclid.org/euclid.ss/1154979824 Refereed journal.

[316] Banks was a WR "reviewer" claimed not to be in Wegman's social network, despite many other contacts [MAS2010 §A.1.5, §A.6.3], at least. www.stat.duke.edu/~banks/banksvitae.pdf

[317] www.washacadsci.org/scans/V.93-n.1.pdf

[318] books.google.com/books?id=IGD1PpypHwkC www.springer.com/statistics/social+sciences+%26+law/book/978-0-387-72836-0 Preview: link.springer.com/book/10.1007/978-0-387-72837-7/page/1 Banks' CV says 2007, so presumably the work was done then.

[319] www.amazon.com/Encyclopedia-Computational-Statistics-Edward-Wegman/dp/0470054077 October 2010 unpublished $1,500 book

~~P409~~**W** 02/16/09 This rejected proposal is covered in **§Z.**
Edward Wegman, Proposal to Army Research Office
'Mathematical and Statistical Foundations of Networks'

Summary of papers fit for 0447 (or 0059)
Paper publication lags actual work, but **§0.3** shows graphically how Wegman focused on 0447-unfit work, starting in 2005, but increasingly in 2006-2008, during which 0447-relevant peer-reviewed journal papers essentially disappeared. Post-0447 years were examined for fit work possibly in the pipeline and published later, but little was found.

From Google Scholar, no more papers were found 2009-present for Wegman or Said beyond those listed.[320] As seen graphically in **§0.1**, and in next section, papers 2009-2011 were limited to:
2009 ***P196***w, P197 Sw☆, P198 ws, P199 Sw
2010 **P201** ws, P505 w, P506 Ws,
2011 P513Ws☆
Thus 4 papers were in their own journal, and 2 had plagiarism issues.

To summarize all the 0447/0059 fit papers (darker blue) from 2006-on:
P172 w□ was book chapter with Moustafa, related to 2001 dissertation.
P173 **w□** was clearly peer-reviewed by *CSDA*, but mostly done before 0447, with only modest revisions, so graphed as 2005. It is included here for completeness, in case people think it should be 2006.
P175 W was a conference paper related to Caudle's dissertation, seemingly done during the grant and **actually relevant to 0447 abstract on p.4.**
P191**w□** had a long publishing delay, by student Reyen for whom Wegman was 2nd supervisor, in a *topic at best slightly-related.*
P196 w was another paper with Caudle, *likely* peer-reviewed by *CSDA*, but **relevant to 0447 abstract.** It was submitted 9 months after 0447.
P199 Sw was a conference paper in Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences, with unknown review.
People can reread the summary of the proposals on p.4 here, or the actual proposals, to assess the results. Good results would not cancel the numerous false claims, *but their sparsity adds evidence that Wegman really was spending much of his grant-relevant time on unfit works.*

[320] Except 1-page editorial introductions to WIREs:CS issues, ignored.

Summary of works relevant to 5876.
P180 Sw①② was the only fit peer-reviewed journal paper, but *Chance* is not a research journal, **§S.6.**

Summary of works where Said's authorship seemed *odd*.[321]
P182 ws①, P184 Sw①, T408 s③, T410 sw③, T412 sw③ , T413 sw③

In the following cases, Said moved to lead author or at least ahead of the student who seemed earlier to have done the work:
P200 Sw①②③☆, T406

Citation analysis.
While the citation analysis was very quick and should not be over-interpreted, the evidence indicates that:
- Peer-reviewed journal productivity evaporated from 2006-onward.
- Citation counts dropped, and the alcoholism papers especially seemed to be ignored by the field.
- The highest citation counts found were for P179 and T126, about 50 and 60 citations, respectively.
  This seems ironic, given the numerous problems identified there.[322]

Conclusion.
Hardly anyone outside Wegman's network cited any of the work derived from grants totaling ~$492K. Very little fit work was ever published in peer-reviewed research journals. *I cannot comment on the quality of work, and* ***low-impact work alone is no allegation of wrongdoing.***[323]

However, this history shows that few people seemed to care about this work. Statisticians who actually spend enough time to learn application areas and engage with their practitioners can bring valuable skills, but this did not seem to happen.

[321] See individual descriptions.
[322] Although beyond the scope of this report, it might be interesting to see how many cited these papers to criticize them and how many used them as support.
[323] After all, most papers are not cited very often.

K.4.2 Papers in Wiley *WIREs:CS*

No allegations of grant fraud depend on this section, provided for context, as Wegman and Said *seemed* to devote much effort to this by 2009.
Concerns have been raised here, **§J,** about Wegman's interpretations of:

Pro Bono	[MAS2010a §A.7]
Plagiarism	[MAS2012c, MAS2010a]
Peer Review	[MAS2010a §A.1]

Wegman and Said persistently attacked the quality of peer review in paleoclimate, in most[324] of the following, **§0.1:**

2006:	T123, T124, T126, T127, T128
2007	P178, P179, T135, T401, T414
2008	T148 (the effort slowed, until "Climategate" in 2009)
2009	T423 CNN "Climategate" Interview [MAS2010a §A.3]
2010	T424, T426, T427, T428, T429

This section raises more concerns *of possible untrustworthy or non-existent peer review*[325] in *WIREs:CS*, **§H.6.**[326] Sampling the overall mix of authors, many seemed to be experts in their topics. *Still, it may be a concern when many (invited) papers were written by close associates,* [MAS2010a §A.6.1].[327] **Under no circumstances should any wrongdoing be attributed to any of the other authors.**
Concern is for peer review quality, editorial process and Wiley's total unwillingness to retract plagiarized papers. *None engender confidence.*

[324] Not in T141, Wegman's talk at NCAR for experts. Personal criticisms of paleoclimate peer review *seemed* to be given mostly to non-expert audiences.
[325] Readers need to trust a journal for **consistent peer review** to weed out obvious problems and avoid any hint that editors simply accept papers from associates. Perhaps most papers are fine, but if readers cannot trust quality control, it causes doubt, especially in a interdisciplinary review journal. Experts easily assesses work in their own specialties, but need to know that reviews are not only correct, but focus on important topics, rather than pet interests or unproven original dissertation research presented as reviews, as perhaps occurred in P509.
[326] onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291939-0068/issues
[327] [MAS2010 §A.6.3] noted frequent appearance of Wegman associates as speakers at ACAS or Interface. Again, association is not guilt, *but concern might be raised about the selection of speakers for government-funded conferences.*

The following lists papers by Said, Wegman or his students.

P197 Sw☆ *was covered in* ***§K.4.1****, as it was listed in Wegman's* resume.

P501[328] [WEG2010] shows many Grand Tour papers.
Rida E. Moustafa and Ali S. Hadi,[329] "Grand tour and the Andrews plot" *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(2) 245-250 Sep/Oct 2009 DOI: 10.1002/wics.30 online 08/28/09. Abstract:
'The relationship between Andrews plot and Grand Tour (GT) stems from the fact that both view multiprojections of hyperdimensional data. The difference, however, is that GT views all possible projections of the data, while Andrews plot views only sets of projections. In this paper, we give a quick introduction to GT and Andrews plot and some visualization examples. '

P502[330]
Winston C. Chow,[331] "Brownian bridge" *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(3) 325-332 Nov/Dec 2009 DOI: 10.1002/wics.38 online 12/01/09.

P503[332] *Likely* ← P168 w☐
David Marchette, [333] "Class cover catch digraphs" *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2) 171-177 March/April 2010 DOI: 10.1002/wics.70 online 03/03/10
'The class cover problem is one of finding a small number of sets covering (containing) points from one class without covering any points from a second class. The class cover catch digraph provides a solution to the class cover problem, which can be extended to a nonparametric classifier, similar in flavor to a reduced nearest neighbor classifier. This article describes the class cover catch digraph and its application to classification.'

[328] onlinelibrary.wiley.com/doi/10.1002/wics.30/abstract
[329] Moustafa was a PhD student of Wegman's, both are principals with him at dMining Technologies, **§H.3.**
[330] onlinelibrary.wiley.com/doi/10.1002/wics.38/abstract
This had an *odd* author contact (as did P510):
'C/o Dr. Edward J. Wegman, Department of Computational and Data Sciences … Email: Winston C. Chow (winstonchow@gmualumni.org)'
[331] Chow did PhD with Wegman, 1994.
[332] onlinelibrary.wiley.com/doi/10.1002/wics.70/abstract
[333] PhD With Wegman 1996, frequent coauthor.

P504[334] ← ***P163*** **w□** ~ A. R. Martinez dissertation, also might ~{❹}

Angel R. Martinez, [335] "Bigram proximity matrix" *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2) 251-25, March/April 2010 DOI: 10.1002/wics.67 online 03/010/10.'

This might be related to P199, **T143**. Abstract:

'A transformation of the text stream called a bigram proximity matrix (BPM) has been developed. The BPM is used to encode free-form text so computational techniques can be applied to this type of information resource. For example, one could classify the encoded documents using *k* nearest neighbor (*k-NN*) discrimination, group the documents according to their topic, search for latent topics, and more. The hope is that encoding text documents using the BPM will preserve the meaning better than the bag-of-words method commonly used in information retrieval and is easier to execute than other natural language processing applications. In this article, we provide a brief introduction to natural language processing and definitions for some of the concepts. Next, we define the BPM and give some examples of how it has been used. We then provide a weighted variant of the BPM and conclude with extensions to this type of document encoding.'

P505 w[336]

Shores, Roger W**.** and **Wegman, Edward J.** (2010) "Bounds on Delaunay tessellations" *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5) 571-580 Sep/Oct 2010 DOI: 10.1002/wics.119 online 08/25/10.

This seemed related to the Mihai/Youn part of 0059.

'Motivated by applications in density estimation and data compression, this article considers the bounds on the number of tiles in a Delaunay tessellation as a function of both the number of tessellating points and the dimension. Results can also be interpreted for the dual of the Delaunay tessellation, the Voronoi diagram. Several theoretical lower and upper bounds are found in the combinatorics and computational geometry literatures and are brought together in this article. We make a comparison of these bounds with several empirically derived curves based on multivariate uniform and Gaussian generated random tessellating points. The upper bounds are found to be very conservative when compared with the empirically derived number of tiles, often off by many orders of magnitude.'

P506 Ws[337]

Wegman, Edward J. and **Said, Yasmin H.** (2010) "Natural homogeneous coordinates" *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010 2(6) 678–685 DOI: 10.1002/wics.122. Abstract:

'The natural homogeneous coordinate system is the analog of the Cartesian coordinate system for projective geometry. Roughly speaking a projective geometry adds an axiom that parallel lines meet at a point at infinity. This removes the impediment to line-point duality that is found in traditional Euclidean geometry. The natural homogeneous coordinate system is surprisingly useful in a number of applications including computer graphics and statistical data visualization. In this article, we describe the axioms of projective geometry, introduce the formalism of natural homogeneous coordinates, and illustrate their use with four applications.'

P507[338]

Rida E. Moustafa, "Pseudograd tour" *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6) 711-718 Nov/Dec 2010 DOI: 10.1002/wics.133 online 10/13/10.

'The pseudograd tour (PGT) is an approximate version of the grand tour (GT), and an example of the dynamic data visualization methods that assist in exploring hyperdimensional data through a continuous sequence of lower dimensional projections. The PGT adds the time dimensionality to the visualization process which in turn increases the human interaction with the data and reveals comprehensive views of the data. The PGT has some clear advantage suitable for the dynamic data visualization—such as the fast and easy computation, can be constructed using many orthogonal bases (interpolation functions), and the flexibility to visualize the projections as a sequence of scatterplot frames changing over time or a static profile plots of interpolated projections. This article focuses on the implementations of the PGT, the use of interpolation functions, the extensions to higher dimensions, and the scalability to large data points.'

[334] onlinelibrary.wiley.com/doi/10.1002/wics.67/abstract
[335] PhD with Wegman, 2002.
[336] onlinelibrary.wiley.com/doi/10.1002/wics.119/abstract
[337] onlinelibrary.wiley.com/doi/10.1002/wics.122/abstract
[338] onlinelibrary.wiley.com/doi/10.1002/wics.133/abstract

Of the 9 articles in the Jan/Feb 2011 issue of *WIREs:CS*, 2 involved Wegman associates (Scheuren, Szewcyzk) and 3 were by past students.

P508[339] *Likely* predecessors: T128 W= P175 W, ***P196*** w.
Kyle A. Caudle,[340] Gary O. Fowler, Leah R. Jager, David M. Ruth,, "Discounting older data" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(1) 30-33 Jan/Feb 2011 DOI: 10.1002/wics.134 online 10/13/10
'This article describes two general methods for discounting older data in the real-time analysis of a data stream. In the first method, the distribution of a data stream is estimated by a series of orthogonal basis functions, and the coefficients of this estimate are updated as new data arrive by combining windowing and exponential smoothing techniques. The second method involves sequential hypothesis testing. When new data arrive, test significance level is adjusted by alpha-investing, which raises or reduces the significance level of subsequent hypothesis tests on the basis of whether the previous hypothesis test rejects or fails to reject the null hypothesis. Both these methods are nonparametric in nature.'

P509[341] ← Extract from 2009 Dissertation
Hadi Rezazad, "Computer network optimization" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(1) 34-36 Jan/Feb 2011
DOI: 10.1002/wics.135 online 11/02/10
'Computer networks are vital in our everyday lives. It is important to design network configurations with special consideration for their various aspects, such as security, integrity, scalability, and cost. It is especially important for a network to be built as robustly as possible to protect against failures, attacks, and intrusions. In this article, I review methods to assess and improve the robustness and efficiency of computer networks. These methods use computer network analysis, social network analysis, evolutionary computing, statistical methods, and graph theory. Specifically, the aim has been to achieve enhanced network robustness and efficiency with a primary focus on architecture and topology of networks. Metrics have been developed for measuring the robustness and efficiency elements of networks and to construct an evolutionary algorithm for the enhancement of these elements. These methods have been applied to various networks, including random networks, biased networks, and real-life networks. These networks have been analyzed and enhanced using the evolutionary algorithm. Using the metrics, it is shown how the robustness and efficiency of the networks improve. In addition, through this evolutionary process, certain network parameters, as well as the network topological configuration converge.'
This was a lightly-edited version of a dissertation chapter, presented here as an "Advanced Review" that did not cite the dissertation.[342]
[WEG2009] used some definitions, **§Z**.

P510[343]
Wendy L. Martinez,[344] "Computational Statistics in MATLAB®" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(1) 69-74 Jan/Feb 2011
DOI: 10.1002/wics.138 online 11/09/10

The March/April issue of *WIREs:CS* had 10 articles, of which:
3 were by past Wegman students, P511, P512, P514
1 was by Wegman and Said, P513
1 was by Scott, the 3rd Editor-in-Chief

P511[345]
Wendy L. Martinez, "Graphical user interfaces" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2) 119-133 March/April 2011
DOI: 10.1002/wics.150 online 02/23/11

P512[346] *Likely* predecessor may have been **P172** w□
Rida E. Moustafa, "Parallel coordinate and parallel coordinate density plots" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2) 119-133 March/April 2011 DOI: 10.1002/wics.145 online 01/28/11

P513Ws☆[347] See [MAS2012c §4.7] for summary of plagiarism issues.
Wegman, Edward J. and **Said, Yasmin H**. (2011) "Color theory and design" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2) 104-117 March/April 2011, first online 02/04/11.

[339] onlinelibrary.wiley.com/doi/10.1002/wics.134/abstract
[340] PhD 2005.
[341] onlinelibrary.wiley.com/doi/10.1002/wics.135/abstract
[342] www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.W.5.10.pdf
[343] onlinelibrary.wiley.com/doi/10.1002/wics.138/abstract
[344] Wendy (Poston) Martinez did PhD with Wegman, 1995, and was a frequent coauthor 1992-2008
[345] onlinelibrary.wiley.com/doi/10.1002/wics.150/abstract
[346] onlinelibrary.wiley.com/doi/10.1002/wics.145/abstract
[347] onlinelibrary.wiley.com/doi/10.1002/wics.146/abstract

P514[348]

Winston C. Chow, "Fractal (fractional) Brownian motion" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(2) 149-162 March/April 2011 DOI: 10.1002/wics.142 online 01/07/11

P515[349] ← likely related to Reeves(2009) dissertation.

D. M. Reeves,[350] G. M. Jacyna, "Support vector machine regularization" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(3) 204-215 May/June 2011 DOI: 10.1002/wics.149 online 03/08/11

P516[351] [WEG2010] has several papers on this topic.

Rida E. Moustafa, "Andrews curves" *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(4) 373-282 Jul/Aug 2011 DOI: 10.1002/wics.160online 03/23/11

P517[352]

David Marchette, "Implicit translation" *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(1) 28-34 Jan/Feb 2012 DOI: 10.1002/wics.181 online 07/15/11

P518[353]

Angel R. Martinez, "Part-of-speech tagging" *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(1) 107-113 Jan/Feb 2012 DOI: 10.1002/wics.195 online 09/30/11

Abstract:

'Presented here is a brief state-of-the-art account on part-of-speech (POS) tagging. POS tagging is an essential preprocessing task for many natural language processing goals and applications. Some POS tagging approaches make use of annotated corpora to train computational models to perform the task with minimal human intervention. Rule-based and stochastic methods have been successful, attaining accuracies of 96–97%. Representative approaches of these two methodologies are discussed.'

[348] onlinelibrary.wiley.com/doi/10.1002/wics.142/abstract
This had the same *odd* author contact as in P501.
[349] onlinelibrary.wiley.com/doi/10.1002/wics.149/abstract
[350] PhD 2009, ack'd for help on WR.
[351] onlinelibrary.wiley.com/doi/10.1002/wics.160/abstract
[352] onlinelibrary.wiley.com/doi/10.1002/wics.181/abstract
[353] onlinelibrary.wiley.com/doi/10.1002/wics.1217/abstract

P519[354]

David Marchette, "Scan statistics in graphs" *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(5) 466-473 Sep/Oct 2012 DOI: 10.1002/wics.1217 online 06/29/12

The last issue they edited was (5)1 Jan/Feb 2013.[355] In a survey of CVs of all authors, not just those here, *WIREs:CS* papers were often called peer reviewed. Most papers *seemed* fine, but concern is raised about the quality or even existence of *WIREs:CS* peer review. P197 and P513 were not only plagiarized, but P197 was so bad it could not possibly have been reviewed. **§J.3** noted Wegman's past interpretations of peer review. Finally, the handling of P509 and the number of papers by Wegman's students raise concerns, not about the authors, but about the editorial process and Wiley.

1	Caudle, etal	P508
2	Chow	P502, P514
3	Marchette	P503, P517, P519
2	Martinez (Angel)	P504, P518
2	Martinez (Wendy Poston)	P510, P511
4	Moustafa	P501, P507, P512, P516
1	Reeves, etal	P515
1	Rezazad	P509
1	**Said** and **Wegman**	P197☆ Peer review unthinkable.
6	Scott	Not numbered here, see CV[356]
1	Shores and **Wegman,**	P505
2	**Wegman** and **Said**	P506, P513☆
26	**Total by Editors and/or Wegman students**	

After a long career with many strong publications, ~2006 Wegman seemed to stop publishing in peer-reviewed research journals. People might compare his recent record with those of a few other current statisticians.[357] or the scientist he criticized so often.[358]

[354] onlinelibrary.wiley.com/doi/10.1002/wics.1217/abstract
[355] onlinelibrary.wiley.com/doi/10.1002/wics.v5.1/issuetoc
[356] "REFEREED PUBLICATIONS" p.12- #12, #14, #15, $16, #17, #18 www.stat.rice.edu/~scottdw/cv.pdf or www.webcitation.org/5yC18xtVJ
All seemed topics for which Scott was well-qualified, but *how refereed were they?*
[357] *WIREs:CS* has covered many topics with obvious experts, so the idea seems good and have added James E. Gentle and Karen Kafadar as Editors, both well-published, who seem likely to run the journal well.
[358] www.meteo.psu.edu/holocene/public_html/Mann/articles/articles.php
Many papers in a range of journals including *Science*, *Nature*, *PNAS*.

K.5 Talk summary table

Year.mo	Auth	Series	# Resume	0447 Code	Ack lack	Ack	Claim	MAS §	F/f	Cites Tot Oth	Papers from Wegman resume, using its number plus others found later, P301- and P401- (Foreign trip or likely flight in USA) §K.5 (a)
2004.12	W	1	T110W	c.1			1				"Visual Data Mining of Streaming Data," Federal Committee on Statistical Methodology (FCSM Statistical Policy Seminar: Achieving Statistical Quality in a Diverse and
2005.01	W		T111W	c.2			1		F		"Cybersecurity on the Internet: Where the Homeland is the World," Public Lecture, International Conference on the Future of Statistical Theory, Practice and Educatic
2005.01	W		T112W①	c.3			①		F		"Ecology of Alcohol and Alcoholism," International Conference on the Future of Statistical Theory, Practice and Education, Hyderabad, India, December 2004-January
2005.03	W		T113W	c.4			1		f		"The Development and Implications of Computational Statistics for Social Science, Health and Other Applications," Keynote Talk, Milestones in 21st Century Science, E
2005.04	W	1	T114W	c.5			1		f		"Strategies for Visual Data Mining," Keynote Talk, SIAM Data Mining Conference 2005, Orange County, California, April, 2005
2005.04	W		T115W	c.6	■	NO	1				"40 Years of Statistics Research: A Personal Perspective," 40 years of Statistical Computing and Beyond, Murray Hill, NJ, April, 2005
2005.06	W	1	T116W	c.7			1		f		"Strategies for Visual Data Mining," Symposium on the Interface, St. Louis, MO, June, 2005
2005.06	W		T117Ws①	c.8			①		f		"Ecology of Alcohol and Alcoholism," with **Y. Said**, Symposium on the Interface, St. Louis, MO, June, 2005
2005.05	Sw		T118Sw						f		"Modeling Alcohol Abuse and its Consequences," with **Y. Said**, (presented by Y. Said),Symposium on the Interface, St. Louis, MO, June, 2005
2005.08	W	1	T119W	c.9			1		f		"Visual Data Mining," Introductory Overview Lecture, Joint Statistical Meetings, Minneapolis, MN, August, 2005
2005.09	W	3	T120W	c.10			1				"Automated Metadata," SAMSI Workshop on Homeland Defense and National Security, Research Triangle Park, NC, September, 2005
2005.10	W	3	T121W	c.11			1		f		"Automated Metadata," Army Conference on Applied Statistics, Monterey, CA, October, 2005
2006.02	W	3	T122W	c.12			1		f		"Automated Metadata for Text Mining," ASA/RAND Conference on Quantitative Methods & Statistical Applications in Defense, Santa Monica, CA, February, 2006
2006.05	W		T123W①	c.13			①		f		"Statistics, Data Mining, and Climate Change," Keynote Talk, Second NASA Datamining Workshop: Issues and Applications in Earth Science, Pasadena, CA, May, 2006
2006.05	W		T124W①	c.14			①		f		"Statistics, Data Mining, and Climate Change," Keynote Talk, Symposium on the Interface, Pasadena, CA, May, 2006
2006.06	W	2	T125W	c.15			1		f		"Text Data Mining with Minimal Spanning Trees," Summer Research Conference on Statistics, Kerrville, TX, June, 2006
2006.07	Ws		T126Ws①③☆★	c.16	■	NO	①③	☆§2,§3			Testimony to House Committee on Energy and Commerce, U.S. House of Representatives, Washington, D.C., July 20 *(sic) a* nd 27, 2006.
2006.08	W		T127W①	c.17			①		f		"The Kyoto Accord, The 2001 IPCC Third Assessment Report and The Academic Papers Underpinning Them," Joint Statistics Meeting, Seattle, WA, August, 2006
2006.08	W		T128W	***E.7***			1		F		**Wegman, Edward J.** and Caudle, Kyle A. (2006) "Density estimation from streaming data using wavelets," in COMPSTAT 2006, (Alfredo Rizzi and Maurizio Vichi, eds.),
2006.08	W		T129W①	c.19			①		F		"Geospatial Distribution of Alcohol-Related Violence in Northern Virginia," COMPSTAT 2006, Rome, Italy, August 2006
2006.09	W	3	T130W	c.20	■	NO	1		F		"On the Extraction of Endogenous Metadata for Text and Image Databases," Keynote Talk, The KNEMO Workshop, Anacapri, Capri Italy, September 2006 (Wegman ar
2006.09	W	1	T131W	c.21	■	NO	1		F		"Computational Statistics – Graphical and Analytic Methods for Streaming Data," Short Course Lectures, Universita Napoli "Federico II", Naples, Italy, September, 200
2006.10	W		T132W①	c.22			①				"The Hockey Stick Controversy: Lessons for Statisticians," Army Conference on Applied Statistics, Research Triangle Park, NC, October, 2006
2007.03	W	1	T133W	c.23			1		F		"Visual Data Mining," Public Lecture, Distinguished Visiting Professor at the American University of Cairo, Cairo, Egypt, March, 2007 03/26
2007.03	W	1	T134W	c.24			1		F		"Visualization of Streaming Data," Public Lecture, Distinguished Visiting Professor at the American University of Cairo, Cairo, Egypt, March, 2007
2007.03	W		T135Ws①★	c.25			①		F		"Reanalysis of the Hockey Stick Paleoclimate Reconstruction," Public Lecture, Distinguished Visiting Professor at the American University of Cairo, Cairo, Egypt, March
2007.03	W	3	T136W	c.26			1		F		"Extraction of Endogenous Metadata," Keynote Talk, Sixth Conference on Statistics in the Social Sciences and Humanities, Cairo Egypt, March, 2007
2007.05	Sw		T137Sw①	c.27			①				"Assessing Interventions Related to the Negative Effects of Ethanol on HIV/AIDS Spread," 39th Symposium on the Interface of Computing and Statistics, Philadelphia,
2007.07	W		T138W①	c.28			①		f		"A Bipartite Graph Model of the Interaction between Alcohol Users and Institutions," Research Society on Alcoholism Annual Meeting, Chicago, IL, July, 2007
2007.08	W		T139Sw①	c.29			①		f		"Assessing Interventions Related to HIV Incidents Under the Influence of Ethanol," Joint Statistical Meetings, Salt Lake City, UT, August, 2007
2007.08	W	4	T140W	c.30			1		F		"Text Mining for Fun and Profit," 3rd International Symposium on Business and Industrial Statistics, Ponta Delgada, Azores, Portugal August, 2007
2007.10	W		T141W①	c.31	■	NO	①		f		"20 Questions a Statistician Should Ask about Climate Change," ASA Workshop on Climate Change, NCAR, Boulder, CO, October 2007
2008.01	W		T142W	c.32			1		F		"Methods for Visualizing High Dimensional Data," Contemporary Frontiers in High Dimensional Statistical Analysis, Cambridge, UK, January, 2008
2008.04	W	4	T143W	c.33			1		f		"Text Mining, Social Networks, and High Dimensional Analysis," Izzet Sahin Memorial Lecture, University of Wisconsin, Milwaukee, WI, April, 2008
2008.06	Ws	4	T144Ws	c.34			1		F		"Text Mining and Social Networks: Some Unexpected Connections," Keynote Address, International Conference on Multivariate Statistical Modeling and High Dimensi
2008.06	Sw	4	T144a Sw 2 ③	c.35			1,2		F		"Approaches to Text Mining that Preserve Semantic Content," **Yasmin H. Said, Lecturer**, International Conference on Multivariate Statistical Modeling and High Dimer
2008.08	W	5	T145W	c.36			1		f		"Mixture Models for Document Clustering," Joint Statistical Meetings, Denver, CO, August, 2008

This part includes all 0447-claimed talks, next page has others.

Year.mo	Auth	Series	# Resume	0447 Code	Ack lack	Ack	Claim	MAS §	F/f	Cites Tot Oth	Papers from Wegman resume, using its number plus others found later, P301- and P401- (Foreign trip or likely flight in USA)	§K.5 (b)
2008.08	W		T146W						F		"A Directed Graph Model of Ecological Alcohol Systems Incorporating Spatiotemporal Effects," COMPSTAT 2008, Porto, Portugal, August 2008	
2008.10	W		T147W						f		"Paleoclimate Temperature Reconstructions: Implications for Climate Change," American Public Health Association Annual Meeting, San Diego, CA, October, 2008	
2008.10	Ws	4	T148Ws③			1,2,③					"Text Mining and Social Networks: Some Unexpected Connections," Army Conference on Applied Statistics, Lexington, VA, October, 2008 10/23/08, ARO_Proposal\T	
2008.11	W		T149W								"A Bipartite Graph Model of the Interaction between Alcohol Users and Institutions," NISS Workshop on Agent-Based Modeling, Research Triangle Park, NC, Novembe	
2009.04	W	4	T150W						F		"Text Mining, Social Networks, and Document Clustering," Keynote Address, UAE Mathematics Day, Sharjah, UAE, April, 2009	
2009.06	W	4	T151Ws③			1,2,③			f		"Document Clustering and Social Networks," Keynote Address, Classification Society Annual Meeting, St. Louis, MO, June, 2009	
2009.06	sw	6	T152ws								"Multi-Mode Social Networks," Human Behavior-Computational Intelligence Modeling Conference, Oak Ridge, TN, June, 2009 (really: Said, Wegman, Sharabati)	
2009.06	sw	6	T153Sw			NO			f		"On Agent-Based Models of Multi-Mode Social Networks," Swarmfest 2009, Santa Fe, NM, June, 2009. 06/28/09 (really: Said, Wegman, Sharabati)	
2009.08	W		T154W								"Multivariate Data Adaptive Compression and Density Estimation," Joint Statistical Meetings, Washington, DC, August, 2009	
2009.08	W	4	T155W						F		"Text Mining, Social Networks, and Document Clustering," ISI 2009 Durban: 57th Session of the International Statistical Institute, Durban, South Africa, August, 2009	
2009.12	W		T156W						F		"Massive Data Streams and Citizen Science," Keynote Address, Tenth Islamic Countries Conference on Statistical Sciences, Cairo, Egypt, December, 2009	
2005.09	S		T400S☐				3		f		**Said, Yasmin H. (2005)** "Agent-based Model Applicable to Homeland Security and Disease Control" SAMSI/NISS, Triangle Park, NC, September	
2005.10	S		T401S☐				3		f		**Said, Yasmin H.** (2005) *"Adaptation of an Alcohol Ecological Agent-Based to Homeland Security", ACAS05 Monterey CA, Oct 19-21*	
2006.08	w		T402w								Amy Braverman, Jet Propulsion Laboratory, **Edward Wegman**, George Mason University; Wendy Martinez, Office of Naval Research; Juergen Symanzik, Utah State Ur	
2006.08	W	3	T403W								**Edward Wegman**, George Mason University, Faleh Alshameri, George Mason University, "Automated Metadata"' JSM2006, August 06-10, 2006, Seattle, WA	
2006.08	Ws		T404ws								**Edward Wegman**, George Mason University, Department of Statistics, ...; **Yasmin H. Said**, Johns Hopkins University; Shabib A. Alhadheri, SUNY Upstate Medical Unive	
2006.08	Ws		T405ws								**Edward Wegman**, George Mason University; **Yasmin H. Said**, Johns Hopkins University, "A Policy Tool for Assessing Alcohol Intervention Strategies" *JSM2006, August*	
2007.01	Ws	3	T406Ws			ARO					**Wegman, Edward J**, with **Said, Yasmin H.**, and Alshameri, Faleh, "On the Extraction of Endogenous Metadata for Text and Image Databases," talk for ARL-Aberdeen 0	
2007.07	s		T407sw③				③				Joost Bottenbley, ...; **Yasmin H. Said**, George Mason University; **Edward Wegman**, et al "Relationship between Competitiveness of Colleges and Levels of Tobacco Use	
2007.07	s		T408s③				③				Fahad Bin Muhaya, Imam University, PO Box 84901, Riyadh, 11681 Saudi Arabia, ...; **Yasmin H. Said**, George Mason University, "Access Control Model for E-Learning S	
2007.07	S		T409S③				③				**Said, Yasmin H.** (2007) *"Pro Bono Statistics and Public Policy" Proc JSM2007, p.176*	
2007.07	S		T410sw③				③				Hadi Rezazad, ... **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason University, "A Statistical Social Network Approach to Computer Network	
2007.07	sw		T411sw③				③				Walid Sharabati, George Mason University, ...; **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason University "Style of author-coauthor social	
2007.07	sw		T412sw③				③				Samer Ellahham, Innovative Medical Institute/Paragon Cardiovascular Foundation; **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason Unive	
2007.07	sw		T413sw③				③				Samer Ellahham, Innovative Medical Institute/Paragon Cardiovascular Foundation; 23044 Winged Elm Drive, Clarksburg, MD 20871 **Yasmin H. Said**, George Mason Ur	
2007.10	S		T415③				③				**Said, Yasmin H.** (2007) Participation in ASA/NCAR workshop in Boulder, CO October 26-27, 2007	
2007.10	S		T414S								**Said, Yasmin H.** (2007) *"Experiences with Congressional Testimony: Statistics and The Hockey Stick" talk at GMU 09/07/07*	
2007.11	S		T416S③				③				**Said, Yasmin H.** (2007) *claim to have spoken at Symposium on Arabic Language and Computers, Riyadh, Saudi Arabia. (November, 2007) Nov 10-12.*	
2008.05	Sw		T417Sw①			①, 3					**Said, Yasmin H.** (2008) "Estimating Spatiotemporal Effects for Ecological Alcohol Systems," Interface 2008, Durham, NC, 05/23/08.	
2008.08	Ws		T418ws①			①, 3					Wieczorek, William F., **Said, Yasmin H. and Wegman, Edward J.** (2008) "Spatial and computational models of risks for alcohol users," Interface 2008. Talk version of F	
2008.08	Sw		T419S			3					**Said, Yasmin H.** (2008) "Estimating Spatiotemporal Effects for Ecological Alcohol Systems," COMPSTAT 2008, Porto, Portugal, August 2008 167-178.	
2008.10	Ws	5	T420Ws③			1,2,③					**Wegman, Edward J.** and **Said, Yasmin H.** (2008) "Mixture Models for Document Clustering" Talk, 10/30/08 @ U MD early version of semantic content talks	
2009.04	S		T421S								Said, Yasmin H. (2009) "Using Text Mining and Social Network Analysis to Study the Quran and Other Sacred Texts", 7th UAE Math Day, Sharjah, UAE, 04/25/09	
2009.10	Sw	4	T422Sw								Said, Yasmin H. (2009) "Approaches to Text Mining that Preserve Semantic Content" ACAS, presumably about same as T425.	
2009.12	W		T423W								**Wegman, Edward J.** CNN interview on "Climategate" 12/10/09	
2010.04	W		T424W			NO					**Wegman, Edward J.** talk at Chapman Conference 04/22/10	
2010.05	Sw	4	T425Sw③			1,2,③					Said, Yasmin H. and Wegman, Edward J. (2009) "Preserving semantic content in text mining using multigrams,"Talk for QMDNS 05/26/10 , T422 likely to be identical.	
2010.06	Sw		T426Sw								Said-organized session at Interface 2010 including speaker Fred Singer	
2010.06	Sw		T427Sw								Said-organized session at Interface 2010: Kueter, Easterbrook, Said	
2010.08	W		T428W								**Wegman, Edward J.** a(2010) "Analysis of Global Warming Data: A Contrarian Data-Based View" JSM2010, 08/04/10	
2010.08	S		T429S								**Said**-organized climate session, cancelled at JSM 2010	
2010.09	ws		T430ws			NO					Sharabati, Walid K., **Wegman, Edward J., Said, Yasmin H.** (2010) "Predicting Edges And Vertices In A Network" 2010 IEEE/WIC/ACM International Conference on Web	
2010.10	W		T431W								**Wegman, Edward J.** (2010) "Understanding Afghanistan" ACAS, Cary, NC , 10/20/10-10/22-10	
2011.08	Sw	4	T432Sw③			1,2,③			F		**Said, Yasmin H.** and **Wegman, Edward J.** (2011) "High Dimensional Vector Space Methods for Characterizing Semantic Content" ISI 2011 Dublin, Ireland, Proc. 58th	

K.6 Talks annotated

Of talks, only T128 was claimed to be peer-reviewed. Given the locations, many required air flights, but only ARO and GMU might know if any were explicitly billed to ARO and if expenses were deemed appropriate.[359] For many talks, it was difficult to find much more than title and perhaps an abstract, so that most acks were unknown.

Wegman 0447-claimed T110-T117, T119-T144, T144a, T145.

Citation counts are usually omitted, as 0/0 would be typical.

***As elsewhere, any of these is alleged to be a clear problem,* §B.2:**

□ 5876-claimed, but given before started, **§B.2**.

①②③ 0447-unfit, 0059-unfit or 5876-unfit, claimed or ack'd

☆ Plagiarism alleged

★ Falsification / misrepresentation alleged

Similar talks often appear multiple times, **not** alleged to be a problem. Unlike journal papers, which should not be duplicative, some talks are naturally given more than once, as discussed in **§K.2.**

Of ~80 talks, 17 were given at IFNA-managed conferences:[360]

T116 W	2005 Interface
T117 Ws①	2005 Interface
T118 Sw①	2005 Interface
T121 W	2005 ACAS
T401 S□	2005 ACAS
T122 W	2006 QMDNS (not yet managed by IFNA)
T124 W①	2006 Interface
T132 W①	2006 ACAS
T137 Sw①	2007 Interface
T417 Sw①	2008 Interface
T418 ws ①	2008 Interface
T148 Ws③	2008 ACAS
T151 Ws③	2009 Interface (Joint with Classification Society)
T422 sw	2009 ACAS
T425 Sw③	2010 QMDNS
T426 Sw	2010 Interface
T427 Sw	2010 Interface
T431 W	2010 Interface

T108 W 0/0 included for context as earliest Wegman talk on alcoholism

'Computational and Mathematical Tools for Alcohol Ecology," Research Society on Alcoholism, Vancouver, BC, Canada, June 2004'

This was about the time Said "completed her dissertation proposal" related to this topic, **§S.1**. No problem is alleged for this.

Wegman was a member of RsoA and not only inspired Said's alcoholism PhD,[361] but he was involved with Said in almost every alcoholism paper or talk listed here.[362] With the exception of {P193, T418}, *most of the alcoholism work seemed* to be variations on Said's dissertation (analysis of alcoholism in Northern Virginia) or extensions to add temporal information to the spatial analysis already done there.[363]

==========

T110 W {❶}

"Visual Data Mining of Streaming Data," Federal Committee on Statistical Methodology (FCSM Statistical Policy Seminar: Achieving Statistical Quality in a Diverse and Changing Environment, Bethesda, MD, December, 2004

Visual data mining had been a major theme for earlier grants.

T111 W

"Cybersecurity on the Internet: Where the Homeland is the World," Public Lecture, International Conference on the Future of Statistical Theory, Practice and Education, Hyderabad, India, December 2004-January 2005

T112 W① {❿}

'"Ecology of Alcohol and Alcoholism," International Conference on the Future of Statistical Theory, Practice and Education, Hyderabad, India, December 2004-January 2005'

Unlike T108, this was 0447-claimed.

T113 W

'The Development and Implications of Computational Statistics for Social Science, Health and Other Applications," Keynote Talk, Milestones in 21st Century Science, Buffalo, NY, March, 2005'

[359] *There may have been no explicit charges and ARO funds may or may not have been used. The general sloppiness of claims and acks makes this unclear.*

[360] *Some of these talks seemed strangely out of place. Would anything by Wegman have been rejected? All this makes quality assessment difficult.*

[361] **§S.1** 'He has been so passionate about issues related to alcohol abuse and alcoholism that he has insired *(sic)* me to want to work in this area as a life career.'

[362] As author, coauthor or mentioned by Said a "joint work with Wegman"

[363] Actual progress would need assessment by alcoholism experts.

T114 W[364] {❶}
"Strategies for Visual Data Mining," Keynote Talk, SIAM Data Mining Conference 2005, Orange County, California, April, 2005'

T115 W[365]
"40 Years of Statistics Research: A Personal Perspective," 40 years of Statistical Computing and Beyond, Murray Hill, NJ, April, 2005
This interesting history[366] may offer insight, especially the last 2 slides:
''2003 ONR abolishes Statistics and Probability Program
Decision Support Program
ARO Statistics Program down to $800,000
2005 John Chambers "retires," Wegman invited to Bell Labs to talk
2006 NAS Committee on Applied and Theoretical Statistics ->
Committee on Data, Informatics, and Statistics
NAS Board of Mathematical Sciences and Applications
80% of their portfolio involves stochastic data'
'Conclusions
Best of times and the worst of times …
Basic research support in statistics by federal government (and industry?) severely eroded,
but statistics … data … is everywhere.
Where is some action?
Social network modeling – homeland security
Streaming data
Text and Image data mining'
This may hint that Wegman's accustomed funding had diminished, as 2005 certainly seemed a break point for behavior.

T116 W {❶}
"Strategies for Visual Data Mining," Symposium on the Interface, St. Louis, MO, June, 2005'

T117 Ws① [367] 0/0 {❿}
"Ecology of Alcohol and Alcoholism," with **Y. Said**, Symposium on the Interface, St. Louis, MO, June, 2005
At this point, Wegman was still consistently using:
Ewegman AT galaxy DOT gmu DOT edu
By Fall 2005, a gmail address was also employed.

T118 Sw① 0/0 {❿}
'"Modeling Alcohol Abuse and its Consequences," with Y. Said, (presented by **Y. Said**),Symposium on the Interface, St. Louis, MO, June, 2005

T119 W {❶}
'"Visual Data Mining of Streaming Data," Federal Committee on Statistical Methodology (FCSM Statistical Policy Seminar: Achieving Statistical Quality in a Diverse and Changing Environment, Bethesda, MD, December, 2004'

T120 W[368] {❸}
'"Automated Metadata," SAMSI Workshop on Homeland Defense and National Security, Research Triangle Park, NC, September, 2005'

T121 W {❸}
'"Automated Metadata," Army Conference on Applied Statistics, Monterey, CA, October, 2005'

T122 W {❸}
'"Automated Metadata for Text Mining," ASA/RAND Conference on Quantitative Methods & Statistical Applications in Defense, Santa Monica, CA, February, 2006' QMDNS

[364] www.siam.org/meetings/sdm05/wegman.htm
[365] stat.bell-labs.com/JMCWorkshop/wegmanTalk.pdf at ect.bell-labs.com/sl/JMCWorkshop/index.html 04/29/05
[366] I used to work there, so this recalls fond memories of decades that Bell Labs supported superb statisticians and widespread use of good statistics.

[367] www.interfacesymposia.org/I2005WebPage/ConferenceAgendaFinal.pdf
Said organized an invited session:T117, T118 and one by Wieczorek.'
[368] www.samsi.info/workshop/2005-06-program-national-defense-and-homeland-security-opening-workshop-and-tutorials Sept 11-14, 2005
This series is *almost certainly y* based on or related to Alshameri's work.
Said spoke at this, as did Rigsby.

T123 W①[369]

"Statistics, Data Mining, and Climate Change," Keynote Talk, Second NASA Datamining Workshop: Issues and Applications in Earth Science, Pasadena, CA, May, 2006' *05/23/06*

The NASA final report includes, p.7:[370]

'Concluding the opening session, Dr. Ed Wegman of George Mason University gave a talk, entitled "*Statistics, Data Mining, and Climate Change",* which showed the danger of using improper data analyses to reach scientific conclusions.'

T124 W①[371]

"Statistics, Data Mining, and Climate Change," Keynote Talk, Symposium on the Interface, Pasadena, CA, May, 2006'

I have been unable to find any trace of this talk. Examination of the program, as of 09/01/06[372] showed Wegman giving a Text Mining talk, but the only Keynote was by Usama Fayyad. Either it was originally planned, then replaced by Fayyad, or Fayyad dropped out and Wegman gave his talk, but the program did not get changed. The last archived copy of the Interface 2006 web page was 08/26/10,[373] and it disappeared sometime later. The Program at IFNA[374] was actually an outdated one from 2005.

T125 W[375] {❷}

'"Text Data Mining with Minimal Spanning Trees," Summer Research Conference on Statistics, Kerrville, TX, June, 2006'

From the title, this seems likely to have been a presentation of an earlier talk, without too much change, since Wegman was busy with T126**.**

T126 Ws①③☆★

This presented the WR to Congress 07/19/06, and included the follow-up testimony the next week. [376] Wegman wrote: [377]

'Testimony to House Committee on Energy and Commerce, U.S. House of Representatives, Washington, D.C., July 20 *(sic)* and 27, 2006, republicans.energycommerce.house.gov/108/home/07142006_Wegman_fact_sheet.pdf republicans.energycommerce.house.gov/108/News/07142006_1990.htm'

The first is [BAR2006] and the second is just another copy of same file.

The authors wrote 91 pages on totally unfamiliar topics, substantial effort, then Wegman claimed credit for it from ARO and Said from NIAAA.[378] As discussed in detail [MAS2010, MAS2010a], *the WR was filled with incompetent science and statistics at best, even ignoring the plagiarisms ☆ and falsifications ★.[379] I alleged that the WR was a major step in a long, well-organized political anti-science campaign to mislead Congress and affect public policy. If so, there may be further consequences beyond alleged grant fraud.*

Misleading Congress can rise to felony, as can conspiracy to do so or obstruction of justice by destroying evidence [MAS2010a].

The citation analysis of the WR itself might make an interesting study. A Google Scholar search[380] yields multiple hits with different spellings, but ~60 total citations.

---

[369] datamining.itsc.uah.edu/meeting06/agenda.html
This illustrates difficulty of assessing travel expenses. This was a trip to Pasadena, for an 0447-unfit talk, but then Interface involvement likely covered the cost.

[370] datamining.itsc.uah.edu/meeting06/docs/2nd_NASA_Data_Mining_D694F.pdf

[371] [MAS2010a] found this at , but that disappeared. Copies were found web.archive.org/web/20060530032004/www.galaxy.gmu.edu/Interface2006/i2006webpage.html

[372] web.archive.org/web/20060530032004/www.galaxy.gmu.edu/Interface2006/i2006webpage.html

[373] web.archive.org/web/20100801000000*/galaxy.gmu.edu/Interface2006/i2006webpage.html

[374] www.interfacesymposia.org/Interface2006/i2006webpage.html

[375] www.sph.emory.edu/srcos/research.htm

[376] A claim for a presentation for which one is the lead author is surely a claim for the report itself, not just the act of presenting it.

[377] archives.republicans.energycommerce.house.gov/108/home/07142006_Wegman_fact_sheet.pdf [BAR2006] URLs have changed.

[378] [WEG2010 p.24] listed the WR separately under Technical Reports, but of course the claim for T126 includes that, because it makes no sense that testimony about the report would be claimed, but not the report itself. Diversion of most 0447 effort in 2005-2006 is strongly supported by fit works' plunge in 2006-2007.

[379] MAS2011b adds much more explicit falsification analysis beyond MAS2010a: deepclimate.org/2010/11/16/replication-and-due-diligence-wegman-style moyhu.blogspot.com/2011/06/effect-of-selection-in-wegman-report.html deepclimate.org/2010/10/25/the-wegman-report-sees-red-noise/
At worst, this includes falsification, since the code used by McIntyre, and apparently re-used by Wegman, contained an obvious 1:100 cherry-pick atop statistical parameters unlike those of the real world.

[380] scholar.google.com/scholar?q=wegman+said+scott+ad+hoc

T127 W①[381]

'**Edward Wegman**, "The Kyoto Accord, The 2001 IPCC Third Assessment Report and The Academic Papers Underpinning Them," Joint Statistics Meeting, Seattle, WA, August, 2006'

Abstract:

'*Key Words:* PCA, paleoclimate reconstruction, climate change, global Warming The Kyoto Accord focused on reducing greenhouse gasses and was supported by the report of the Intergovernmental Panel on Climate Change, "2001 Third Assessment Report." This report featured the "hockey stick" millennium temperature reconstruction based on a number of proxy variables. The academic papers that developed the reconstruction used a principal components analysis, which has been challenged by critics due to inappropriate use of PCA and a lack of independent verification of the findings. **We discuss the methodology used, the use of potentially problematic data, and the social network of authors in temperature reconstruction.** The implications of global warming are important for the financial and human dimensions, and public policy decisions must be made on a statistically sound, uncontroversial basis. **D. Scott, D. Brillinger, Y. Said, J. Rigsby, D. Reeves, and A. Braverman contributed to this report.**'[382]

The session was reviewed by Richard Smith,[383] and his notes summarized later [MAS2010a **§A.3**]. He described Wegman's talk:

'The three speakers were Ed Wegman, J. Michael Wallace of the Department of Atmospheric Sciences, University of Washington, and myself. Ed and Mike both talked about the hockey stick reconstruction. **Ed focused on statistical flaws that, in his view, render much of the current literature on this subject of doubtful validity.** Mike presented the broader findings of a recent NRC panel that, while acknowledging the statistical issues of Wegman's report, defended the hockey stick curve based on a broader scientific context.'

T128 W= P175 W

'**Wegman, Edward J.** and Caudle, Kyle A. (2006) "Density estimation from streaming data using wavelets," in COMPSTAT 2006, (Alfredo Rizzi and Maurizio Vichi, eds.), 231-244 (Rome, Italy)'

T129W①=P174Sw① 1/0 {❿}

'"Geospatial Distribution of Alcohol-Related Violence in Northern Virginia," COMPSTAT 2006, Rome, Italy, August 2006'

T130 W[384] {❸}

This was a talk + 4-page paper in Anacapri, Capri, IT.

'"On the Extraction of Endogenous Metadata for Text and Image Databases," Keynote Talk, The KNEMO Workshop, Anacapri, Capri Italy, September 2006 (**Wegman** and Faleh Alshameri)'

A talk of the same name was given 01/23/07, but by Wegman, with Said and Alshameri.

T131 W[385] {❶}

This short course was given in Naples, IT.

'"Computational Statistics – Graphical and Analytic Methods for Streaming Data," Short Course Lectures, Universita Napoli "Federico II", Naples, Italy, September, 2006'

T132 W①[386]

'"The Hockey Stick Controversy: Lessons for Statisticians," Army Conference on Applied Statistics, Research Triangle Park, NC, October, 2006'

See [MAS2010a **§A.3**].

T133 W[387] {❶}

'"Visual Data Mining," Public Lecture, Distinguished Visiting Professor at the American University of Cairo, Cairo, Egypt, March, 2007' *03/26/07*

T134 W {❶}

'"Visualization of Streaming Data," Public Lecture, Distinguished Visiting Professor at the American University of Cairo, Cairo, Egypt, March, 2007' *~03/26/07*

[381] www.amstat.org/meetings/jsm/2006 (slightly mis-cited) www.amstat.org/meetings/jsm/2006/PDFs/JSM06AbstractBook.pdf

[382] Brillinger and Braverman were not listed as authors or ack'd in the WR.

[383] lv-twk.oekosys.tu-berlin.de//project/lv-twk/images/pdfs/ENVR_9_1.pdf

[384] www.stat.unipg.it/iasc/Proceedings/2006/COMPSTAT_Satellites/KNEMO/Lavori/Papers%20CD/Wegman%20Alshameri.pdf www.webcitation.org/6D8OK5Bb2

[385] www.novauniversitas.it/System/2447/Wegman-NapoliShortCourse2006.ppt

[386] www.armyconference.org/ACAS06/default.htm Proceedings not online.

[387] web.archive.org/web/20090403115008/www1.aucegypt.edu/academic/math/events.html

T135 Ws①★[388]

"'Reanalysis[389] of the Hockey Stick Paleoclimate Reconstruction," Public Lecture, Distinguished Visiting Professor at the American University of Cairo, Cairo, Egypt, March, 2007' 03/21/07

The AUC event calendar billed this as:

'A Reanalysis of the "Hockey Stick" Paleoclimate Reconstruction Seminar by DVP Dr. Edward Wegman and Dr. Yasmin Said.'[390]

The abstract said:[391]

'A Reanalysis of the "Hockey Stick" Paleoclimate Reconstruction Edward J. Wegman, Yasmin H. Said, and David W. Scott Abstract: One of the most interesting visual graphic to emerge in the last decade was the so-called "Hockey Stick" paleoclimate reconstruction. The papers published in 1998 and 1999 by Mann et al. used tree ring data and other temperature proxies to estimate the temperature over a 1000 year period beginning approximately 1000 C.E. Their reconstruction showed a gradual decrease in temperature from 1000 to approximately 1850 and then a rapid rise. The conclusion of their paper was that the decade of the 1990s was probably the hottest decade in a millennium and that 1998 was probably the hottest year in a millennium. The graphic was exploited by the Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report published in 2001 and was widely used to support assertions of anthropogenic climate change. Their methodology incorrectly used a principal component-like analysis. **When the PCA methodology is correctly used, the hockey stick essentially disappears (i.e. the rapid rise from 1850 disappears.** The change in mean global temperature is approximate .7 degrees centigrade over 150 years. We discuss this and other statistical faults with the paleoclimate reconstructions. The analysis of these three authors was presented as testimony in July 2006 before the United States Congress.'

The ~1850-onward temperatures were determined by measurement and were irrelevant to the PCA issues. The WR failed to redo the calculations with the correct (centered) PCA methodology. Wahl and Ammann had already shown it made no significant difference, [MAS2010a **§A.3**]. This gets tagged ★ as an objectively-false statement, an example of false information given in non-expert venues. See also ~~P407~~ Ws☆★.

[388] web.archive.org/web/20090403115008/www1.aucegypt.edu/academic/math/events.html

[389] The WR did no reanalysis, just reran McIntyre' code [DEE2010r].

[390] Since Wegman presented, it is unclear whether or not Said attended.

[391] web.archive.org/web/20090403115008/www1.aucegypt.edu/academic/math/documents/Dr.Edward_Abstract.doc

T136 W {❸}

This was another lecture on that trip.

'"Extraction of Endogenous Metadata," Keynote Talk, Sixth Conference on Statistics in the Social Sciences and Humanities, Cairo Egypt, March, 2007

T137 Sw① 0/0 {⓫}

"Assessing Interventions Related to the Negative Effects of Ethanol on HIV/AIDS Spread," 39th Symposium on the Interface of Computing and Statistics, Philadelphia, PA, May, 2007

Testa explained why this was unlikely, **§S.3.2**-05/08/07.

T138 W①[392] 0/0 {⓬}

'A Bipartite Graph Model of the Interaction between Alcohol Users and Institutions," Research Society on Alcoholism Annual Meeting, Chicago, IL, July, 2007'

I am unable to find this, but possibly it was a poster. P186 and P187 were posters at the 2008 RSoA conference and with T138 comprised the only alcoholism-conference presentations listed here.[393]

[392] onlinelibrary.wiley.com/doi/10.1111/acer.2007.31.issue-s2/issuetoc

[393] Bipartite graphs are normally discussed in undergraduate graph theory courses. I still have 3 textbooks from the mid-1960s, and the topic was hardly new then. en.wikipedia.org/wiki/Bipartite_graph

Two-mode networks (as in SNA, [WAS1994 pp29-30, 298-300]) naturally form bipartite graphs if the only relationships are those between the two modes (sets of actors). This is as familiar as a spreadsheet whose rows represent alcohol users and columns represent institutions.

T139 Sw① [394] 0/0 {⓫}
"'Assessing Interventions Related to HIV Incidents Under the Influence of Ethanol," Joint Statistical Meetings, Salt Lake City, UT, August, 2007'
Said may have given the talk, but the Abstract Book has:
'**Yasmin H. Said**, George Mason University**; Edward Wegman**, George Mason University'
Abstract:
Key Words: social networks, alcohol abuse, disease mechanism
Alcohol abuse leads to serious individual and societal outcomes. Among these, we identify irresponsible behavioral outcomes notably risky sexual contacts and other violence related outcomes. Risky sexual contacts can lead to infections with HIV and other STDs. Social, structural, and environmental factors are major influences on HIV-related behaviors. Our work provides a policy tool for evaluation of interventions using a dynamic agent-based simulation based on estimates of conditional probabilities. Alcohol abusers are embedded in a social network that is reminiscent of an ecology system. We formulate a model of this network to explore interventions that reduce the overall probability of negative outcomes. Historically, modeling attempts have focused on specific negative outcomes. The unique feature of this work is that we explore the simultaneous reduction of all negative outcomes.'
Testa explained why this was unlikely, **§S.3.2-**05/08/07.

T140 W[395] {❹}
'Text Mining for Fun and Profit," 3rd International Symposium on Business and Industrial Statistics, Ponta Delgada, Azores, Portugal August, 2007'

T141 W①[396]
'20 Questions a Statistician Should Ask about Climate Change,' ASA Workshop on Climate Change, NCAR, Boulder, CO, October 2007
This was given at a small workshop of climate scientists and climate-experienced statisticians and *was not well-received* [MAS2010a §A.3]. *It was far below the standard set by the other, highly-professional presentations.* Wegman usually gave such talks for audiences unlikely to have the same level of knowledge. Said also attended, **T415**.
Who paid for time and travel? ARO? NCAR? ASA? "pro bono"?

[394] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.320
[395] isi.cbs.nl/Nlet/NLet072-sections.htm Aug 18-20, 2007.
[396] www.image.ucar.edu/Workshops/2007/ASAclimate/index.shtml Oct 26-27

T142 W
This might belong with **T143**, or have evolved from earlier work.[397]
"Methods for Visualizing High Dimensional Data," Contemporary Frontiers in High Dimensional Statistical Analysis, Cambridge, UK, January, 2008

T143 W {❹}
"Text Mining, Social Networks, and High Dimensional Analysis," Izzet Sahin Memorial Lecture, University of Wisconsin, Milwaukee, WI, April, 2008
This is *very likely* the same as 04/25/08 talk at GMU,[398] with abstract:
'A traditional approach to text mining has been to represent a document by a vector. In the bag-of-words representation binary vectors are used and two documents are regarded as similar if the angle between their corresponding vectors is small (i.e., correlation between the vectors is high). The document vectors may be assembled into a term-document matrix (TDM). A more satisfying representation of a document can be formulated in terms of bigrams or trigrams, because these have a better chance of capturing semantic content Bigram vectors ran be assembled into bigram document matrices (BDM). The TDM and BDM resemble the two-mode adjacency matrices associated with social network analysis (SNA). Using cues from SNA, we formulate the one-mode social network adjacency matrices to form document-document matrices (DD) and bigram-bigram matrices (BB). In this talk I outline the basics, discuss the connection between text mining and social networks and, by example, illustrate the dimensionality issues raised by such vector space methods.'

T144 Ws {❹}
"Text Mining and Social Networks: Some Unexpected Connections," Keynote Address, International Conference on Multivariate Statistical Modeling and High Dimensional Data Mining, Kayseri, Turkey, June, 2008.'

[397] Edward J. Wegman, Jeffrey L. Solka, "On some mathematics for visualizing high dimensional data," Indian J. of Statistics (2002). www.jstor.org/discover/10.2307/25051404?uid=3739560&uid=2&uid=4&uid=3739256&sid=21101746302131
[398] tech.dir.groups.yahoo.com/group/SSPK/message/10766?var=0

T144a Sw③ {❹}
This was 0447-claimed, but not in [WEG2010]. Said claimed it for 5876.
"Approaches to Text Mining that Preserve Semantic Content," Yasmin H. Said, Lecturer, International Conference on Multivariate Statistical Modeling and High Dimensional Data Mining, Kayseri, Turkey, June, 2008.
Said *likely* gave the talk in Kayseri. Her presentations sometimes showed her as sole author, but credited Wegman in an annotation, as in T417. The conference clearly happened[399] and was referenced by others, but searches failed to find the actual Proceedings, so perhaps it was unpublished.

T145 W {T145, T420}
"Mixture Models for Document Clustering," Joint Statistical Meetings, Denver, CO, August, 2008

Talks T146-T156 were listed in resume, but not 0446-claimed. Several inappropriate acks (① or ③) occurred.

T146 W=***P192*** Ws① 2/0 {⓭}in session with T419Sw and P193sw①.
'"A Directed Graph Model of Ecological Alcohol Systems Incorporating Spatiotemporal Effects," COMPSTAT 2008, Porto, Portugal, August 2008'
The abstract[400] was:
'Users of alcohol are incorporated into a societal system, which for many purposes resembles an ecological system. We have previously modeled such systems using a directed graph with acute outcomes reflecting undesirable individual and societal outcomes. In this paper, we expand the model to a hybrid social network directed graph model for modeling the acute outcomes associated with alcohol use and abuse. We describe the approximate estimates of conditional probabilities based on available data. In the present model, we also approximate geospatial effects related to transportation as well as temporal effects. Acute outcomes include assault, murder, suicide, sexual assault, infection with STDs or HIV, domestic violence, child abuse, and DWI and alcohol related fatal crashes. The model is calibrated using demographic, crime, alcohol-related, and alcohol outlet data from Virginia and Fairfax County in Virginia. We propose proxy data and methods for capturing temporal and geospatial effects. The goal is to investigate methods for simultaneous suppression of multiple negative public health consequences. The model may be used as a public policy decision tool by adjusting conditional probabilities in order to investigate the effect of interventions.'

[399] wayback.archive.org/web/*/hdm2008.erciyes.edu.tr
[400] www.zentralblatt-math.org/stmaz/en/?q=an%3A1152.91763

T147 W[401]
'"Paleoclimate Temperature Reconstructions: Implications for Climate Change," American Public Health Association Annual Meeting, San Diego, CA, October, 2008'
This seems really strange. Wegman gave many alcoholism talks in venues likely to have relatively few health professionals, but given a venue filled with health professionals, talked instead about the hockey stick.

T148 Ws③[402] {❹}
Wegman and Said authored this talk:
'"Text Mining and Social Networks: Some Unexpected Connections," Army Conference on Applied Statistics, Lexington, VA, October, 2008 10/23/08.'
The ack was:
- 'Dr. Walid Sharabati
- Dr. Angel Martinez
- Army Research Office (Contract W911NF-04-1-0447)
- Army Research Laboratory (Contract W911NF-07-1-0059)
- National Institute On Alcohol Abuse And Alcoholism (Grant Number F32AA015876)
- Isaac Newton Institute
- Patent Pending'

Alcoholism appeared once, p.14, in a minor way in 43 pages, and the presentation was really about text mining and cluster analysis, hence ③.
That might be arguable, but people should study the talk.

T149 W[403] 0/0 {⓬}
'"A Bipartite Graph Model of the Interaction between Alcohol Users and Institutions," NISS Workshop on Agent-Based Modeling, Research Triangle Park, NC, November, 2008' *November 20-21.*

[401] *Would the American Public Health Association care about paleoclimate? How many people in that audience were able to assess the credibility of this talk?*
[402] "ARO_Proposal\Text Mining and Social Networks Wegman ACAS" www.documentcloud.org/documents/550203-text-mining-and-social-networks-wegman-acas.html
[403] nisla05.niss.org/affiliates/agent-based-modeling200811/abm_home.html

T150 W [404] {❹}
‘“Text Mining, Social Networks, and Document Clustering,” Keynote Address, UAE Mathematics Day, Sharjah, UAE, April, 2009’

T151 Ws③[405] {❹}
“Document Clustering and Social Networks,” Keynote Address, Classification Society Annual Meeting, St. Louis, MO, June, 2009’[406]
The ack was:
‘This is joint work with Dr. Yasmin Said and Dr. Walid Sharabati.
Dr. Angel Martinez
Army Research Office (Contract W911NF-04-1-**0447**)
Army Research Laboratory (Contract W911NF-07-1-**0059**)
National Institute On Alcohol Abuse And Alcoholism (Grant Number F32AA01**5876**)
Isaac Newton Institute
Patent Pending’[407]

T152 ws[408] {6} this might have been orange, but got benefit of doubt
‘“Multi-Mode Social Networks,” Human Behavior-Computational Intelligence Modeling Conference*(sic)*, Oak Ridge, TN, June, 2009’
The correct conference title was
‘Computational Modeling and Interoperability Conference.’
Authorship was Sharabati, Wegman, Said. The abstract[409] had:
‘Network theory has become an important tool in the analysis of societies and entities. There has been discussion in the literature about dichotomous two-mode networks and methods to derive the one-mode network, but no research have yet contributed to the expansion of this problem and its applications. In this paper, we present a novel advanced mathematical methodology to manipulate multi-mode (higher dimensional) networks. In particular we study in detail three mode networks and present a generalized multi-mode model for networks. In network theory, a network may be represented by a matrix or graph. We focus our study on the matrix approach; we incorporate multi-dimensional tensor analysis to expand on vertex attributes and infinite vector space analysis to address network expansion in terms of vertices and edges. Finally, we utilize linear algebra to derive relational networks from the N’ (?)

T153 Sw[410] {6} this might have been colored orange.
‘“On Agent-Based Models of Multi-Mode Social Networks,” Swarmfest 2009, Santa Fe, NM, June, 2009.’ 06/28/09
The actual authorship was **Said, Wegman, and Sharabati**.
On p.35, it does have the ack:
‘This is joint work with Dr. Walid K. Sharabati and Dr. Yasmin H. Said. This work is based on Dr. Sharabati's Ph.D. dissertation.’
Indeed, most of the talk *seemed* directly from [SHA2008]. They also note:
‘Look for the Third International Conference on Social Computing, Behavioral Modeling and Prediction, SBP(2010)
To be held in Washington, DC March-April 2010
However, neither was in that program, nor 2011.

T154 W
‘“Multivariate Data Adaptive Compression and Density Estimation,” Joint Statistical Meetings, Washington, DC, August, 2009’

T155 W[411] {❹}
‘“Text Mining, Social Networks, and Document Clustering,” ISI 2009 Durban: 57th Session of the International Statistical Institute, Durban, South Africa, August, 2009’

[404] www.webcitation.org/6EXO8TYaL older.sharjah.ac.ae/English/Conferences/uaemathday08/Documents/The_7th_UAE_Math_Day_Program_2NC.pdf
[405] https://www.ideals.illinois.edu/bitstream/handle/2142/12608/WegmanKeynote.ppt.pdf?sequence=5
[406] www.classification-society.org/clsoc/meeting.php Joint with Interface.
[407] Use of this term in the US requires that the patent application has been filed: en.wikipedia.org/wiki/Patent_pending
Nothing was found at USPTO, but it could have been missed or withdrawn.
[408] This appears not to be online anywhere.
[409] https://par.cos.gmu.edu/node/1114

[410] ftp.swarm.org/pub/swarm/swarmfest/2009/MultiMode_Swarmfest.pdf
[411] www.stat.auckland.ac.nz/~iase/conferences.php?show=isi57 www.stat.auckland.ac.nz/~iase/publications/isi57/IASE%20IPM%27s%20for%20the%2057th%20ISI%20Session_updated_July_09.pdf
I cannot find conference materials that list Wegman, but incomplete.

T156 W[412]

'"Massive Data Streams and Citizen Science," Keynote Address, Tenth Islamic Countries Conference on Statistical Sciences, Cairo, Egypt, December, 2009'

Proceedings were edited by Zeinab Amin and Ali. S. Hadi.[413] Only the abstract seems available:

'Science has traditionally been built on two major paradigms: Theory and Experiment. These are often believed to be complementary one feeding into the other. More recently, Computation has become a third paradigm, i.e. exploring the possibility space computationally in a way that could never be done without modern computing resources. In some way Computation partially replaces, but also augments Experiment. The Google experience suggests that there is yet another paradigm emerging. This is a data centric perspective, where masses of data almost replace Theory, or, more precisely Massive Data partially replaces, but also augments, Theory with much the same relationship between Computation and Experiment. In this talk, I will give some examples of massive data streams including data coming from the Large Synoptic Survey Telescope. This instrument is expected to generate 100 petabytes of images of 50 billion astronomical objects. Classification of such objects is a task beyond both any single team and beyond the most sophisticated machine learning algorithms. There is no substitute for the human eye-brain classifier. Using the Internet, enthusiastic citizens can be drawn into the process of classifying these objects. Some examples of citizen science will be illustrated.'

That was the last talk found in the Wegman Resume, but various searches found more, assigned codes T400 and up.

T400 S☐[414] 2005.09 0/0 {⓾}

Said, Yasmin H. (2005) "Agent-based Model Applicable to Homeland Security and Disease Control" SAMSI/NISS, Triangle Park, NC, September.

Wegman was involved in organizing this.[415] *Said's talk seemed a real stretch for relevance to homeland security.* See ⑧ in **§S.3.2** and **§S.3.8.**

T401 S☐[416] 2005.10 0/0 {⓾}

Said, Yasmin H. (2005) "Adaptation of an Alcohol Ecological Agent-Based *(sic)* to Homeland Security" ACAS05 Monterey CA, Oct 19-21

See ⑧ in **§S.3.2** and **§S.3.8.** [417]

Abstract:

'This research is to instigate the prototype for developing a model framework for ethyl alcohol, which will provide an assessment for interventions, which are meant to minimize acute outcomes (intentional and unintentional injuries/death) that are influenced by the consumption of ethanol, without causing a financial or social burden, and imposing interventions that are ultimately ineffective (or even simply not cost effective). Our framework is ecological (individual agents and interactions are represented), stochastic (neither individual behavior nor consequences of interventions are certain) and very flexible. Constructing the framework raises deep issues in the domain science of alcohol, statistics, mathematics, and computer science. We have developed a time and space dependent stochastic digraph model of alcohol use and abuse. The model is intended as a social network model, which captures the dynamics of alcohol abuse and in particular the acute outcomes associated with alcohol abuse. The intent is to study potential interventions and investigate their effectiveness at reducing the overall prevalence of acute outcomes. Current interventions focus on one outcome at a time rather than simultaneously considering all outcomes. The work involves both sophisticated mathematics (stochastic digraphs) as well as intensive data collection. **It is clear that a similar model structure of social networks can be applied to terrorists' networks with the same ability to examine interventions in order to assess their effectiveness.'**

The reader might study that abstract carefully, especially the strong claim in the last sentence for the only alcoholism talk at that conference. This preceded the 5876 grant, so one might ask who paid for the travel. 0447? IFNA/ACAS? Said?

In **§S.3.1**, she claimed this for 5876 as having happened in 2006.

[412] www.scribd.com/doc/56159596/Volume-1-Conference-ICCS-X
www.scribd.com/doc/56160595/Volume-2-conference-ICCS-X

[413] Hadi is Principal with Wegman and Moustafa in dMining Technologies, **§H.3.**

[414] www.webcitation.org/6DtYmrzYr
www.samsi.info/workshop/2005-06-program-national-defense-and-homeland-security-opening-workshop-and-tutorials

[415] www.samsi.info/programs/2005-06-program-national-defense-and-homeland-security

[416] www.webcitation.org/6Dsb14iwo
www.armyconference.org/ACAS05/Abstract_Booklet_05.pdf p.13

[417] She listed this under Progress, so I treated it as a claim, *perhaps arguable*.

T402 w[418]

'Amy Braverman, Jet Propulsion Laboratory, **Edward Wegman**, George Mason University; Wendy Martinez, Office of Naval Research; Juergen Symanzik, Utah State University; Brad Wallet, Automated Decisions, "Using Image Grand Tour To Explore Multiangle, Multispectral Satellite Imagery"' *JSM2006, August 06-10, 2006, Seattle, WA*

Abstract:

'***Key Words:*** remote sensing, visualization, exploratory data analysis, grand tour

Remote sensing data are spatial arrays of p-dimensional vectors where each component corresponds to one of p variables. Applying the same R^p to R^d projection to all pixels creates d new images, which may be easier to analyze than the original because d < p. Image grand tour (IGT) steps through the space of projections, and d=3 outputs a sequence of RGB images, one for each step. In this talk, we apply IGT to multiangle, multispectral data from NASA's MISR instrument. MISR views each pixel in four spectral bands at nine view angles. Multiple views detect photon scattering in different directions and are indicative of physical properties of the scene. IGT allows us to explore MISR's data structure while maintaining spatial context; a key requirement for physical interpretation. We report results highlighting the uniqueness of multiangle data and how IGT can exploit it.'

T403 W[419] {❸}

'**Edward Wegman**, George Mason University; Faleh Alshameri, George Mason University, "Automated Metadata"' *JSM2006, August 06-10, 2006.*

Abstract:

'Homeland security implies searching massive databases for information involving possible terrorists and the threats they are likely to bring. Many of these databases include free-form text such as intercepted emails and transcripts of phone calls. The implication is that these massive databases are sufficiently large that they cannot be examined thoroughly by humans. Generally, metadata involve information about the format of the data, but not necessarily the actual content. Th e concept of automated metadata is to use data mining tools to extract features from the data and attach the features to the data as digital objects in the form of metadata. Thus, an investigator could search for specific datasets having some desired features. We have employed this notion with a dataset involving 16,000 articles gathered from CNN and Reuters. This work is joint with Faleh Alsham.' *(sic)*

T404 Ws[420]

'**Edward Wegman**, George Mason University, Department of Statistics, …; **Yasmin H. Said**, Johns Hopkins University; Shabib A. Alhadheri, SUNY Upstate Medical University, "Experiences with a Virtual Regionalization Model for Cardiac Surgery" *JSM2006, August 06-10, 2006, Seattle, WA*

Abstract:

'Key Words*:* hypoplastic left heart syndrome, risk factors, pediatric cardiology

Serious medical conditions are often redirected from regional medical facilities to more experienced and larger central facilities. However, additional risk is incurred by such transfers. Two university-based medium- sized referring pediatric cardiac surgical centers have implemented a virtual regionalization model in which one full-time surgeon operates at both institutions using a common management protocol. In this paper we analyze the risk factors for Hypoplastic Left Heart Syndrome patients using the Virtual Regionalization Model.'

Wegman did not 0447-claim this, so it was not tagged ①. No Said claim was found, so it was not tagged③, but may have been redacted from reports.

T405 Ws[421] 0/0 {❿}

'**Edward Wegman**, George Mason University, Department of Statistics, …; **Yasmin H. Said**, Johns Hopkins University, "A Policy Tool for Assessing Alcohol Intervention Strategies" *JSM2006, August 06-10, 2006, Seattle, WA*

Abstract:

'Key Words*:* acute outcomes, DALY, alcohol intervention, stochastic digraph

This paper establishes a modeling framework for alcohol abuse that allows evaluation of interventions meant to reduce adverse effects of alcohol overuse without the financial, social and other costs of imposing interventions that are ultimately ineffective (or even simply not cost effective). The framework is ecological (individual agents and their activities are represented), stochastic (neither individual behavior nor consequences of interventions are certain) and flexible. Constructing the framework involves interactions among the domain science of alcohol studies, statistics, and computer science.'

Wegman did not 0447-claim this, so it was not tagged ①.

[418] www.amstat.org/meetings/jsm/2006 p.225
[419] www.amstat.org/meetings/jsm/2006 p.263
[420] www.amstat.org/meetings/jsm/2006 p.280
[421] www.amstat.org/meetings/jsm/2006 p.423

T406 Ws[422] {❸}

'**Wegman, Edward J, with Said, Yasmin H.,** and Alshameri, Faleh**,** "On the Extraction of Endogenous Metadata for Text and Image Databases,"

This was a talk for ARL 01/31/07, similar to T130 or T136, but with Said included as 2nd author ahead of Alshameri. They ack'd:

- 'The Text Data was supplied courtesy of Dr. Angel Martinez and Dr. Wendy Martinez.
- The Image Data was supplied by NASA LaRC with the assistance of Dr. Amy Braverman.
- Dr. Jeffrey Solka and Mr. Avory Bryant were exceptionally helpful with ideas and tools related to the text mining.
- The work of Dr. Wegman was supported by the U.S. Army Research Office.' *(ARO was at least ack'd, if imprecisely)*

T407 sw③[423]

'Joost Bottenbley, ...; **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason University; Samer Ellahham, Innovative Medical Institute/Paragon Cardiovascular Foundation; David Anderson, *(sic)* "Relationship Between Competitiveness of Colleges and Levels of Tobacco Use" JSM2007 *(July29-August 2, 2007, Salt Lake City)*

Abstract:

'This paper describes the relationship between the compositeness of a school and the level of smoking and makes recommendations on which category of smokers anti-smoking campaigns can target to reduce the level of smoking the most for the least amount of money. Data was collected from 15 schools over a period of three years (f2003 to 2005) in the state of Colorado.'

This appeared in same session as P184, with some of same authors. Although arguably vague, it is labeled 5876-claimed via **§S.3.2,** although with the redactions, it is unclear whether she claimed this specifically**:**

'JSM2007, Salt Lake City, UT. I participated heavily in JSM 2008, (sic) the Joint Statistical Meetings. I organized four sessions. I gave a talk on Experiences with Congressional Testimony and co-authored several other talks listed below. (August 2007) …

. A list of my papers is given below:

■■ *(redacted)*

3. Wieczorek, William, Said, Yasmin, and Wegman, Edward (2008) …'

T408 s③[424]

'Fahad Bin Muhaya, Imam University, PO Box 84901, Riyadh, 11681 Saudi Arabia, ...; **Yasmin H. Said**, George Mason University, "Access Control Model for E-Learning System"

JSM2007 (July29-August 2, 2007, Salt Lake City)

Abstract:

'The educational system has gradually shifted from a face-to-face to an elearning system, which has become prevalent in advanced countries with the advance of information technology, and connection of global networks. Therefore, in this paper, we develop an access control model particularly for e-learning management system.'

This was in same session as P182. Comments in T407 also apply.

T409 S③[425]

The JSM2007 Abstract Book (July 29-August 2, Salt Lake City) shows

'Pro Bono Statistics and Public Policy

Yasmin H. Said, George Mason University, 4400 University Drive – Mail Stop 682, Department of Computational and Data Sciences, Fairfax, VA 22030

Key Words: pro bono statistics, volunteerism, public policy, global warming, climate change, hockey stick

Efforts to persuade Congress to enact legislation that affects public policy are constantly being made by lobbyists who are paid by special interests. While this mode of operation is frequently extremely effective for achieving the goals of the special interest groups, it often does not serve the public interests in the best possible way. **As counterpoint to this mode of operation, pro bono interaction with individual legislators and especially testimony in Congressional hearings can be remarkably effective in presenting a balanced picture. The debate on anthropogenic global warming has in many ways left scientific discourse and landed in political polemic.** In this talk I will discuss our positive and negative experiences in formulating testimony on this topic.' *JSM2007 (July29-August 2, 2007, Salt Lake City)*

This was given in the middle of her alcoholism postdoc and the evidence of serial plagiarism on her part is quite strong. Presumably, this was a subset of T414 S, given about a month later.

[422] "ARL Presentations\Aberdeen-ARL-Talk.ppt" www.documentcloud.org/documents/550204-aberdeen-arl-talk.html

[423] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.20

[424] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.54

[425] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.176

T410 sw③[426]

'Hadi Rezazad, ... **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason University, "A Statistical Social Network Approach to Computer Network Optimization"
JSM2007 (July29-August 2, 2007, Salt Lake City)

This was Rezazad's dissertation topic. Abstract:

'Computer network design involves consideration of both fault tolerance and efficiency. In many settings these design issues are antithetic. We review work on using social networks for optimizing LAN networks. We consider vulnerabilities of highly centralized and highly decentralized servers and use evolutionary computing techniques to scale the social network optimization methods to enterprise level networks.'[427]

T411 sw③[428]

'Walid Sharabati, George Mason University, **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason University, " Style of Author-Coauthorship Social Networks: Statisticians of Prominent U.S. Universities"
JSM2007 (July29-August 2, 2007, Salt Lake City)

Abstract:

'Key Words: social networks, coauthorship, cliques, blockmodeling, MDS clustering, preferential attachment In the past 20 years social networks have been used to analyze relations and ties among individuals of the same network and similarities between different networks in an attempt to obtain a better understanding on how societies interact. One of the applications of social networks is the author coauthor networks also known as the citation networks. This branch of social networks tries to answer the question of "who-wrote-with-who" and with what frequency. It also investigates other important features such as cliques, structural equivalence, MDS and CONCOR clustering. Social networks can be treated as directed graphs in which actors (individuals) are represented by vertices (nodes) while interactions between actors are represented by edges (ties) which may have weights. In this paper, we study in depth coauthorship social networks of statisticians from prominent U.S. universities.'[429]

This was in same session as T410. The abstract fits Sharabati's dissertation [SHA2008 §5.1-§5.3], some of which appeared in Wegman's Congressional testimony in reply to Rep. Stupak.[430] Some later appeared in P179, sometimes mis-cited with this title, and in which Said became 1st author and Sharabati moved to 3rd.

T412 sw③[431]

'Samer Ellahham, Innovative Medical Institute/Paragon Cardiovascular Foundation; **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason University; Rima Allaham, Paragon Cardiovascular Foundation. "Effect of Intravenous Insulin Infusion on Coronary Artery Disease" JSM2007 *(July29-August 2, 2007, Salt Lake City)*

Abstract:

'Key Words*:* clinical manifestations, CAD
Coronary Artery Disease (CAD) is the leading cause of death in adults in the United States. National health and nutrition surveys suggest that at around 14.0 million people in the United States have CAD. CAD can cause myocardial ischemia, which can be clinically stable or acute coronary syndrome. Atherosclerosis is an inflammatory state and the inflammatory markers; hs-CRP and SAA have been shown to be predictive of clinical manifestations and complications of coronary artery disease. Pro-atherogenic stimuli increase generation of reactive oxygen species and transcription factors. These phenomena occurring in the coronary artery present clinically as acute coronary syndrome. Insulin infusion have been demonstrated in clinical trials of acute myocardial infraction in diabetics and non-diabetics to be safe and efficacious.'

This appeared in same session as T413, **§Q.3, §S.3.1.**

[426] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.303
[427] www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.W.5.10.pdf 'DC wrote a full side-by-side and discussed the topic in [DEE2010p]: deepclimate.files.wordpress.com/2010/09/rezazad-wegman-social-network.pdf This seems an irrelevant injection of SNA terminology into computer networks.
[428] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.304
[429] www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.W.5.7.pdf
[430] www.webcitation.org/6Dv8qSu6B, called [WEG2006] elsewhere.
[431] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.419

T413 sw③[432]

'Samer Ellahham, Innovative Medical Institute/Paragon Cardiovascular Foundation; 23044 Winged Elm Drive, Clarksburg, MD 20871, [433] ...; **Yasmin H. Said**, George Mason University; **Edward Wegman**, George Mason University.. "Update on Cardiometabolic Risk Reduction: Role of the Endocannabinoid System Disease" JSM2007 *(July29-August 2, 2007, Salt Lake City)*

Abstract:

'Key Words: clinical effectiveness, rimonabant
Obesity is the most common nutritional disorder in western industrialized countries and arises from the accumulation of excess fat in the body from over consumption of fatty foods. The prevalence of obesity and diabetes is increasing dramatically in the Sates and worldwide. Obesity is a major risk factor for diabetes, cardiovascular disease, and metabolic syndrome. The metabolic syndrome is a precursor to cardiovascular disease and diabetes. The endocannabinoid system controls food intake via both central and peripheral mechanisms. The cannabinoid receptor antagonist effect has a broad impact on metabolism and is associated with several sites of action and mechanisms. The administration of CB1 antagonists has several clinical implications. Blockade of the endocannabinoid system has therapeutic potential in the management of obesity and cardiometabolic risk'

T414 S[434]

Said gave this unexpectedly-informative talk at GMU 09/07/07:

'Experiences with Congressional Testimony: Statistics and The Hockey Stick'

In August 2010 the file not only disappeared, but even its mention was deleted from the GMU symposium record as though it had never occurred [MAS2010a §A.11]. [435] This was not tagged ③ because it was unclaimed, unlike the earlier T409, *which presumably used a subset of this talk.*

T415 S③[436]

Said attended an ASA/NCAR workshop at NCAR in Boulder, CO in October 2007, along with Wegman, who gave T141, discussed in [MAS2010a §A.3]. *The others were well-published senior climate scientists and climate-experienced statisticians. As a young postdoc with neither relevant funding nor experience, her inclusion seemed very odd. Who paid for her time and travel? NCAR? ASA? ARO? NIAAA?*

T416 S③[437]

§S.3.2 included:

'and co-authored several other talks listed below. …
Symposium on Arabic Language and Computers, Riyadh, Saudi Arabia. (November, 2007'

I could find no mention of her in the program, but she claimed it, thus the ③ tag was added. **§S.3.2** included:

'Interest in **social network** and **text mining** work has allowed me to spend approximately two months in Saudi Arabia last Fall. There I made contact with a number of physicians concerned with drug and alcohol use in Saudi Arabia.'

SNA has many applications, but most have nothing to do with alcoholism. Text mining might be fit for 0059, but not 5876, and it is hard to understand why visiting Saudi Arabia applied to 5876. The value of contact with unnamed physicians is difficult to calibrate.

Alcohol, of course, is banned in Saudi Arabia.[438] *Was this 2-month trip done on personal funds? Charged to 5876? Charged to 0559? Or did someone else pay?*

[432] www.amstat.org/meetings/jsm/2007/pdfs/jsm2007abstractbook.pdf p.419
[433] Said's home address, as seen in [WEG2010a, p.10], but *likely* mis-edit.
[434] www.webcitation.org/5mtVnmG6W
[435] *18USC§1519 might eventually be relevant.*

[436] www.image.ucar.edu/Workshops/2007/ASAclimate/index.shtml Oct 26-27
[437] www.iscal.org.sa
www.webcitation.org/6DtnhjpHM
www.iscal.org.sa/files/iscal_program_english.pdf Nov 10-12
[438] news.bbc.co.uk/2/hi/middle_east/1160846.stm
Of course, alcohol still exists.
wikitravel.org/en/Saudi_Arabia 'In Riyadh, and many other places in Saudi, women are expected to be accompanied by a male relative in public, …'
The logistics of this trip are curious.

T417 Sw① [439] 0/0 {⓭}

'**Said, Yasmin H.** (2008) "Estimating Spatiotemporal Effects for Ecological Alcohol Systems," Interface 2008, Durham, NC'

This was presented 05/23/08, annotated:

'Joint work with Edward J. Wegman'

This was thus coded as "Sw". The ack was:

- 'The work of Dr. Said is supported in part by National Institutes of Alcohol Abuse and Alcoholism under grant 1 F32 AA01**5876**-01A1.
- The work of Dr. Wegman is supported in part by the Army Research Office under contract W911NF-04-1-**0447**.
- I gratefully acknowledge the assistance of students and colleagues:
  - Dr. Rida Moustafa
  - Mr. Walid Sharabati
  - Mr. Byeonghwa Park and
  - Mr. Peter Mburu.'

T417 and T418 were the two papers in a session at Interface 2008, organized and chaired by Said.

T418 ws ① [440] 0/0 {⓮}

This was presented by Wegman, although Wieczorek was lead author.

'Wieczorek, William F., **Said, Yasmin H.** and **Wegman, Edward J.** (2008) by Wegman "Spatial and computational models of risks for alcohol users," Interface 2008.'

The presentation was annotated:

'Joint work with Yasmin H. Said and William F. Wieczorek'

[439] niss.org/sites/default/files/Said_Interface_2008_Talk.ppt

[440] niss.org/sites/default/files/Wegman_Spatial%20and%20Computational%20Models%20of%20Risks%20for%20Alcohol.ppt \

In Wegman's resume, many papers/talks are listed as "with" other authors, so one needs to check the actual papers to find lead authorship.

T419 S [441] 0/0 {⓭}

'Said, Yasmin H. (2008) "Estimating Spatiotemporal Effects for Ecological Alcohol Systems," COMPSTAT 2008, Porto, Portugal, August 2008 167-178.'

Abstract:

'In this paper, I consider data on fatal automobile crashes, DWI arrests, and alcohol addiction admissions in Virginia, USA and use these as a basis for estimating the hourly, weekly, monthly, and annual cycles associated with alcohol consumption. In addition, I use surveys carried out by the Department of Alcoholic Beverage Control in Virginia to establish geospatial patterns of purchases of distilled spirits. This data analysis allows me to conjecture spatiotemporal patterns that can be incorporated into calibration of a more complex ecological alcohol systems model.'

This *seems identical to* T417,[442] ack'd:

'First, I would like to acknowledge Dr. EdwardWegman, my frequent coauthor and mentor, for his continuing support and stimulating discussions. I would like to acknowledge the contributions of Dr. Rida Moustafa and **my students, Peter Mburu and Walid Sharabati**,[443] who assisted me with the development of various figures and computations in this paper. I also acknowledge the contributions of John Rigsby whose ideas on allegiance in social networks provided me with insight in the analysis of the geospatial component of my discussion. My work is supported in part by Grant Number F32AA01**5876** from the National Institute on Alcohol Abuse and Alcoholism. The content is solely the responsibility of the author and does not necessarily represent the o±cial views of the National Institute on Alcohol Abuse and Alcoholism or the National Institutes of Health. I am a Visiting Fellow at the Isaac Newton Institute for Mathematical Sciences at University of Cambridge in Cambridge, England. I am in debt for the support provided by the Newton Institute, which has contributed to the successful completion of this work.'

[441] link.springer.com/chapter/10.1007/978-3-7908-2084-3_14
"Paper 193\COMPSTAT2008_paper_Said.pdf"
www.documentcloud.org/documents/550100-compstat2008-paper-said.html
This was slightly confusing, as P192 was #192 in Wegman Resume, but P193 was not this. T406 was described as joint work with Wegman, so this is tagged "Sw."

[442] Same talks were often given at Interface and then COMPSTAT.

[443] Sharabati was working on his PhD, co-supervised by Said.
Mburu seems not to have gotten his PhD and is not found at GMU.
Is it normal practice at GMU to for postdocs to supervise PhD students?

T420 Ws③ [444] {❺}

Wegman, Edward J. and Said, Yasmin H. (2008) "Mixture Models for Document Clustering" *Talk at U MD 10/30/08*

This was ack'd:

- 'Dr. Walid Sharabati
- Dr. Angel Martinez
- Army Research Office (Contract W911NF-04-1-**0447**)
- Army Research Laboratory (Contract W911NF-07-1-**0059**)
- National Institute On Alcohol Abuse And Alcoholism (Grant Number F32AA01**5876**)
- Isaac Newton Institute
- Patent Pending'[445]

T421 sw③[446]

'Said, Yasmin H. (2009) "Using Text Mining and Social Network Analysis to Study the Quran and Other Sacred Texts", 7th UAE Math Day, Sharjah, UAE, 04/25/09

This occurred a month before the end of 5876.

T422 sw[447] {❹}

Said gave this at ACAS 10/20/09. The abstract was unavailable.[448]

'Said, Yasmin H. (2009) "Approaches to Text Mining that Preserve Semantic Content"

T423 W

CNN Interview 12/07/09 on "Climategate"[449] is annotated [MAS2010a §A.3]. *Wegman spoke strongly without being very well-informed.*

T424 W★[450]

Wegman spoke 04/22/10 at a Chapman conference.[451] He repeated the same false statements(★) as in the WR, complained about blogs[452] and repeated common cherry-picked quotes from "Climategate," contradicting 2 of the formal[453] inquires that had already occurred.

Wegman's talk is 22:00-35:30, with a few quotes transcribed:

22:39 'Based on **tree-rings**'
Misrepresentation: it was multi-proxy.[454]

23:00 'Blue related to **tree-rings**, red supposed to be the actual temperature measurement'
Misrepresentation: multiproxy, and the red uptick was indeed derived from actual measurements, not proxies.

23:11- 'This is the zero-temperature anomaly. … According to this **tree-ring** reconstruction, … the anomaly was negative, in other words, it was colder than average. That doesn't make sense, if it's colder than average, and the reason it was colder than average in this reconstruction was that they took the centerpiece for this from … about 1902 … until 1995.'

The last statement above is false. MBH98/99 used the 1902-1980 average as zero-baseline, but Wegman's IPCC graph used 1961-1990, as shown on his own slide. Choice of display baseline is arbitrary and has zero input to any reconstruction. Most of the millennium was below the 1961-1990 average. Wegman makes no sense and seemed not to read his own slide. This talk has numerous issues, too many to discuss here.

---

444 www.statconsortium.umd.edu/hot_topics/DocumentClusWegman.ppt

445 In the US, this requires application to have been filed.

446 www.webcitation.org/6EXO8TYaL older.sharjah.ac.ae/English/Conferences/uaemathday08/Documents/The_7th_UAE_Math_Day_Program_2NC.pdf

447 www.armyconference.org/ACAS09/brief.htm It is tagged "sw" because similar talks named Wegman as coauthor or claimed joint work.

448 www.armyconference.org/ACASAbstracts.pdf

449 transcripts.cnn.com/TRANSCRIPTS/0912/07/ltm.02.html

450 ibc.chapman.edu/Mediasite/Play/80725145abc64c90817027331d2d35f4 *Microsoft Silverlight required.*, comments 01:16:30-01:18:30, argued against land footprint of solar and wind, advocated small, local nuclear. Wegman can express opinions, but is neither a climate scientist nor an energy-systems specialist.

451 www.chapman.edu/scst/news-and-events/past-events/beyond-copenhagen-conference.aspx April 21-23, 2010.

452 Deep Climate was the obvious blog, and comment about felony was likely an allusion to [MAS2010], but [MAS2010a] was still 5 months' away.

453 en.wikipedia.org/wiki/Climatic_Research_Unit_email_controversy
03/31/10 House of Commons Science and Technology Committee
04/14/10 Science Assessment Panel
Of course, the later inquiries more-or-less disagreed with Wegman, who seemed to continue channeling complaints from others.

454 "The network includes (Fig. 1a) the collection of annual resolution dendroclimatic, ice core, ice melt, and long historical records used by Bradley and Jones combined with other coral, ice core, dendroclimatic, and long instrumental records." Mann, Bradley, Hughes (1999).

T425 Sw③[455] {❹}
Said gave this at QMNDS 05/26/10:
"Preserving semantic content in text mining using multigrams."
She noted:
'This is joint work with Edward J. Wegman'
She ack'd:
'Dr. Angel Martinez Dr. Jeff Solka and Avory Bryant Dr. Walid Sharabati
Funding Sources
National Institute on Alcohol Abuse and Alcoholism (Grant Number F32AA01**5876**)
Army Research Office (Contract W911NF-04-1-**0447)**
Army Research Laboratory (Contract W911NF-07-1-**0059**)
Isaac Newton Institute'

T426 Sw was not a talk, but my label for a Said-organized session at Interface 2010, 06/06/10 that included Fred Singer [MAS2012a, §A.6.4].

T427 Sw was another Said-organized session at Interface 2010, whose 3 speakers were Jeff Kueter (George Marshall Institute), Don Easterbrook and Said herself [MAS2012a, §A.6.4] Her talk was
'Climate Change Policy and the Climategate Scandal
Yasmin H. Said, George Mason University
The release of emails from the East Anglia University Climate Research Unit just before the Copenhagen Climate summit has had a damaging effect on public support for action on global warming. **The lack of transparency by some climate researchers, the willingness to bend the peer review process, and the willingness to destroy data rather than share it with researchers of a different perspective all raise fundamental issues of climate change policy.** Perhaps the best thing to come from the climategate scandal is the formal recommendation of engaging statisticians. In this talk I will discuss some of the implications of climategate on climate change policy.'

T428 W
"Analysis of Global Warming Data: A Contrarian Data-Based View"
This was an 08/04/10 talk by Wegman at JSM2010.[456]

T429 S
Said organized this hockey-stick/climate session for JSM2010, but it was canceled just before the conference. The proposed speakers were then mostly Wegman GMU students (Bold), none with any obvious relevant expertise.[457] It was canceled just before the conference and it is unclear if it was ever real and if the speakers actually even knew about it.[458]
'Environmental Statistics—Topic-Contributed
Section on Statistics and the Environment, Scientific and Public Affairs Advisory Committee
Organizer(s): **Yasmin H. Said**, George Mason University
Chair(s): **Yasmin H. Said**, George Mason University
2:05 p.m. Analysis of the Hockey Stick—Nan Chen, **GMU**
2:25 p.m. Statistician's Perspective on Paleolimate *(sic)* Reconstruction—**Roger Shores**, **GMU** *(PhD 2011)*
2:45 p.m. Analysis of Climategate—**Duane King**, **George Mason University** *(PhD 2012)*
3:05 p.m. Analysis of Wegman Report, National Academies, and Others—Sukhaynah Said, **George Mason University** *(Yasmin Said's sister?)*
3:25 p.m. National Academies Analysis of the Hockey Stick and Global Warming—**Byeonghwa Park**, **George Mason University** *(PhD 2011)'*

[455] www.qmdns.org/2010/Talks/IS5_Said.ppt
www.webcitation.org/6Ckanykzm

[456] www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.A.6.6.pdf
[457] www.desmogblog.com/sites/beta.desmogblog.com/files/strange.scholarship.A.6.6.pdf
This group included an apparent relative of Said's and several Wegman students.
[458] Someone with subpoena power might ask them.

T430ws[459]

This conference paper was presented 08/31/10 in Toronto, ON

'Sharabati, Walid K., **Wegman, Edward J., Said, Yasmin H.** (2010) "Predicting Edges And Vertices In A Network" 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, DOI 10.1109/WI-IAT.2010.317'

It was an extract of Chapter 3 of Sharabati's dissertation[SHA2008]: "I" was changed to "we," "chapter" was changed to "paper," references and figure numbers were reformatted. A few introductory paragraphs were added, minor rewordings were done, typically in section introductions.

- II ← [SHA2008 3.1], minus a few paragraphs
- III ← [SHA2008 3.3], "method" ← "mechanism"
- IV ← [SHA2008 3.4], most of the paper

Most of the text and diagrams came from unattributed [SHA2008], but no allegation of (self)-plagiarism is made here. Sharabati was lead author, co-supervised by Wegman and Said.[460] *People can argue whether or not it would have been better form to ack the dissertation, but that is minor. Doing minor edits on a student's dissertation and then adding supervisors as authors may fall in an academic grey zone.*

T431 W[461]

This talk by Wegman *seemed unconnected with other work.*

"Understanding Afghanistan" ACAS, Cary, NC , 10/20/10-10/22-10

This was also found at JSM2011[462], but not noticed early enough to be given a code and added to graphs.

T432Sw③[463] {❹}

Said gave this talk at the ISI 2011 conference in Dublin 08/25/11:

"High Dimensional Vector Space Methods for Characterizing Semantic Content,"

This was identical to T425 ("Preserving semantic content in text mining using multigrams," presented a year earlier) except for deletion of pp. 14, 15, 22, 23, 24. This is an example of cases where the use of different names for similar talks impeded comparisons.

That ends the annotations of talks. Obviously many could have been missed. The striking characteristics were:

- Irrelevance of many talks to any of the grants.
- Seeming repetition of talks with little obvious progress.
- Many talks at "friendly" conferences like Interface or ACAS, managed by Wegman's IFNA.

[459] www.computer.org/csdl/proceedings/wi-iat/2010/4191/03/4191c071-abs.html

[460] In P200, a Sharabati chapter was converted to a paper that listed him 3rd.

[461] www.armyconference.org/ACAS10/brief.htm

[462] www.amstat.org/meetings/jsm/2011/onlineprogram/ActivityDetails.cfm?SessionID=206214 andrewgelman.com/2011/09/16/groundhog-day-in-august comments on talk

[463] isi2011.congressplanner.eu/showabstract.php?congress=ISI2011&id=790 PDF is attached there, but has no simple URL. The abstract calls this "Approaches to Text Mining that Preserve Semantic Content"

These Final Reports describe *substantial work with various coauthors that either clearly fit the grants or was at least plausible.*[464] *They do display some possible double-counting,* including 5 papers claimed for 3 separate grants. Some work might legitimately be funded by several concurrent grants, so this may or may not be a problem.

M. ARO DAAH04-94-G-0267– 07/15/94 – 10/01/97

M.1 Wegman Proposal – 11/12/93[465]

'Visualization Methods for the Exploration of High Dimensional Data'

M.2 ARO Award $240K – 07/05/94[466]

M.4 Wegman Final Report – 08/07/98[467]

Its 534 pages included many papers, although one problem did surface:

> **p.11** 'TR 128 Edward J. Wegman, Daniel B. Carr, R. Duane King, John J. Miller, Wendy L. Poston, Jeffrey L. Solka and John Wallin, Statistical software, siftware and astronomy, May, 1996, Published with discussion, *in Statistical Challenges in Modern Astronomy II*, (Babu, G. J. and Feigelson, E. D., eds.) New York: Springer-Verlag, (1997), 185-206.'

[MAS2012c §4.2] alleged plagiarism of several paragraphs from an earlier PhD dissertation. *Compared to most of the other plagiarism allegations, this seemed minor*, but it was the chronologically-earliest plagiarism found, and the text got re-used in Wegman student Alshameri's dissertation and their patent. Thus, plagiarism has been documented to GMU for 4 Wegman PhD students, and casual use of others' text as early as 1996.

N. ARO DAAG55-98-1-0404– 01/05/98 – 10/30/01

'Massive Data Sets: Visualization and Analysis'

N.4 Wegman Final Report – 08/05/02[468]

The claimed peer reviewed journal papers included:

7 Wegman was lead author
8 5 different lead authors with Wegman as coauthor
15 Total

Perhaps some of these were not actually in peer-reviewed research journals, from just a cursory check. Still by contrast, as *fit* papers 0447 claimed as peer-reviewed journal papers (***A***) only the following

1 ***P157*** w☐ This really was a book chapter and published in 2003.
1 ***P158*** w☐ This appeared in 2004
1 ***P160*** w☐ This also appeared in 2004
1 ***P173*** w☐ published within grant, but originally submitted 07/28/04
1 ***P191*** w☐ published after grant, but submitted in late 2005
Wegman was 3rd author.
5 Total, of which 3 appeared before 0447 started, and the work on the other 2 seemed to have been done or mostly done earlier.

The contrast seems striking between the earlier grant, which produced many journal papers that seemed peer-reviewed, and 0447, which produced none except possibly ***P191.***

[464] *Some of these topics are familiar*. I had managed some data mining applications at Bell Labs, networking software development at several other companies, and as a Chief Scientist at Silicon Graphics was heavily involved in Big Data applications. bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/

[465] www.desmogblog.com/sites/beta.desmogblog.com/files/DAAH04-94-G-0267-1-Proposal.pdf

[466] www.desmogblog.com/sites/beta.desmogblog.com/files/DAAH04-94-G-0267-2-Award_0.pdf

[467] www.desmogblog.com/sites/beta.desmogblog.com/files/DAAH04-94-G-0267-4-Final.pdf

[468] GMU, 08/05/02, 278 days after end of contract
www.dtic.mil/dtic/tr/fulltext/u2/a414241.pdf
www.desmogblog.com/sites/beta.desmogblog.com/files/DAAG55-98-1-0404-4-Final_0.pdf

O. ARO DAAD19-99-0314 – 01/08/99 – 12/31/02

O.4 Wegman Final Report – 07/21/03[469]

'Nonparametric Function Estimation and Visualization Application to C2' This may also be worth examining for contrast with 0447. Wegman claimed 13 papers in peer-reviewed journals. *At least from a cursory inspection, most papers and talks plausibly fit that proposal.*

> 'This project focused on the development of fast, accurate density estimation procedures. The methods raised basic research issues as to the implementation, computational complexity, visualization, and optimization of estimators in this class. In addition to being useful in a direct role, it is argued that density estimation plays a crucial role in clustering algorithms, discriminant methods and pattern recognition. All of these methods are used extensively in Situation and Informational Awareness and Understanding and in Monitoring and Discovery Processes. In addition, because of their intuitive appeal and ease in understanding, visually rendered density and function estimators provide a natural format for human-computer interactions with decision makers. This report describes results related implementation, computational complexity, visualization, optimization and application of recursive orthonormal density estimators.'

On p.2, 'Manuscripts Submitted but not yet Published' claimed papers also claimed by 0447:

P157 'Solka, J. L., Wegman, E. J., and Marchette, D. J. "Data mining strategies for detection of chemical warfare agents," to appear *Statistical Data Mining and Knowledge Discovery'*
P158 'Marchette, D. J. and Wegman, E. J. "Statistical analysis of network data for cybersecurity," to appear *Chance*.'
P159 'Chow, Winston, Wegman, E. J. "Modeling continuous time series driven by fractional Gaussian noise," to appear *Institute for Mathematics and its Applications Monographs*
P165 'Wegman, E. J. "On some statistical methods for parallel computation," to appear *Handbook of Parallel Computing and Statistics'*
P168 'Marchette, D. J., Wegman, E. J., and Priebe, C.E. "A fast algorithm for approximating the dominating set of a class cover digraph," submitted to *Journal of Computational and Graphical Statistics'*
P176 'Dorfman, A. H., Lent, Janice, Leaver, S. G. and Wegman, E. J. "On sample survey designs for consumer price indexes," with submitted.' *(sic)*[470]

[469] www.desmogblog.com/sites/beta.desmogblog.com/files/DAAD19-99-0314-4-Final_0.pdf 07/21/03 (202 days after completion)
[470] This paper's relationship to the ARO contract is unclear.

P. AF OSR F49620-01-1-0274 - 04/15/01-10/14/03[471]

P.4 Wegman Final Report – 01/31/04[472]

'Intrusion Detection Using Data Mining Techniques'
On pp.12-14 it claimed papers as follows, also 0447-claimed**,** and all but P165 published before (or within a month) of 0447's start.

P157 #12, already published
P158 #16, already published
P159 #15, to appear[473]
P160 #14 to appear
P162 #17 to appear
P165 #13, to appear
P166 #18 to appear (published in book the month after 0447 started)
P168 #19, submitted.

P157, P158, P159, P165, P168 were thus claimed for at least 3 grants.[474]

With multiple overlapping grants, and work in progress, it is not unreasonable to list forthcoming work, especially if known to be accepted. However, it certainly makes it harder to assess the work actually done.

The topic of overlapping grants is complex.[475]
Seeming overlap may be improper or may be just fine.

*Work done **after** the grant period might fairly ack that grant, to give credit appropriately, even if a paper was unmentioned for credit in the Final Report. Of course work completed **before** a grant started should not be claimed.*

[471] www.dtic.mil/dtic/tr/fulltext/u2/a421061.pdf
[472] www.desmogblog.com/sites/beta.desmogblog.com/files/F49620-01-1-0274-4-Final_0.pdf
[473] This is Chow and Wegman, correct here in **§O.4** and **§P.4**, but then wrong in Wegman resume and 0447 final report.
citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.87.8501
[474] The 0314 and 0274 pair overlapped, but 0447-claim is *odd*.
[475] www.nature.com/news/funding-agencies-urged-to-check-for-duplicate-grants-1.12317

Q. ARO W911NF-04-1-0447 (1) 11/01/04-04/30/08

Q.1 Wegman Proposal – 07/29/03[476]

'Analytical and Graphical Methods for Streaming Data with Applications to Netcentric Warfare' *submitted by GMU on behalf of Edward Wegman.*

Abstract (p.C-1)

'This proposal focuses on a **new data structure**,[477] namely massive streaming data. We propose to develop recursive algorithms and evolutionary graphics for handling massive streaming data. Streaming data is essentially a new data acquisition paradigm, in which data becomes constantly available. Older data has less value and therefore must be discounted. Strategies for discounting are proposed as well as strategies for multi-scale resolution of data streams. The particular example we have in mind is streaming Internet packet headers, although theoretical and practical results will not be limited to this form of data. However, network traffic data are especially important to military and the U. S. Army in particular as Netcentric Warfare and joint operations between services and with allies become increasingly important. The ultimate goal of these techniques is to detect intrusion and fraud in streaming data systems.'

pp.D-1 to D-13 described the work in detail, [478] ending with:

'• I propose what I like to call *evolutionary graphics*, i.e. graphics, which evolve as a function of new data being added. The combination of recursive algorithms and evolutionary graphics will provided *(sic)* a fundamental approach for analyzing streaming data.

• I propose to develop *scalable graphics devices*[479] (pan and zoom) for discrete data that has many more points than screen resolution would allow.

• I propose to investigate the theoretical properties and implementation of the quantizing algorithm for truly massive streaming data .

• I propose to investigate recursive kernel density estimators in the critical multidimensional case.

• I propose to investigate adaptive mixtures density estimation algorithms using orthonormal bases such as wavelets to reduce or eliminate spurious terms.

• I propose to investigate classical exponential smoother with adaptive time scaling and their potential use as multiscale data analysis.

• I propose to develop evolutionary graphics tools including waterfall diagrams, transient geographic mapping, and related multivariate pan and zoom methodology.' *This seems redundant with the first two bullets.*

These were specific proposals. *It is unclear how many of the papers actually attacked these problems, but such assessment is ARO's purview.* However, about 40% of the works *(and likely much more than 60% of the effort, given that needed for T126)* were *obviously 0447-unfit***.**

p. G-1 lists other grants, but without numbers.

- "Intrusion Detection Using Data Mining Techniques"
  Air Force Office of Scientific Research (AFOSR)
  $470,275 04/15/01-10/14/03
- "Adaptive Data Cube for Integrated Sensing and Processing"
  Johns Hopkins University (Prime: DARPA)
  $450,000 07/01/01-06/30/04
- "Analytic and Graphical Methods for Streaming Data with Applications to Netcentric Warfare"
  Office of Naval Research (ONR)
  $284,601 PENDING *This seems duplicate or not awarded.*

p. 25 describes proposed $217K budget for 3 years, essentially:

$107,218	Wegman, Summer/Part-time, 0.22 FTE,
$18,000	Doctoral student
$5,996	**for "Domestic travel 1 Trip/Yr, 3 trips total**
	(other costs)
$153,643	total direct costs, including the above, fringe, travel, students
$63,358	indirect costs (45% of modified direct costs implies on-campus research by comparison with p33.)
$217,000	total request from sponsor, direct plus indirect

Thus, ~29.2% of the funds were paid to GMU for overhead.
Did GMU thus have any role in monitoring use of the funds?

---

[476] www.desmogblog.com/sites/beta.desmogblog.com/files/W911NF-04-1-0447-1-Proposal_0.pdf

[477] *It is unclear why experienced people would think this a new data structure in 2003. This kind of application was one of those for which Silicon Graphics(SGI) built especially powerful input/output systems 10 years earlier.*

[478] Via a quick read, *this research proposal seems at least plausible*. I used to work with companies doing commercial fraud detection work on SGI parallel servers, examining streaming data, from mid-1990s onward, during which "Big Data" was a major theme at SGI: en.wikipedia.org/wiki/Talk:Big_data
For example Google: silicon graphics fraud detection 1998. The CIA, NSA, etc, had similar problems. Whether this proposal was groundbreaking or not, *it seemed at least plausible*, especially given Wegman's past work.

[479] The novelty of this is unclear, given the state of 3D graphics 10 years before.

Q.2 ARO Award $217K – 09/29/04[480]

p.2 showed $70K, ($72K, $75K options) for 3 years, total $217K.
pp. 3, 6, 9 show the payment schedules for the grant and renewals:

	Original Award		Award MOD 1		Award MOD 2	
1	11/08/04	$17,500	12/15/05	$18,000	12/01/06	$18,750
2	02/08/05	$17,500	03/15/06	$18,000	03/01/07	$18,750
3	05/08/05	$17,500	06/15/06	$18,000	06/01/07	$18,750
4	08/08/05	$15,750	09/15/06	$10,800	09/01/07	$16,875
5	on approval	$1,750		$7,200		$1,875

p.11 extended the end date from 10/31/07 to 04/30/08, at zero extra cost:
'in accordance with the Recipient's letter dated 17 October'
That gives the following comparison:

Original	Revised	
11/01/04	11/01/04	Start
10/31/07	04/30/08	End
01/29/08	07/29/08	Nominal Final Report
-	12/10/08	Actual Final Report

Grant extension is not unusual, but the timing might be compared to accomplishments at original end date, when the only fit peer-reviewed journal paper actually published after the start date was ***P173***, and it had originally been submitted before the start. Year 2007 produced no relevant papers, only a few talks**.** Year 2008 saw a burst of possibly 0447-fit activity, but only ***P191*** was published in a peer-reviewed journal, and it had actually been submitted originally in late 2005, placed there on **§0.1.** Almost no fit peer-reviewed journal work was done in 3.5 years.

Work started on the WR in September 2005 and continued through July 2006 as part of T126. Wegman several times mentioned concerns about paying his mortgage, so perhaps this regular funding was important to him. Charts **§0.1** and **§0.3** show substantial diversion from fit work (blue) into unfit (green, orange, red) during 2006-2007. The original award seemed to be used for plausible activities, but MOD 1 and MOD 2, not at all, as they were totally dominated by unfit works.

[480] www.desmogblog.com/sites/beta.desmogblog.com/files/W911NF-04-1-0447-2-Award_0.pdf

Q.3 Wegman Progress Report 05/07/07[481]

The 2010 FOIA included a 3-page progress report:

'Accomplishments
- Graduated six Ph.D. students focused on these topics, 2004-present.
- Graduated 1 M.S. student, 2005.
- Published two books, 2005, 2006.
- Filed new patent application, 2007.
- Published 10 papers, 2005-2006.[482]
- Chaired Committee on Applied and Theoretical Statistics, National Academy of Science, 2004-2007.
- Member of the Board of Mathematical Sciences and their Applications, National Academy of Science, 2004-2007
- Elected to Board of Directors, American Statistical Association, 2006-2008.
- Assumed financial management of the Army Conference on Applied Statistics, 2006.[483]
- Program Chair, Quantitative Methods in Defense and National Security Meeting, 2007.

That sounds find, but would anyone at ARO realize that Wegman had spent a substantial amount of time on T126 and other unfit work during this time?

One expects a senior researcher to be involved in a variety of activities.[484] *Still, many of them had little to do with 0447.* Wegman also was:

- Treasurer of IFNA and involved with INFA conferences, **§H.1**
- a principal of dMining Technology, **§H.3**
- setting up *WIREs:CS* for Wiley, **§H.6**
- apparently working on a hockey-stick book for Wiley, **§H.7**

[481] "ARO PSS Report\AROReport_5-6-2007.doc" www.documentcloud.org/documents/550096-aroreport-5-6-2007.html
[482] P168-P176 totals 9, so he was *likely* counting one of the talks, making 10 reasonable. However, 3 were in a book co-edited by Wegman (P168, P169, P170), 3 were unfit. Of the rest (P172, P173, P175), only P173 was published in a peer-reviewed research journal.
[483] This was done via IFNA, **§H.1.**
[484] If a senior researcher gets funds from several contracts, and some of the money goes towards the overall research discipline, *I personally would not begrudge some such usage*, but it is a gray area for funders to evaluate. *In this case, some activities seemed reasonable, some seemed not, but many were clearly unfit.*

Q.4 Wegman Final Technical Report – 12/10/08[485]

Abstract, p.1

'This project began with a strong interest in streaming data with particular attention to intrusion detection in computer networks. The netcentric battlefield communications was a particular motivation. Several different techniques for univariate and multivariate probability density estimations were developed with recursive updating. An ability to detect subtle shifts in Internet traffic patterns using streaming Internet headers and the recursive density estimators was demonstrated. We also investigated text streaming data and developed methods for topic identification using mathematical representations of text documents. Finally we have noted the connection between two-mode social network analysis and latent semantic indexing.'

pp.5-10 started:

'List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:'

As shown in **§K**, this Final Report:

- Made at least 30 false claims of 0447-unfit work, and Wegman later ack'd 0447 in 4 more, **totaling 34 false claims**. . **→§1.1**
- Claimed T126 and related papers and talks, which were not only 0447-unfit, but contradicted "pro bono" claims by Wegman and Barton, show them to have been **false statements**. **→§1.2**
- Made at least **25 false statements**, claiming that papers or talks ack'd 0447, but did not. **→§1.3.**
- In claiming the WR, another Wegman statement was shown to be a **false statement** to Rep. Henry Waxman. **→§1.4**

For simplicity, 0447-fit and 0059-fit were treated as a combined category, as one part of 0059 might be considered to cover 0047 topics.

[485] Final Technical report (224 days after project completion) www.desmogblog.com/sites/beta.desmogblog.com/files/W911NF-04-1-0447-4-Final_0.pdf

- p.10 showed:

Graduate Students

NAME	PERCENT_SUPPORTED
Yasmin Said	0.25
Eun Noh	0.50
FTE Equivalent:	**0.75**
Total Number:	**2**

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Yasmin Said	0.25
FTE Equivalent:	**0.25**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Edward Wegman	0.33	No
FTE Equivalent:	**0.33**	
Total Number:	**1**	

Support for Wegman had risen from 0.22 (in proposal) to 0.33.
Yasmin Said received her PhD Spring 2005 [SAI2005][486] on alcoholism.

10/31/04-05/04/05	Grad student, 0447 work implied by above
==========	Postdoc – at least one of (a), (b) or (c)
05/05/05-Summer 05	(a) Some postdoc work at GMU 0447?
Fall-05-Spring 06	(b) At JHU, most of WR work done, 0447?[487]
05/26/06	Kirschstein Fellowship started
Summer 06-	(c) Back at GMU, 0447?

She coauthored none of the fit 0447-claimed papers, and just a few of the talks that were found,[488] *so her 0447 work is unclear, as is any use of 0447 funds in conjunction with the WR,*

[486] AGENT-BASED SIMULATION OF ECOLOGICAL ALCOHOL SYSTEMS, web.archive.org/web/20060905150733/http:/www.galaxy.gmu.edu/stats/syllabi/IT871/MasterCopyDissertation.pdf This disappeared from server in August 2010.
[487] She has claimed her work on the WR was unpaid, but that may or may not be true.
[488] Many talks could not be found and some may have said "joint work" rather than listing Said as coauthor.

R. ARO(ARL) W911NF-07-1-0059 (2) 12/15/06-12/15/07

This proposal to ARL was administered by ARO and maybe ONR.[489] Earlier discussion is found in [MAS2010a §A.7], revised here.

R.1 Wegman Proposal – 11/27/06[490]

p.3 'Adaptive Multi-modal Data Mining and Fusion For Autonomous Intelligence Discovery

My proposal addresses the challenges of autonomous discovery and triage of contextually relevant information in massive, complex, dynamic text and imagery streams. I will develop a prototype system to mine, filter and fuse multi-modal data streams and dynamically interact with the analysts to improve their efficiency through feedbacks and autonomous adaptation of the algorithms. I plan to implement four core capabilities:

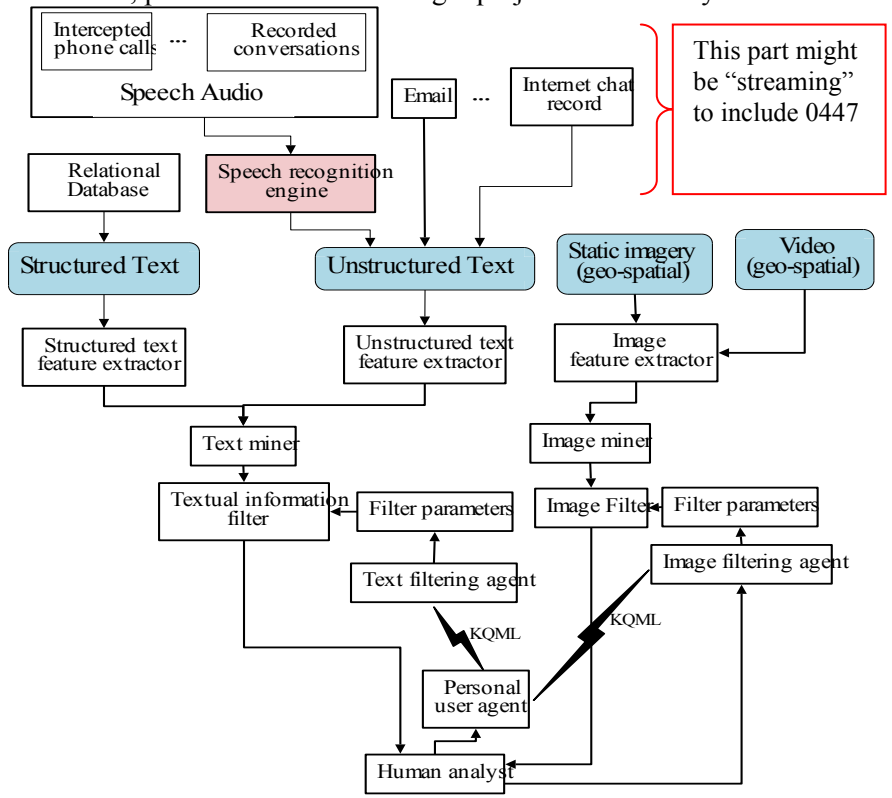
- Text and image mining for feature extraction[491]
- Multi-modal data fusion[492]
- Agent-based adaptive information filtering[493]
- Cognitively friendly information visualization.[494],

The 2010 FOIA found another proposal,[495] by the Space Dynamics Laboratory at Utah State University Research Foundation for the National Geospatial-Intelligence Agency (NGA), in response to HM1582-05-BAA-0004 solicitation[496] in May 2005. Wegman had coauthored at least 6 papers with Jürgen Symanzik,[497] whose work was relevant. They[498] *very likely* had written 7 pages of the proposal in 2005.[499]

The 0059 proposal copied pp.2-3 and 1st paragraph of p.4 of the NGA proposal, edited "We" to "I" and removed section numbers. It deleted 7 references, including 3 to Braverman, added another, but failed to edit out NGA:

p.4 'A special case of interest to NGA is that found in persistent …'

The proposal did not change the block diagram, introduction, or headings. As a result, p.3 described a much larger project than actually in the text.



People often legitimately rework their own earlier rejected proposals, so no misconduct is alleged here.[500] *For an extensive project, it is good to design an overall architecture, but then specify the parts actually expected to be done. When much of the staffing is graduate students, unpredictability can be even higher than with regular employees.*

The later progress report was more specific, **§R.3.**

---

[489] The FOIA request was originally directed to ARL, but they said that this belonged with ARO and kindly sent the request there.

[490] www.desmogblog.com/sites/beta.desmogblog.com/files/W911NF-07-1-0059-1-Proposal_0.pdf
Proposed start date was 09/01/06. That was typo or this was earlier version.

[491] Described in the 0059 text.

[492] Partly described, but "Data fusion of text and imagery" onward was gone

[493] Agent framework was in section 2.2.2 of earlier proposal, gone

[494] Visualization was given several paragraphs in earlier proposal, gone.

[495] "ARL Presentations\ARL Project Report\NGA_proposal_version_final.doc"
www.documentcloud.org/documents/541257-nga-proposal-version-final.html

[496] www.fbodaily.com/archive/2005/05-may/28-may-2005/FBO-00816616.htm

[497] www.math.usu.edu/~symanzik

[498] Of 10 references, 3 included Wegman, one with Symanzik.

[499] That proposal was incomplete. It may never have been finished, or it may have been finished but not accepted. No evidence was found of it elsewhere.

[500] However, readers might want to evaluate the quality of care in proposing this.

R.2 ARO Award $100K –12/22/06[501]

p.1 The ARO grants specified GMU as the contractor, saying also:
'This grant award supports the research identified in the Recipient's proposal'

p.2 listed Terms and Conditions document, **§B.1**, and said that payment would be by reimbursement, as opposed to the recurring payments in 0447. ARO would have those records, but they are not public via FOIA.

p.3 showed that $30K of the $100K went to GMU overhead:

EXHIBIT A – Budget

Item Description	12 Month Period
Direct Costs	
Salary	64,713.00
Other Direct Costs	
Travel	5,022.00
Indirect Costs	
Overhead	30,265.00
Total	100,000.00

The proposal shown in **§R.**1 just requested $100K, but had no budget details and seemed to have been created before 09/01/06. *Most likely* some later revision provided the details from which this exhibit was derived. Neither that nor this one mentioned Said explicitly, but she did claim several works for 0059, so perhaps some salary money went to her, amidst her alcoholism-modeling fellowship.

[501] www.desmogblog.com/sites/beta.desmogblog.com/files/W911NF-07-1-0059-2-Award_0.pdf

R.3 Wegman Progress Report 05/07/07

A presentation with modify date 0505/07[502] mentioned 6 grad students, of whom 4were included in Final Report.[503] Only Hohman got her PhD under Wegman.

Name[504]	Final	PhD	Advisor(s)
① Alshammari, Eiman Tameh	Y	2010 (IT)	Lin (CS)
② Mihai, Felix	Y	2012 (?)	Griva (CDS), Seshaiyer(Math)
③ Youn, In-ja	Y	not yet (CS)	Mark (CS), Richards (EE)
④ Mburu, Peter	N	not found	Claimed as Said student(?)
⑤ Hohman, Elizabeth Leeds	Y	2008 (CS&I)	Wegman (CDS)
⑥ Alsumait, Loulwah	N	2010 (CS)	Barbará,Domeniconi (CS)

p.1 'Adaptive Multi-modal Data Mining and Fusion For Autonomous Intelligence Discovery - **Edward J. Wegman, Ph.D. Yasmin H. Said, Ph.D.**'

p.2 'Outline of Presentation
•Problem Description
• Background in Text Mining
• Outline of System
• Arabic Language Tool
• Geospatial Tool
• Integration of Text and Images
• Streaming Documents'

p.22 'Four core capabilities:
•Text and image mining for feature extraction
• Multi-modal data fusion
• Agent-based adaptive information filtering
• Cognitively friendly information visualization'

p. 24 highlights (red) the actual work expected (not just the overall architecture), starts first discussion found in 0059 documentation of Arabic tool not mentioned in Proposal).

[502] "ARO PSS Report\Report4ARL.ppt" www.documentcloud.org/documents/550202-report4arl.html
[503] This often happens, but *some oddities raise possible concerns.*
[504] en.wikipedia.org/wiki/Al- Al-names are often used inconsistently.

p. 25 'Arabic Language Tool
•Our fundamental premise is that Arabic language documents, open source and otherwise, provide valuable insight.
•Open source documents are streaming.
•Not enough Arabic language experts are available to translate everything.
•We need a system for English language queries to an Arabic language text database.'

p. 28 'Arabic Language Tool Status
•**Native Arabic speaker, Eiman Alshammari①, is our graduate student developing tool.**
•We met with the Arabic Language Data Mining Group in Cairo and secured cooperation and an Arabic language corpus.
– Professor Aly Fahmy, Dean of the Faculty of Computers and Information, Cairo University.
– Dr. Amir Atiya, Associate Professor of Computer Engineering, Cairo U…
– Dr. Ahmed S. Moussa, Program Manager, Smart Village.
• We met with representatives of King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.
– Dr. Turki Saud Mohammed Al-Saud, Vice President Research Institutes
– Dr. Mansour M. Alghamidi, Director, Computers and Electronics
– Dr. Ibrahim A. Al-Kharashi, Arabic Language Projects
Project is underway … Eiman is anxious to graduate'[505]

p. 29 'Geospatial Tool
•Status
• **Felix Mihai② and In-ja Youn③** are graduate students developing this tool.
• The basic map functionality is available
• IP locator is underway
• Geospatially located satellite image database is also available (MISR imagery)
• Graduate student funding is a problem for Felix in particular

[505] *This 2007 comment seems odd in light of PhD in 2010 under Lin.*

p.38 'Integration of Text and Images
•Status
– **Peter Mburu④** is the graduate student identified to work on this part of the project
• Work has just begun … this is a hard problem.[506]
• Peter is very bright, but not yet in candidacy.'

p.45 'Streaming Documents
•Status
• Graduate students, **Elizabeth Leeds Hohman⑤ and Loulwah Al-Samait⑥,** *(sic, Al-Sumait)* are separately working on streaming documents.
  • **Elizabeth is developing a visual representation using graph theory of streaming document clusters.**
  • **Loulwah** is developing a method for understanding evolving sense of documents.
• Theory development is relatively advanced, system development is less so.
• Project has been underway about 4 months.

p.46 'Work Left to Be Done!
•Lots!
•Progress is good and a number of bright students are working on the project
•The Arabic Text Tool should be in hand by December.
•The Geospatial Tool is fairly advanced, but Felix has no funding and is fragile.
•The Text and Image Integration is at early stages and is probably the most difficult conceptually.
•The Streaming Text Tools are advanced theoretically, but system development is not yet underway
•Filtering tasks and system integration has not yet begun.
•But, we have only been at it for four months.'
p.47 'Contact Information[507]
Edward J. Wegman, Ph.D. … ewegman@gmail.com
Yasmin H. Said, Ph.D. ysaid99@hotmail.com'

[506] *Yes, very hard*. This has grown into a major area of research and development in academe, government and industry. Large teams have spent years working on this. Google: research integration text image.
It seems *slightly odd* that work had just begun, halfway through the grant.
[507] Said seemed to use only her the hotmail account for everything, but it is *curious* that Wegman used gmail.com for official reports to ARL, rather than gmu.edu, which he had used consistently through 2004.

Mburu and Alsumait are covered here, the others in the next section.

④ **Peter Mburu+**
A search by GMU PeopleFinder for "Mburu" found no one.[508] A search of GMU [509] gave 6 hits, showing him as teaching assistant in Computer Science in 2002, and an instructor in 2004, 2005. He was also listed as a research assistant for the Kellar Instructional Handheld System.[510] P185 was Mburu, Peter K., Said, Yasmin H. and Wegman, Edward J. (2007) "Temporal statistics for consequences of alcohol use," Proceedings of the Joint Statistical Meetings, 2005-2009.[511] That was the only Mburu in Wegman's resume. The GMU Computer Science Department listed PhDs in CS and Information Technology 2008-2011, but he did not appear.[512]

T417 was Said, Yasmin H. (2008) "Estimating Spatiotemporal Effects for Ecological Alcohol Systems," Interface 2008, Durham, NC, 05/23/08:
> 'I would like to acknowledge the contributions of Dr. Rida Moustafa **and my students, Peter Mburu and Walid Sharabati**, who assisted me with the development of various figures and computations in this paper.'

T419 (2008) was talk of same title, by Wegman and Said, with similar ack.

It seems slightly strange that a young postdoc would be supervising students to work on her alcoholism papers.

⑥ **Loulwah Alsumait**
She got her PhD in CS in 2010:[513]
> 'Dissertation Title: Online Topic Detection, Tracking, and Significance Ranking Using Generative Models
> Codirector: Daniel Barbará, PhD' Codirector: Carlotta Domeniconi, PhD'

LinkedIn gives some history:[514] BS CS 1989-1994, MS CS 1995-1999, from Kuwait University, PhD 2004-2009 GMU. She gave a lecture in 2009,[515] showing affiliation with Kuwait University, and wrote a book chapter with Domeniconi.[516]

[508] peoplefinder.gmu.edu/index.php?search=mburu&group=all&x=0&y=0
[509] search1.gmu.edu/search?q=mburu&site=mason_test&client=mason_test&proxystylesheet=mason_test&output=xml_no_dtd&as_dt=i
[510] masonlife.gmu.edu/assets/docs/kihd/kihd_system/KIHdSYSTEM_white_paper.pdf White paper, September 2007. cehd.gmu.edu/assets/docs/kihd/kihd_system/KIHdBrochureFinal.pdf
[511] So, he did get a presentation, but not on anything 0059-relevant.
[512] www.cs.gmu.edu/news/report/2011.pdf pp.6-7.
[513] www.cs.gmu.edu/news/report/2011.pdf p.6.
[514] www.linkedin.com/in/lalsumait
[515] videolectures.net/loulwah_alsumait; with Barbará, Domeniconi, James Gentle
[516] www.scribd.com/doc/53770544/4/Loulwah-AlSumait-and-Carlotta-Domeniconi

R.4 Wegman Final Technical Report, 03/18/09[517]

Wegman and Said claimed or ack'd at least 3 0559-unfit papers, of which: P178, P179, (retracted for plagiarism), P200 (plagiarism alleged).
This totals **3 false claims➔§2.1**.

Claims

The claims at end were a bit ambiguous, unlike the specifics for 0447:

> **p.10** 'Several papers were developed based on research carried out under this project. These papers exploited aspects of the developments here. They are Said et al. (2007❷, 2008❶), Said and Wegman (2009)❸, and Wegman and Said (2007) ❹. Presentations were given in a number of forums that credited this contract.❺'

These claims, including the vague ❺, fail ARO requirements, **§B:**

> 'b. "The Report Documentation Page (SF298) Continuation Sheet (Enclosure 2)" or a plain piece of paper **must include brief, but complete, information for each of the following categories:**
> (1) Submissions or publications under ARO sponsorship **during this reporting period**. **List the title of each and give the total number for each of the following categories**: …'

An October 2010 FOIA file[518] included one more page (labeled p.11), which added important references beyond those provided by ARO:

> TRIANGLE, A Two-Dimensional Quality Mesh Generator and Delaunay Tiangulator, http: //www.cs.cmu.edu/~quake/triangle.html
> Said, Y.H., Wegman, E.J., Sharabati, W.K., and Rigsby, J.T. (2008) "Style of author-coauthor social networks," *Computational Statistics and Data Analysis*, 52, 2177-2184, **2008;** ❶ doi:10.1016/j.**csda**.2007.07.021, **2007**. ❷ (?)
> Said, Y.H. and Wegman, E.J. (2009) "Approaches to text mining that preserve semantic content," to appear *Proceedings of HDM-2008* ❸
> Wegman, E.J. and Marchette, D.J. (2003) "On some techniques for streaming data: A case study of Internet packet headers," *Journal of Computational and Graphical Statistics*. 12(4), 893-914.
> Wegman, E.J. and Said, Y.H. (2008) "Text mining with application to fraud discovery," submitted.❹ *This was P301, but seems unpublished.*.
> WinPcap: The Wondows*(sic)* Packet Capture Library, www.winpcap.org.

[517] GMU, Final Technical Report, 450 days after completion; www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA495346&Location=U2&doc=GetTRDoc.pdf www.desmogblog.com/sites/beta.desmogblog.com/files/W911NF-07-1-0059-4-Final_0.pdf The same file was provided by ARO.
[518] "ARL Presentations\ARL Project Report\ARL Final Report.pdf" www.documentcloud.org/documents/554594-arl-final-report.html

The missing page clarifies a few errors in [MAS2010a §A.3], which had to speculate on the more-ambiguous claims.

Said et al (2007, 2008) were 0059-unfit, assumed to be P178 and P179:

❶ Said etal(2008) was P179 [called SAI2008], retracted from *CSDA*, mis-titled here, and not really peer-reviewed [MAS2011a §5].

❷ The only other paper from Wegman's resume that could be called Said et al(2007) was P178, more charitable than assuming they counted the same paper twice for print and online, although the reference makes it seem that way.[519] It is *likely* that if P178 were found, ☆ would be added. In any case, it is a problem for 0059.

❸ Said and Wegman (2009) this is a bit ambiguous, assumed T144a:
T144a Sw (c.35 in 0447) Said gave a talk of this title June 2008 in Kayseri, but it is unclear if the Proceedings were published in 2009 (or ever). T144a resembles the following, so might be P199 or T422.

P199 Sw	(Dec 2009)	no 0059-ack
T422 sw	(Oct 2009)	unknown, but *likely* had 0059-ack
T425 Sw③	(May 2010)	0059-ack.
T432 Sw③	(Aug 2011)	0059-ack, subset of T425, minus 5 pages.

❹ Wegman and Said (2007) was called ~~P301~~. Fraud detection is a well-established topic that *seems* 0059-unfit, but was charitably labeled fit on the assumption that anything that might be 0447-fit could be included.

Other parts of this report just summarize this, without the caveats, as:

P178 Sw①②③	0059-unfit, *likely* similar to P179, if so add ☆
P179 Sw①②③☆	0059-unfit, retracted for plagiarism
~~P301~~ Ws	0059-fit (maybe), *seems unpublished so hard to tell*
T144a Sw	0059-fit, as others in same set above
P200 Sw①②③☆	0059-unfit, ack'd, but not claimed

Some later talks ack'd 0059 and actually *seemed* 0059-fit, but also claimed 5876, for which they were unfit:
T148 Ws③, T151 Ws③, T420 Sw③, T425 Sw③, T432 Sw ③.

[519] Careless referencing in Wegman 's group was found often [MAS2010].

0059 report, pp.2-9.

The following 3 sections comprise the bulk of 0059 Final Report:

> **pp.2-4 '1.1 Mixed Language Text Database Search**
> A particularly useful component that was under development was on a mixed language text database search of open literature and intelligence documents. Because of the impetus from the fighting in Iraq, we initially developed a prototype for an Arabic and English mixed database. This project was directed at building a bilingual Arabic/English Web-Based Text Processing system that provides analysts with search results in Arabic based on English language queries. …
> **The structure of Arabic is quite different from Western languages not only because of the alphabet, but also because of conceptual difference in construction of words and sentences.** ….
> Figure 4 is a screen shot from the system.[520] This system, while not implemented is extensible to a Korean-English system. **The work reported here was implemented by Eiman Alshammari**, a Ph.D. student who is a native Arabic speaker, **under the supervision of Dr. Edward J. Wegman and Dr. Yasmin H. Said**,[521] who is also a native Arabic speaker and a U.S. citizen.

This raises several questions or concerns:

- The 0059 proposal never mentioned Arabic text.
- [MAS2010a §A.6.1] had listed Wegman's students and co-authors, but Alshammari was never mentioned, unsurprisingly, as she was in Computer Science by 2008 and her dissertation supervisor was Lin. *Was that true during 2007 or did she move in 2008?*
- Although Wegman's papers included some earlier text-mining (such as P170), and later P189 and P190, the first paper that might have involved Arabic was P199, 2 years after completion of the grant.
  No talks *seemed* possibly Arabic-related until 2008 - **T143**, **T144**, T144a.
  Much of the Arabic language-handling expertise seems to have been that of Alshammari's.[522]

> **pp.5-6 '1.2 Streaming Text Data Classification**
> Text processing is usually performed on a fixed corpus of text. The natural question to ask is what to do in the case of streaming news articles, weblogs, military reports, or even research articles when you want to examine them as they evolve in time? …This work was carried out by **Dr. Elizabeth Leeds Hohman** under the supervision of Dr. Edward J. Wegman. It was part of her Ph.D. dissertation.'

That seems straightforward.

> **pp.7-9 '1.3 Transient Geographic Mapping System**
> The geolocation tool is designed to enable real-time identification of incoming threats and attacks. … This work was carried out by **In-ja Youn** and **Felix Mihai** under the direction of Dr. Edward Wegman. It is based on an idea suggested by Wegman and Marchette (2003).'

This *seems slightly odd*. Both students ended up doing PhDs with other professors, and the work seems closer to the research efforts of those in ECE and CS. *Were they ever Wegman's students, or just helping out?*

Hohman 's section seems well-described as part of the original proposal, but the other sections do not. Wegman was clearly the main supervisor of her PhD and this work, whereas *that is less clear of the others*. Each section *seemed* like plausible work, from a cursory look, but it not obvious that this all fit together very well.[523]

Of the claimed papers, only T144a/ P199 *seemed* connected with the bulk of the Final Report, which described the work of 4 students, covered next:
① Eiman Alshammari
⑤ Elizabeth Leeds Hohman
② In-ja Youn
③ Felix Mihai

None coauthored a paper/talk with Wegman, through February 2010, but their work seemed to be included in a patent proposal by Wegman and Said that did not mention any of them.

[520] The screenshot includes Alshammari.

[521] On 12/15/06, Said was a year-and-a-half post-PhD, working on an alcoholism postdoc. Although an Arabic speaker, *there was little else in her background to that point that suggested she would be an appropriate co-supervisor of work on this topic. One might think she would be busy on **5876***.

[522] Alshammari grew up and lived in Kuwait, except for some college time in US. Said was certainly an Arabic speaker, but born in Puerto Rico, lived there until age 5, moved to Palestine for grades 1-3 and then to the US, **§S.7.1.**

[523] Grad student labor may not work well when attempting to actually build a complex hardware and/or software system, even a prototype.. I have worked for 2 companies derived from grad-student-staffed Stanford projects, but they had coherent groups that worked together for years with the same professors.
I've seen the same at other universities. *The 0059 staffing seems different.*

① **Eiman Tamah Alshammari**

She worked on 0059 but was in Computer Science by 2008 and defended her PhD in Information Technology in 2010.[524]

Web searches at GMU located: [525]

'**Ph.D. in Information Technology Final Defense for Eiman Tamah Al-Shammari**
PhD in Information Technology Doctoral Dissertation
Candidate: Eiman Tamah Al-Shammari
Bachelor of Science, Kuwait University, 1999
Master of Science, Kuwait University, 2001
Committee
Jessica Lin,[526] Chair *(Asst. Prof, Computer Science)*
Carlotta Domeniconi *(Assoc. Prof, Computer Science)*
Daniel Barbará *(Prof, Computer Science)*
Duminda Wijesekera *(Assoc. Prof, CSIS, Computer Science)*
Jeremy Allnutt *(Prof, Electrical and Computer Engineering)*
Title: **Improving Arabic Text Processing via Stemming With Application to Text Mining and Web Retrieval**
Friday, July 2, 2010, 1:00pm 3:00pm'

A search for Eiman Alshammari found a page last modified 11/26/08:[527]

'George Mason University
2005-Current PhD Candidate
Walton School of Business- University of Arkansas
2001-2002 M.S in Management of Information Systems
College of Engineering- Kuwait University
1999-2002 M.S in Electrical and Computer Engineering
College of Engineering- Kuwait University
1994-1999 B.S in Electrical and Computer Engineering'
'Publications

Al-Shammari, Eiman. Towards an Error Free Stemming. *In proceedings of the 2008 International Conference on Data Mining (IADIS'08)*. Netherlands, Amsterdam. July 24-27.
Al-Shammari, Eiman. & Lin, Jessica. (2008). A Novel Arabic Lemmatization Algorithm. *In proceedings of the 2008 2nd Workshop on Analytics for Noisy Unstructured Text Data (AND)*. Singapore. July 24.
Al-Shammari, E. & Lin, J. (2008). A New Arabic Stemming Algorithm. *In proceedings of the 2008 ISCA Workshop on Experimental Linguistics (ExLing)*. Athens, Greece. August 25-27.[528]
Al-Shammari, Eiman. & Lin, Jessica. (2008). Towards an error-free Arabic stemming. *In Proceeding of the 2nd ACM workshop on Improving non English web searching (INEWS)*. Napa Valley, California, USA. '[529]

The following were also found:

'Virtual Learning Environment in Primary Schools – Seamless Integration of Pedagogical Learning Values through an Integrated Virtual Learning Model
Dr. Eiman Tamah Al-Shammari
Department of Information Sciences, Kuwait University Kuwait'[530]

Finally, several presentations were found:
10/20/07 Eiman Alshammari, "Bilingual Arabic/English Web-based Text Processing System"[531] 19 pages
pp.5-6 "The Team" is Edward Wegman and Eiman Al-Shammari
p.7 (notes)

'Dr. Wegman & Dr.Said , Made modules ..student will use various models
They gave me the idea as a black box
I am not attempting automatic language translation, rather suggesting to the analyst a reduced set of Arabic language documents he or she might consider. Arabic is one of the most complex languages, both spoken and written. However, it is also one of the most common languages in the world.'

[524] www.cs.gmu.edu/news/report/2011.pdf p.6 lists her for 2011.
[525] volgenau.gmu.edu/graduates/graduate_news.php?start_from=70&ucat=&archive=&subaction=&id=&
[526] www.cs.gmu.edu/news/report/2011.pdf p.24.
'Lin has taught at Mason since 2005.Her areas of interest and expertise are temporal, spatiotemporal, multimedia, and stream data mining. She recently served as chair for the International Workshop on Data Mining for Geoinformatics in cooperation with SIGSPATIAL GIS in 2010. Her work focuses on the discovery of anomalies, frequent patterns, contrasting patterns, and structures in massive time series data.' *This background seems highly relevant.*
[527] mason.gmu.edu/~ealshamm; www.webcitation.org/6DDU5wNb7
[528] www.isca-speech.org/archive_open/exling2008/exl8_013.html
[529] www.informatik.uni-trier.de/~ley/pers/hd/l/Lin_0001:Jessica.html
The papers with Alshammari seem to be Lin's only ones on Arabic, i.e., unsurprisingly, the former was the source of Arabic expertise.
[530] cerc.wvu.edu/download/WORLDCOMP%2711/2011%20CD%20papers/EEE4379.pdf This seemed to show she had returned to Kuwait, as did a website: dreiman.net/page7.php with email contact information.
[531] "ARL Presentations\ARL Project Report\version_original.ppt" www.documentcloud.org/documents/550205-version-original.html

10/20/07 Eiman Al-Shammari, Edward J. Wegman, "Bilingual Arabic/English Web-based Text Processing System"[532] p.30 shows a screen shot of a document search, including her name at upper right, apparently using the WR as a test document.[533]

⑤ **Elizabeth Leeds Hohman**

She worked at NSWC and gave "Graphs for Streaming Text" 03/31/06: [534]

'Work supported by the Office of Naval Research In-house Laboratory Independent Research (ILIR) program'

She also gave "Measuring Word Frequencies for an Evolving Lexicon" 02/08/07.[535] Thus, her work had been well under way before 0059 began.

'Thanks to David Marchette for countless hours of help Work supported by the Office of Naval Research In-house Laboratory Independent Research (ILIR) program'

Her PhD (Computational Sciences and Informatics), with committee:[536]

'Dr. Edward J, Wegman Dissertation Director
Dr. David J. Marchette, Committee Member *(NSWC)*
Dr. Daniel B. Carr, Committee Member
Dr. Jeffrey L. Solka, Committee Member *(NSWC)*
Dr. Clifton D. Sutton, Committee Member'

② **Inja Youn** is a Computer Science PhD student, later published[537] with Brian L. Mark[538] and Dana Richards.[539] According to Mark's page,[540]

'Research Faculty:
Brian L. Mark, Professor of Electrical and Computer Engineering …
Ph.D. Students:
Inja Youn, Ph.D. student in Information Technology.
Co-advised with Prof. Dana Richards, Dept. of Computer Science
Topic: **"Internet Geolocation."**

③ **Felix Mihai** recently defended his dissertation:[541]

'Computational Methods for Coupled Electromagnetic Fluid-Structure Interaction Models
Defense Date:
Monday, December 3, 2012 - 10:00am …
Dissertation Director:
Dr. Igor Griva[542] and Dr. Padmanabhan Seshaiyer[543]
Committee:
Dr. Dimitrios A. Papaconstantopoulos [544]
Dr. Daniel Carr[545],

Students sometimes switch advisors, and some of these may not actually have been seeking a PhD with Wegman, but *it seems odd* that 5 of the 6 students mentioned in **§R.3** did not get their PhDs with Wegman.

[532] "ARL Presentations\ARL Project Report\Arabic Language Presentation.ppt" www.documentcloud.org/documents/550206-arabic-language-presentation.html
[533] "ARL Presentations\ARL Project Report\text" directory contains a mixture of Arabic materials and most of the files of the WR.
[534] "ARL Presentations\TCS2-2-Hohman.pdf" 02/08/07 www.documentcloud.org/documents/550207-tcs2-2-hohman.html
[535] "ARL Presentations\Talk31.pdf" 03/31/06 www.documentcloud.org/documents/550208-talkmarch31.html
[536] https://docs.google.com/file/d/0B2BDKYetoG59OTkyM2Q4ODMtYWQwYi00MGE3LTkwMDYtN2IwNjJlZjhlZDM0/edit?num=50&sort=name&layout=list Hohman is at NSWC, as are Marchette and Solka, both of whom got PhDs at GMU under Wegman. Carr and Sutton are in Statistics: statistics.gmu.edu/people_pages/sutton.html

[537] "Statistical Geolocation of Internet Hosts" napl.gmu.edu/pubs/CPapers/YounMarkRichards-ICCCN09.pdf
Inja Youn – Dept of Computer Science, GMU
Brian L. Mark – Dept of Electrical & Comp. Eng (ECE), GMU
Dana Richards – Dept. of Computer Science
[538] ece.gmu.edu/~bmark "His main research interests lie in the design, modeling and performance evaluation of communication network architectures and protocols." napl.gmu.edu
[539] cs.gmu.edu/~richards
[540] napl.gmu.edu/people.html
[541] cos.gmu.edu/node/12011
[542] cds.gmu.edu/node/38 Computational and Data Sciences → SPACS mason.gmu.edu/~igriva
[543] math.gmu.edu/faculty_staff/seshaiyer.htm Mathematical Sciences
[544] cds.gmu.edu/node/36 Chair Computational and Data Sciences → SPACS
[545] researchfocus.gmu.edu/node/688 Statistics

R.5 Invention – patent proposal[546]

One file in Wegman's 2010 FOIA reply to Vergano was a GMU Invention Disclosure Form, i.e., proposing to patent some work, modified 12/07/07. Excerpts:

'**1. Descriptive Title:**
A multi-modal system tool for aiding autonomous discovery

2. Name of Individual Completing this form:
Edward J. Wegman

4. Summary of the Intellectual Property …

1. System for retrieval of relevant documents from a mixed languages (such as English, Arabic, Korean) database using words and bigrams as metadata. Words/bigrams only are translated. System does not attempt to translate documents. *(Al-Shammari?)*

2. Use of bigram-document matrix to capture semantic content. *(Al-Shammari? Or may have been earlier from someone else.)*

3. Method of clustering documents using two-mode social network analysis on term-document and bigram-document matrices. Convert two-mode network to one mode network and use allegiance criterion to cluster documents relative to the document database. *(Sharabati? [SHA2008, §5.5])*

4. Using said method listed in 3 as a means of synonym generation.

5. Using an efficient O(n) algorithm on lexicon vectors to cluster documents. *(Hohman)*

6. System for querying files (text, images, videos) to determine geospatial location. System for querying geospatial locations for finding relevant files (text, image, videos).

7. System for geolocating emails by IP addresses. *(Youn & Mihai)*

8. System for geolocating other files (text, images, videos) by using cross-corpus discovery techniques to associate documents with geolocated emails.

9. Use of streaming data techniques on 1-8. *(Hohman)*

10. Use of text data mining techniques such as cross corpus discovery, clustering, author identification to authenticate religious texts including Qur'an, Hadiths, Bible, Book of Mormon, and apocryphal books and scrolls.

11. Use of these techniques in Arabic, Aramaic and Greek. *(Al-Shammari?)*

12. Using two-mode social network analysis on religious texts to discover relationship between concepts or persons within a sacred text.

5. List any Grants, Gifts or Contracts which supported the work …
Edward J. Wegman, Army Research Laboratory, W911NF-07-1-0059

6. Has this work been disclosed to anyone outside the University community or is a disclosure planned? If yes, please provide details including to whom, dates and whether there is a confidentiality agreement.
1 and 6 briefed to Army Sponsors on May 7, 2007.

1 through 9 briefed to Army Sponsors on October 23, 2007.

10 through 12 briefed to **potential sponsors in Saudi Arabia in April, August, October and November, 2007**; There have been no US disclosures of 10 through 12.

…inventor(s) Share of the revenues …
Edward J. Wegman
Yasmin H. Said'

Some of this seems a bit odd, but presumably acceptable to GMU:

- Some sections *seem* to have covered 0059 work by PhD students. *One might have thought they would have been mentioned.*
- Using Army funds to build software, and first briefing potential funders in Saudi Arabia *seems unusual*, *but maybe this is normal practice*.

Readers might compare with the various email excerpts of next section. The August 2007 Saudi Arabia visit is mentioned.

Said was involved with the Army activity for years, starting with ACAS in 2003. She ack'd 0059 in some papers/talks, although that work was rather unrelated to her 5876 fellowship. *That might have been acceptable to NIAAA, or not* – she did ask about part-time work, **§S.5**-08/30/06.

It is unclear whether or not this proposal ever turned into an actual patent filing. Various talks do mention "Patent pending," which requires that a patent application have been submitted to the US Patent Office, but of course, that might have described some other patent.

[546] www.documentcloud.org/documents/541254-invention.html

R.6 Emails

Messages are excerpted here in chronological order from [WEG2010a] on this grant and related discussions. **§I** lists travel dates.

11/27/06 p.1 GMU to various people, incl. Wegman and Said

'Currently, I am working with Drs. Wegman and Said to finalize all documents for a resubmission to Department of the Army - Army Research Lab. ARL is asking for George Mason University to transfer our budget to their format (ARO Form 99) and submit a Cover Sheet (ARO 51).'

➔ *Said was helping get Army grant, 6 months into 5876.*

01/18/07 p.24 Said to ARL 09:45AM

>I hope this message finds you well. Happy New Year! **I have been on >extensive travel in Europe and Asia** and just got back and totally >jetlagged.
> **I did not have Internet connection during my extensive travel**.'

01/18/07 p.2 Said to ARL 3:56PM

'Good Morning Yasmin,
Thank you for your emails. Dr. Wegman and I will come next Tuesday at 10:30 AM. My understanding is that you would like us to talk about text data mining and image data mining.'

01/22/07 pp.1-2 ARL to Said

'Good Morning Yasmin,
We look forward to your visit. I believe the primary focus of the meeting tomorrow should be the topic area of your proposal that was submitted and funded under a grant.'

08/05/07 p.73 Wegman to Elsevier, others *(on P179)*

'I **am in Saudi Arabia at the moment**. … . I have Internet access until Wednesday and then will be out of touch for about 3 days.'

08/28/07 p.13 Wegman to *Chance (on P180)*

'Attached are two pdf files. I have marked up the pdf files with Acrobat type sticky notes. The changes are minor. Thanks for your patience. **We were at the ISBIS[547] meeting in Azores** when your email arrived and **so I hope being 4 days late is forgivable. I am usually much more late.**'

[547] www.isbis.org/history.html; webservices.itcs.umich.edu/drupal/isbis/node/29

03/13/08 p.16 GMU to Wegman on 0059

'> >I am reviewing your ARO grant entitled "Army/Adaptive Multi-Modal
> >Data Mining and Fusion for Autonomous Intelligence Discovery",
> GMU
> >fund 201194, for closeout. I will be submitting the final
> financial
> >report this week and I am working with the technology transfer
> >office to submit the final patent report by this week or next as
> >well. **The award document also states that a final technical**
> report
> >is due by 3/15/08. Have you submitted the final report to ARO?'

03/17/08 p.15 Wegman to GMU on 0059

'> I am confused. This grant was from ARL, not ARO. I just returned
> last
> night from England. **So the answer is no, we have not sent in a**
> final
> report. But will do so hopefully this week. We did have a patent
> application filed based on work proposed, for this project. The
> sponsors had hoped to have a follow-on, but funds are tight at the
> moment.'

03/17/08 p.15 GMU to Wegman on 0059

'Welcome back, I hope you had an enjoyable trip!
The award documents reference an ARO grants officer, but a grants specialist with an ARL/ARO title, so I believe both must be involved. Also, the tech transfer office contacted me a few months ago because they were unable to report your invention under ARL, but could do so under ARO. At any rate, thank you for the status update on your final report. **If you wouldn't mind, could you let me know when the final report has been submitted so that I may document that information in the file.'**

05/06/08 p.14 GMU to Wegman on 0059

'I just wanted to follow up with you regarding this final report. **Did you have a chance to complete it yet?** The Office of Naval Research was delegated administrative closeout authority by the sponsor for this grant and they have requested a copy of the report. If the report is complete could I get a copy so that I may forward it along to ONR?'

➔ *GMU reminded Wegman, 3rd time.* ***Final Report was 03/18/09.***

GMU at least tried to monitor in this case, albeit unsuccessfully, as Wegman seemed to simply ignore due dates.

S. NIAAA F32AA015876 (3) 05/26/06-05/29/09 [548]

This had a complex history, of which some steps had to be *inferred.*
At least 3 5876 proposal versions were identified
(A) original proposal, rejected by NIAAA, so unavailable
(B) interim versions from files[549], close to (C), not identical
(C) final proposal eventually accepted by NIAAA[550]
The assembled chronology was, with inferences in *italics*:

2004 (?) *(A) sent to NIAAA, late 2004 or early 2005,*
11/17/04 date for signature remained in (B.1) p.14.
One might guess it suggested Summer/Fall 2005 start date.
02/28/05 Said wrote to NIAAA, inquired of status, **§S.5**
03/02/05 NIAAA needed more signatures, review date not set, **§S.5**
05/04/05 Said defended dissertation, **§S.5**
06/01/05 Said emailed to NIAAA, adding accomplishments, **§S.5**
(?) *(*A) rejected*, likely June or July*
(?) (A) was edited into (B), in response to critiques
(?) *(C) mostly seemed same as (B), but with a few more edits*
08/05/05 (C) received by NIAAA, 41 pages
Both (B) and (C) proposed 01/01/06-12/31/08 duration, *implying plan to keep Said at GMU,* but she instead went to Johns Hopkins to teach statistics 2005-2006, then returned to GMU, *perhaps an odd choice*.[551]
12/22/05 Reviewers commented on strengths and weaknesses, **§S.5**
02/24/06 Said sent more material to NIAAA, **§S.5**
05/17/06 NIAA granted Award to Said, **§S.2**

Concerns might be raised regarding the whole process of getting this grant and then fulfilling it:

- To get the grant, claims were made about existing work, against some of which plagiarism is alleged, P401 and P402.
- Reviewers thought Said's statistics background was fine, but were worried about her and Wegman's actual alcoholism research field expertise. She proposed a training program to allay their concerns, and then essentially ignored it, merely "sitting in" on a few courses, not even the ones most needed. She claimed insufficient funds to actually take classes, **§S.3.1**, although funds were found for extensive travel, **§I.**
- The final proposal emphasized working closely with (very credible) health researcher William Wieczorek of Buffalo State U, but the identically-titled works in 2008 were the only visible joint works:
  T418 ws① presented by Wegman at Interface conference
  P193 sw① paper at COMPSTAT 2008, presenter unclear
- The proposal also mentioned working with GMU's David Anderson, also a member of Said's dissertation committee with Wieczorek. No joint works with Anderson were found.
- Said's many statements, and some of Wegman's, tagged 0 1 etc, described in **§S.8,** *might generate concerns about over-sell.*

§S.1 is based primarily on (C), as its official provenance is clear, but includes some clearly-distinguished (B) text in gray, to show text likely to have been included in (C), but redacted, or interim text that was changed by (C) in ways that were inconsistent. *(B) and (C) were fairly close, although with a few instructive changes.* (B) contained some relevant contextual information that NIAAA redacted from (C).

The following page shows excerpts from an NIH policy that described the conditions for a Kirschstein award. This specific copy was provided by GMU to Vergano, **§T.**[552] *Some parts raise concerns, indicating either contradictions or perhaps stretching the rules.*

[548] projectreporter.nih.gov/project_info_history.cfm?aid=7491624
[549] FOIA reply from Wegman, October 2010, to Vergano. pp.41-42
B.1 www.documentcloud.org/documents/564094-f32-form-yasmin-revised-redacted.html
B.29 www.documentcloud.org/documents/603354-416-1cont-29-revised.html
B.30 www.documentcloud.org/documents/603247-416-1cont-30-revision.html
B.36 www.documentcloud.org/documents/524549-416-1cont-36-revised.html
[550] www.desmogblog.com/sites/beta.desmogblog.com/files/aa15876-1a1-1-Proposal.Redact_0.pdf FOIA from NIAAA,
[551] Said lived in MD, about same driving time from GMU or JHU. In FY2006, JHU got $598M in 1309 awards, of which 13 were from NIAAA ($3.1M), alone more GMU's $2.8M in 13 awards, including 5876, the only from NIAAA. report.nih.gov/award/index.cfm?ot=&fy=2006&state=&ic=&fm=&orgid=819801&distr=&rfa=#tab5 JHU's School of Public Health is huge.

[552] GMU reply to FOIA by Vergano, February 2013 included a copy of an NIH document, implying that the supplement was attached to the NIAAA grant. www.documentcloud.org/documents/604588-pa-03-067-kirschstein.html

Excerpts from NIH PA-03-067:

p.2 'Sponsor. Before submitting a fellowship application, the applicant must identify a sponsoring institution and an individual who will serve as a sponsor (also called mentor or supervisor) and will supervise the training and research experience.

The applicant's sponsor should be an active investigator in the area of the proposed research[553] who will directly supervise the candidate's research. The sponsor must document the availability of research support and facilities for high-quality research training. **In most cases, the F32 supports research training experiences in new settings in order to maximize the acquisition of new skills and knowledge.**[554] **However, in unusual circumstances, applicants may propose postdoctoral training experiences at their doctorate institution** or at the'

p.3 'institution where they have been training for more than a year**. In such cases, the applicant must carefully document the opportunities for new research training experiences specifically designed to broaden their scientific background.** [555]…

ALLOWABLE COSTS

Stipends. Kirschstein-NRSA awards provide stipends to postdoctoral fellows as a subsistence allowance to help defray living expenses during the research training experience. …

For fellows sponsored by domestic non-federal institutions, the stipend will be paid through the sponsoring institution. …

Tuition and Fees**. The NIH will offset the combined cost of tuition and fees at the following rate: 100 percent of all costs up to $3,000 and 60 percent of costs above $3,000**.[556] Costs associated with tuition and fees are allowable only if they are required for specific courses in support of the research training experience supported by the fellowship. A full description of the tuition policy is contained within the NRSA section of the Grants Policy Statement at grants.nih.gov/grants/policy/policy.htm.

Institutional Allowance. At the time of publication, fellows receive an institutional allowance of $5,500 per 12-month period to nonfederal, nonprofit, or foreign sponsoring institutions to help defray such'

p.4 'awardee expenses as research supplies, equipment, health insurance (either self-only or family as appropriate), and travel to scientific meetings. …

SUPPLEMENTATION OF STIPENDS, COMPENSATION, AND OTHER INCOME

The sponsoring institution is allowed to provide funds to the fellow in addition to the stipends paid by the NIH. Such additional amounts either may be in the form of augmented stipends (supplementation) or in the form of compensation, such as salary or tuition remission for services such as teaching or serving as a laboratory assistant, provided the conditions described below are met. **Under no circumstances may the conditions of stipend supplementation or the services provided for compensation interfere with, detract from, or prolong the fellow's approved Kirschstein-NRSA training program**.[557]

Stipend Supplementation: Supplementation or additional support to offset the cost of living may be provided by the sponsoring institution. Supplementation does not require additional effort from the fellow. DHHS funds may not be used for supplementation under any circumstances. **Additionally, no funds from other Federal agencies may be used for supplementation unless specifically authorized by the NIH and the other Federal Agency.**'[558]

p.5 'Compensation: The sponsoring institution may provide additional funds to a fellow in the form of compensation (as salary and/or tuition remission) for services such as teaching or serving as a research assistant. **A fellow may receive compensation for services as a research assistant or in some other position on a Federal research grant, including a DHHS research grant. However, compensated services should occur on a limited, part-time basis apart from the normal research training activities, which require a minimum of 40 hours per week**.[559] In addition, compensation may not be paid from a research grant supporting the fellow's research training experience.

…

p.5 'Data Sharing. It is the policy of the DHHS that **the results and accomplishments of all funded activities should be made available to the public.** This policy also applies to individuals supported by individual Kirschstein-NRSA postdoctoral fellowships. …'

[553] Wegman was to be the sponsor, but reviewers were concerned, **§S.1** p.8 and again in **§S.5**-12/22/05**:** 'One concern is that, although the proposal states Dr. Wegman has been working with NIAAA since 2001, his letter and biosketch do not discuss or list any alcohol-related publications, grant funding, or contracts.'

[554] *Could they not find a sponsor at the nearby and far more appropriate JHU? The entire process seemed designed to keep her at GMU with Wegman.*

[555] Many claims were made of frequent contact and close work with Wieczorek in Buffalo, *but few results seem to have been produced.*

[556] Said stated she could not actually take courses from lack of funds, **§S.3.1.**

[557] 5876-fit work was far outweighed by unfit work, **§0.2**.

[558] Presumably ARO and NIAAA approved any use of 0059 funds?

[559] 40 hrs/week? Not even a single peer-reviewed research journal paper?

S.1 Said Proposal – 08/03/05[560]

p.2 18. GOALS FOR KIRSCHSTEIN NRSA FELLOWSHIP TRAINING AND CAREER' My background has been principally in mathematics, statistics, and computer science. **Dr. Wegman has been working with NIAAA since 2001. In addition, his brother died of alcohol-induced cardiomyopathy**. **He has been so passionate about issues related to alcohol abuse and alcoholism that he has insired** *(sic)* **me to want to work in this area as a life career.**[561] My immediate desires for the postdoctoral fellowship is to broaden my background so that I can bring to bear a wide variety of interdisciplinary skills to problems. **To this end I am seeking to develop skills in social networks, geographic information systems, neuroscience, biology**,0[562] and alcohol studies in general, to complement my background in mathematics, statistics, and computer science. I believe that this will position me to be a very significant contributor to the field of alcohol studies. **Again, it is my strong desire to continue working in this health related area for my entire career.**' 1

p.2 22. DESCRIPTION

'We recognize that alcohol abuse leads to acute outcomes for both society and individuals. Among these, we identify DWI crashes with fatalities, assault and battery, suicide, murder, sexual assault, domestic violence, and child abuse. Alcohol abusers are embedded in a social network that involves the user, family and friends, producers and distributors of alcohol products, law enforcement, the judiciary, remediation, education, and intervention facilities such as Alcohol Safety Action Programs, and detox and treatment facilities, which are, of course, coupled to insurance and managed-care programs. This complex network is reminiscent of more traditional biologic ecology systems, hence the name. The basic idea is to formulate a model of this network with the goal of exploring short- and long-term interventions that reduce the overall probability of acute outcomes. The unique feature of this proposed work is that we are attempting to explore the simultaneous reduction of all acute outcomes.'

The next section had been redacted from 5876(C).

Her grades at Trinity College and American University were quite good, and at GMU they were spectacular, especially considering her many other concurrent activities. Her lack of courses relevant to the research domain itself worried reviewers, in both 5876(A) and revision (C), **§S.5.**

p.4 23. SCHOLASTIC PERFORMANCE, 5876(B.1), redacted in (C)

23. SCHOLASTIC PERFORMANCE: ***Predoctoral*** applicants: List by institution and year all undergraduate and graduate courses with grades. ***Postdoctoral*** applicants: List by institution and year all undergraduate courses and graduate scientific and/or professional courses germane to the training sought under this award with grades. Complete block at bottom of page, if applicable. ***Senior*** applicants: Omit this page.

SCIENCE			OTHER		
YEAR	COURSE TITLE	GRADE	YEAR	COURSE TITLE	GRADE
	GEORGE MASON UNIVERSITY:				
2005	Ph.D. in Computational Statistics		1999	STAT 690: Adaptive Design	A
	4400 University Drive; Fairfax, VA 22030		1999	STAT 515: Regression	B+
2005	IT 971: Probability Theory	A			
2004	CSI 876: Measure & Linear Spaces	A+	1998	41-590: Probability	A
2004	CSI 996: Mathematical Statistics	A+	1998	42-524: Data Analysis	A
2004	CSI 996: Real Analysis & Statistics	A+	1998	*41-322: Advanced Calculus*	A
2004	CSI 776: Time Series Analysis	A+	1997	42-202: Basic Statistics	A
2004	CSI 776: Numerical Methods	A+			
2004	CSI 776: Scientific Computation	A+		TRINITY COLLEGE	
2004	STAT 753: Computer Intrusion Detection	A+		A.B. Pure Mathematics	
2004	CSI 776: Scientific Databases	A		125 Michigan Ave., NE; Washington, DC	
2003	CSI 776:	A+	1998	MAT 490: Math Senior Seminar	A
2003	CSI 776: Genetic Algorithm	A+	1998	MAT 301: Abstract Algebra I	A
2003	CSI 877: Geometric Methods in Statistics	A+	1997	MAT 140: C++ Programming	B+
2003	CSI 996: Topics in Probability	A+	1997	PHI 315: Moral Psychology	A
2003	CSI 898: Seminar in Bioinoformatics	A	1997	MAT 247: Linear Algebra	A-
2003	CSI 899: Computational Science Colloq	A	1996	PHY 115: General College Physics I	B+
2003	CSI 991: Computational Statistics	A	1995	BIO 101: Intro. to Biology	A
			1995	MAT 126: Calculus II	A
	AMERICAN UNIVERSITY:		1995	MAT 125: Calculus I	A
	M.S. Computer Information Systems		1994	MAT 123: Pre-Calculus	A
	4400 Mass. Ave. NW, Washington, DC		1994	INT 115: First Yr Sem-Environmental Sci.	A
2003	CSC 565: Operating Systems	A			
2003	CSC 546: Intro to Computer Networks	A			
2002	CSIS 760: Information Systems Seminar	A-		University of Maryland at College Park	
2002	CSIS 665: Analysis & Design Workshop	B		College Park, MD 20742	
2001	CSIS 511: Concepts in Systems and Info.	A-			
2001	CSIS 560: Systems Analysis and Design	A-	1998	MAT 111: Intro to Probability	A
2001	MKTG 756: MKTG Strategy for High Tech	A-			
2001	CSIS 635: Database Management	B	1998	MAT 211: Elements of Geometry	A
2000	CSIS 606: Quant Analysis for Info System	A			
2000	MATH 690: Theory of Probability	B+			

The instructions say:

'***Postdoctoral*** applicants**:** List by institution and year **all undergraduate courses and graduate scientific and/or professional courses germane** to the training sought under this award with grades.'

The experience for writing P402 S☆ seemed to be the course marked

If she was dedicated to alcoholism research, why did she not take a few courses more directly related to that field, before she did a dissertation on the topic?

---

[560] www.desmogblog.com/sites/beta.desmogblog.com/files/aa15876-1a1-1-Proposal.Redact_0.pdf FOIA from NIAAA with some elements included from (B), shown in gray.

[561] *Wegman has my sympathy for* ***this****. It may help explain some strange behavior.*

[562] The 0 1 ... tags consolidate explanations enumerated once each in **§S.8.**

p.6 RESEARCH TRAINING PLAN, 5876(C)

30. RESEARCH TRAINING PLAN

a. Approximate percentage of proposed award time in activities identified below. *(See instructions.)*

Year	Research	Course Work	Teaching
First	50%	50%	
Second	65%	25%	10%
Third	90%		10%

GMU says:[563]

'Graduate students …A full-time academic load is nine semester hours.'

Thus, full-time was 3 3-hour courses, but Said's plan at right treated 4 3-hour courses as only half-time.

No time was allocated in 3rd year. Later, Wegman wrote:

'She was a brilliant and devoted student . …In her first semester at George Mason University (GMU), **she took double overload, more than 18 graduate credits,** and the same amount of graduate credits the subsequent semesters … I have had 25 Ph.D. students, and only two have been able to complete their doctorate in 1 year and 9 months. **In addition to studying full time**, for much of her time as a graduate student **she was simultaneously working full time at a private high school**…'

It might seem unlikely that she could spend only half-time taking 4 graduate courses/semester when 3 was called a full-time load, but she had been taking 6 courses/semester and got an A+ in 11 of 16 courses.

She also got "Best Statistics Dissertation – 2005" award.

So, maybe she was one of the most brilliant statistics PhDs at GMU and quite capable of doing all this at once, while doing full-time alcoholism research, working on 0059 and numerous other projects.

It does seem strange, and any case, it seems that no courses were actually taken for credit.

This plan seems hastily and inconsistently constructed. Perhaps next page's 5876(B) version helps explain some of the oddities. It appears that sections of (B) were edited to (C) without making them consistent.

[563] www.gmu.edu/academics/catalog/9798/registra.html

p.10 30b. RESEARCH TRAINING PROPOSAL, 5876(C)

'SPECIFIC AIMS: My training in statistics, mathematics, and computer science is very strong. **However, I have not focused in my previous academic career on the social, psychological, and geospatial dimensions necessary for my desired career in health-related research.** [564]**My goal is to augment my statistical, mathematical and computer science skills with other skills aimed at providing a well-rounded background to complete the research agenda outlined above.** Specific aims are

1. To increase my direct knowledge of social science, health policy, and alcohol epidemiology topics in support of developing a career in modeling these issues.
2. To expand the proof of principle simulation model developed in my dissertation to a fully agent based simulation as described below.'

p.11 'I plan to take the following courses at George Mason University.

From the Psychology Department
PSYC 561 - Behavioral Biology of Substance Abuse
PSYC 822 - Psychopathology

From the Computational Social Science Program
CSS 645 - Spatial Agent-Based Models of Human-Environment Interactions
CSS 692 - Social Network Analysis

From the School of Public Policy
PUBP 713 - Policy and Program Evaluation

From the Earth Systems Program
GEOG 553 - Geographic Information Science

From the Program in Health Science
HSCI 712 - Epidemiology and Health Service Research

And From the Data Sciences Program
DATA 789 - Spatial Statistics

This broad interdisciplinary background is intended to focus specifically of tools necessary to carry out the proposed research, but also enrichment background that will allow me to focus my career on broader health related research. **My planned distribution of courses is four each semester for the first year, two each semester for the second year and one each semester for the third year.**[565] … All of the listed courses are graduate courses and generally are sequenced so that lower numbered courses are prerequisites for higher numbered courses. My highest priorities are PSYC 561, CSS 645 and 692, GEOG 553, PUBP 713, and HSCI 712.'

[564] Correct, she had taken almost no relevant coursework.
[565] She lists 8, but 4+4+2+2+1+1 totals 14.

The earlier version was slightly more consistent, (B) changes in gray.

p.1 30b. RESEARCH TRAINING PROPOSAL, 5876(B.30)

'SPECIFIC AIMS: My training in statistics, mathematics, and computer science is very strong. However, I have not focused in my previous academic career on the social, psychological, and geospatial dimensions necessary for my desired career in health-related research. My goal is to augment my statistical, mathematical and computer science skills with other skills aimed at providing a well-rounded background to complete the research agenda outlined above.[566] Specifically, I plan to take the following courses at George Mason University.

From the Psychology Department
PSYC 561 – Behavioral Biology of Substance Abuse
PSYC 616 – General Psychopathology
PSYC 669 – Social and Emotional Development
PSYC 822/823 – Psychopathology I and II

From the Computational Social Science Program
CSS 640 – Land-Use Modeling Techniques and Applications
CSS 645 – Spatial Agent-Based Models of Human-Environment Interactions
CSS 692 – Social Network Analysis

From the School of Public Policy
PUBP 713 – Policy and Program Evaluation

From the Earth Systems Program
GEOG 553 – Geographic Information Science
GEOG 563 – Advanced GIS
GEOG 653 – Geographic Information Analysis

From the Program in Health Science
HSCI 712 – Epidemiology and Health Service Research

And From the Data Sciences Program
DATA 789 – Spatial Statistics'

This broad interdisciplinary background is intended to focus specifically of tools necessary to carry out the proposed research, but also enrichment background that will allow me to focus my career on broader health related research. **My planned distribution of courses is four each semester for the first year, two each semester for the second year and one each semester for the third year.** … All of the listed courses are graduate courses and generally are sequenced so that lower numbered courses are prerequisites for higher numbered courses. My highest priorities are PSYC 561 and 616, CSS 645 and 692, GEOG 553, PUBP 713, HSCI 712 and DATA 789.'

[566] Through here, (B) is the same as (C), then diverges, with (B) text in gray.

p.6 RESEARCH TRAINING PLAN, 5876(B.1)

30. RESEARCH TRAINING PLAN
a. Approximate percentage of proposed award time in activities identified below. *(See instructions.)*

Year	Research	Course Work	Teaching
First	40%	60%	
Second	60%	30%	10%
Third	80%	10%	10%

Compared to 5876(C), this was more consistent and perhaps more sensible, although it still seems more heavily-weighted with statistics work than really needed to address reviewers' critiques.

- Prerequisites *seemed* better covered.
- (B) had a stronger PSYC component than (C), with 3 more courses.
- It proposed 14 courses, not 8, calling them 60%, or 15%/course, higher than the 12.5% of (C), still less than the 33% specified by GMU, but.

This part of (B) seemed hastily and inconsistently edited into (C).
But none of this mattered, because the plan simply was not executed.
In the first 2 years, she just "sat in on" only CSS 645 and 692, not even other more relevant courses. Reviewers were still worried these were insufficient, **§S.3.1.** At least during the first 2 years, no courses were actually taken for credit 2. *That is no surprise,* given her busy travel schedule. The first month or two had to have been focused on T126, and within 6 months, she was helping Wegman get 0059.

Intent is unknowable, but the pattern raises concerns. NIAAA reviewers offered critiques, the proposal was changed to address them somewhat, promises were made, *but little was actually done to fulfill them.* Said spent much time on unfit work, **§S.4.** *This whole process tried to get money to let her stay at GMU, without much effort to follow through. Johns Hopkins would have been far more appropriate. Wegman had students there and she even had 2004 contact with people at JHU,[567] whose public health school, including biostatistics, is highly-rated and well-funded.[568]*

[567] www.interfacesymposia.org/I04/index.html
Said co-edited proceedings of Interface 2004, held in Baltimore, organized by Wegman PhDs Solka (then at GMU) and Marchette (then at JHU). At least 5 other JHU people spoke, and Wegman student Cary Priebe is a Professor there.
[568] grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-health-schools/public-health-rankings
chronicle.com/article/NRC-Rankings-Overview-Public/124702

p.6 28. RESEARCH EXPERIENCE

'a. Summary: I have had fairly extensive experience in independent research as an undergraduate and graduate student at the M.S. Level. These reports include in 1998, "Cardinality of Infinite and Finite Sets, (an undergraduate thesis of 75 pages length), in 1999, "Statistical Analysis of Star Measurements," (under guidance of Professor David Crosby**, approximately 980 pages**, report submitted to NOAA), in 1999, "Adaptive Design: Optimal Drug Dosage for Patient Treatment - The Migrating Urn," (under guidance of Professor Nancy Flournoy, **276 pages**, material incorporated in "A birth and death urn for randomized clinical trials: Asymptotic methods," by A. Invanova, W. Rosenberger, S. Durham, N. Flournoy, (2000) Sankhya: The Indian Journal of Statistics 2000, Volume 62, Series B, Pt. 1, 104-118), in 2001, "Video Conferencing TVProduct Development and Marketing Strategy," (under guidance Professor David Eppright, approximately 100 pages), in 2001, "Automated Prescription System," (under guidance of Professor Anita La Salle, approximately 200 pages), and in 2002, "Demand for Video Conferencing Television," (under guidance of the late Professor Thomas Slavinski, approximately 100 pages). I am a member of the following professional/research societies: American Statistical Association, Institute of Mathematical Statistics, Institute of Electrical and Electronic Engineers, American Association for the Advancement of Science, American Mathematical Society, Society for Industrial and Applied Mathematics, The Interface Foundation of North America, Royal Statistical Society (Fellow), and Research Society on Alcoholism (Elected).

b. Doctoral Dissertation: Agent-Based Simulation of Ecological Alcohol Systems, under the direction of Professor Edward J. Wegman P401 S☆

'c. Publications[569]

2005 "On genetic algorithms and their applications," *Handbook on Statistics: Data Mining and Data Visualization,* 24, (Rao, C. R., Wegman, E.J. and Solka, J.L., eds.), Amsterdam: Elsevier North Holland, 359- 390. P402 S☆

2005 "Alcohol tree simulator," www.alcoholecology.com

2005 *Agent-Based Simulation of Ecological Alcohol Systems,* Ph.D. dissertation, School of Computational Science, George Mason University, 144 pages. P401 S☆

2006 "Statistics," to appear *Encyclopedia of the Modern World,* (Stearns, Peter N., ed.), New York: Oxford P403 S'

[569] All she really had was her dissertation. P403 was *very likely* written in 2006/2007, **§X, §Y.** P402 S☆ written for Wegman, note plagiarism allegation.

p.7 PRESENTED PAPERS:[570]

2005 *Ecology of Alcohol and Alcoholism,* Invited, **Edward Wegman** and **Yasmin Said**, CSNA/Interface 2005, St. Louis, MO, June, 2005. T117 Ws①

2005 *Modeling Alcohol Abuse and Consequences,* Invited, **Yasmin Said** and **Edward_Wegman**, CSNA/Interface 2005, St. Louis, MO, June, 2005. T118 Sw

2005 *Comparing Nonlinear Approaches for Classification,* Invited, Carlos Alzola and **Yasmin Said**, CSNA/Interface 2005, St. Louis, MO, June, 2005.

2005 *A Digraph Model of Alcohol Ecology,* Contributed, **Yasmin Said** and **Edward Wegman**, JSM, Minneapolis, MN, August, 2005.

2005 *Tutorial on Regression Splines,* Contributed, Jill McCracken, Pier Bobys, **Yasmin Said**, Carlos Alzola, JSM, Minneapolis, MN, August, 2005.

2005 *Statistical Analysis of Star Data,* Contributed, John Rigsby and **Yasmin Said**, JSM, Minneapolis, MN, August, 2005.

INVITED PRESENTATIONS:

2005 *Geo-Spatial Analysis of Alcohol Ecology,* Invited, **Yasmin H. Said**, Lawrence Livermore National Laboratory, Livermore, CA, July 5, 2005.

2005 *Spatial Topologies of Complex Social Systems,* Invited, **Yasmin H. Said**, Jet Propulsion Laboratory, Pasadena, CA, July 22, 2005.

2005 *Agent-Based Model for Homeland Security and Disease Control,* Invited, **Yasmin H. Said**, SAMSI Workshop on Homeland Security, Research Triangle Park, NC, September 11-14, 2005. T400 S☐

2005 *Agent-Based Model Applicable to Food and Drug Administration,* FDA, Rockville, MD, August 22, 2005.

2005 *Applications to Agent-Based Model,* WESTAT, Rockville, MD, August 25, 2005.

2005 *Adaptation of an Alcohol Ecological Agent-Based Model to Homeland Security,* Invited, **Yasmin H. Said**, Bureau of Labor Statistics, Washington Statistical Society Workshop on Homeland Security, Washington, DC, September 28-29, 2005.' Seems same as T400 S.

[570] Most of these preceded 5876 and were not claimed for it, so were not given T*nnn* codes. None of this activity appeared in 5876(B), but perhaps it was truncated. Many were to occur after the 08/05/05 date on which NIAAA received the proposal. *Few alcoholism talks were presented to obviously-relevant and knowledgeable audiences, but rather in other venues, claiming widespread applicability of her model. She was included in 3 JSM talks led by others, on very different topics unrelated to her dissertation. All this seems like resume-padding.*

p.7 'PROFESSIONAL ACTIVITIES:[571]
Member, Program Committee, SIAM Conference on Data Mining (SDM06), Washington, DC, April, 2006.[572]
Co-Chair (with Dr. Amy Braverman), Interface 2006: Massive Data Sets and Streams, Pasadena, CA, May, 2006.
Editor, Interface Publications, 2005-2007.
Interface Webmaster, The Interface Foundation of North America, 2005-2007
Editor, US Army Conference on Applied Statistics Publications, 2005-2007.
Army Conference Webmaster, US Army Conference on Applied Statistics, 2005-2007.
Invited Session Organizer, Symposium on Interface -Joint Meeting of the Interface and Classification Society of North America 2005: "Modeling Alcohol Abuse and Consequences," St. Louis, MO, June 8-12, 2005.
Invited Session Organizer, Symposium on Interface -NASA Data Mining Workshop/Interface 2006: "Alcoholism," Pasadena, CA May 23-27, 2006.'

p.8 29. REVISED APPLICATION 5876(B.29), redacted from (C)
'This is a revised application of 1 F32 AA015876-01. I have tried to take into account all of the comments from the study panel. Specifically Section 30 is completely revised. **The lack of detail in the training plan has been noted and is now very specific as to courses planned.** They have been planned for specific contributions to the immediate research project as well as training aimed a development of capabilities aimed at longer term health related research. **With this I have tried to convey a clearer deliniation of the mentor's role based on the technical math/stat/computing role of Dr. Wegman and the alcohol studies/social policy role of Dr. Wieczorek. I have also noted their symbiotic relationship**. **I have dropped the direct mentoring role of Dr. Anderson.** I have also made it clearer how the mentoring process will be carried out with respect to interaction with Dr. Wieczorek. I have also tried to give a clearer discussion of the model design with a description of what was done in the dissertation and how the proposed research expands on the dissertation research. I also mention the website that I developed in connection with the dissertation, which illustrates the functionality of the relatively simple model developed in my dissertation. I allude to the some of the data sources developed for my dissertation, which were quite extensive, and indicate that the more comprehensive model will require development of additional data sources. I make it clear that I am trying to develop a more sophisticated model that is a major expansion of that developed in my dissertation. To this end, I have re-titled the application to reflect the expanded scope when compared with my dissertation. Finally in connection with the summary discussion, I have attempted to make it more clear why this is an agent-based model rather than just a micro-simulator. **Some concern was expressed in connection with my previous work being focused on problems outside biomedical and behavioral work.[573] I have energetically pursued giving conference papers and professional talks on my dissertation research.[574] I have also added additional letters of reference from biomedical and behavioral researchers.[575]** In my discussions, I have **emphasized my commitment to a career in biomedical / behavioral / health related research, [576] which I sensed from the reviews may have not come across as strongly as I would have liked**. In connection with the Summary and Recommendations of Critique 1, (1) the mechanisms have been clarified, specifically how I will interact with Dr. Wieczorek at Buffalo State, (2) **the coursework has been detailed and the recommendation to drop the chemistry has been taken**, (3) the current state of the dissertation model and planned expansion have been articulated, (4 and 5) I have taken the recommendation to narrow the scope under advisement. Indeed I do not disagree that the chains leading to DWI may be quite different than that leading to for example child abuse. However, I can envision a setting where additional law enforcement may prevent asssaults *(sic)* at a night club, but lead to DWI and even heavier drinking at home increasing domestic violence and child abuse. I contend that these events are not independent. The proposed research may very well indicate some outcomes are relatively indpendent *(sic)* of others and that some have much more serious societal and individual consequences than others. Thus I am certainly open to revision and down-sizing of the model as appropriate. Indeed, underage drinking appears to have far reaching consequences so may very well be a focused target for my modeling efforts. Based on NIH guidelines, I have concluded that the ambiguity as to human subject research in my previous application was poorly articulated and that Exemption 4 is appropriate.

Critique 2 echoes essentially the same concerns Critique 1 had and so are addressed above. Additional changes are reflected in an updated Academic Performance, updated Background, and an updated Facilties *(sic)* and Commitment page. As mentioned earlier, I have asked for additional letters of reference.

[571] The first appears false and the rest were all under Wegman's control.
[572] www.webcitation.org/6GVJjGOlk Her name appears nowhere in huge PC list.
[573] *The problem was clear from the courses shown earlier.*
[574] Talks were presented at Interface, JSM, JPL, SAMSI Homeland Security, FDA and Bureau of Labor Statistics, not obviously alcoholism-oriented venues.
[575] From 5876(B.1), references were by Wegman, Wieczorek, Anderson, Jeffrey Solka (NSWC) and Reza Khorramshahgol (American University.
[576] *Given lack of relevant courses, it was unsurprising they were skeptical.*

p.11 30. RESEARCH TRAINING PLAN (starts at p.9)
'It is anticipated that Professor Wegman will become the founding Chair of the new Department of Computational and Data Sciences. Coupled with his strong interest in accelerating his efforts in alcohol research, he will be an ideal mentor focused on the mathematical, statistical and computer science side of this proposed effort. Dr. Wegman and Dr. Wieczorek have been discussing the approach I outlined above for more than a year. **They[577] are planning an R01 to complement the effort represented by this F32. …**

Dr. Wieczorek is also intended to be my other mentor on this F32. The strong interaction Dr. Wegman and Dr. Wieczorek have makes this joint mentorship very natural. (They have each visited each other Centers.) I will meet with Dr. Wieczorek face-to-face quarterly during at least the first year and we will have telephone meetings on a bi-weekly basis. Dr. Wieczorek's mentoring will focus on the etiology and epidemiology of drinking and driving, spatial aspects of alcohol use and problems, and the general epidemiology of alcohol use including measurement, data sources, and overviews of treatment and prevention activities. We will do

p.12 this through directed readings and continuing conversations and feedback. I plan to visit his Center **at least twice a year**[578] to see projects in the field and to meet with him and his staff to discuss any issues related to our work.

Dr. Wegman has also been appointed to the George Mason University Taskforce on Underage Drinking on University Campuses. As additional complementary education, **Dr. Wegman and I are anticipating working with Dr. David Anderson in his activity doing data analysis and reduction on his long-term surveys.**[579] This will complement the mentoring of Dr. Wieczorek by dealing directly with data on underage drinking.'

Wegman gave her a glowing reference, 3 but in this case the same (unredacted) words were found elsewhere.[580] In this case, the gray shows the redactions made by NIAAA, but found with the exact same formatting:

p.23 '36. She was a brilliant and devoted student. To demonstrate her proficiency, determination, and qualifications for this post-doctorate position, I will describe a synopsis of her recent accomplishments. In her first semester at George Mason University (GMU), she took double overload, more than 18 graduate credits, and the same amount of graduate credits the subsequent semesters.[581] **In less than ten months, she was able to complete her entire doctoral course work, to take her written comprehensive exam and her oral comprehensive exam, and to complete her dissertation proposal**. With respect to her written and oral comprehensive exams, she received high distinction (high pass) on every exam. In addition, her cumulative grade point average is a 4.0 on a 4.0 scale. Hence, in less than a year, she have *(sic)* been able to advance to candidacy and progress to her dissertation research. **She completed her dissertation in and graduated with Ph.D. in the next nine months.** 4 I have had 25 Ph.D. students, and only two have been able to complete their doctorate in 1 year and 9 months**. In addition to studying full time, for much of her time as a graduate student she was simultaneously working full time at a private high school in Alexandria, Virginia, where she taught high-school-level and college-level mathematics and statistics. She has published a chapter on genetic algorithms in the Handbook on Statistics: Data Mining and Data Visualization.** 5 She was also a student member of the Interface and American Statistical Association; the latter is the nation's leading professional association for statistics and statisticians. 6 She is also a student member of the Research Society on Alcoholism. 7 Being able to successfully accomplish all this in less than 21 months leads to the conclusion that she is extremely devoted and has a burning internal desire and determination to suceed *(sic)* in both research and academia.'[582]

[577] Wieczorek had a major R01 in 2007-2012, at least, which generated ~8 peer-reviewed papers in relevant journals, but never with Wegman or Said. projectreporter.nih.gov/project_info_history.cfm?aid=8077451 projectreporter.nih.gov found Said's 5876, but no mention anywhere of Wegman.
[578] In **§S.3.1**, she noted visiting Buffalo once.
[579] Anderson was on her committee, but his C.V. shows zero papers coauthored with Wegman or Said. cehd.gmu.edu/assets/files/cv/14.pdf

[580] FOIA reply from Wegman, October 2010, to Vergano. B.36 www.documentcloud.org/documents/524549-416-1cont-36-revised.html
[581] That would be 6 classes per term, 16.7% per class.
[582] She claimed T126 and *very likely* wrote much of the WR [MAS2010a], which had zero relevance to her teaching at JHU or any grants.

[WEG2010a pp.40-44] either extended the proposal or was a different proposal more focused on HIV/AIDS and alcohol. However, the proposal in the NIH database seems more like the original one, so I am not sure what this means. If it matters, NIAAA may have emails and series of proposals unavailable to me.

'Yasmin Said to kbryant, gbloss, mhilton' 2/24/06

p.42 'Alcohol abuse leads to serious individual and societal outcomes. Among these, we identify irresponsible behavioral outcomes notably promiscuous sexual contacts as well as other violence related outcomes including DWI crashes with fatalities, assault, suicide, murder, sexual assault, domestic violence, child abuse and other. The risky sexual contacts in turn can lead to infections with AIDS, HIV, and other STDs. Social, structural, and environmental factors are major influences on HIV-related behaviors yet the dearth of evaluation tools impedes progress in intervention development. My research is intended to provide a policy tool for evaluation of interventions using a dynamic agent-based simulation. Alcohol abusers are embedded in a social network that includes the user, family and friends, producers and distributors of alcohol products, law enforcement, remediation and intervention facilities, and treatment facilities, which are coupled to insurance programs. This complex network is reminiscent of more traditional biologic ecology systems. The basic idea is to formulate a model of this network with the goal of exploring short- and long-term interventions that reduce the overall probability of negative outcomes. Intervention approaches will target networks and behavioral settings and provide alcohol users with socially meaningful and rewarding behavioral options that are consistent with valued prosocial identities. Historically, modeling attempts have focused on specific negative outcomes. The unique feature of this proposed work is that we are attempting to explore the simultaneous reduction of all negative outcomes.

Primary Objective:

Improve the quality of behavioral and social science methodology in HIV research: Support research to advance innovative quantitative and qualitative methodologies to enhance behavioral and social science on HIV prevention and care, and to address pressing ethical issues in the conduct of such research. …(more)'

[WEG2010a pp.49-51] has a reply 05/07/07 to Wegman from an alcoholism researcher in Buffalo, followed by Wegman's request 05/03/07.

Wegman and Said seemed less connected with the relevant literature than they might have been, consistent with concerns of the reviewers. But, only experts would know for sure.

S.2 NIAAA Award $145K 05/26/06

The purpose of this type of award is:[583]

> 'To provide postdoctoral research training to individuals to broaden their scientific background and **extend their potential for research in specified health-related areas**.'

The award is summarized:[584]

> **'A Social Network Model of Ecological Alcohol Systems**
> We recognize that alcohol abuse leads to acute outcomes for both society and individuals. Among these, we identify DWI crashes with fatalities, assault and battery, suicide, murder, sexual assault, domestic violence, and child abuse. Alcohol abusers are embedded in a social network that involves the user, family and friends, producers and distributors of alcohol products, law enforcement, the judiciary, remediation, education, and intervention facilities such as Alcohol Safety Action Programs, and detox and treatment facilities, which are, of course, coupled to insurance and managed-care programs. This complex network is reminiscent of more traditional biologic ecology systems, hence the name. The basic idea is to formulate a model of this network with the goal of exploring short- and long-term interventions that reduce the overall probability of acute outcomes. The unique feature of this proposed work is that we are attempting to explore the simultaneous reduction of all acute outcomes. Historically, modeling attempts have focused on specific acute outcomes. The framework we are pursuing is a dynamic agent-based simulation.'

NIH expects reports:[585]

> '**Application for Continued Support.** When multiple years are involved, awardees will be required to submit the Kirschstein National Research Service Award Individual Fellowship Progress Report for Continuation Support (PHS 416-9) annually as required in the *NIH Grants Policy Statement.* The report is due two months before the beginning date of the next budget period and must include information describing the **current year's progress as well as the research and training plans for the coming year**. **Note that the instructions request a listing of all courses and publications completed during the past year.**' 2

The 5876 main database entry was found.[586] Although Said authored or coauthored some conference papers, the database had nothing, starting from the main entry and checking each of the 3 periods. All stated 'no publications available..' Her interim reports mentioned papers, excerpted later, so that may just have been a database entry problem, **§S.6.** Perhaps the database counts only peer-reviewed papers, i.e., none relevant.

A few excerpts from the award[587] are

> 'NATIONAL RESEARCH SERVICE AWARD Issue Date:**05/17/2006** …
> Grant Number: 1 F32 AA015876-01A1 (Revised)
> Principal Investigator: SAID, YASMIN H MS
> Project Title: A Social Network Model of Ecology Alcohol Systems
> Financial Officer
> GEORGE MASON UNIVERSITY
> CTR FOR COMPUTATIONAL STATISTICS …
> The National Institutes of Health hereby revises this award (see "Award Calculation" in Section I and "Terms and Conditions" in Section III) to GEORGE MASON UNIVERSITY in support of the above referenced project. …
> SPREADSHEET
> GRANT NUMBER: 1 F32 AA015876-01A1 (Revised)
> P.I.: SAID, YASMIN H
> INSTITUTION: GEORGE MASON UNIVERSITY

	YEAR 01	YEAR 02	YEAR 03
Stipends	38,976	41,796	43,428
Institutional Allowance	7,000	7,000	7,000
TOTAL FEDERAL DC	45,976	48,796	50,428
TOTAL FEDERAL F&A	0	0	0
TOTAL COST	45,976	48,796	50,428'

The 3-year totals were:

Stipends (to Said)	124,200
Institutional (to GMU)	21,000
TOTAL COST	145,200

583 grants.nih.gov/grants/funding/ac_search_results.htm?text_curr=f32&Search.x=0&Search.y=0&Search_Type=Activity
584 projectreporter.nih.gov/project_info_description.cfm?aid=7491624&icde=0
585 grants.nih.gov/grants/guide/pa-files/PA-10-110.html#SectionV

586 projectreporter.nih.gov/project_info_history.cfm?aid=7491624 accessed 11/17/12 and includes 1F32AA015876-01A1 Fiscal year 2006, 5F32AA015876-02 Fiscal year 2007, 5F32AA015876-03 Fiscal year 2008
587 www.desmogblog.com/sites/beta.desmogblog.com/files/aa15876-1a1-2-Award.Redact_0.pdf

S.3 Said Interim Reports

NIAAA redacted some text (■) in accordance with NIH rules.[588]
The page numbers here use those of the PDF, not the text.

S.3.1 5F32AA015876-02 07/27/07[589]

This report essentially covered 05/25/06 -05/25/07, the first third.

p.3 (Said) 'A: CHANGES

The major change has been associated with classroom coursework. **Because of a limitation of funds available to take formal coursework**, 2 **I have had to take coursework informally. I have sat in on CSS 692, and CSS 645 as planned. I have written two papers on social network analysis.** 0 I have also spent three weeks at Buffalo State with Dr. Wieczorek as planned and have made extensive use of training at conference venues. The theme of my research project and my academic status are unchanged**. I have spent somewhat more time on research and somewhat less time on formal classroom instruction. I have not been doing any formal teaching.** 2

B: PROGRESS

TRAINING: I have been intensively studying materials relevant to the F32 proposal. **I spent the period September.17 through October 1 in the Center for Health and Social Research at Buffalo State University**. Specifically I worked with my mentors, Drs. William Wieczorek, Edward Wegman. I worked with Dr. William Wieczorek's .deputy director, Dr. Kelly Marczynski, learned more about their operation at Buffalo State. Part of the effort was to develop connections between alcohol use and HIV/AIDS. They were conducting a survey to establish such connections in drivers that had DWI arrests. I have been **sitting· in on** 0 and **working with students in social network courses** and have developed the applications mentioned above.[590] Working on the time-of-day analysis has enhanced my training in time series analysis and general linear models. I have been attending many conferences, statistical, computer science, as well as alcohol related including the 2006 RSA meeting in Baltimore, MD and the 2007 RSA meeting in Chicago . In most of the conferences, I have either spoken or organized one or more sessions. The conferences are as follows:

-SAMSI/NISS 2005-The NSF-sponsored Statistics and Applied Mathematics Institute held in Research Triangle Park, NC (September). I gave one of four young investigator talks on my NIAAA related research T400 S□ 8

-COMPSTAT 2006 - Meeting of the International Association of Statistical Computing, (held even numbered years), held in Rome, Italy. I gave a talk relevant to my NIAAA research. ***P174***Sw□①

-Interface 2006 - Symposium on the Interface of Computing Science and Statistics, held in Pasadena, California. I organized a session and gave an invited talk on my NIAAA-related research.[591]

-JSM 2006 -Joint Statistical Meetings (sponsored by the American Statistical Association; the Institute of Mathematical Statistics, the International Biometric Society, and the Statistical Society of Canada, 6000 attendees), held in Seattle, Washington. Again I gave a talk on my NIAAA research. T405 Ws

-ACAS 2006 *(sic)* -Army Conference on Applied Statistics, held in Monterey, California. I gave a talk on NIAAA related research T401 S□[592]

-Interface 2007 - (sponsored as above) Systems Biology is the theme. I have organized two invited Sessions, held in Philadelphia, Pennsylvania.[593]

The ACAS and QMDNS although having a military focus in general also have sessions related to drinking behaviors of young enlisted men. 9 I was also invited to give a talk on my alcohol research to the Washington Statistical Society (WSS). Finally, I am one of **three Editors-in-Chief of the Wiley Encyclopedia of Computational Statistics and have arranged for a number of methodology articles related to alcohol studies.** ~~P408~~ Ws

RESEARCH: I completed several reports that are directly related to the topic of the F32 proposal and have been working on additional material. One paper, a refereed invited paper, is focused on a geospatial model of the acute consequences of alcohol use in Northern Virginia. This paper was an expansion of my dissertation work and focuses on the geospatial distribution of acute outcomes, including assault, suicide, murder, domestic violence, child abuse, sexual assault, and driving while intoxicated (DWI) crashes with fatalities. I have been attempting to find sources of data relating international classification of disease (ICD) codes for alcohol and HIV simultaneously, but this does not seem to be so easy to find. Another paper has been accepted in one of the most read outlets in the statistics community and focuses on the use of disability adjusted life years (DALYs) as a measure of both mortality and morbidity. I specifically focused on the consequences for North America compared to the world. Using data from the World Health Organization (WHO), we showed that alcohol is the number two risk factor for DALYs in

[588] According to NIAAA, they also redact not-yet-published works.
[589] www.desmogblog.com/sites/beta.desmogblog.com/files/aa15876-2-3-Progress.Redact_2.pdf
[590] She noted taking an **introductory** 1.5-day SNA course October 2007, **§S.3.2.**
[591] www.interfacesymposia.org/Interface2006/i2006webpage.html gives no title.
[592] This was really in 2005, before 5876 started.
[593] www.interfacesymposia.org/I2007WebPage/documents/second_announcement.pdf She gave T137 Sw① in one of the sessions she organized.

North America following risks associated with tobacco. The interesting aspect of our study is that 34% of the DALYs can be attributed to violence-related outcomes and that these violence-related outcomes are often visited upon the non-drinking bystanders. This suggests a major societal issue. Incidentally, we briefed these results to staffers at the House Subcommittee on Oversight and Investigations and developed in them an encouraging interest, where a hearing on this matter is planned.

Another significant aspect of my work is the analysis of time-of-day, day-of-week, and month-of-year effects. The larger scale model we proposed in the F32 application suggested that we needed to model these temporal effects. We have assembled data on time of day of alcohol related crashes, on time of DWI arrests, and time of admissions for alcohol treatment. I have collected time-related mortality data from the Virginia Department of Health and Virginia OMV crash data both for the entire Commonwealth of Virginia. We have evolved graphical time series plots over a 5-year period and are in the process of building a more complex random effects model of the data at the three time scales. These will eventually be built into the alcohol ecology model.

The F32 proposal also is built on the premise that the model is a social network model. I have been intensively studying and working on developing applicable social network model. In particular, I have under development a bipartite alcohol-related social network model with individuals as one part of the social network and with institutions as the other part. I am developing the theory behind bipartite social network analysis for alcohol studies. I have been exercising this model on co-author networks (one set of actors being authors, the other set being papers). **The social network research was reported in the Congressional testimony during July 19 and July 27, 2007 *(sic)* and in the report that was submitted the U.S. House Subcommittee on Oversight and Investigations.** [594] T126 Ws①③☆★

■ *(NIAAA told me they redacted papers that had not yet been published)* **I have also traveled extensively** and **have gained great insight** into research and professionalism by giving talks and meeting distinguished professors and other professionals.

PROFESSIONAL NETWORKING: I have attempted to develop my network with scholars associated with alcohol research through direct contact, visits and invitations to participate in conferences. I do this to learn of methods and gain insights into alcohol studies. Some of my contacts include Drs. William Wieczorek and Kelly Marczynski of the Center for Health and Social Research at Buffalo State College, Dr. Maria Testa[595] of the Research Institute on Addictions of the University of Buffalo, Dr. Paul Gruenewald of the Prevention Research Center, Gregory Bloss of NlAAA, Kendall Bryant of NIAAA, and Dr. Robert Wilson of the University of Delaware, Dr. Richard Scribner of the Louisiana State University School of Medicine, Drs. Linda and Bengt Muthen of UCLA, Dr. Charles Branas of the Center for Clinical Epidemiology and Biostatistics at University of Pennsylvania, Dr. Bridget Freisthler of the. Department of Social Welfare at UCLA, and Dr. David Anderson of the Center for the Advancement of Public Health at George Mason University. I have been working with Hani Ali, M.D. of George Washington University and with Samer Ellaham, M.D. of Georgetown University respectively in the areas of neonatal pediatrics and cardiology especially with respect to effects of alcohol use in their respective specialty areas. Moreover, I met with Drs. Al Azmi, Kahn and Al Mazrou of the King Fahad Medical City to discuss alcohol and HIV/AIDS in Saudi Arabia.

PAPERS:

Said, Yasmin H. and Wegman, Edward J. (2006) Geospatial distribution of alcohol-related violence in Northern Virginia," in Compstat 2006, (Alfredo Rizzi and Maurizio Vichi, eds.), 197-208' P174 Sw□①[596]

■ ■ ■ ■ *(more papers redacted)*

'Wegman, Edward J, Said, Yasmin H., and Scott, David W. (2006) "Ad hoc Committee Report on the 'Hockey Stick' Global Climate Reconstruction,• A Report to Chairman Barton, House Committee on Energy and Commerce and to Chairman Whitfield, House Subcommittee on Oversight and Investigations: Plaeoclimate *(sic)* Reconstruction, to United States House of Representatives, 2006. T126 Ws①③☆★ again

RESEARCH TRAINING PLANS: I plan to continue with formal and Informal training. **Specifically, I am planning coursework on behavioral biology of substance abuse, psychopathology, geographic information science and epidemiology.** 0 My research plans will follow the logical extension of existing work. Most importantly we are focusing on inclusion of HIV/AIDS. as an additional acute outcome in our model. Also, Dr. Wegman and I are working on the .inclusion of time-of-day effects into the model. We have also recognized that we are dealing with a two-mode social network (individuals and institutions} and are incorporating this dimension into our model. The two-mode network results based on zip code and ABC stores were presented at the 2007 RSA meeting and we were able to develop some mobility information from these results. I am. also working to include and upgrade additional demographic information including gender and age distributions in my model.

[594] **Said thus claimed the specific (and incompetent, plagiarized) SNA work,** of which a subset caused retraction of P179.

[595] **§S.5**-05/08/07 Testa replied to a request from Wegman on HIV.

[596] The paper was almost entirely from her dissertation, done a year before 5876.

In summary, the research in the next year will focus on including HIV/AIDS, temporal and geospatial effects, mobility, two-mode social network, and gender and age-distribution demographics.'

p.6 (Wegman): 'Let me comment first on the George Mason University and me vis-a-vis pr. Said. George Mason University and I have not had much experience with previous F32 awards. We are on a learning curve as well and hope that NIAAA management will take into account our inexperience. That being said, ■ The **original proposal called for a series of courses as well as on-site training with both mentors. Dr. Said has audited courses on Spatial Agent-Based Models of Human-Environment Interactions and on Social NetWork Analysis.** 0 In addition she has spent approximately 2 1/2 weeks at the Center for Health and Social Research at Buffalo State under the mentorship of William Wieczorek, Ph.D. as well as time with the Center for the Advancement of Public Health under the leadership of David Anderson, Ph.D. The coursework was provided by the Center for Social Complexity. Dr. Said 's training previously had been focused on mathematics, computer science and statistics. We have made a great effort this past year in beefing up her background in the quantitative social science dimension. Dr. Wieczorek's Center has a strong geospatial focus while Dr. Anderson's has a strong focus on underage drinking. ■

In conjunction with Dr. Wieczorek's Center as well as with Virginia state agencies, she has been actively seeking data on the connection between HIV/AIDS and alcohol use. She has **undertaken a literature search** along these lines as well. ■ she established additional contacts with epidemiologists at U. Penn medical school, which is **a very promising potential collaboration**. This past year has seen work by Dr. Said published or accepted for publication on geospatial directed graph models, on impact of alcohol on disability adjusted life years in North America, and on two-mode social network analysis. In addition to these papers, Dr. Said has given talks on her work in national and international venues including Italy and Saudi Arabia. She has organized sessions at a variety of meetings and through these sessions, has given additional exposure to the work of NIAAA in general, to the statistics community. ■ Dr. Said's energetic outreach is simultaneously exposing statisticians to alcohol-related research as well as exposing alcohol researchers to more sophisticated statistical modeling methods. In addition to the focus on HIV/AIDS data collection, she has been developing additional data and models for the significance of time-of-day and other temporal effects, and introducing bi-partite graph models (two-mode social networks) into the alcohol modeling framework. In my opinion, ■

S.3.2 5F32AA015876-03 05/30/08[597]

p.3 (Said) 'A: CHANGES

The major change has been associated with classroom coursework. Because of unprecedented travel opportunities, **I have taken coursework informally.** 0 **Interest in social network and text mining work has allowed me to spend approximately two months in Saudi Arabia last Fall.** There I made contact with a number of physicians concerned with drug and alcohol use in Saudi Arabia. This Spring, I was invited as a Visiting Fellow of the Isaac Newton Institute for Mathematical Sciences (INI), Cambridge University, England. The theme of my research project and my academic status are unchanged. **I have spent somewhat more time on research and somewhat less time on formal classroom instruction.**[598] 0

B: PROGRESS

TRAINING: I have been intensively studying materials relevant to the F32 proposal. I have continued my studies with the group (Drs. William Wieczorek and Kelly Marczynski) at Buffalo State College, especially working with the Erie County 2005 Indicators database. Their work also involves interviewing DWI convicted drivers asking them questions concerning the links between alcohol and STDs, especially HIV/AIDS. I have spent several months in Saudi Arabia working with Dr. Abdulrahman Al-Mazrou and Dr. Faroque Khan of the King Fahad Medical City and Dr. Abdullah Al-Ammari of the King Abdul Aziz Medical City learning about their problems with alcohol and drugs and their procedures for dealing with issues. **I also studied social network analysis under Dr. Stanley Wasserman whose book is widely regarded as the authoritative treatment of social network analysis.**[599] **Dr. Wasserman gave the short course at the Army Conference on Applied Statistics in October, 2007.** [600] 0 Beginning January 2 until the present time, I was honored to be appointed as Visiting Research Fellow at the Isaac Newton Institute, Cambridge University. There I was able to study under some of the most distinguished statisticians internationally. The theme of the program at the Institute was Statistical Theory and Methods for Complex, High-

[597] www.desmogblog.com/sites/beta.desmogblog.com/files/aa15876-3-3-Progress.Redact.pdf

[598] Compare with proposal, **§S. 1:** 50% course works 1st year, 25% 2nd year. In first 2 years, she seemed to have done nothing but "sit in on" a few courses.

[599] Wasserman and Faust(1994), often plagiarized by Wegman's group.

[600] Stan Wasserman and Ann McCranie gave this 1.5-day short-course 10/15/07 by, 3 months after P179 and 15 months after T126, for whose social networks part Said had claimed credit in **§S.3.1**. The T126 SNA efforts had started in 2005, 2 years before this introductory short course www.armyconference.org/ACAS07/course2007.htm

Dimensional Data. Of particular interest was a subtheme of combinatorics and statistical mechanics. It turns out that combinatorial methods used in statistical mechanics can be adapted to agent-based simulation. Based on discussions at the INI, the Statistics and Applied Mathematical Sciences Institute (SAMSI) will be developing a program in Fall 2008 to which I have been invited. I have been involved with many conferences during the past year. Details follow.

- Interface 2007: Theme was Systems Biology, Philadelphia, PA. I organized two sessions: 1) Modelling Ethanol Abuse and HIV/AIDS and 2) Systems Approaches for Alcohol Abuse Modelling. Dr. Wegman presented a joint paper with me entitled, Assessing Interventions Related to Negative Effects of Ethanol on HIV/AIDs Spread. T137 Sw① I also collaborated with several others to author papers listed later in this report. (May 2007)
- JSM2007, Salt Lake City, UT. I participated heavily in JSM 2008, *(sic)* the Joint Statistical Meetings. I organized four sessions.[601] **I gave a talk on Experiences with Congressional Testimony** and co-authored several other talks listed below. (August 2007) T401 S③
- Symposium on Arabic Language and Computers, Riyadh, Saudi Arabia. (November, 2007) T416 S③
- ASA Climate Change Workshop, Boulder, CO. I was facilitator for a session. (November, 2007) T415 S③

Interface 2008: Theme was Risk: Reality, Durham, NC. I organized a session and co-organized a second session with Dr. Wegman. My session was Alcohol-Related Public Health Risks and our co-organized session was Text Mining Applications. (May 2008) T417 Sw①, T418 Ws ①

- Statistical Theory and Methods for Complex, High Dimensional Data, Cambridge UK. Semester long program (January to June 2008). I gave an invited talk entitled, Text mining and High Dimensional Statistical Analysis. Upcoming events include:
- HDM-2008 (High Dimensional Statistical Analysis and Data Mining Conference), Kayseri, Turkey. I ·have an invited talk on text mining. (June 2008) T144a Sw③
- RSA Annual Meeting, Washington DC. Dr . Wegman and I have two posters. (July, 2008) P186 Sw①, P187 Ws①
- JSM2008, Denver, CO. I have a contributed paper. (August, 2008)

- COMPSTAT2008, Porto, Portugal. Dr. Wegman, Dr. Wieczorek and I have an invited session on Alcohol Studies. Each of us will give a paper in the session. ***P192*** Ws①, ***P193*** sw①

[601] Mostly 5876-unfit: P182 ws①, P183 ws①, P184 Sw①, P185 sw①, T407 sw③, T408 s③, T409 S③, T410 sw③, T411 sw③, T412 sw③, T413 sw③.

RESEARCH: I have been very active in alcohol studies as well as what have emerged as closely-related fields of *(sic)* social network analysis and text mining. Perhaps most exciting for me is the fact that I was approached by VDM-Verlag to publish an expansion of my dissertation. ⑰ The title of the book will be Intervention to Prevention: A Policy Tool for Alcohol Studies. The publisher has already indicated that they wish to translate the book into German. A list of my papers is given below:

■ ■ *(more papers redacted, here and below)*

3. Wieczorek, William, Said, Yasmin, and Wegman, Edward (2008) Spatial and Computational Models of Alcohol Use and Problems, COMPSTATS2008, Invited Talk and Invited Paper, Porto, Portugal (Refereed paper) ***P193*** sw①

■ ■ ■

7. Said, Yasmin and Wegman, Edward (2008) Agent-Based Simulation of the Alcohol Ecological System, Submitted to JASA Applications ~~P302~~ Sw①

■

9. Wegman, Edward J., Said, Yasmin H .. Scott, David W. (2009) Encyclopedia of Computational Statistics, New York: John Wiley and Sons. ~~P408~~ Ws

■ ■ ■ ■ ■ ■ ■ ■ ■ ■ *(10 works redacted)*

• 20. Said, Yasmin H. (2007) On the Eras in the History of Statistics and Data Analysis, Journal of Washington Academy of Sciences, 93(1), 17-35. P405 S③☆

21. Said, Yasmin H., Edward J. Wegman, Walid K. Sharabati, John T. Rigsby (2007) Implications of co-author networks on peer review, in Classification and Data Analysis, Macerata, Italy: EUM-Edizoni Universita di Macerata, 245-248. P178 Sw①②③

22. Said, Yasmin H .. Edward J. Wegman, Walid K. Sharabati, John (2008) Style of author-coauthor social networks, Computational Statistics and Data Analysis, 52, 2177-2184; doi:10.1016/j.csda.2007.07.021 , 2007 P179 Sw①②③☆

■ ■

PROFESSIONAL NETWORKING: I continue to develop my network with scholars associated with alcohol research through direct contact, visits and invitations to participate in conferences. I do this to learn of methods and gain insights into alcohol studies. Some of my contacts include Drs. **William Wieczorek** and Kelly Marczynski of the Center for Health and Social Research at Buffalo State College, Dr. Maria Testa of the Research Institute on Addictions of the University of Buffalo, Dr. Paul Gruenewald of the Prevention Research Center, Gregory Bloss of NIAAA, Kendall Bryant of NIAAA, and Dr. Robert Wilson of the University of Delaware, Dr. Richard Scribner of the Louisiana State University School of Medicine, Drs. Linda and

Bengt Muthen of UCLA, Dr. Charles Branas of the Center for Clinical Epidemiology and Biostatistics at University of Pennsylvania, Dr. Bridget Freisthler of the Department of Social Welfare at UCLA, and Dr. David Anderson of the Center for the Advancement of Public Health at George Mason University. I have been working with Hani Ali, M.D. of George Washington University and with Samer Ellaham, M.D. of Georgetown University respectively in the areas of neonatal pediatrics and cardiology especially with respect to effects of alcohol use in their respective specialty areas. I have been developing relationships with Ph.D. students as well. **I have worked with a student, Byeonghwa Park, who has completed his comprehensive examinations and will soon go into candidacy for the Ph.D**. He is working on implementing the extended model that we have described in several papers. **I have also been working with another student, Walid Sharabati**, who is developing algorithms for two-mode networks. Most network theory deals only with binary networks. Mr. Sharabati's work deals with frequency count adjacency matrices, i.e. matrices that measure the strength of association. Finally, I have been working with **Dr. David Banks**, a statistician from Duke University, on new grid models for agent-based simulation.[602]
RESEARCH TRAINING PLANS: **I plan to continue with formal and informal training. Specifically, I am planning coursework on behavioral biology of substance abuse, psychopathology, geographic information science and epidemiology.** My research plans will follow the logical extension of existing work. Most importantly we are focusing on inclusion of HIV/AIDS as an additional acute outcome in our model. Also, Dr. Wegman and I are working on the inclusion of time-of-day effects into the model. We have also recognized that we are dealing with a two-mode social network (individuals and institutions) and have incorporated this dimension into our model. The two-mode network results based on zip code and ABC stores were presented at the 2007 RSA meeting and we were able to develop some mobility information from these results. I am also working to include and upgrade additional demographic information including gender and age distributions in my model. In summary, the research in the next year will focus on including HIV/AIDS, temporal and geospatial effects. mobility, two-mode social network, and gender and age-distribution demographics.'

p.7 'For the second year, there was a $30,000 supplement from George Mason University funds supplied by the Commonwealth of Virginia.' §T

p.7 (Wegman) '■ We have begun to exploit the two-mode social network, individuals and institutions in the case of alcohol studies.. **During the last year we came to appreciate that the two-mode social network adjacency matrix has exactly the same structure as the term-document matrix used in text mining.** Because of this Dr. Said has been able to exploit developments in both discipline areas, which accounts for the seeming emphasis on text mining.[603] **Dr. Said has audited courses on Computational Linguistics and Text Mining, on Statistical Methods for High Dimensional Data, and on Social Network Analysis**.⓪ In addition she has spent time with the Center for the Advancement of Public Health under the leadership of David Anderson, Ph.D. Some of the coursework was provided by the Center for Social Complexity. Dr. Sa id's training previously had been focused on mathematics, computer science and statistics. We have made a great effort this past year in beefing up her background in the quantitative social science dimension. Dr. Wieczorek's Center has a strong geospatial focus while Dr. Anderson's has a strong focus on underage drinking. quantitative social science dimension. Dr. Wieczorek's Center has a strong geospatial focus while Dr. Anderson's has a strong focus on underage drinking. ■

Dr. Yasmin H. Said is currently a Visiting Fellow at the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge in England. She earned her A.B. in pure mathematics, her M.S. in computer science and information systems, and Ph.D. in computational statistics. She is the editor of Computing Science and Statistics, ⓫ is an associate editor of the journal, Computational Statistics and Data Analysis, ⓬ serves on the board of the Washington Statistical Society, ⓭ and the Interface Foundation of North America. ⓮ She was chosen to serve on the American Statistical Association Presidential Task Force on Science Policy⓯ and currently serves on the American Statistical Association Presidential Advisory Committee on Scientific and Public Affairs. ⓰ She is one of the editors in chief and serves as managing editor of the multi-volume Wiley Encyclopedia of Computational Statistics. ⓾ She is finalizing a book, Intervention to Prevention: A Policy Tool for Alcohol Studies, to be published by VDM Verlag. ⓱ The publisher has indicated that they wish to translate this work into German. This past year, Dr. Said was elected as a member of the International Statistical Institute (ISI). ⓲ ■ Society on Alcoholism, ⓳ an elected member of Sigma Xi, the Scientific Research Society, ⓴ an elected member of the Washington Academy of Science, ⓪ and is a **member of more than 15 scientific, research, and professional societies**. ① ■'

[602] How many peer-reviewed research journal papers resulted from all this? **Zero**.

[603] *They may share the same basic underlying graph theory, but one might ask alcoholism experts of their views on the applicability of text mining.*

S.4 Summary in lieu of Final Report

NIAAA said that a Final Report was not required, and I did not inquire further, given the number of issues already accumulated, summarized here:

In progress reports (and works found), Said claimed (or ack'd) 5876 for unfit works: **18 false claims** by topic ③:

P178, P179, P200, P405, T407, T126, T148, T151, T408, T409, T410, T411, T412, T413, T415, T420, T425, T432.

She claimed P174, done a year earlier and T400, T401, done ~9 months early □, adding 3 more for a total of **21 false claims.**

Claiming T126 and P179 falsified her assertions of "pro bono" work made in T409 and T415, but those statements were not made to the government, so were not counted as false statements in the legal sense.

Searches via Google Scholar for papers from 2009 onward for {yasmin said or yh said} and {alcohol or alcoholism} found P194, P195 and a patent, but nothing later.

After the fellowship ended in 2009, no more alcoholism works were found.

A curiosity was her apparent use of an off-campus email (ysaid99 AT hotmail DOTcom) address for all work, even grant-related, throughout this entire time.

S.5 Emails

[WEG2010a] contains some back history on the NIAAA **5876** grant, and also shows the acks for ***P180***. It includes some possible distractions (red), as well. Some messages are excerpted here in chronological order.

02/28/05 p.31 Said wrote to NIAAA**,** asking how postdoc is coming

03/02/05 p.28 Proposal for grant Fedexed to NIAAA

03/02/05 p.30 NIAA found several missing signatures, review date not set

03/21/05 p.31 Said wrote to NIAA, noting problems with email

05/04/05 p.30 Said dissertation defense

06/01/05 p.30 Email from Said to NIAAA, noting accomplishments.

'I have officially graduated with a Ph.D.[604]
The website that I created is www.alcoholecology.com. [605]I have won couple of recent awards for my dissertation and for my academic accomplishments. The Daily Mason Gazette wrote an article about the alcohol research that I did. You may see the article at gazette.gmu.edu/articles/index.php?id=6811**. I have also done six conference papers since graduation**[606] and have been invited to write a **chapter in the Encyclopedia of the Modern World** to be submitted this summer and will be published by Oxford University Press**. In addition, the Genetic Algorithm chapter that was written in the Handbook of Statistics is now available and in print.**' P402 S☆

09/xx/05-07/xx/06 Wegman Report, much effort

12/22/05 p.32 NIAA summarized current status

'Reviewers indicated that this resubmission addressed most of the concerns raised in the original review, including a better explanation of the mentors' roles. **Particular strengths include a candidate who is a great student** and has a lot of potential, and an ambitious exciting model with a lot of promise for the field. **Remaining weaknesses include no inclusion of sociological and economic theory issues in the training plan; a need for greater demonstration of understanding about the alcohol policy content area (e.g., the proposal contains no alcohol references); and a need for greater detail about the research methods (including some preliminary findings from the model drawn from her dissertation work)**. Reviewers were divided on this last item, with some noting that the proposal may be a little thin in this regard largely because where the applicant is taking this concept is an evolving paradigm in the field.'

12/22/05 p.33-36 part of Critique 1

'Sponsor and Training Environment: As noted in the original review of this application, and not repeated here, Drs. Wegman, Wieczorek and Anderson are eminently qualified to mentor Dr. Said in her traineeship. **This resubmission further explains her working relationship with Dr. Wieczorek**,[607] illuminating his contribution to her training program. **Dr. Wieczorek's long association with ecological modeling activities in alcohol studies eminently qualifies him for this role. As noted in the original review, the training environments at the Center for Computational Statistics, and in Dr. Wieczorek's research group at the State University of New York at Buffalo, are excellent.'**

'2. The role of Dr. Anderson has been de-emphasized and that of **Dr. Wieczorek expanded** to reinforce Dr. Said's specific training with regard to alcohol epidemiology, policy and ecological issues.'

'4. Dr. Said has focused …. **During her traineeship, she will study the current state of systems dynamic models of alcohol and drug ecologies with Dr. Wieczorek.**

5. With expanded association and training with Dr. Wieczorek, including regular contact with him at his site in Buffalo, and including his now regular trips to work with Dr. Wegman in Delaware,[608] it is clear that Dr. Said will get the requisite guidance needed to move into this area of alcohol studies.'

[604] [MAS2012c §3.4] summarized evidence of plagiarism in her dissertation.
[605] wayback.archive.org/web/*/www.alcoholecology.com
This was first archived May 2005, but not after February 2006.
[606] I am not sure which these might be, but 6 papers in a month is *impressive*.

[607] [MAS2010a §A.1.4, §A.1.5] Wieczorek was on Said's committee and one of the 8 "reviewers" of the draft WR, claimed by Wegman "None of these people have actively collaborated with me in writing research papers." and so were not part of his social network. He is a well-published health researcher.
[608] So, far, no references have surfaced for studies in Delaware.

'7. The human subjects section has been corrected to reflect the fact that Dr. Said will be using secondary data sources throughout her work. Although these responses do address, to some extent, the concerns of the original reviewers, the application is weakened by the lack of information in three domains: (1) **There is a substantial literature which bears upon the social ecology of alcohol use. Acknowledgement of this literature, and recognition of the major scientific questions that drive this research should be provided.** What will the model contribute to this literature? How will it do so? (**The fact that this literature was not reviewed in the application reduces confidence that the applicant will receive adequate mentoring during the training period.**) (2) Although a preliminary model was developed as part of the applicant's dissertation, no results from this development step, nor any primary scientific questions emerging from This development, are noted. The model itself remains vaguely described. (Presumably, some of the relevant alcohol literature was reviewed as part of the candidate's dissertation effort. Again, the absence of a review of this literature remains problematic.) (3) As described, mentor participation in the training program is not sufficient to fill the necessary gaps in knowledge evidenced by this application.'

Summary and Recommendation: This is an adequate response to the reviewer's initial concerns regarding the first submission. It would appear that Dr. Said's training could place her squarely at the center of current developing research on mathematical ecological models of alcohol use, abuse and related problems. **However, the lack of program support for acquiring the substantial amount of specific information needed to build and successfully test the proposed model detracts from enthusiasm for this work. General courses in psychology, public health, public policy, etc. will not provide the very focused content necessary to conduct this work. To this end, the training program will need extensive integration with relevant resources in the alcohol literature.**'

12/22/05 p.36-39 part of Critique 2

'2) *Sponsor and Training Environment.* The main sponsor, Dr. Wegman, has a prestigious set of accomplishments, honors, memberships, and publications in applied statistics. …**One concern is that, although the proposal states Dr. Wegman has been working with NIAAA since 2001, his letter and biosketch do not discuss or list any alcohol-related publications, grant funding, or contracts.**'

'However, there are several concerns about the research plan as written. A major concern is the absolute absence of a background and significance section that includes a review of the literature relevant to the research plan and that summarizes prior alcohol policy work of others and states how her model addresses gaps or improves upon things already done. **Strangely, the proposal does not cite any literature at all. This makes it very difficult to gauge: a) her understanding of the alcohol policy literature; b) her understanding of the relevance of her work; and c) the amount of help she is getting from her mentors on alcohol issues.** This is all the more important because the model she intends to construct is very complex.'

'This revised F32 application has a number of strengths. The candidate appears to be a talented, motivated, and hard-working individual. She was clearly a star graduate student. **She has established mentoring relationships with the sponsors who are very committed to her success**. She has initiated novel work in an important area of alcohol related research that is matched to her skills in statistics and computer science. **However, the application is weakened by the absence of an involved alcohol expert at her home institution and by several significant deficiencies in the research training plan. These include the absolute absence of cited alcohol literature and poorly detailed methodology**.'

02/24/06 p.40- Said sent more material.

05/19/06 p.46- NIAA congratulated Said on her recent fellowship award, specified stipends.

(finish Wegman Report by 07/13/06)
07/19/06 Wegman presented to Congress, Said accompanied him
07/27/06 Wegman presented to Congress

08/30/06 p.46- NIAA wrote to Said:
'**Please call me regarding part time work while having an F32 Date**:
Wed, 30 Aug 2006 15:37:10 -0400
I've gathered information about this issue. Please give me a call.'

09/01/06 p.48 Said wrote to NIAAA
'I just got your email and I am currently in Italy at a conference presenting my alcohol work. **All the feedback that I am getting regarding my research has been exceptionally positive.**'[609]

[609] This seems the same trip in which Wegman claimed P174-P175, T129-T130, so Said *likely* delivered P174.

11/27/06 p.1 §R.6 Army-related
12/15/06-12/15/07 0059 grant, Said heavily involved, §R.
01/22/07 pp.1-2 §R.6 ARL to Said for visit next day

05/03/07 p.50 Wegman to Testa
Sought information on HIV/alcohol.

05/08/07 p.49 Testa to Wegman
'Sorry for the delay in responding. Unfortunately, don't think I can be terribly helpful here since this is really quite removed from my own work. But here are a couple of thoughts for what they are worth
1) If I am understanding correctly, I don't think that you can calculate risk of contracting HIV when intoxicated from ICD codes. Those 7 people seem to be people whose cause of death was a combination of alcoholism and AIDS. This is very different than having sex with an infected person when intoxicated.
2) The risk of contracting HIV when intoxicated is going to be really, really small. The chance of contracting HIV in a single sexual encounter with an infected person is very low (I know you can find the numbers out there, they are much much *(sic)* lower than the chances of getting an STI). And in the general population rates of HIV infection are very low. I'm not sure how much being intoxicated when having sex would alter these rates...
The vast majority of people in the general population have 1 sexual partner who is not infected so they are at low risk and having sex while intoxicated does not affect their risk. In young heterosexuals, acute intoxication has been linked to greater likelihood of having sex with casual partners/ partner not well known but not with a lower likelihood of condom use. In fact, many studies suggest the opposite - more likely to use condoms when inoxicated *(sic)* because more likely to use condoms with casual partners.'

07/19/07 p.52 NIAAA wrote to Said
'**I am extremely concerned that you have not yet filed your progress report/renewal application for your fellowship grant**, "A Social Network Model of Ecological Alcohol Systems" (5-F32-AA-015876-02). As you know, this report was due on March 1 of this year, and **your first year of support ended on May 25**. I understand that you have spoken to Heather Gill-Denny, Grants Management Specialist for your award, on several occasions and that she understood that you intended to submit your renewal application in short order. **However, an extended time has now passed and your renewal application still has not been received. This is highly unusual and I am not certain what procedures may apply at this point.**'

07/26/07 p.54 GMU wrote to NIAAA
'Attached please find a signed PDF file progress report related to NIH 1F32-AA015876-01A1. This is in addition to original documents sent by express mail as well as a faxed document *(sic)* sent to Kathrine Gill-Denny, this afternoon.' (this was filed

07/27/07 p.53 GMU wrote to NIAAA
'I did send by Fedex, the entire packet to to *(sic)* the central mailing address and it should be there sometime today. **Also, on another note, when I prepared the cover letter cover letter *(sic)* I indicated that Dr. Wegman is the PI on this award and note *(sic)* Dr. Said. This due to the fact that in our process Dr. Wegman must be listed as PI on our internal paperwork.**[610] Dr. Said brought this to my attention. I hope this does not hamper the review process.' ***(§S.3.1** shows the filed report.)*

07/13/07 p.68 *CSDA* Editor Azen wrote to Wegman[611]
'I personally reviewed your very interesting (and unique) manuscript. I think the paper is very interesting, and I could not identify any errors. So, I am pleased to inform you and your colleagues that your paper "Social Networks of Author-Coauthor Relationships" has been accepted for publication in Computational Statistics and Data Analysis. address.'

08/31/07 p.12-13 Said to *Chance* (Amstat), about ***P180***:
'Please include the following acknowledgements: *(Usual words for F32AA015876 for Said, W911NF-04-1-0447 for Wegman , W911NF- 07-1-0059 for both.)*

12/15/07 0059 grant ends

03/05/08 Authors Guide for *WIREs:CS*, Said was Managing Editor
WIREs:CS had been in planning since December 2005, **§H.6.**

[610] This may explain why Wegman listed this, **§K.1**, *although it makes little sense, given that this sort of award is supposed to be for a postdoc, with mentoring from others, as a step towards being a PI.*
[611] [MAS2011a §5] Wegman and Said frequently criticized peer review in climate science. Editor-in-Chief Azen, who had no publications on SNA, accepted it in less than a week, although *CSDA* generally did not publish SNA papers.

04/21/08 p.54 Said wrote to NIAAA

'I hope this message finds you well. I have spoken to Dr. Wegman and he said that he will look into the status of the F32 this week.[612] He was in Wisconsin and had just gotten back. I will immediately inform you the moment I get any information. Once again, thank you very much for your patience and for looking after me. I honestly and truly appreciate it. I hope you have a wonderful and productive day.'

02/03/09 p.55 Said to NIH(?)

'Dear Dr. Yum,

My advisor, Edward Wegman, has informed me that you gave me permission to submit my application by email. The first two attempts (made I believe on January 27 and 28) gave me error messages so I very much appreciate the opportunity to use the email. I will emulate the electronic form as much as possible..

When I met my advisor, mentor, and colleague, Dr. Edward J. Wegman, his discussion of the data sciences, holding a data centric viewpoint for his work, drew my interest in working with him. **He suggested the alcohol research**, which blended my interests in health-related applications and novel mathematical and statistical tools. My Ph.D. dissertation involved developing an agent based microsimulation of all 1,000,000 residents of Fairfax County, Virginia with respect to acute outcomes related to alcohol use and abuse. I am currently adding social network model to the microsimulation. I believe my participation in the workshop will add a depth of understanding to the research I am currently doing and planning for my career. …

How did I hear about the Institute?

My project manager for my F32 NIAAA post doctoral award is Gregory Bloss. He sent me an email strongly encouraging me to apply for this Institue *(sic)* training program. **I currently have a Career Transition K22[613] application pending**. Dr. Bloss told me that not only would participation in this Institute be good for my research ambitions, but also that it would be a good credential to bolster my K22 application.'

04/02/09 p.24 Army to Said:

'Professor Ed Wegman requested that I forward you the invitation below from the **Army Research Office**.

On behalf of Dr. Janet Spoonamore, Army Research Office, please join us for the Decision Sciences Workshop at the United States Military Academy, West Point, NY during 16-17 June 2009.'

04/02/09 p.24 Said to Army

'Thank you for your email. I accept your kind invitation.'

04/xx/09 Said interviewed at Oklahoma State U, offered an Assistant Professorship in Statistics, accepted for Fall Semester.

She later negotiated less teaching time.

05/25/09 5876 ended

08/xx/09 Said requested start in January 2010, refused by OSU, who withdrew the offer, so she never was employed there.

She labeled herself a Professor at OSU for 2 years on *WIREs:CS* masthead, **§H.6**

10/01/09 p.57 NIAAA to Said

'Please submit a termination notice for the above referenced grant application. The grant ended on May 25, 2009 but we have not received a termination notice. Please submit one as soon as possible as per the terms of the Notice of Grant Award:

Upon completion of NRSA support, policy requires the submission of a Termination Notice (PHS 416-7), available at the web address grants.nih.gov/grants/forms.htm. **Please submit this form immediately following the termination of support to the above address.'**[614]

[612] *It is not yet clear why Wegman would be looking into this, rather than Said.*

[613] 'The following Career Transition Awards provide support to an individual postdoctoral fellow in transition to a faculty position:' grants.nih.gov/training/careerdevelopmentawards.htm
She apparently did not get this.

[614] NIAAA said no Final Report was required, so perhaps I did not ask the right question to get this. I did not pursue it further, given the large number of problems already documented. Finding that, if it existed, could only add more problems, not remove any seen so far.

S.6 Alcoholism works, rarely in alcoholism-relevant venues

Although many could not be examined, they seemed to fall into a few groups (especially ⑩ and ⑬) in which the same research results were given more repetitively than seems typical for research progress.[615]

=====2005=====

⑩T112W①"Ecology of Alcohol and Alcoholism," Int'l Conf. on the Future of Statistical Theory, Practice and Education, Hyderabad, IN, Dec '04-Jan '05'

⑩P401S☆ 0/0 "Agent-Based Simulation of Ecological Alcohol Systems" PhD dissertation, May 2005

⑩T117Ws①"Ecology of Alcohol and Alcoholism," Wegman, with Y. Said, **Interface** 2005. June, St. Louis

⑩T118Sw "Modeling Alcohol Abuse and its Consequences," with Y. Said, (presented by Y. Said**), Interface** 2005. June, St. Louis.

⑩T400S□"Agent-based Model Applicable to Homeland Security and Disease Control" SAMSI/NISS, Triangle Park, NC, September.

⑩T401S "Adaptation of an Alcohol Ecological Agent-Based *(sic)* to Homeland Security" **ACAS05**, Monterey, CA October 2005

=====2006=====

⑩T405 Ws[616] "A Policy Tool for Assessing Alcohol Intervention Strategies" *JSM2006,* August, Seattle, WA

⑩P174 Sw□①=T129 W① 1/0 "Geospatial distribution of alcohol-related violence in Northern Virginia" COMPSTAT 2006, Aug-Sep
Mostly a direct extract from P401S, although often changing "we" ← "I."

=====2007=====

⑪T137 Sw① "Assessing Interventions Related to the Negative Effects of Ethanol on HIV/AIDS Spread," **Interface** 2007, Phila., PA, May

⑫T138 W① "A Bipartite Graph Model of the Interaction between Alcohol Users and Institutions," Research Soc. on Alcoholism Meeting, Jul '07 Chicago.

⑪T139 W① "Assessing Interventions Related to HIV Incidents Under the Influence of Ethanol,"JSM200, Salt Lake City, UT, August

? ***P180*** Sw①② 5/0 "Quantitative assessments of alcohol-related outcomes" *Chance* 2007. *(peer-reviewed, but not research journal)*

⑬P185 sw 1/0 "Temporal statistics for consequences of alcohol use" *Proceedings of the Joint Statistical Meetings*, *JSM2007, Salt Lake City, August'*

=====2008=====

⑬T417 Sw① "Estimating Spatiotemporal Effects for Ecological Alcohol Systems," **Interface** 2008, May *annotated: joint work with Wegman.*

⑭T418 Ws ① "Spatial and computational models of risks for alcohol users," **Interface** 2008. May 2008, Wegman presented, Wieczorek was lead.

⑬P186 Sw① 0/0 "Using administrative data to estimate cyclic effects of alcohol usage (refereed abstract)," *Alcoholism: Clinical and Experimental Research*" June 2008. Washington, DC *According to Said, a poster*, **§S.3.2.**

⑬P187 Ws① 0/0 "Modeling spatiotemporal effects for acute outcomes in an alcohol system (refereed abstract)," *Alcoholism: Clinical and Experimental Research*" June 2008. Washington, DC *According to Said, a poster*, **§S.3.2.**

⑬***P192*** Ws① = T146 W 2/0 "A directed graph model of ecological alcohol systems incorporating spatiotemporal effects," COMPSTAT 2008, August, Porto, Portugal. *(counted twice)*

⑬T419 Sw "Estimating Spatiotemporal Effects for Ecological Alcohol Systems," COMPSTAT 2008, August

⑭***P193*** sw① "Spatial and computational models of alcohol use and problems," COMPSTAT 2008, August. This was led by Wieczorek

⑫T149 W "A Bipartite Graph Model of the Interaction between Alcohol Users and Institutions," NISS Workshop on Agent-Based Modeling, RTP NC Nov'08

~~P302~~ Sw① "Agent-based simulation of the alcohol ecological system," submitted *Journal of the American Statistical Association*' unpublished.

=====2009=====

⑬P194 Sw① 0/0 "Estimating cyclic and geospatial effects of alcohol usage in a social network directed graph model," in *Social Computing and Behavioral Modeling* 2009 (workshop)

⑬P195 Ws① 1/0 "A social network model of alcohol behaviors," in *Social Computing and Behavioral Modeling* 2009 (workshop)

[615] It is *fine* to present similar material at different conferences, and then hopefully once in a peer-reviewed journal. *At some level of duplication, it is just recycling.*

[616] www.amstat.org/meetings/jsm/2006 p.423

Similar papers/talks *seemed* to be given with different titles, sometimes by either Wegman or Said, whether or not other was listed as coauthor. Discovery of this pattern led to the more extensive discussion in **§K.2**. All this generated zero papers in peer-reviewed research journals.

A confusing2008 set of references is consolidated here.

Organizer: Said		**Organizer: Juergen Symanzik**[617]	
Interface[618]	**Presenter**	**COMPSTAT**	**Presenter**
-		***P192***Ws = T146W	Wegman
T417	Said	T419 S	Said
T418 ws	Wegman	***P193*** sw	Wieczorek

[WEG2010] listed T146W (no Said), but [WEG2010a p.60] had:[619]

> 'Yasmin Said to Juergen, ewegman, me, wieczowf, mpbrito
> show details 1/13/08
> Dear Juergen,
> You are most welcome and thank you very much for your email. The order of talks and speakers are the following:
> 1. Edward J. Wegman, title of talk: "A Directed Graph Model of Ecological Alcohol Systems Incorporating Spatiotemporal Effects" ***P192***Ws①
> 2. Yasmin H. Said, title of talk: "Estimating Spatiotemporal Effects for Ecological Alcohol Systems" T419S
> 3. William F. Wieczorek, title of talk: "Statistical and Computational Models of Alcohol Use and Problems" ***P193***sw①
> Please change the session title to: "Computational Statistics and Data Mining Methods for Alcohol Studies" (Interface Session)

That was in reply to an 01/12/08 message from Symanzik, reminding them of 01/20/08 due date for papers. On pp.61-62, Wegman noted uploading Paper 192 on 02/18/08, and then Said sent paper 193[620] 02/19/08, a month late. On p.62, 04/28/08, Conference Chair Paula Brito told them they needed to convert colored figures to black and white, as required.

P193 used parallel coordinate displays (on which Wegman had written) and conditioned choropleth maps (CCmaps), to analyze Buffalo-area data, **but not Said's software on which so many of the other Wegman/Said works were based.** *That is a bit curious. Perhaps the model was not generally very useful?*

A domain expert might examine the whole set and assess the amount of real progress made beyond P401S and its connection to the field, a reviewer's concern,**§S.5**-12/22/05.

Google Scholar searches found very few citations to any of these beyond a few self-referential ones.

To try to understand Wegman and Said's interaction with the alcoholism research field,[621] I did quick survey, on the assumption that a known senior researcher would likely publish in relevant journals.

Wieczorek published alcoholism studies in journals[622] such as *Am J Drug Alcohol Abuse*(2), *Addict. Behavior*(2), *J. Crim. Justice, Subst. Use Misuse*(4), *Soc. Sci. Med, J Stud Alcohol*(2), *Alcohol Clin Exp Res, Alcohol Health Res World, Alcohol Clin Exp Res, J Consult Clin Psychol*.

Nothing by Wegman or Said was found in any of those.

The NIH provides a powerful search engine for research grants.[623] From Said's grant,[624] to Similar Projects [625] and selecting NIAAA gave ~100 projects. I checked 3 with roughly similar cost

[617] One of Wegman's earlier postdocs and occasional coauthor.
[618] www.webcitation.org/6EhvC60ce
[619] [WEG2010a pp.58-67] shows COMPSTAT2008 correspondence.
[620] The emails were confusing, because they mentioned papers 192, 193 and 194, but while 192 was indeed ***P192 Ws, 193 was really T419S, and 194 was P193sw.***

[621] Unlike some of the other topics, I know nothing about this one.
[622] www.ncbi.nlm.nih.gov/pubmed?dispmax=25&db=pubmed&pmfilter_EDatLimit=added+to+PubMed+in+the+last+0+i&cmd_current=Limits&orig_db=PubMed&cmd=Search&term=Wieczorek+WF&doptcmdl=DocSum
In an unfamiliar field, it helps to see where a known serious researcher publishes.
[623] projectreporter.nih.gov/reporter.cfm
[624] projectreporter.nih.gov/project_info_history.cfm?aid=7491624
[625] projectreporter.nih.gov/project_info_Like.cfm?aid=7491624

Each shows (# peer-reviewed journal papers) and lists the journals found:

COMPARING SOCIAL NETWORK INFLUENCE ON ALCOHOL USE USING AFFILIATION DATA,[626] (4), in journals:
The Journal of adolescent health : official publication of the Society for Adolescent Medicine, Social science & medicine, Child development, Social networks[627]

PREDICTING HIV-RELATED BEHAVIORS OF SOCIAL NETWORKS IN A MPS ENVIRONMENT[628] (2) in:
Social networks, *International journal of medical informatics*

ECONOMIC EVALUATION OF ADOLESCENT ALCOHOL USE AND THE IMPACT OF SOCIAL NETWORKS[629] (6) in:
Applied economics, Drug and alcohol dependence, Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention, Health affairs, BMC pediatrics, Academic pediatrics.

Wegman and Said published nothing in any of these journals, either.

This is only a small sample, but ***argues against evidence of much real engagement of Said and Wegman with the broader alcoholism field.***

As one last check, I examined the NIAAA newsletters to see if relevant meetings were available to Wegman and Said.

The NIAAA newsletters[630] listed relevant meetings for 5876's duration:

Dates	**# meetings**	**# nearby (Washington, DC area)**	
2006.06-2006.10[631]	10	2	
2006.11-2007.02[632]	2	1	
2007.03-2007.06[633]	7	3	
2007.07-2007.11[634]	8	2	T138 W① RSoA Chicago
2008.02-2008.03[635]	6	3	
2008.04-2008.07[636]	6	6	P186 Sw①, P187 Ws①[637] Washington
2008.09-2008.10[638]	6	3	
2009.02-2009.04[639]	6	2	
Totals	**51**	**22**	

Despite the wealth of relevant USA meetings, even just the local ones, Wegman spoke at 2 such, and Said only 1. P186 and P187 were posters. Their alcoholism works, starting with T117 were actually given at:

6	COMPSTAT: Rome(2), Portugal(4)
4	Interface: St. Louis (2), Philadelphia, Durham,
3	**Research Society on Alcoholism: Chicago, Washington(2)**
2	Social Computing and Behavioral Modeling, Phoenix, AZ
3	JSM: Seattle, Salt Lake City (2)
2	SAMSI, NISS at RTP, NC
2	ACAS: Monterey, CA, Lexington, VA(2)
1	*Chance* (peer-reviewed, but not a research journal)
23	**Total**

They simply did not seem very engaged with the most relevant conferences, but spoke often about alcohol models to others.

[626] projectreporter.nih.gov/project_info_description.cfm?aid=8326357
[627] MAS2010a §W.5.6.1. This is a sibling journal of *CSDA*.
[628] projectreporter.nih.gov/project_info_description.cfm?aid=8309291
[629] projectreporter.nih.gov/project_info_results.cfm?aid=8311015
[630] www.niaaa.nih.gov/news-events/niaaa-newsletter
Some of these may not have been open meetings.

[631] www.niaaa.nih.gov/sites/default/files/newsletters/Newsletter_Number10.pdf
[632] www.niaaa.nih.gov/sites/default/files/newsletters/Newsletter_Number11.pdf
[633] www.niaaa.nih.gov/sites/default/files/newsletters/Newsletter_Number12.pdf
[634] www.niaaa.nih.gov/sites/default/files/newsletters/Newsletter_Number13.pdf
[635] www.niaaa.nih.gov/sites/default/files/newsletters/Newsletter_Number14.pdf
[636] www.niaaa.nih.gov/sites/default/files/newsletters/Newsletter_Number15.pdf
[637] T138 was at 2007 RSoA conference. P186 and P187 were presented at the 2008 RSoA-cosponsored conference. Wegman labeled the latter as peer-reviewed journal papers, but they were at most peer-reviewed abstracts.
Said called them posters, for which the bar is much lower.
[638] www.niaaa.nih.gov/sites/default/files/newsletters/Newsletter_Number16.pdf
[639] www.niaaa.nih.gov/sites/default/files/newsletters/NIAAA_Newsletter_17.pdf

S.7 Other biographies with annotations

See also **§S.1** for comparison**, but this is just for context.**

S.7.1 The Spring 2005 dissertation [SAI2005][640] offered:

'**Yasmin Said was born in Arecibo, Puerto Rico and was resident there until age five. She began elementary school in Tulkarem, Palestine (West Bank) from first through third grade. She came to the United States to finish her education. By the time she finished her elementary education, Yasmin was trilingual in Spanish, Arabic and English.** She attended Rockville High School, where she was a journalist for her high school newspaper, an editor of the yearbook, and a photographer for both the newspaper and the yearbook. She graduated with an honors certificate of merit. She attended Trinity College where she earned an A.B. in pure mathematics with minor in **bioethics** and philosophy. She was awarded the Trinity College Leadership Award for four years, was on the National Dean's List, and the Trinity College Dean's List. She received an M.S. in Computer and Information Systems at American University in Washington, DC. At American University, she placed first in the regional mathematics competition, was awarded the American University Leadership Award and the AU Women's Leadership Award.

Subsequent to her graduation from Trinity College, she began a five-year high school teaching career in Washington, DC and Northern Virginia, teaching calculus, statistics, algebra, and finite mathematics. She enrolled in the Ph.D. Program in Computational Sciences in Fall 2003 and pursued the Ph.D. with a specialization in Computational Statistics with a 4.0 grade point average. Her dissertation, *Agent Based Simulation of Ecological Alcohol Systems*, introduces a new framework designed to simultaneously mitigate the acute effects of alcohol abuse. Her research can be viewed at the website http: //www.alcoholecology.com.

During her time at George Mason University, she has published a research paper entitled "On Genetic Algorithms and their Applications" in the *Handbook of Statistics: Data Mining and Data Visualization*. She has organized sessions at both the Symposium on the Interface of Computing Science and Statistics and the Joint Statistical Meetings. She is a fellow of the Royal Statistical Society, 10 an elected member of the Research Society on Alcoholism, 19 and a member of the American Statistical Association, Institute of Mathematical Statistics, Institute of Electrical and Electronic Engineers, American Association for the Advancement of Science, American Mathematical Society, Society for Industrial and Applied Mathematics, and The Interface Foundation of North America.'

[640] web.archive.org/web/20060905150733/www.galaxy.gmu.edu/stats/syllabi/IT871/MasterCopyDissertation.pdf original disappeared in August 2010.

S.7.2 The 03/05/08 *WIREs:CS* Guide for Authors biography showed:[641]

'*Yasmin H. Said,* Ruth L. Kirschstein National Fellow, George Mason University.

Dr. Yasmin H. Said is a Visiting Fellow at the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge in England and is a National Research Fellow from the National Institutes of Health. She earned her A.B. in pure mathematics, her M.S. in computer science and information systems, and Ph.D. in computational statistics. She does **alcohol modeling, agent-based simulation modeling, social network analysis, text, image, and data mining**,[642] and **major public policy work trying to minimize negative acute outcomes, including HIV/AIDS, related to alcohol consumption.** Dr. Said is also the Statistical Methodology Director of the Innovative Medical Institute, LLC, and Co-Director of the Center for Computational Data Sciences in the College of Science at George Mason University. She is the editor of *Computing Science and Statistics*, 11 is an associate editor of the journal, *Computational Statistics and Data Analysis*, serves on the board of the Washington Statistical Society, 13 and serves on the American Statistical Association Presidential Task Force on Science Policy. 16 Dr. Said is an elected member of the International Statistical Institute, an elected member of the Research Society on Alcoholism, 19 and an elected member of Sigma Xi, the Scientific Research Society. 20 She is currently writing a book, *Controversies in Global Warming*[643] and another, *Statisticians of the Twentieth Century*.[644] She has published a book, *Intervention to Prevention: A Policy Tool for Alcohol Studies.* 17 **With colleagues she has developed testimonies on global warming for the House Committee on Energy and Commerce and to the House Subcommittee on Oversight and Investigations**.[645] She has also taught probability and statistics at The Johns Hopkins University in Baltimore, MD.'

Both bios seemed extremely impressive. It takes effort to check all this.

[641] media.wiley.com/assets/2205/94/WIREs_comp_stats_author_guide.pdf www.webcitation.org/6DGCsA8Pp p.6
Various people commented on this during 2011 at:
deepclimate.org/2011/05/15/retraction-of-said-wegman-et-al-2008-part-1 and
deepclimate.org/2011/10/04/said-and-wegman-2009-suboptimal-scholarship

[642] As of that date, she had published little besides alcoholism work.

[643] This has not been published, **§H.7.**

[644] I can find no trace of this either.

[645] *It would not surprise me if she had spent more time on the WR than Wegman, but given this was both 0447- and 5876-claimed, some questions might be asked.*

S.7.3 Her biography for 08/xx/11 ISI 2011[646] was

Biography: Dr. Yasmin H. Said was a Ruth L. Kirschstein Research Professor from the National Institutes of Health at George Mason University and was a Visiting Fellow at the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge in England. She earned her A.B. in pure mathematics, her M.S. in computer science and information systems, and Ph.D. in computational statistics. She does alcohol modeling, agent-based simulation modeling, social network analysis, text, image, and data mining, and major public policy work trying to minimize negative acute outcomes, including HIV/AIDS, related to alcohol consumption. Dr. Said is also the Statistical Methodology Director of the Innovative Medical Institute, LLC, ② and Co-Director of the Center for Computational Data Sciences[647] in the College of Science at George Mason University. She is the[648] editor-in-chief of *Wiley Interdisciplinary Reviews: Computational Statistics, editor of Computing Science and Statistics*, ⑪ was an associate editor of the journal, *Computational Statistics and Data Analysis*, ⑫ serves as the President of the Interface Foundation of North America,[649] serves on the executive board of the U.S. Army Conference on Applied Statistics,[650] serves on the executive board of the Washington Statistical Society, ⑬ serves on the executive board of directors of the Interface Foundation of North America, ⑭ serves on the American Statistical Association Presidential Task Force on Science Policy, ⑮serves on the American Statistical Association Presidential Advisory Committee on Scientific and Public Affairs, ⑯ serves on the Executive Committee of the American Statistical Association Section on Defense and National Security,[651] and serves on the Scientific Steering Committee for the U.S. Army Conference on Applied Statistics.[652]

She has published a book, *Intervention to Prevention: A Policy Tool for Alcohol Studies*. ⑰ She is currently one of the editors in chief of the multi volume *Wiley Encyclopedia of Computational*. ~~P408~~ Ws Dr. Said has a patent on *Policy Analysis and Action Decision Tool*,[653] two patents pending, *Automated Generation of Metadata*,[654] and *A Multimodal System Tool for Aiding Autonomous Discovery*. [655] Dr. Said is an elected member of the International Statistical Institute⑱,[656] an elected member of the Research Society on Alcoholism, ⑲ an elected member of Sigma Xi, the Scientific Research Society, ⑳ an elected member of the Washington Academy of Science.' ⓪[657]

[646] isi2011.congressplanner.eu/showabstract.php?congress=ISI2011&id=790

[647] This seemed *odd* for a postdoc who was not listed as a faculty member. In any case, it seems to have disappeared a while ago. www.webcitation.org/6Ct6xusOl

[648] She was the Managing Editor, but there were 3 E-i-C's.

[649] IFNA is Wegman's organization.

[650] IFNA has managed ACAS for the Army.

[651] www.amstat.org/sections/officers.cfm?txtComm=SSDNS
2012: Wegman was Chair
2011: David Marchette was Chair, Said was Program Chair
2010: Lara Schmidt (RAND) was Chair, Solka was Program Chair
2009: Wendy Martinez was Chair
2008: David Banks was Chair

[652] This is ACAS again, managed by IFNA.

[653] Said, Wegman, Patent 7,800,616, 09/21/10, filed Feb 2006. Via patft.uspto.gov/netahtml/PTO/search-adv.htm "IN/said and IN/wegman"

[654] patft.uspto.gov/netahtml/PTO/search-adv.htm
"IN/Said and TTL/metadata" and "IN/said and TTL/tool" do not find these.
"IN/Said and TTL/metadata" - 4 hits, none involving her or Wegman
In any case, via patft.uspto.gov/netahtml/PTO/search-bool.html, Patent 8145677 was by Faleh Al-Shameri, Filed 03/27/08, granted 03/27/12, "Automated generation of metadata for mining image and text data." That seems to be a later version of **§C**-[c], www.freepatentsonline.com/y2010/0223276.html, noted in deepclimate.org/2010/12/02/wegman-et-al-miscellany, originally by Wegman and Said.

[655] appft.uspto.gov/netahtml/PTO/search-adv.html Same searches used:
"IN/said and TTL/tool" - also 4 hits, including 7,800,616, 09, not this.
The second patent application is described in **§R.4,** but *apparently* was not filed.

[656] www.isi-web.org
www.isi-web.org/images/about/Declaration-EN2010.pdf Ethics declaration
www.isi-web.org/membership/join-the-isi-or-an-isi-section-/29-membership/membership/58-how-to-apply-for-isi-elected-membership
To be an elected member, one essentially needs 3 members as sponsors, who present the candidate to the elections committee. David Banks, David Marchette, Wendy Martinez, Carey Priebe, David Scott, and Wegman all are members.
isi.cbs.nl/ISImembers/b.asp?Country=United+States she is not on list.

[657] www.washacadsci.org
www.washacadsci.org/join-us
'The qualification for regular/student membership is simple. The applicant must show an interest in science/technology. This may be indicated by university degree (student members major in a technical field) or membership in professional scientific societies.'

S.8 Concerns?

The following enumerates issues or contradictions that might raise concerns about resume inflation or perhaps even misrepresentation to the NIAAA.[658] The following tags are used throughout **§S.**

0 'to develop skills in social networks, **geographic information systems**, neuroscience, biology' **§S.1**
'Because of a limitation of funds available to take formal coursework '[659]
'I am planning coursework on behavioral biology of substance abuse, psychopathology, **geographic information science** and epidemiology.' **§S.3.1**
'Because of unprecedented travel opportunities, I have taken coursework informally.'[660] **§S.3.2**
'Dr. Said has audited courses on Computational Linguistics and Text Mining, on Statistical Methods for High Dimensional Data, and on Social Network Analysis'.
Most of that did not seem to happen, ***§J.4.***
§S.3.1 *notes time spent at Buffalo State U (Wieczorek), relevant to the geographic information systems, but other topics seem not to have been covered, a concern for reviewers of the original proposal.*

1 'it is my strong desire to continue working in this health related area for my entire career.' **§S.1**
Wegman certainly was interested in alcoholism research, but it was unclear that Said really was, given all the other distractions and the discontinuance of alcoholism work after 5876 ended. A year's work on PhD and a 3-year postdoc produced zero peer-reviewed journal papers and only 2 poster sessions at a local conference, P186, and P187.

2 Research Training Plan, 50%+25%+10% course work **§S.1**
These simply did not happen, **§S.3.1** and **§S.3.2.**

3 Wegman recommendation was exceptionally strong, **§S.1**
Was this recommendation justified? Did other faculty agree?

4 'completed her dissertation in and graduated with Ph.D **§S.1**
Her PhD [SAI2005 or P401] had plagiarism , summarized [MAS2012c §3.4].

5 'chapter on genetic algorithms' **§S.1**
§K.4-P402, edited by Wegman. Plagiarism was summarized [MAS2012c §4.4].. She had a course in genetic algorithms in 2003, **§S.1,** *but no obvious other background in this.*

6 'student member of the Interface and American Statistical Association; the latter is the nation's leading professional association for statistics and statisticians.' **§S.1**
Interface has long been Wegman's conference. Grad students (and sometimes undergrads) normally join professional societies.

7 'student member of the Research Society on Alcoholism' **§S.1**
Anyone planning a career in alcoholism research would join RSoA.

8 'SAMSI/NISS 2005-The NSF-sponsored Statistics and Applied Mathematics Institute held in Research Triangle Park, NC (September). I gave one of four young investigator talks on my NIAAA related research' **§S.3.1** T400 S☐
'ACAS 2006 -Army Conference on Applied Statistics, held in Monterey, California. I gave a talk on NIAAA related research' **§S.3.1** T401 S☐
This talk was given twice ~9 months before 5876 started.
The first was
"Agent-based Model Applicable to Homeland Security and Disease Control"
and the second, at Monterey, but ACAS **2005**[661] [MAS2010a §A.6.3]
"Adaptation of an Alcohol Ecological Agent-Based to Homeland Security".
Read the T401abstract and see if it makes any sense and if so, what relationship it bears to homeland security.

9 'The ACAS and QMDNS[662] although having a military focus in general also have sessions related to drinking behaviors of young enlisted men' **§S.3.1**
She did not actually mention QMDNS before. The reader might search the ACAS and QMDNS websites and see how often drinking behavior talks have occurred. On the other hand, IFNA runs Interface, ACAS and QMDNS.

[658] Many may be just fine, alone. Submitted papers get rejected, plans for books do not work out, career goals change. *The overall pattern may be a concern.*
[659] But there were funds for much travel in 2007, **§0.3, §I.**
[660] Again, formal course work got completely lost.

[661] www.webcitation.org/6Dsb14iwo ACAS05 agenda
www.webcitation.org/6DsaBSMsf ACAS05
www.webcitation.org/6DsZab7MS ACAS06
www.webcitation.org/6DsaJWbME ACAS07
[662] www.qmdns.org

⑩ 'fellow of the Royal Statistical Society'
www.rss.org.uk/site/cms/contentCategoryView.asp?category=46
'The main category of membership - open to anyone with an interest in statistics - is Fellowship. Fellows receive the full package of member benefits listed above.'

⑪ 'She is the editor of Computing Science and Statistics' **§S.3.2**
That sounds as though it were a journal, but is really the proceedings of the Interface conference. As per [MAS2010a §A.6.3], during 1997-2008, she co-edited it twice, once with Amy Braverman in 2003 and once with David Marchette in 2004. She may have done so later.

⑫ 'associate editor of the journal, Computational Statistics and Data Analysis' **§S.3.2** *She had been an Associate Editor at CSDA (a rare role for a postdoc), and might still have been one in 2008, but was not by 2010.* **§H.5.**

⑬ 'serves on the board of the Washington Statistical Society' **§S.3.2**
washstat.org/wssinfo.html She was a non-voting member, on the Social Arrangements Committee in 2010. She was not mentioned in 2012-2013.

⑭ 'Interface Foundation of North America.' **§S.3.2**
IFNA

⑮ 'American Statistical Association Presidential Task Force on Science Policy' **§S.3.2** *An ASA "Science Policy Task Force" existed in 2007, when she joined. www.amstat.org/misc/2007_03_02_Full_Report_of_Task_Force.pdf www.amstat.org/education/caucuspowerpoints/SKMsciencepolicy.ppt This was 2006-2007 task force, not continuing, and not "Presidential": www.amstat.org/committees/commdetails.cfm?txtComm=ABTBOD02*

⑯ 'serves on the American Statistical Association Presidential Advisory Committee on Scientific and Public Affairs' **§S.1**
www.amstat.org/committees/commdetails.cfm?txtComm=CCNPRO06 She was a member 2009-2011, but not 2012. There was no "Presidential."

⑰ 'She is finalizing a book, Intervention to Prevention: A Policy Tool for Alcohol Studies, to be published by VDM Verlag.' **§S.3.2**
This was her dissertation, published via an *interesting* group:
www.amazon.com/Intervention-Prevention-Policy-Alcohol-Studies/dp/3836478625 "VDM Verlag Dr. Müller (January 23, 2009)" en.wikipedia.org/wiki/VDM_Publishing accrispin.blogspot.com/2009/09/victoria-strauss-vdm-verlag-dr-mueller.html expatacademic.wordpress.com/2010/11/23/who-in-the-world-is-dr-muller

⑱ 'elected as a member of the International Statistical Institute (ISI)'
www.isi-web.org www.isi-web.org/images/about/Declaration-EN2010.pdf Ethics www.isi-web.org/membership/join-the-isi-or-an-isi-section-/29-membership/membership/58-how-to-apply-for-isi-elected-membership To be an elected member, one needs 3 members as sponsors, who present the candidate to the elections committee. David Banks, David Marchette, Wendy Martinez, Carey Priebe, David Scott, and Wegman all are members. isi.cbs.nl/ISImembers/b.asp?Country=United+States she is not on list..

⑲ '(Research) Society on Alcoholism' **§S.3.2**
www.rsoa.org; People become regular members by submitting applications to the membership committee.

⑳ 'elected member of Sigma Xi, the Scientific Research Society' **§S.3.2**
www.sigmaxi.org/member/join/index.shtml
https://www.sigmaxi.org/member/join/qualification.html
'This noteworthy achievement must be evidenced by publication as the primary author (defined in the manner appropriate to the discipline) on at least two different articles published in a refereed journal, patents, or refereed monographs.''

⓪ 'elected member of the Washington Academy of Science' **§S.3.2**
www.washacadsci.org; www.washacadsci.org/join-us 'The qualification for regular/student membership is simple. The applicant must show an interest in science/technology. This may be indicated by university degree (student members major in a technical field) or membership in professional scientific societies.'

① '**member of more than 15 scientific, research, and professional societies**' **§S.3.2.** *A single peer-reviewed journal paper might have been more credible.*

② deepclimate.org/2011/05/15/retraction-of-said-wegman-et-al-2008-part-1/#comment-8890 In addition, see affiliations in P.184 in **§K.6.**
www.bizapedia.com/dc/INNOVATIVE-MEDICAL-INSTITUTE-LLC.html
Whatever it was, there is little trace of it.

T. Commonwealth of Virginia/GMU (4) 07/01/07-12/15/07 $30K

GMU found no documentation regarding any request for the funds or acknowledgement requirements, if any, or any reports.[663]
This included a *somewhat confusing* comment:

> **p.2** 'According to University records, Dr. Yasmin Said did not receive a $30,000 supplementary grant during the time specified by Dan Vergano in his FOIA request, but, as Dr. Said's progress report to the NIAAA Kirschstein Fellowship indicates, she did receive a $30,000 supplementary fellowship as allowed by PA-03-067 (see attached under "Supplementation of Stipends, Compensation, and Other Income"). Also attached are the associated payment request forms.'

It did clarify ambiguity and confirm Wegman's note, **§S.3.2:**

> p.7 'For the second year[664], there was a $30,000 supplement from George Mason University funds supplied by the Commonwealth of Virginia.'

In reply to Vergano's FOIA, a 13-page file[665] listed 12 $2,500 payments to Yasmin Said, twice/month, 07/01/07 through 12/15/07.[666]
These were signed by D. Papaconstantopoulos, Chair of Comp & Data Sciences in the GMU College of Science.

A page from NIH policy PA-03-067+Kirschstein was also given to Vergano, [667] showing the GMU supplement was related to the NIAAA.

What did the Commonwealth of VA and GMU expect for this grant? Does GMU just give supplement money automatically to fellowship recipients? That is certainly possible, but one might expect some monitoring or reports or something as a result.

---

[663] GMU reply to FOIA request by Vergano, February 2013:
www.documentcloud.org/documents/604587-emails-from-staff.html
[664] This meant $30K in the second year, not $30K in each year of 2 years.
[665] GMU reply to FOIA request by Vergano, February 2013:
www.documentcloud.org/documents/604589-yasmin-said-gmu-fellowship-pmts-2007.html
[666] Fund/Org: 182101 Account: 78515.
[667] GMU reply to FOIA request by Vergano, February 2013:
www.documentcloud.org/documents/604588-pa-03-067-kirschstein.html
PA-03-067_ Kirschstein.pdf - a page from pp.4-5 was included.

X. Plagiarism in another 5876-claimed paper

Said claimed P405 S③☆ for 5876 in **§S.3.2**[668]:

'20. Said, Yasmin H. (2007) On the Eras in the History of Statistics and Data Analysis, *Journal of Washington Academy of Sciences*, 93(1), 17-35'[669]

This was 5876-unfit, but also *seemed* an odd effort for a busy postdoc, less than 2 years after PhD. The paper started, p.17:

'Yasmin H. Said
Center for Computational Statistics
George Mason University
Abstract
In this paper, **we**[670] present a view of the evolution of statistical thinking through eras we designate as Pre-modern, Classical, Recent Past, and Future. **We** argue that modes of thinking about data and statistical inference are noticeably different from one era to the next. **We** discuss some of the leading figures in each of these eras.'

The paper, p.18 describes this as:

'Pre-modern Period prior to 1900
(pp.18-19, many)
Classical Period 1900 to 1985
(pp.19-24,K. Pearson, Gossett, Fisher, E.Pearson, Neyman, Kolmogorov, Mahalanobi, Hotelling, Cramér, Rao, Wilks)
Recent Past Period 1962 to 2005
(pp.25-26, Tukey and colleagues)
Future Period after 1981.'
(pp. 26-30, Wegman, Efron, Friedman, David Scott)
Statistical Thinking in Government, Science, and Law, Conclusions, etc
(pp. 3135-, Wegman, Efron, Friedman, David Scott)

A few hours were spent documenting obvious examples of mosaic plagiarism, with no pretense of thorough search. Similar phrases appear in many Internet sources, so one cannot be sure of the adaptation flow, just the existence of at least one plausible prior antecedent. However, it is *almost certain* that such existed in various websites before this paper.

The side-by-side comparison style is the same as used in earlier reports:

- Cyan for identical, in-order text between P405 and antecedents, with some reformatting for alignment.
- Yellow for trivial edits P405 ← antecedents.
- ~~Cross out~~ is used for obviously-deleted antecedent text.
- No highlight implies plausible paraphrase, original work, or unfound.

Said wrote, p.13:

'Acknowledgement
The author gratefully acknowledges the long discussions with Professor Edward J. Wegman, whose contact and experience with both the early contributors and the evolution of statistics as a discipline over the last 40 years provided valuable insight that made this discussion possible.'

Substantial text of "striking similarity" was quickly identified, in the same copy-paste-edit style of which 90+ pages have been documented so far.

P405. ← Plausible antecedent

Page	antecedent source
18	Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900*, Belknap Press of Harvard University Press (March 1, 1990)[671]
20	www-history.mcs.st-andrews.ac.uk/Biographies/Pearson.html
20-21	www-gap.dcs.st-and.ac.uk/~history/Biographies/Fisher.html
22	www-history.mcs.st-andrews.ac.uk/Biographies/Pearson_Egon.html
22	statprob.com/encyclopedia/PrasantaChandraMAHALANOBIS.html
22	www.amstat.org/about/statisticiansinhistory/index.cfm?fuseaction=biosinfo&BioID=7
23	www-history.mcs.st-andrews.ac.uk/Biographies/Cramer_Harald.html
24	www-history.mcs.st-andrews.ac.uk/Biographies/Wilks.html
25	**P169** Wegman, Edward J. and Solka, Jeffrey L.(2005) "Statistical data mining," *Handbook of Statistics: Data Mining and Data Visualization* (Rao, C. R., Wegman, E. J. and Solka, J. L., eds.), 1-46'[672]

This paper has 17 pages of text, plus 2 of endnotes and references. A few hours' work easily found ~6 pages of text copy-paste-edited from unacknowledged sources, casting doubt on Said's qualification to write a statistics history article for an encyclopedia, i.e., P403, **§Y**.

[668] www.desmogblog.com/sites/beta.desmogblog.com/files/aa15876-3-3-Progress.Redact.pdf , PDF p.5.

[669] www.washacadsci.org/scans/V.93-n.1.pdf
The article PDF itself was not online, but the Washington Academy of Sciences kindly and promptly sent me a copy.

[670] The "we" usage might be stylistic or it might imply this was originally planned as a paper by Said and Wegman.

[671] The hardback was (1986), paperback (1990). This seems more paraphrased than the other copy-paste sections.. Stigler's book was well-known, heavily-cited, and 2 decades old, so intermediate antecedents were easily possible.

[672] That book also contained P402 S☆.

P405 p.18

In the Pre-modern Period, one of the most interesting early examples of the recognition of variability is the so-called Trial of the Pyx. The Trial of the Pyx is a procedure for maintaining the integrity of newly minted coins in the United Kingdom (England). From shortly after the Norman Conquest (1066) in a procedure that has been essentially unchanged since 1282, the London (later Royal) Mint

selects a sample of each day's coins that are reserved in a box called the Pyx. The earliest agreements between the mint and the monarchy stated that a certain tolerance would be allowed in the weight of a single coin and by linear extrapolation in the aggregate weight of the contents of the Pyx. Thus, earlier than 1100, there was a formalized methodology for allowance of uncertainty and a method by which the integrity of the entire coinage could be judged based on a sample in the presence of uncertainty in the production process.[i]

--- (note on p.33)

[i] The use of linear extrapolation is a flawed procedure by modern standards. If a tolerance of 2 units per coin is allowed, then for 100 coins, the Trial of the Pyx would allow 200 units tolerance, whereas modern theory would dictate a tolerance of $2\sqrt{100} = 20$ units tolerance.

P405 p.20 (on Pearson)

lectures on statistics,[673] dynamics and mechanics, completed the unfinished first volume of Clifford's *The Common Sense of the Exact Sciences* (published in 1885), completed and edited the half-written first volume of Todhunter's History of the *Theory of Elasticity*, began working on the second volume published many papers on applied mathematics, lectured on *The Ethic of Free Thought*, and undertook research on a number of historical topics, including the evolution of Western Christianity. …

Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900*, Belknap Press of Harvard University Press (March 1, 1990)[674]

p.3

One dramatic early instance of a numerical assessment of accuracy that was not given in terms of explicit probabilities was the Trial of the Pyx.

From shortly after the Norman Conquest up to the present, the London (later Royal) Mint maintained the integrity of its coinage[675] through a routinized inspection scheme in which a selection of each day's coins was reserved in a box ("the Pyx") for a later trial. Even in the earliest indentures between the mint and the king ~~the contract~~ stated that the trial would allow a tolerance in the weight of a single coin and, by linear extrapolation, in the aggregate weight of the entire contents of the Pyx. Thus as early as 1100 an economic necessity had led to an institutionalized numerical allowance for uncertainty, uncertainty in how the value of the entire coinage could be judged by that sample, in the presence of unavoidable variability in the production process.[2]

---

2. The Trial of the Pyx was not without its flaws. The use of linear extrapolation was a major one. If a coin was allowed a tolerance of 5 grains, an aggregate of 100 coins would be allowed a tolerance of 500 grains, rather than the $\sqrt{100}$ X5 = 50 grains modern theory might suggest. The story of the Pyx, including Isaac Newton's connection to it, is told in Stigler(1977b).

www-history.mcs.st-andrews.ac.uk/Biographies/Pearson.html

lectures on statics, dynamics and mechanics, ~~he~~ completed the unfinished first volume of Clifford's The Common Sense of the Exact Sciences (published in 1885), completed and edited the half written first volume of Todhunter's History of the Theory of Elasticity, began working on the second volume ~~which had hardly been started by Todhunter, and~~ published many papers on applied mathematics. ~~He also~~ lectured on The Ethic of Free Thought, and undertook research on a number of historical topics such as the evolution of Western Christianity. …

[673] Google: karl pearson lecture statics dynamics
Statics is a specific topic within physics, as per "Statics of Rigid Bodies" as in Chapter 14 in my sophomore college physics book Halliday and Resnick (1963). **Statics** is very definitely not **statistics**, but this sort of error has been found often in works involving Said, including P401 and T126: amidst a block of obviously-copied text, trivial edits were made, but some introduced errors. Some were really silly, indicating lack of understanding that went far beyond poor proofreading.

[674] www.amazon.com/The-History-Statistics-Measurement-Uncertainty/dp/067440341X

[675] This likely got rearranged into "maintaining the integrity of newly minted coins," but the (manual) comparison algorithm does not try to track movements of text, i.e., it approximates the old UNIX *diff(1)* command.

P405 p.20 (cont)

'Sir Ronald Fisher (1890-1962) is widely recognized as the third and probably most important of the first modern statisticians. He studied mathematics and astronomy at Cambridge, but was also interested in biology. He graduated with distinction in the Mathematical tripos of 1912. He continued his studies at Cambridge on the theory of errors. Fisher's interest in the theory of errors eventually led him to investigate statistical problems. After leaving Cambridge, Fisher worked for several months on a farm in Canada, but soon returned to London and took up a position as a statistician in the Mercantile and General Investment Company. When war

P405 p.21

broke out in 1914 he tried to enlist in the army, having already trained in the Officers' Training Corps while at Cambridge. He was rejected for military service because of his eyesight. He became a teacher of mathematics and physics, teaching at Rugby and other similar schools between 1915 and 1919. Fisher gave up being a mathematics teacher in 1919 when he was offered two posts simultaneously. Karl Pearson offered him the post of chief statistician at the Galton laboratories, but he was also offered the post of statistician at the Rothamsted Agricultural Experiment Station, which was the oldest agricultural research institute in the United Kingdom. It was established in 1837 to study the effects of nutrition and soil types on plant fertility, and this appealed to Fisher's interest in farming. He accepted the post at Rothamsted. Here he made many contributions to statistics, in particular the design and analysis of experiments, and also to genetics. He studied the design of experiments by introducing the concept of randomization and the analysis of variance, procedures now used throughout the world.
In 1921 he introduced the concept of likelihood. The likelihood of a parameter is proportional to the probability of the data, and it gives a function that usually has a single maximum value, which he called the maximum likelihood.
Fisher published a number of important texts; in particular, *Statistical Methods for Research Workers* (1925) ran to many editions that he extended throughout his life.

Pearson and Fisher had a long, bitter, and very public dispute. At first they exchanged friendly letters after Pearson received a manuscript from Fisher in September 1914 of a paper he was submitting for publication to *Biometrika*. Pearson's initial response was to offer his hearty congratulations and, if correct, offered to publish the paper
Later, having read the paper fully he indicated that it marked a distinct advance.

By May 1916 they were still corresponding in a friendly manner. However, Pearson misunderstood the assumptions of Fisher' maximum likelihood method, and criticized it in his May 1917 *Cooperative Study,* a paper that he co-authored with his staff concerning tabulating the frequency curves. Fisher, believing that Pearson's criticism was unwarranted, responded with a paper that criticized examples in the *Cooperative Study* to the extent of ridiculing them. Fisher had looked again at his earlier correspondence with Pearson, noticed that many of his papers had been rejected, and concluded that Pearson had been responsible. Thus began one of the most famous feuds in the history of statistics.'

www-gap.dcs.st-and.ac.uk/~history/Biographies/Fisher.html

'Although he studied mathematics and astronomy at Cambridge, he was also interested in biology. … He graduated with distinction in the mathematical tripos of 1912. Awarded a Wollaston studentship, he continued his studies at Cambridge under Stratton on the theory of errors reading Airy's manual the *Theory of Errors.* It was Fisher's interest in the theory of errors that eventually led him to investigate statistical problems. After leaving Cambridge, Fisher had no means of financial support and worked for a few months on a farm in Canada. He returned to London, taking up a post as a statistician in the Mercantile and General Investment Company. When war

broke out in 1914 he enthusiastically tried to enlist in the army, having already trained in the Officers' Training Corps while at Cambridge. His medical test showed him A1 on all aspects except his eyesight, which was rated C5, so he was rejected. He became a teacher of mathematics and physics, teaching at Rugby and other similar schools between 1915 and 1919. Fisher gave up being a mathematics teacher in 1919 when he was offered two posts simultaneously. Karl Pearson offered him the post of chief statistician at the Galton laboratories and he was also offered the post of statistician at the Rothamsted Agricultural Experiment Station. This was the oldest agricultural research institute in the United Kingdom, established in 1837 to study the effect of nutrition and soil types on plant fertility, and it appealed to Fisher's interest in farming. He accepted the post at Rothamsted where he made many contributions ~~both~~ to statistics, in particular the design and analysis of experiments, and to genetics. ~~There~~ he studied the design of experiments by introducing the concept of randomisation and the analysis of variance, procedures now used throughout the world. …
In 1921 he introduced the concept of likelihood. The likelihood of a parameter is proportional to the probability of the data and it gives a function which usually has a single maximum value, which he called the maximum likelihood …
Fisher published a number of important texts; in particular *Statistical Methods for Research Workers* (1925) ran to many editions which he extended throughout his life.

www.educ.fc.ul.pt/icm/icm2003/icm14/Pearson.htm

Pearson had a long, bitter, and very public dispute with Fisher. At first they exchanged friendly letters after Pearson received a manuscript from Fisher in September 1914 of a paper he was submitting for publication. Pearson's initial response was to say (see [18]): *I congratulate you very heartily on getting out the actual distribution form ... if the analysis is correct which seems highly probable, I should be delighted to publish the paper in Biometrika.* Again a week later [18]:- *I have now read your paper fully and think it marks a distinct advance.* … By May 1916 they were still corresponding in a friendly manner. However Pearson misunderstood the assumptions of Fisher's maximum likelihood method, and criticised it unfairly in the May 1917 *Cooperative Study* paper which he co-authored with his staff concerning tabulating the frequency curves. Fisher, believing that Pearson's criticism was unwarranted, responded with a paper which criticised examples in the *Cooperative Study* to the extent of ridiculing them. Fisher had looked again at his earlier correspondence with Pearson, noticed that many of his papers had been rejected, and concluded that Pearson had been responsible.'

P405 p.22

'Egon Pearson (1895-1980) was the son of Karl Pearson
In 1921 he joined his father's Department of Applied Statistics at University College London as a lecturer. However, his father kept him away from lecturing. Egon attended his father's lectures and began to produce a stream of high quality research publications on statistics. In 1924, Egon became an assistant editor of *Biometrika*, but perhaps one of the most important events for his future research happened in the following year.

Jerzy Neyman (1894-1981) was stimulated by a letter from Egon Pearson, who sought a general principle from which Gosset's tests could be derived. Neyman went on to produce fundamental results on hypothesis testing and, when Egon Pearson visited Paris in the spring of 1927, they collaborated in writing their first paper
Between 1928 and 1933, they wrote a number of fundamental papers on hypothesis testing, the best-known result being the Neyman-Pearson Lemma. Neyman moved to the University of California, Berkeley in 1938 and remained there until his death in 1981. He was reputed to have been working on a research paper in the hospital where he died.'

Andrei Nikolaevich Kolmogorov (1903-1987) laid the axiomatic foundations for probability theory in 1933 and also in 1938 laid out the foundations for Markov random processes.

Prasanta Chandra Mahalanobis (1893-1972) undertook work on experimental designs in agriculture. In 1924, he made some important discoveries about the probable error of results of agricultural experiments, which put him in touch with Fisher. Later in 1926, he met Fisher at the Rothamsted Experimental Station and a close personal relationship was immediately established that lasted until Fisher's death.
In 1927, Mahalanobis spent a few months in Karl Pearson's laboratory in London. During this period he performed extensive statistical analyses of anthropometric data and closely examined Pearson's Coefficient of Racial Likeness (CRL) for measurement of biological affinities. He noted several shortcomings of the CRL and in 1930 published his seminal paper on the *D-square* statistic, which is now recognized as the Mahalanobis Distance.'

www-history.mcs.st-andrews.ac.uk/Biographies/Pearson_Egon.html

'In 1921 Pearson joined his father's Department of Applied Statistics at University College London as a lecturer. However, ~~despite being a lecturer~~, his father ~~seems to have~~ kept him away from lecturing. ~~Instead~~ Pearson attended ~~all of~~ his father's lectures and began to produce a stream of high quality research publications on statistics. In 1924 Pearson became an assistant editor of *Biometrika* but perhaps one of the most important events for his future research happened in the following year.
www-history.mcs.st-and.ac.uk/Biographies/Neyman.html
However his interest in statistics was stimulated ~~again~~ by a letter from Egon Pearson, who sought a general principle from which Gosset's tests could be derived. Neyman went on to produce fundamental results on hypothesis testing and, when Egon Pearson visited Paris in the spring of 1927, they collaborated in writing their first paper. …
Between 1928 and 1933 Neyman and Egon Pearson had written a number of important papers on hypothesis testing …'

statprob.com/encyclopedia/PrasantaChandraMAHALANOBIS.html

Prasanta Chandra MAHALANOBIS b. 29 June 1893 - d. 28 June 1972
Some of the early statistical studies he undertook were on experimental designs in agriculture. In 1924, he made some important discoveries pertaining to the probable error of results of agricultural experiments, which put him in touch with R.A. Fisher (q.v.). Later in 1926, he met Fisher at the Rothamsted Experimental Station and a close personal relationship was immediately established which lasted until Fisher's death.
In 1927, Mahalanobis spent a few months in Karl Pearson's (q.v.) laboratory in London, during which period he performed extensive statistical analyses of anthropometric data and closely examined Pearson's Coefficient of Racial Likeness (CRL) for measurement of biological affinities. He noted several shortcomings of the CRL and in 1930 published his seminal paper on the D^2-statistic entitled ``Tests and measures of group divergence".'

P405 p.22

'Harold Hotelling (1895-1973) earned a Ph.D. in mathematics from Princeton University, and began teaching at Stanford University that same year, 1924. Hotelling realized that the field of statistics would be more useful if it employed methods of higher mathematics, so in 1929, he went

P405 p.23

to England to study with R. A. Fisher.
When Hotelling returned to the United States, he began developing some of his techniques at Stanford University. His early applications involved the diverse fields of journalism, political science, population, and food supply.
Hotelling was a pioneer in the field of mathematical statistics and economics in the 20th century, with contributions to the theory of demand and utility, welfare economics, competition, game theory, depreciation, resource exhaustion, and taxation. His work in mathematical statistics included his famous 1931 paper on the Student's *t*-distribution for hypothesis testing, in which he laid out what has since been called *confidence intervals*.

P405 p.23

Carl Harald Cramér (1893-1985) entered the University of Stockholm in 1912 and worked as a research assistant on a biochemistry project before becoming firmly settled on research in mathematics.
He earned a Ph.D. in 1917 for his thesis, *On a class of Dirichlet series*.
In 1919 Cramér was appointed assistant professor at the University of Stockholm. He began to produce a series of papers on analytic number theory. It was through his work on number theory that Cramér was led towards probability theory. He also had a second job, namely as an actuary with the Svenska Life Assurance Company. This led him to study probability and statistics that then became the main area of his research. Cramér became interested in the rigorous mathematical formulation of probability in work of the French and Russian mathematicians, in particular the axiomatic approach of Kolmogorov. By the mid 1930s Cramér's attention had turned to the approach of the English statisticians such as Fisher and Egon Pearson as well as contemporary American statisticians.
During World War II, Cramér was cut off from the rest of the academic world. By the end of World War II he had written his masterpiece *Mathematical Methods of Statistics*. In addition to his seminal book, Cramér is known for his work on stationary stochastic processes and for the Cramér-Rao inequality.'

www.amstat.org/about/statisticiansinhistory/index.cfm?fuseaction=biosinfo&BioID=7

Harold Hotelling 1895-1973 … In 1924, he earned a PhD in mathematics from Princeton University, and began teaching at Stanford University that same year. Hotelling soon realized that the field of statistics would be more useful if it employed methods of higher mathematics, so in 1929, he went

to England to study with R. A. Fisher, a very prominent statistician.
When Hotelling returned to the United States, he began developing some of his techniques at Stanford University. His early applications involved the diverse fields of journalism, political science, population, and food supply. …
Hotelling was considered a pioneer in the field of mathematical statistics and economics in the 20th century, with contributions to the theory of demand and utility, welfare economics,
and taxation. His work in mathematical statistics included his famous 1931 paper on the Student's t distribution for hypothesis testing, in which he laid out what has since been called "confidence intervals." His economics papers throughout the 1920s and 1930s discussed
competition, game-theory, depreciation, and resource exhaustion. He also covered topics in mathematical statistics such as hypothesis testing and confidence intervals.'

www-history.mcs.st-andrews.ac.uk/Biographies/Cramer_Harald.html

Harald Cramér entered the University of Stockholm in 1912. …
worked as a research assistant on a biochemistry project before becoming firmly settled on research in mathematics. …
resulted in the award of a PhD in 1917 for his thesis *On a class of Dirichlet series*.
In 1919 Cramér was appointed assistant professor at the University of Stockholm. He began to produce a series of papers on analytic number theory ,… It was not only through his work on number theory that Cramér was led towards probability theory . He also had a second job, namely as an actuary with the Svenska Life Assurance Company. This led him to study probability and statistics which then became the main area of his research. Cramér became interested in the rigorous mathematical formulation of probability in work of the French and Russian mathematicians ~~such as Paul Lévy , Sergei Bernstein , and Aleksandr Khinchin in the early 1930s, but~~ in particular the axiomatic approach of Kolmogorov . By the mid 1930s Cramér's attention had turned to look at the approach of the English and American statisticians such as Fisher , ~~Neyman~~ and Egon Pearson ~~(Karl Pearson 's son)~~ …
During World War II Cramér was ~~to some extent~~ cut off from the rest of the academic world. … By the end of World War II Cramér had written his masterpiece *Mathematical Methods of Statistics*'.

P405 p.24

'Samuel Wilks (1906-1964)
began to study mathematics at the University of Texas in 1926 where he was taught set theory and other courses in advanced mathematics.
Wilks received an M.A. in mathematics in 1928.

Wilks was awarded a fellowship to the University of Iowa where he studied for his doctorate under H. L. Rietz. Rietz introduced him to Gosset's theory of small samples and R. A. Fisher's statistical methods. After receiving his doctorate in 1931 on small sample theory of 'matched' groups in educational psychology, he continued research at Columbia University in the 1931-1932 session.
In 1932, Wilks spent a period in Karl Pearson's department in University College, London. In 1933 he went to Cambridge where he worked with John Wishart, who had been a research assistant to both Pearson and Fisher. He was appointed instructor of mathematics at Princeton in 1933. He was to remain there for the rest of his career, being promoted to professor of mathematical statistics in 1944.
Wilks's work was all on mathematical statistics. His early papers on multivariate analysis were his most important, one of the most influential being, *Certain generalizations in the analysis of variance*. He constructed multivariate generalizations of the correlation ratio and the coefficient of multiple correlation and studied random samples from a normal multivariate population.
He advanced the work of Neyman on the theory of confidence-interval estimation. In 1941, Wilks developed his theory of 'tolerance limits.' Wilks was a founder member of the Institute of Mathematical Statistics (1935). There are obviously many other important contributors to the development of statistical theory in this Classical Period, but the ones mentioned here will suffice to give a flavor of the group. Much theory and methodology in the sense of the Classical Period still continues to be developed.

P405 p.25

In the landmark 1962 paper of Tukey entitled, "The future of data analysis," and later in the 1977 book, *Exploratory Data Analysis*,iii Tukey sets forth a new paradigm for statistical analysis. In contrast to confirmatory analysis in which a statistical model is assumed and inference is made on the parameters of that model, exploratory data analysis (EDA) is predicated on the fact that one does not

necessarily know that model assumptions actually hold for data under investigation. Because the data may not conform to the assumptions of the confirmatory analysis, inferences made with invalid model assumptions are subject to (potentially gross) errors. The idea then is to explore the data to verify that the model assumptions actually hold for the data in hand.'

P405 p.33

'The author gratefully acknowledges the long discussions with Professor Edward J. Wegman, whose contact and experience with both the early contributors and the evolution of statistics as a discipline over the last 40 years provided valuable insight that made this discussion possible.'

www-history.mcs.st-andrews.ac.uk/Biographies/Wilks.html

'During session 1926-27 Wilks taught at a school in Austin, Texas and at the same time he began to study mathematics at the University of Texas. Here he was taught set theory and other courses in advanced mathematics by Robert Moore and he took courses in probability and statistics with E L Dodd. Wilks received an M.A. in mathematics in 1928 and during this time, in fact from 1927 until 1929, he was an instructor in mathematics.
Wilks was awarded a fellowship to the University of Iowa where he studied for his doctorate. Here H L Rietz, who supervised his doctorate, introduced him to Gosset's theory of small samples and R A Fisher's statistical methods. After receiving his doctorate in 1931, on small sample theory of 'matched' groups in educational psychology, he continued research at Columbia University in session 1931-32.
In 1932 Wilks ~~went to England where he~~ spent a period in Karl Pearson's department in University College, London. In 1933 he went to Cambridge where he worked with John Wishart, who had been a research assistant to both Pearson and Fisher. He was appointed instructor of mathematics at Princeton in 1933. He was to remain there for the rest of his career, being promoted to professor of mathematical statistics in 1944.
Wilks's work was all on mathematical statistics. His early papers on multivariate analysis were his most important, one of most influential being *Certain generalizations in the analysis of variance*. He constructed multivariate generalisations of the correlation ratio and the coefficient of multiple correlation and studied random samples from a normal multivariate population. …
He advanced the work of Neyman on the theory of confidence-interval estimation. In 1941 Wilks developed his theory of 'tolerance limits'. Wilks was a founder member of the Institute of Mathematical Statistics (1935). ….

P169[676] Wegman and Solka (2005) p.2

The landmark paper of Tukey (1962) entitled, "The future of data analysis," and later in the book, *Exploratory Data Analysis* John Tukey(1977) sets forth a new paradigm for statistical analysis. In contrast to confirmatory analysis in which a statistical model is assumed and inference is made on the parameters of that model, exploratory data analysis (EDA) is predicated on the fact that we do not

p.3

necessarily know that ~~the~~ model assumptions actually hold for data under investigation. Because the data may not conform to the assumptions of the confirmatory analysis, inferences made with invalid model assumptions are subject to (potentially gross) errors. The idea then is to explore the data to verify that the model assumptions actually hold for the data in hand.'

[676] www.amazon.com/s/ref=nb_sb_noss?url=search-alias%3Daps&field-keywords=handbook%20of%20statistics%20wegman%20solka

Y. "Statistics" in Encyclopedia, possible plagiarism

P403 was a 2-page article listed by Said in support of her 5876 proposal:[677]

'2006 "Statistics" to appear *Encyclopedia of the Modern World*, (Stearns, Peter N., ed.), New York: Oxford University Press.

Externally-visible chronology was:

2005.08	Said gave 2006 as publication for *Encyclopedia, P403* .
2007.spring	19-page P405was published, **§X.**
2008.03	*Encyclopedia*[678] was actually published (Mar 28, 2008)

Given the inherent long creation time for the 8-volume *Encyclopedia*, real ordering is hard to know, but 3 alternatives seem possible:

- *P403 might have been written first, then expanded into P405.*
- *P405 might have already existed in 2005, and then been edited to P403*
- *Draft P405 was written in 2006/2007, either by Wegman and Said or by Said drawing heavily on Wegman knowledge. Said then extracted a shorter version for the Encyclopedia, where it was further edited.*

The last alternative *seems likeliest*, from the evidence:

- P403's structure defines 4 historical periods and covers each. P405 retains some of this structure, but *oddly* devotes paragraphs to material long before the 1750-present period of the *Encyclopedia*.
- P403 is more crisply copy edited than P405.
- P403 keeps odd vestiges of P405, but also cites a 2006 book not mentioned by it, making P403 unlikely to have been done in 2005.
- Stearns likely asked Wegman, and obvious choice. *He may have suggested Said, who could then could add P403 as further support for 5876. I'd defer serious assessment to statisticians, but P403 seemed strange* - **the only 20th-century statistician discussed was Wegman.**

It may have been misleading for Said to list this in support for a grant proposal and it may have been a distraction, but plagiarism is less clear. P403 was clearly derived from P405, but most of the obvious P405 plagiarism text was deleted, except the "Pyx" discussion.

Copyedits make more sense in this direction, English improved.

P403 p.135

'**STATISTICS is** at once an academic discipline, a tool for analyzing and inferring conclusions from data, and a collection subjected to the application of statistical tools. Statisticians generally think of the word statistics as ~~referring~~ either ~~to~~ the discipline or ~~to~~ the body of statistical methods whereas the general public more often thinks of statistics in the third sense, as a collection of numerical data, as in 'sports statistics.' The word "statistics" has its origins in the Latin *statisticum collegium* meaning "council of state." Similarly, the Italian word *statista* means "statesman" or "politician." Generically, statistics refers to data about the state. The modern English term derives from the German word *Statistik,* popularized and perhaps coined by the political scientist Gottfried Achenwall (1719-1772) in his *Vorbereitung zur Staatswissenschaft* (1748). The word seems to have been introduced into the English language by Sir John Sinclair (1754-1835). Sinclair was the supervisor of the twenty-one-volume *Statistical Account of Scotland,* published in the 1790s, which was the first systematic attempt to compile social and economic data on every parish in the country. In the *Statistical Account of Scotland*, Sinclair describes where he had come across the word *statistics* and why he translated and used it as an English word.'

P405 p.17

'Abstract

In this paper, we present a view of the evolution of statistical thinking through eras we designate as Pre-modern, Classical, Recent Past, and Future. We argue that modes of thinking about data and statistical inference are noticeably different from one era to the next. We discuss some of the leading figures in each of these eras.'

THE WORD "STATISTICS" refers at once to an academic discipline, to a powerful tool for inference on data and to results of the collection and application of statistical tools to data. Statisticians generally think of the word statistics as either the discipline or the body of methods comprising the tool while the general public more often thinks of statistics in the third sense, that is, a collection of numerical data as in 'sports statistics.' The word *statistics* is derived from the Latin *statisticum collegium* meaning the *council of state.* Similarly, the Italian word *statista* means *statesman* or *politician.* Thus, generically statistics means data about the state. The more modern term seems to have been the German word *Statistik,* popularized and perhaps coined by the German political scientist Gottfried Achenwall (1719-1772) in his *Vorbereitung zur Staatswissenschaft* (1748). The word *statistics* seems to have been introduced as an English language word by Sir John Sinclair (1754-1835). Sinclair was the supervisor of the *Statistical Account of Scotland* (1791- 1799), which was published in 21 volumes and was the first systematic attempt to compile social and economic statistics for every parish in the country. In the *Statistical Account of Scotland*, Sinclair describes where he had come across the word *statistics* and why he translated and used it as an English word.'

[677] www.desmogblog.com/sites/beta.desmogblog.com/files/aa15876-1a1-1-Proposal.Redact_0.pdf p.6

[678] www.amazon.com/Oxford-Encyclopedia-Modern-World-Present/dp/B007MXUUSU/ref=sr_1_1?ie=UTF8&qid=1360035163.

P403 p.135 (cont)

'The development of statistics as an academic discipline parallel the ever-increasing amounts of data generated by states and institutions.

The first U.S. Census was taken under the authority of Secretary of State Thomas Jefferson in 1790, when U. S. Marshals on horseback counted 3.9 million people. By 1810, the U. S. census was expanded to obtain information on manufacturing, including the amount and value of products. By 1840,

the American Statistical Association had been founded.

In the United States,[679] Abraham Lincoln established the United States Department of Agriculture (USDA) in 1862 to collect and analyze information pertaining to the country's agrarian economy.

A major development took place in Europe in 1953 with the development of the European Statistical System

(Eurostat), which for'

P403 p.136 (cont)

'the first time, integrated statistics across all of Western Europe. The discipline's close association with the state continues to facilitate advances in survey research and sampling theory.

The set of methodologies that constitute statistics includes mathematical, computational, and graphic methods that may be applied to a wide variety of data types including traditional numerical data, categorical data, image data, and even text data.'

P405 p.31

'The first U.S. Census was taken under the authority of Secretary of State Thomas Jefferson in 1790. U. S. Marshals on horseback ~~took the Census and they~~ counted 3.9 million people. By 1810, the U. S. Census was expanded to obtain information on manufacturing including the amount and value of products. By 1839, the American Statistical Society was formed to be renamed shortly the American Statistical Association because of an unfortunate acronym. In England, William Farr (1807-1883), an early medical statistician, was the compiler of abstracts in the office of the Registrar General. Using data that he compiled along with methods earlier attributed to John Snow, he identified the source of the 1866 cholera epidemic as water from a particular well of the London Water Company. Meanwhile his contemporary, Ernst Engel (1821-1896) served from 1860 as Director of the Royal Prussian Statistical Bureau.

Back in the United States, Abraham Lincoln establishes the United States Department of Agriculture (USDA) in 1862. Lincoln refers to USDA as "the people's department." In 1863, the first crop report appears and the USDA Division of Statistics is established. U. S. Census Bureau employee Herman Hollerith invented tabulating card machines, which were first used in the 1890 census, which counted nearly 63 million people. In 1913, the U. S. Department of Labor is established along with the Bureau of Labor Statistics.

A major development took place in Europe in 1953 with the development of the European Statistical System'

P405 p.32

'(EUROSTAT), which, for'

'the first time, integrated statistics across all of Western Europe. In short, the roots of statistics as a *state science* continues to stimulate and motivate statisticians with continuing advances in survey research and sampling theory associated with survey research.

P405 p.18

'Generally for statisticians, the set of methodologies that comprise statistics include mathematical, computational, and graphical methods and may be applied to a wide variety of types of data including traditional numerical data, categorical data, image data, and even text data.'

[679] In this case, P403 seems extracted from P405 by removal of the European events, but leaving the redundant "In the United States."

P403 p.136 (cont)[680]

'**Premodern Period**.[681] One of the most interesting early examples of the application of statistical methods before 1900 is the so-called Trial of the Pyx, a procedure developed in England beginning in the twelfth century to test newly minted coins for adherence to a quality standard. Uncertainty in the production process meant that some measure of error, or variation, had to be allowed for.
The London (later Royal) Mint selected a sample of each day's coins to be reserved in a box called the Pyx for later trial. The trial allowed for

a tolerance in the weight of a single coin and, by linear extrapolation in the aggregate weight of the contents of the Pyx. In this way

the integrity of the entire coinage could be judged based on one sample.

The roots of modern statistical methodology can be traced to the mid-seventeenth century.

John Graunt's (1620-1674) *Natural and Political Observations upon Bills of Mortality* published in 1662, used spatial data and map layouts to make inferences about sex ratios and disease types based on death records.

Toward the end of the Pre-modern period, **Sir Francis Galton** (1822-1911), in his study of heredity, developed the concept of regression toward the mean, described as early as the 1870s, and in 1888 he established the concept of correlation. In 1889 he published *Natural Inheritance,* in which he formally described the notions of regression and correlation.

The Classical Period.
The Classical Period (1900-1985) is characterized by a shift from descriptive methods to an increasingly mathematical formulation of methodologies. Computation was a tedious procedure and data collection a relatively costly process. Thus in the classical period there was considerable emphasis on optimality so that data were used efficiently, and on mathematical simplicity so that computation could be done rapidly. Hallmarks of theory developed in this era include small data sets, manual computation and strong and often unverifiable assumptions.'

P405 p.18

'In the Pre-modern Period, one of the most interesting early examples of the recognition of variability is the so-called Trial of the Pyx. The Trial of the Pyx is a procedure for maintaining the integrity of newly minted coins in the United Kingdom (England). From shortly after the Norman Conquest (1066) in a procedure that has been essentially unchanged since 1282,
the London (later Royal) Mint selects a sample of each day's coins that are reserved in a box called the Pyx.
The earliest agreements between the mint and the monarchy stated that a certain tolerance would be allowed in the weight of a single coin and by linear extrapolation in the aggregate weight of the contents of the Pyx. Thus, earlier than 1100, there was a formalized methodology for allowance of uncertainty and a method by which the integrity of the entire coinage could be judged based on a sample in the presence of uncertainty in the production process.

The roots of modern statistical methodology can be traced to the mid-seventeenth century. The earliest inferences are to a large extent based on graphical methods that are later echoed in what is labeled above as the Future Period.
John Graunt's (1620-1674) *Natural and Political Observations upon Bills of Mortality* published in 1662 gathered and used spatial data and map layouts to make inferences about sex ratios and disease types based on the bills of mortality.'

P405 p.19

'Towards the end of the Pre-modern period, Sir Francis Galton (1822-1911), cousin to Charles Darwin, developed the concept of regression toward the mean, described as early as the 1870s, and in 1888 he established the concept of correlation. In 1889, he published *Natural Inheritance,* in which he formally described the notions of regression and correlation.

The Classical Period
The Classical Period (1900-1985) is characterized by a shift from descriptive methods to an increasingly mathematical formulation of methodologies. ~~It must be remembered~~ that computation was a tedious procedure and data collection a relatively costly process. For this reason, in the classical period there was considerable emphasis on optimality so that data were used efficiently, and on mathematical simplicity so that computation could be done rapidly. Hallmarks of theory developed in this era include small data sets, manual computation, and strong and often unverifiable assumptions.'

P405 pp.19-24 discussed Karl Pearson, William S. Gossett, Sir Ronald Fisher, Egon Pearson, Jerzy Neyman, Andrei Nikolaevich Kolmogorov, Prasanta Chandra Mahalanobis, Harold Hotelling, Carl Harald Cramér, Calyampudi Radhakrishnan Rao, Samuel Wilks, i.e., major statisticians, none of whom were mentioned in P403.
The only statisticians named were Graunt and Galton for the 19th century and Wegman for the 20th.

[680] The red-bracketed section was almost certainly edited from the text at right, for which a plausible antecedent was Stigler's *The History of Statistics,* **§X.** The ideas remained, but several rounds of editing reduced the amount of identical text.
[681] It seems *odd* to spend several paragraphs on pre-1750 events and then cover almost nothing in the immediately-following "Classical Period."

P403 p.136 (cont)

'**Modern Period**. The Modern Period, from 1962, was marked by a major transition in thinking. Prior to 1962 in the Classical Period the focus was on the development of what is now called confirmatory analysis. Hypothesis testing, estimation, regression analysis, and variants of these were the major methodologies.

In contrast to confirmatory analysis in which a statistical model is assumed and inference is made on the parameters of that model, exploratory data analysis (EDA) is predicated on the fact that one does not necessarily know that model assumptions actually hold for data under investigation. Because the data may not conform to the assumptions of the confirmatory analysis, inferences made with invalid model assumptions are subject to (potentially gross) errors. The idea is to explore the data to verify that the model assumptions actually hold for the data in hand. This concept sparked a major revolution in the thought processes of statisticians.

Computers dramatically increased the statistician's ability to work with large quantities of data at greater levels of complexity and to analyze an interpret data faster and more efficiently.
The mid-1970s saw the emergence of integrated circuits and their use in primitive microcomputers,

but it was not until the IBM personal computer was introduced in 1981, and the Apple Macintosh was introduced in 1984, that serious computer power was in the hands of individual users. The introduction of personal computers and workstations in the 1980s dramatically increased access to computational resources, resulting an explosion of new statistical methods.'

P405 p.24-25

'**The Recent Past Period** The Recent Past Period (1962-2005) was marked by a major transition in thinking. Prior to 1962 in the Classical Period the focus was on the development of what is now called confirmatory analysis. Hypothesis testing, estimation, regression analysis, and variants of them were the major methodologies. ~~As mentioned earlier, these methods usually required strong and often unverifiable assumptions. John Tukey (1915-2000) represents a bridge between the Classical Period and the Recent Past Period. In the landmark 1962 paper of Tukey entitled, "The future of data analysis," and later in the 1977 book, *Exploratory Data Analysis*,iii Tukey sets forth a new paradigm for statistical analysis~~. In contrast to confirmatory analysis in which a statistical model is assumed and inference is made on the parameters of that model, exploratory data analysis (EDA) is predicated on the fact that one does not necessarily know that model assumptions actually hold for data under investigation. Because the data may not conform to the assumptions of the confirmatory analysis, inferences made with invalid model assumptions are subject to (potentially gross) errors. The idea ~~then~~ is to explore the data to verify that the model assumptions actually hold for the data in hand. This concept sparked a major revolution in the thought processes of statisticians and stimulated an outpouring of new methods.'

P405 pp.25-26 covered John Tukey, but also mentioned his many colleagues at Bell Labs and elsewhere. Tukey got no mention in P403.

P405 p.26-27

'The Future Period

' The introduction of personal computers and workstations circa 1981 sparked the beginnings of the Future Period (1981 onwards). In some ways it seems strange to date the Future from 1981, but the access to computational resources became so dramatically different, that literally an 'explosion of new methods resulted. … The placement of computer power in the hands of the end user made an enormous change in productivity. It should be noted that in the EDA table above the 1980-1984 and 1985-1989 period saw an explosion in papers in these two periods directly attributable to the introduction of personal computing.
The mid-1970s saw the emergence of integrated circuits and their use in primitive microcomputers. Indeed the first widely distributed microprocessor-based computer, Altair 8800, was announced in December of 1974. By July of 1976, the Apple I computer is introduced. Clearly a revolution was afoot, but it was not until the IBM personal computer, the SUN and Apollo Workstations in 1981 and the Apple Macintosh in 1984, that serious computer power was in the hands of individual users.'

P405 pp.27-30 then spent a page on Wegman, followed by paragraphs on Bradley Efron, Jerome Friedman, David W. Scott, with mentions of others.

P403 p.136 (cont)

'The contemporary period has also witnessed a clear change in research emphases. The post-*Sputnik* era (1957-1979) saw relatively lavish funding of basic research in statistics and an increasing emphasis on the development of methodology. However, the post-1981 era saw a significant shift in emphasis to applications. Computers allowed for new data structures, or methods of organizing information, many of which did not follow traditional statistical models. **Edward J. Wegman** called for the statistical profession to become more data-centric rather than methodology-centric: that is, to take on challenges of the new data structures even though they did not fit conveniently within the framework of existing statistical models. Some emerging data structures and future directions for the profession include streaming data, image data, text data, and data available in the form of random graphs. No longer is basic research money easily available for research in statistical methodology alone. Increased emphasis on real problems cannot help but be good for the discipline, because

P403 p.137 (cont)

'virtually every significant advance in statistics has been motivated by addressing some real problem.

At the beginning of the twenty-first century, forensics continued to emerge as an important new focus of statistics. Statistical methods have been used to discredit to a large extent the use of polygraphs for lie detection and exam results are rarely admisable as evidence in U.S. courts. Similarly, the National Research Council of the National Academies studied, using statistical methods, the use of bullet lead analysis by the Federal Bureau of Investigation resulting i increased legal challenges to this type of evidence. Other forensic science evidence likely to come under statistical and other technical scrutiny in the future includes what is now called friction ridge evidence and blood alcohol concentration evidence. Though DNA evidence has been shown to be valid from a statistical perspective, the statistical certainty of these other forms of forensic evidence is far less clear and is likely to lead to additional significant

adjustment in legal procedures and less aggressive pursuit of convictions based on these methods.'

P403 p.137 (cont)

'Assessment. The rise of statistics as a research field and as a vital component of statecraft is an important aspect of modern history, first in the West, then more globally. The importance of the discipline for key institutions like insurance companies, where actuarial work began to expand from the late nineteenth century onward, ensured its development. Arguably, popular training in statistics – as opposed to more conventional mathematics-has lagged somewhat in many societies, creating gaps in interpreting the results of data, and in some instances, significant disagreements over calculations of risk.

P405 p.30 (cont)

'The Future Period[682] is clearly changing the research emphases. The post-Sputnik era (1957-1979) saw relatively lavish funding of basic research in statistics with only some lip service being paid to applications. This substantial funding of undirected basic research saw also increasing emphasis on the development of methodology. However, the post-1981 era saw a significant increase in emphasis on applications. The availability of computing also resulted in new and novel data structures, many of which did not follow traditional statistical models. Wegman (2000) called for the statistical profession to become more data centric rather than methodology centric, i.e. to take on challenges of the new data structure even though they did not fit conveniently within the framework of existing statistical models. Some emerging data structures and future directions for the profession include streaming data, image data, text data, and data available in the form of random graphs. No longer is basic research money easily available for research in statistical methodology alone. Increased emphasis on real problems cannot help but be a good feature for academic research because

virtually every significant advance has been motivated by addressing some real problem.'

P405 p.32

An interesting new direction has been emerging with respect to forensics in the courtroom. Statistical methods have been used to discredit to a large extent the use of polygraph for lie detection and such testimony is no longer allowed (National Research Council, 2003). Similarly, the National Research Council of the National Academies (2004) has considered bullet lead analysis used by the Federal Bureau of Investigation using statistical methods and has increased legal challenges to this type of evidence. Other *forensic science* evidence likely to come under statistical and other technical scrutiny in the future include what is now called friction ridge evidence and blood alcohol concentration evidence. While DNA evidence has been vetted from a statistical perspective, the statistical certainty of these other forms of forensic evidence is far less clear and is likely to lead to additional significant

P405 p.33

'adjustment in legal procedures and less aggressive pursuit of convictions based on these methods.'

==========

P403 cited 5 sources[683], of which 3 were found in P405: NRC(2004), NRC(2003) and Wegman(2000). The following were new:

'David, H.A., and A. W. F. Edwards. Annotated Readings in the History of Statistics. New York: Springer, 2001.'

'Schweber, Libby. Disciplining Statistics: Demograph and Vital Statistics in France and England, 1830-1885. Durham, N.C. Duke University Press, 2006.'

[682] Historians might be *uncomfortable* labeling a time as a "Future Period."

[683] Stigler was not cited.

Z. Wegman Proposal to ARO – 02/16/09

'Mathematical and Statistical Foundations of Networks'

Z.1 Proposal, $529K direct, labeled WEG2009

Federal agencies generally make proposals public if they are accepted, but (properly) not if they are rejected. In response to Dan Vergano's October 2010 FOIA request, Wegman sent many files related to an apparent proposal[684] by him to ARO, with several folders:

ARO_Proposal

This working folder had many files, including a copy of Sharabati's dissertation, "MULTI-MODE AND EVOLUTIONARY NETWORKS" (2008), as "Dissertation.pdf," [SHA2008], as in [MAS2012c §4.4].

ARO Proposal

This folder's files were created 02/16/09, with the various pieces of a detailed proposal, including Table of Contents,[685] abstract, [686] biography[687], budget,[688] bibliography[689] and the 27-page technical proposal itself.[690]
Wegman listed 8 publications. Of the "5 publications most closely related," none were peer-reviewed, and he led only one, on alcoholism:
***P163* w** (2004), Martinez, A.R., **Wegman,** and Martinez, W.L.
P170w (2005), Solka, Bryant, Avory C., and **Wegman**
P178 Sw (2007), Said, **Wegman,** Sharabati, and Rigsby.
P179 Sw (2007, retracted), Said, **Wegman,** Sharabati, and Rigsby.
P192 Ws (2008). **Wegman** and Said.

[684] Wegman just included it. Vergano neither knew about this nor asked for it.
[685] www.documentcloud.org/documents/524443-toc-network-science-proposal.html
[686] www.documentcloud.org/documents/524438-abstract-network-science-proposal.html
[687] www.documentcloud.org/documents/524440-biskchf-network-science-proposal-1.html
[688] www.documentcloud.org/documents/524435-masonbudget-aro-wegman-federal.html
[689] www.documentcloud.org/documents/524439-bibgraphyf-network-science-proposalpdf.html
[690] www.documentcloud.org/documents/524550-descriptf-network-science-proposal.html

The 5-year budget specified $322K to Wegman and $134K to Said, each 0.33 FTE, plus $37K for domestic travel, and other costs for total $529K. That did not include the usual ~48% added by GMU for indirect costs.

When funders evaluate proposals, they consider past performance, and *it is possible that the results and late reports of 0447 and 0059 were not plusses. However, from emails, ARO was generally encouraging and helpful to Wegman, perhaps given the long association.*
Some chronology may be worth reviewing:

11/01/04	0447 start
12/15/06	0059 start (of last ARO grant to Wegman)
12/15/07	0059 completion
04/30/08	0447 completion, after 6-month extension
12/10/08	0447 final report, 224 days after completion
12/15/08	Wegman: "I will be drafting something in the next week"
02/16/09	This proposal, [WEG2009]
03/08/09	0059 final report, 450 days after completion, cursory **§R.4**
05/07/09	negative feedback from ARO
05/29/09	5876 completion (Said)

§Z.2 gives some of the email history that shows the proposal was written in a month or two. He wrote that lack of support over previous summer caused him to have to take a second mortgage.
The emails also show some review comments, which were *as a group were strongly negative.*[691] One criticized a flaw in basic graph theory. Another said that many of the ideas had been explored before and that this proposal's lack of references to earlier work was not encouraging.[692]
In any case, the proposal was rejected, likely fortunate for Wegman, given the NSF Career Writing Workshop at GMU, 2009, p.66:[693]

> 'Plagiarism – material copied without citation and quotation – if you copy it, cite it and off-set it: **if you accept an award based on a proposal that includes plagiarism, you may have committed a felony**.'

[691] They are polite, but anyone familiar with reviews would understand the strength of the negatives, especially for a well-known, experienced researcher, not a proposal neophyte. Even one of the negatives would likely sink a proposal.
[692] This was an example of a pattern seen elsewhere, as Wegman and students *seemed to jump into unfamiliar areas* without studying them deeply first.
[693] grants.soe.ucsc.edu/sites/default/files/2%20George.ppt

Z.2 Emails

[WEG2010a][694] contains some back history on this proposal.
Some messages are excerpted here in chronological order.

10/01/08 p.17 Wegman to ARL

'Great to hear from you. Yes I am sure he would be willing to do so. He is my Ph.D. student and will be defending his dissertation shortly. The final defnse *(sic)* will be in about four weeks. **I am still hoping we can get into a new contract arrangement with ARL. This past summer I had no research support, which is very hard for me.**'

12/10/08 pp.21-22 ARO to Wegman on 0447

'Your Final Report has been received. …
U.S. Army Research Office …
DATES COVERED: 1-Nov-2004 to 30-Apr-2008 …
PROPOSAL TITLE: Analytic and Graphical Methods for Streaming Data with Applications to netcentric Warfare'

12/15/08 p.20 Wegman to ARO

'**Now that I have finally turned in my final report**,[695] I'd like a little advice. I want to submit a new proposal. I believe the last effort was very fruitful and I have some good ideas. I want to work on the mathematical foundations of network science and have several ideas on how to detect missing nodes and edges and also how to deal mathematically with dynamically expanding networks. I hope this is of interest to ARO. **Last summer, I went without research support and wound up having to take a second mortgage on my home in order to make ends meet.** I have been very supportive of the Arny nission *(sic)* over the years and ARO has returned the favor by being supportive of me. Without extracting a commitment from you ahead of time, I wonder if you might give me some guidance on possible funding levels. **I will be drafting something in the next week**, but I'd like to have some sense of what might be feasible. **For myself I would like summer support and perhaps a little released time during the year**. **Would it be possible to include some support for Dr. Yasmin Said who has been working closely with me over the last 2 ½ years?**'

02/19/09 p.23 GMU to Wegman

'Your application has been submitted to U.S Army Research Office on Grants.gov. Attached is the submission confirmation.'

05/07/09 pp.27-28 ARO to Wegman (entirely quoted, 4:1 *negative*)

'Prof Wegman,
It was good to have a conversation about the proposed project Mathematical and Statistical Foundations of Networks. I did contact the Army reviewer yesterday. She assures me she will work on the review.
Here are a comments from reviewers that I find helpful. Please do not take these comments as negative. They are intended to help improve the proposal.

"Very important topic, prominent investigators, however, **the description of the ideas is rather vague and still needs to be developed into a sound research plan. It is not clear how addressing the proposed research tasks will advance the existing theory behind network science."** *negative*

"Some of the proposed ideas … For example, the authors propose a method to estimate the probability of missing edges. Then they propose to use this method for estimating the probability of missing nodes for G by estimating the probabilities for corresponding edges in the line graph representation of G. **The problem is that, unless G is claw-free,[696] there may not be a one-to-one correspondence between vertices in G and edges in its line graph** representation (even though there is a one-to-one correspondence between edges in G and vertices in its line graph, this is not what one needs in this case)." *negative (claw-free comment applies to section 2.2.3 of the proposal)*

"The proposal outlines some potential new capabilities that can arise from the proposed study. **Some of those are very interesting**, like task **2** focusing on conversion of multimode non-binary adjacency tensors and matrices into lower degree networks or evolutionary algorithms for optimizing network assessment metrics. **Other tasks,** like **3** (evolving social networks), **4** (missing edges), 7 (connection between text mining and social networks), or **8** (limiting behavior of agent-based systems) **have been already studied and lack of references to the relevant work[697] makes this reviewer doubtful of the likelihood that the proposed research will lead to new capabilities in this tasks."** *negative*

"**With better description of the methods to be applied and perhaps some initial results showing the promise of these methods, as well as with clearly defined expected results, the proposal will be much stronger**." *negative*

"The PI is the distinguished scientist with strong past experience and publication record of papers relevant to the proposed study."' *positive*

[694] www.documentcloud.org/documents/527437-relevant-emails-redact-addresses.html
[695] That was the **0447** final report. The **0059** final report was **03/18/09,** ~year late.
[696] en.wikipedia.org/wiki/Claw-free_graph Well-known graph theory.
[697] [MAS2010a §W.5] noted this issue for Wegman group's SNA.

Z.3 Pages of proposal, side-by-side comparisons

1-2 Most of this text is quoted (properly) from an NRC report, but of course provides no information on the proposal itself.

2-9 Most is taken from Walid Sharabati's dissertation [SHA2008], co-supervised by Wegman and Said. About 2 pages were plagiarized from Wikipedia, [DEN2005] and especially [WAS1994].

10-15 Adds Iraq War, evolving networks, but ~half is from [SHA2008].

15-17 Most is almost identical to text and definitions from Hadi Rezazad's dissertation [REZ2009], Spring 2009, a few months later. However, this text must have been done much earlier.

18-21 No/few antecedents were found.

22- Mostly unknown, some passages from Wikipedia.

24-25 "Ising Models" is nearly identical to English in Chinese Wiki. Both *likely* came from Binder or Brush below, which Wegman cited as "See Brush (1967) and Binder (2001)," but with no hint that the text might have been copied from them.

26-27 Summary of tasks, no antecedents found.

Total About half the text appears to have been copied without proper attribution,[698] leaving half as possible new content. *Some may ignore the use of Wegman-supervised PhD dissertations,* but ~4 pages were copied from others besides. The dissertations were not mentioned.

The bibliography[699] seems a *strange* mix of textbooks and sometimes-obscure references, with few relevant, recent papers to demonstrate field familiarity:

'Binder, K. (2001) "Ising model," in Encyclopaedia of Mathematics (Hazewinkel, Michiel, ed.), Kluwer Academic Publishers.
Vaguely cited in the Summary,

Brush, Stephen G. (1967) "History of the Lenz-Ising model," Reviews of Modern Physics (American Physical Society) 39, 883–893. doi: 10.1103/RevModPhys.39.883 *Vaguely cited in the Summary,*

Committee on Network Science for Future Army Applications (2005) Network Science, Washington, DC: The National Academies Press. *Properly cited.*

de Nooy, W., Batageli, V., and Mrvar, A. (2004) Exploratory Social Network Analysis with Pajek, Cambridge, UK: Cambridge University Press.
[DEN2005] cited "for more details refer to de Nooy et al. (2004)"

Epstein, Joshua and Axtel, Robert (1996) Growing Artificial Societies: Social Science from the Bottom UP (Complex Adaptive Systems), Washington, DC: The Brookings Institution. *Cited once, vaguely.*

Marchette, David J. and Priebe, Carey E. (2008) "Predicting unobserved links in incompletely observed networks," Computational Statistics and Data Analysis, 52(3), 1373-1386. *Cited once.*

Martinez, Angel (2002) A Framework for the Representation of Semantics, Ph.D. Dissertation, School of Computational Sciences, George Mason University, Edward J. Wegman, Dissertation Director. *Cited 2 times.*

Martinez, Wendy, Martinez, Angel, and Wegman, Edward (2008) "Classification and clustering using weighted text proximity matrices," Computing Science and Statistics, 36, 600-611.
Interface proceedings, *Cited once.*

North, Michael and Macal, Charles (2007) Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation, New York: Oxford University Press. *Cited once, vaguely.*

Said, Yasmin (2009) Intervention to Prevention: A Policy Tool for Alcohol Studies, Saarbrücken, Germany: VDM Verlag. Dissertation, *See **§S.8*** 17.

Solka, Jeffrey L., Bryant, Avory C., and Wegman, Edward J. (2005) "Text data mining with minimal spanning trees," Handbook of Statistics: Data Mining and Data Visualization, (Rao, C. R., Wegman, E. J. and Solka, J. L., eds.), 133-170. Amsterdam: Elsevier/North Holland. P170w *Cited once.*

van Rooij, A. and Wilf, H. (1965) "The interchange graph of a finite graph," Acta Math. Acad. Sci. Hungar. 16, 263-269.

Wasserman, Stanley and Faust, Katherine (1994) Social Network Analysis: Methods and Applications, Cambridge, UK: Cambridge University Press.'
[WAS1994] This is actually cited 3 times, but with no quotations. Indeed, much text came from there but in 2011, Wegman was claiming he'd thought it had been original text by his student Reeves. The sequence was as follows:

[698] If text is copied, it must be quoted and cited, not vaguely mentioned nearby.
[699] www.documentcloud.org/documents/524439-bibgraphyf-network-science-proposalpdf.html

[WEG2009] ← [SHA2008] ← Sharabati ← Wegman ← Reeves ← {sources} Wegman wrote to Elsevier about the origin of the SNA text[700] used in the WR and [SAI2008], [MAS2011a, p.6-7]:

> 'I asked her *(Reeves)* to write up a short description we could include in our summary. She provided that within a few days, **which I of course took to be her original work.** Neither Yasmin, Walid Sharabati, John Rigsby nor I did know that she had basically copied and pasted this into her MS Word file. We included her boilerplate in our Congressional testimony and acknowledged Denise's contribution in that testimony. …'[701]
> '**thinking that the page and ½ Denise had given me was original work** that had not been formally published, I gave it as reading material to Walid as background material along with a number of other references. **Walid included it as background material in his dissertation with only minor amendments**.'

Some of the SNA text appeared in WR, [SAI2008], [SHA2008] and [REZ2009]. [WEG2009] was the 5th known re-use of Reeves' text, **§G.**

Wegman seems to have re-used Rezazad's work later. The reader can compare Rezazad[702] last modified 09/30/09, on material from his dissertation, with him as contact and Wegman, Rezazad, Shores[703] dated 10/22/09 with Wegman as contact. It edited "we" ← "I" on a few pages, and added pp.34-40, i.e., graph theory that *seems a bit extraneous*.

Wegman was entirely responsible for this proposal to the ARO. No grad students were ack'd or involved. Hence, there can be no doubt as to the authorship.

[VER2011] quoted Milton Johns and Wegman:

> '"**Neither Dr. Wegman nor Dr. Said has ever engaged in plagiarism**," says their attorney, Milton Johns, by e-mail. In a March 16 e-mail to the journal, Wegman blamed a student who "had basically copied and pasted" from others' work into the 2006 congressional report, and said the text was lifted without acknowledgment and used in the journal study. "**We would never knowingly publish plagiarized material**" wrote Wegman, a former *CSDA* journal editor.'

§Z.3 gives evidence of plagiarism, in the side-by-side comparison style used elsewhere by DC and [MAS2010a], adapted as needed for 3-way:

- Cyan for identical, in-order text between [WEG2009] and antecedents
- Yellow for trivial edits [WEG2009] ← [SHA2008]
- Yellow for trivial edits [SHA2008] ← antecedents, shown only in antecedents, unlike in two-way comparisons.

Page numbers use the document's own numbers, not those of the PDF. When mosaic plagiarism[704] is suspected in document X, if closely-matching text in Y can be found for a section of X then either:

- Y is the antecedent, either directly or indirectly OR
- X and Y both have another hidden antecedent OR
- Possibly X is original and was a source for Y. This always needs checking, but can usually be resolved via publication dates and/or authors' expertise and other publications.

If no matching text can be found:

- X is original or well-paraphrased from Y OR
- A hidden antecedent source has not yet been identified.

[700] Deep Climate, —A comparison of Said, Wegman, et al and Unattributed Sources, 09/08/10. deepclimate.files.wordpress.com/2010/09/said-et-al-social-networks-2.pdf

[701] The SNA introduction in the WR was ~5.5 pages, *likely* more than the work of the 2nd author, Scott, but she was only vaguely Ack'd, not labeled an author.

[702] "ACAS 2009\GMU-Presentation-092509.ppt" www.documentcloud.org/documents/550209-gmu-presentation-092509.html

[703] "ACAS_2009\ACAS_Wegman_Rezazad\Shores.ppt" www.documentcloud.org/documents/550210-acas-wegman-rezazad-shores.html

[704] isites.harvard.edu/icb/icb.do?keyword=k70847&pageid=icb.page342054#a_icb_pagecontent732741_mosaic for example.

[WEG2009, pp.1-2]
Most of this quotes excerpts from a 2005 NRC report on Network Science, acceptably referenced, but not original.

[WEG2009, p.2[
2 SOCIAL NETWORKS
Social Network Analysis (SNA) or Network Theory is becoming an important tool used to analyze, model, and simulate the behavior of groups of people or entities both on the global level (how two or more groups interact with other group(s)) and on the local level (how individuals interact with each other within the same network.) In the past two decades, SNA has been used to analyze relations and ties among individuals of the same network and similarities between different networks to obtain a better understanding on how societies interact.

The basic mathematical structure for visualizing the social network is a *graph*. A graph is a pair {V;E} where V is a set of nodes or vertices and E is a set of edges or links
Social network analysis

has emerged as a key technique and a topic of study in modern sociology, anthropology, social psychology and organizational theory.
The shape of the social network helps determine a network's usefulness to its individuals. Smaller, tighter networks can be less useful to their members than networks with lots of loose connections (weak ties) to individuals outside the main network. More "open" networks,

[WEG2009, p.3[
with many weak ties and social connections, are more likely to introduce new ideas and opportunities to their members than closed networks with many redundant ties.

←

[SHA2008, p.1]

Social Network Analysis (SNA) or Network Theory is becoming important tools used to analyze, model, and simulate the behavior of groups of people or entities both on the global level (how two or more groups interact with other group(s)) and on the local level (how individuals interact with each other within the same network.) In the past two decades, SNA has been used to analyze relations and ties among individuals of the same network and similarities between different networks to obtain a better understanding on how societies interact.

[SHA2008, p.2]
The basic mathematical structure for visualizing the social network is a graph. A graph is a pair V;E where V is a set of nodes or vertices and E is a set of edges or links.
Social network analysis ~~(also called network theory)~~

has emerged as a key technique and a topic of study in modern sociology, anthropology, social psychology and organizational theory.
The shape of the social network helps determine a network's usefulness to its individuals. Smaller, tighter networks can be less useful to their members than networks with lots of loose connections (weak ties) to individuals outside the main network. More "open" networks,

with many weak ties and social connections, are more likely to introduce new ideas and opportunities to their members than closed networks with many redundant ties.

←

Original antecedents, most from [WAS1994], some from Wikipedia or [DEN2005].

Wikipedia – Social Networks – 01/02/06[705]
[Head section 1]

Social network analysis (also ~~sometimes~~ called network theory)
has emerged as a key technique in modern sociology, anthropology, Social Psychology and organizational studies, ~~as well as a popular topic of speculation and study.~~
The shape of the social network helps determine a network's usefulness to its individuals. Smaller, tighter networks can be less useful to their members than networks with lots of loose connections (weak ties) to individuals outside the main network. More "open" networks,

with many weak ties and social connections, are more likely to introduce new ideas and opportunities to their members than closed networks with many redundant ties.

[705] en.wikipedia.org/w/index.php?title=Social_network&oldid=33590649 called [WIK2006a] elsewhere.

[WEG2009, p.3 continued]

Social network analysis is concerned with understanding the linkages among social entities and the implications of these linkages. The social entities are referred to as actors that are represented by the vertices of the graph.

Most social network applications consider a collection of actors that are all of the same type. These are known as one-mode networks.

Social ties link actors to one another.
The range and type of social ties can be quite extensive.

Linkages are represented by edges of the graph. Examples of linkages include the
evaluation of one person by another (such as expressed friendship, liking, respect),
transfer of material resources (such as business transactions, lending or borrowing things),
association or affiliation (such as jointly attending the same social event or

belonging to the same social club),
behavioral interaction (talking together, sending messages),
movement between places or statues
(migration, social or physical mobility),
physical connection (a road, river, bridge connecting two points),
formal relations such as authority, and
biological relationships such as kinship or descent.

The tie is an inherent property of a pair.

Many kinds of network analysis are concerned with understanding ties among pairs and are based on the *dyad* as the unit of analysis.

The "statues" typo (or bad original OCR by Reeves) got carried through all re-uses of [WAS1994] except [SAI2008].

[SHA2008, p.2]

Social network analysis is concerned with understanding the linkages among social entities and the implications of these linkages. The social entities are referred to as actors that are represented by the vertices of the graph.

Most social network applications consider a collection of actors that are all of the same type. These are known as one-mode networks

Social ties link actors to one another.
The range and type of social ties can be quite extensive.
~~A tie establishes a linkage between a pair of actors.~~

Linkages are represented by edges of the graph. Examples of linkages include the
evaluation of one person by another (such as expressed friendship, liking, respect),
transfer of material resources (such as business transactions, lending or borrowing things),
association or affiliation (such as jointly attending the same social event or

[SHA2008, p.3]

belonging to the same social club),
behavioral interaction (talking together, sending messages),
movement between places or statues
(migration, social or physical mobility),
physical connection (a road, river, bridge connecting two points),
formal relations such as authority and
biological relationships such as kinship or descent.
~~A linkage or relationship establishes a tie at the most basic level between a pair of actors.~~

The tie is an inherent property of the pair.

Many kinds of network analysis are concerned with understanding ties among pairs and are based on the dyad as the unit of analysis.

[WAS1994, p.17]

Actor.
social network analysis is concerned with understanding the linkages among social entities and the implications of these linkages. The social entities are referred to as actors. ~~Our use of the term "actor" does not imply that these entities necessarily have volition or the ability to "act". Further~~
most social network applications focus on collections of actors that are all of the same type …. We call such collections one-mode networks…

[WAS1994, p.18]

Relational tie. Actors are linked to one another by social ties. …the range and type of ties can be quite extensive. The defining feature of a tie ~~is that it~~ establishes a linkage between a pair of actors. Some of the more common examples of ties employed in network analysis are:

- Evaluation of one person by another (for example expressed friendship, liking, or respect)
- Transfers of material resources (for example business transactions, lending or borrowing things)
- Association or affiliation (for example jointly attending a social event, or belonging to the same social club)

- Behavioral interaction (talking together, sending messages)
- Movement between places or statuses (migration, social or physical mobility)
- Physical connection {a road. river, ~~or~~ bridge connecting two points}
- Formal relations (for example authority)
- Biological relationship (kinship or descent)

Dyad. At the most basic level, a linkage or relationship establishes a tie between two actors.
The tie is inherently a property of the pair ~~and therefore is not thought of as pertaining simply to an individual actor.~~

Many kinds of network analysis are concerned with understanding ties among pairs. All of these approaches take the dyad as the unit of analysis

[WEG2009, p.3 cont]

A social network consists of a finite set or sets of actors and the relation or relations defined on them. The presence of relational information is a significant feature of a social network.

A partition of a network is a classification or clustering of the vertices in the network so that each vertex is assigned to exactly one class or cluster; that is,

partitions divide the vertices of a network into a number of mutually exclusive subsets. Partitions may specify some property that depends on attributes of the vertices.
Partitions are also sometimes called blocks or block models. These are essentially a way to cluster actors together in groups that behave in a similar way.

In a network setting, actors or entities have several attributes to identify their role, behavior, background, and/or assets; some of which are unique to that actor and some are common among other actors. These attributes are the node's properties such as gender, age, political affiliation, ethnicity, race, nationality, religion, spoken languages, scientific field, income, education level, job class, and geographic location.

2.1 PRELIMINARIES

[WEG2009, p.4]
Networks can be treated as directed graphs.

There are three basic representations of a network: the planar graph visualization, the adjacency matrix, and the sparse-graph representation.
There are several algorithms to study interactions within the network include centrality measures (node degree and closeness), network partitioning (cliques and clique overlapping), network connectivity (cut-points and bridges), structural *equivalence, structural holes, brokerage roles, and block-modeling.*

[SHA2008, p.3 cont]

A social network consists of a finite set or sets of actors and the relation or relations defined on them. The presence of relational information is a significant feature of a social network.
"significant" ← *"critical&defining" not same.*

A partition of a network is a classification or clustering of the vertices in the network so that each vertex is assigned to exactly one class or cluster. Partitions may specify some property that depends on attributes of the vertices.
Partitions divide the vertices of a network into a number of mutually exclusive subsets. ~~That is, a partition splits a network into parts.~~
Partitions are also sometimes called blocks or block models. These are essentially a way to cluster actors together in groups that behave in a similar way.

In a network setting, actors or entities have several attributes to identify their role, behavior, background, and/or assets; some of which are unique to that actor and some are common among other actors. These attributes are the nodes' properties such as gender, age, political affiliation, ethnicity, race, nationality, religion, spoken languages, scientific field, income, education level, job class, and geographic location.

[SHA2008, p.4]
1.2 Preliminaries

Networks can be treated as directed graphs in which actors (individuals) are represented by vertices (nodes) while interactions between actors are represented by edges (ties), which may have weights.
There are three basic representations of a network {the planar graph visualization, the adjacency matrix, and the sparse-graph representation.
There are several algorithms to study interactions within the network include centrality measures (node degree and closeness), network partitioning (cliques and clique overlapping), network connectivity (cut-points and bridges), structural equivalence, structural holes, brokerage roles and block-modeling, ~~which will all be defined shortly.~~

[WAS1994, p.20]

Social Network. Having defined actor, group, and relation we can now give a more explicit definition of social network
A social network consists of a finite set or sets of actors and the relation or relations defined on them. The presence of relational information is a critical and defining feature of a social network …

[DEN2005, p. 31]
A partition of a network is a classification or clustering of the vertices in the network so that each vertex is assigned to exactly one class or cluster.
[DEN2005, p. 36]
Partitions divide the vertices of a network into a number of mutually exclusive subsets. In other words, a partition splits a network into parts.

From just these short passages alone, it would not be obvious that [DEN2005] was the original antecedent.

However, [MAS2010a §W.2.3] showed these as a few of many [DEN2005] passages used in the WR, of which only a small subset got re-used.

Such material is an example of approach seen elsewhere by Wegman and students, in which a dissertation, paper or proposal is bulked up with standard definitions whose text is copy-paste-edited from elsewhere, without clear attribution, rather than just writing "using standard definitions from …"

[WEG2009, p.4 cont]

Definition 2.1. A graph, G, is a collection of vertices V and edges E; G = {V, E} …
Definition 2.2. An adjacency matrix, E, associated with a graph, G
(slightly simplified notation)

A key insight of these definitions is that there is a fundamental duality between the graph and its adjacency matrix. That is, if one is given the adjacency matrix, one is able to construct the graph, and similarly, if one is given the graph, one can construct the adjacency matrix. The computationally-oriented social scientists tend to focus on the graph because it is a descriptive way of visualizing the social network. The adjacency matrix is used, but in SNA, the properties of the network are rarely explored in the context of a deeper mathematical analysis of the matrix representation. Because much is known about linear spaces, it is the theme of this proposal that we can understand much about networks in terms of this matrix representation. There are a number of metrics that describe quantitative aspects of a social network. In particular, there is much discussion of dyads, triads and cliques in social network analysis.

2.1.1 CENTRALITY MEASURES
There are three main centrality measures defined in Wasserman and Faust (1994); namely, degree centrality, closeness centrality, and betweenness centrality.

Degree of a vertex is the number of edges that connect it to other nodes. Degree can be interpreted as measure of power or importance of a node, or measure of workload. The

actor with most ties is the most important actor in a network. It has been shown that in a simple random graph, degree centrality has the Poisson distribution. Nodes with high degree are likely to

[WEG2009, p.5]
be at the intuitive center. Deviations from a Poisson distribution suggest non-random processes, such processes form *scale-free networks*.

[SHA2008, p.4 cont]

Definition 1.2.1. A graph G, is a collection of vertices V and edges E; G = {V;E}, ….
Defnition 1.2.2. An adjacency matrix A associated with a graph G
(standard definitions of graphs, no problem for Sharabati)

No antecedent found.

1.2.1 Centrality Measures
There are three main centrality measures defined in [60]; namely, degree centrality, closeness centrality and betweenness centrality. ~~To serve the purposes of this research, I will define degree and closeness centrality measures only.~~
Degree of a vertex is the number of edges that connect it to other nodes, ~~see Figure 1.1~~. Degree can be interpreted as measure of power or importance of a node, or measure of workload. The
[SHA2008, p.5]
actor with most ties is the most important figure in a network. It has been shown that in a simple random graph, degree centrality has the Poisson distribution. Nodes with high degree are likely to

be at the intuitive center. Deviations from a Poisson distribution suggest non-random processes, such processes form "scale-free" networks

The reader might wonder if well-published SNA experts would agree with the opinions at left.
Google: "social network analysis" "adjacency matrix" OR
"social network analysis" "adjacency matrix" multimode OR
"social network analysis" "adjacency matrix" cuboid

It is curious that references to [WAS1994] and [DEN2005] appeared in [SHA2008], but no one seemed to worry that the original SNA text in the WR might have been plagiarized. In any case, Wegman had claimed they all thought this was original work of Reeves.

Wegman and some of his students seemed to have a habit of copying text that gave fairly-standard basic definitions, many of which were not even used in the remainder of their text. It seems that this was just intended to convey the impression of expertise, and add bulk, because the results were useful neither to the general public (as in the WR) nor to experts. The latter would likely write "we use the standard terminology of <source>, notation summarized below"

[WEG2009, p.5 cont]
Definition 2.3. Degree of a vertex …
Definition 2.4. Closeness; …
(slightly simplified notation, but standard)
Closeness centrality measure is based on the inverse of the distance of each actor to every other actor in the network.
Distance in this context is defined to be the number of steps a vertex v_i needs to reach a vertex v_j . If an actor is close to all other actors then this actor is considered important.

Definition 2.5. The geodesic is the length of the shortest path between any two vertices.

2.1.2 COHESIVE SUB-GROUPS: CLIQUES
Definition 2.6. A dyad is a pair of vertices and the edge connecting them.

Definition 2.7. A triad is a set of three vertices and the edges connecting them.

A triad is identified by a M-A-N number system of three digits and a letter; for more details refer to de Nooy et al. (2004)
One of the interesting features in a network that caught structural analysts' attention is secondary sub-structures such as network cohesion.
[WEG2009, p.6]
Researchers interested in cohesive subgroups gathered and studied sociometric data on affective ties

in order to identify "cliquish" subgroups (face-to-face group).

The clique is the foundational idea for studying and analyzing cohesive subgroups in social networks.

Definition 2.8. A clique in a graph is a maximal complete subgraph of three or more nodes, mutual dyads are not considered to be cliques (Wasserman and Faust, 1994).
It consists of a subset of nodes all of which are adjacent to each other, and there are no other nodes that are also adjacent to all of the members of the clique.
A clique is a very strict definition of cohesive subgroups.
Cliques are a subset of the network in

which the actors are more closely and intensely tied to one another than they are to other members of the network and if one actor disappears for any reason, the others are still directly connected to each other.

[SHA2008, p.5 cont]
Definition 1.2.3. Degree of a vertex …
Definition 1.2.4. Closeness; …

Closeness centrality measure is based on the inverse of the distance of each actor to every other actor in the network.
Distance in this context is defined to be the number of steps a vertex v_i needs to reach a vertex v_j . If an actor is close to all other actors then this actor is considered important.
[SHA2008, p.6]
Definition 1.2.5. The geodesic is the length of the shortest path between any two vertices.

1.2.2 Cohesive Sub-Groups: Cliques
Definition 1.2.6. A dyad is a pair of vertices and the edge connecting them.

Definition 1.2.7. A triad is a set of three vertices and the edges connecting them.

A triad is identified by a M-A-N number system of three digits and a letter; for more details refer to [14].
One of the interesting features in a network that caught structural analysts' attention is secondary sub-structures such as network cohesion.

Researchers interested in cohesive subgroups gathered and studied sociometric data on affective ties

in order to identify "cliquish" subgroups (face-to-face group).

The clique is the foundational idea for studying and analyzing cohesive subgroups in social networks.

Definition 1.2.8. A clique in a graph is a maximal complete subgraph of three or more nodes, mutual dyads ~~(2 nodes)~~ are not considered to be cliques [60].
It consists of a subset of nodes all of which are adjacent to each other, and there are no other nodes that are also adjacent to all of the members of the clique.
A clique is a very strict definition of cohesive subgroups.
Cliques are a subset of the network in
[SHA2008, p.7]
which the actors are more closely and intensely tied to one another than they are to other members of the network and if one actor disappears for any reason, the other two can still write/talk to each other.

Standard definitions.

[WAS1994, p.253]
Researchers interested in cohesive subgroups gathered and studied sociometric data on affective ties,
such as friendship or liking in small face-to-face groups,
in order to identify "cliquish" subgroups …
[WAS1994, p.254]
The clique is the foundational idea for studying and analyzing cohesive subgroups in social networks.

A clique in a graph is a maximal complete subgraph of three or more nodes … mutual dyads are not considered to be cliques.
It consists of a subset of nodes all of which are adjacent to each other, and there are no other nodes that are also adjacent to all of the members of the clique.
A clique is a very strict definition of cohesive subgroups.

[SHA2008] at least had a vague citation to [WAS1994], but still was plagiarized.

[WEG2009, p.6 cont]
2.1.3 BLOCKMODELS
Definition 2.9. A blockmodel is the process of identifying positions in the network.
A block is a section of the adjacency matrix consisting of a group of actors that are structurally equivalent. It consists of two things according to Wasserman and Faust (1994):

A partition of actors in the network into discrete subsets called positions.
For each pair of positions a statement of the presence or absence of a tie within or between the positions on each of the relations.
A blockmodel is thus a hypothesis about a multirelational network. It presents general features of the network, such as the ties between positions, rather than information about individual actors.

A blockmodel is a simplified representation of multirelational network that captures some of the general features of a network's structure. Specifically, positions in a blockmodel contain actors who are approximately structurally equivalent. Actors in the same position have identical or similar ties to and from all actors in other positions.

Thus, the blockmodel is stated at the level of the positions, not individual actors.

2.1.4 TWO-MODE AND MULTIMODE NETWORKS
Some social network relationships can be treated as a two-mode "bipartite" networks, or three-mode "tripartite" networks. As an example, consider the author-paper networks,

[WEG2009, p.7]
there are two types of vertices, one class of vertices represents authors, while the other class of vertices represents papers. There is one relationship type; "person A authored/coauthored paper P". It is understood that in this case the adjacency matrix is binary, i.e. the entries in the matrix are only 0s and 1s. (Later we shall see situations where the entries in the adjacency matrix can represent strength of relationship, e.g. measured as frequency counts or probabilities.)
One can perform matrix operations such as the product of matrices to obtain interesting results.

[SHA2008, p.8]
1.2.4 Blockmodel
Definition 1.2.10. A blockmodel is the process of identifying positions in the network.
A block is a section of the adjacency matrix "a group of people" structurally equivalent. It consists of two things according to Wasserman and Faust [60]:

●A partition of actors in the network into discrete subsets called positions.
●For each pair of positions a statement of the presence or absence of a tie within or between the positions on each of the relations.
A blockmodel is thus a hypothesis about a multirelational network. It presents general features of the network, such as the ties between positions, rather than information about individual actors.

A blockmodel is a simplified representation of multirelational network that captures some of the general features of a network's structure. Specifically, positions in a blockmodel contain actors who are approximately structurally equivalent. Actors in the same position have identical or similar ties to and from all actors in other positions.

Thus, the blockmodel is stated at the level of the positions, not individual actors.

[SHA2008, p.10]
1.3.1 Relational Networks
Some social network relationships can be treated as a two-mode "bipartite" networks, or three-mode "tripartite" networks. As an example, consider the author-paper networks,

there are two types of vertices, one class of vertices represents authors while the other represents papers. There is one relationship type; "person A authored/coauthored paper P". ~~This two-mode relational socio-network can be concluded from the PCANS model [33], [9]. The PCANS model is presented in Table 1.1:~~

I can perform matrix operations such as the product of matrices to obtain interesting results given that the two-mode matrix is binary.

[WAS1994, p.395]

A blockmodel consists of two things:

(i)A partition of actors in the network into discrete subsets called positions.
(ii)For each pair of positions a statement of the presence or absence of a tie within or between the positions on each of the relations.
A blockmodel is thus a hypothesis about a multirelational network. It presents general features of the network, such as the ties between positions, rather than information about individual actors.
[WAS1994, p.395]
A blockmodel is a simplified representation of multirelational network that captures some of the general features of a network's structure. Specifically, positions in a blockmodel contain actors who are approximately structurally equivalent. Actors in the same position have identical or similar ties to and from all actors in other positions.
For example, all actors in position B_k have similar ties to actors in positions B_l, B_m, and so on.
Thus, the blockmodel is stated at the level of the positions, not individual actors.

After this point, no obvious text was found from [WAS1994] or elsewhere, so presentation switches to 2-column.

[WEG2009, p.7 cont]

2.1.4.1 EXAMPLE

Consider a bipartite "coauthor-by-paper" social network. Let A be the adjacency matrix of size m x n representing the graph of the network, with m = number of coauthors, and n = number of papers. Then ,

$C_{m\times m} = A_{m\times n} \cdot A^T_{n\times m}$ = the one-mode coauthorship adjacency matrix, and

$P_{n\times n} = A^T_{n\times m} \cdot A_{m\times n}$ = the one-mode paper-by-paper adjacency matrix

where

$c_{ii} = \sum_{j=1}^{n} a_{ij}$ = the number of papers author i published

$p_{jj} = \sum_{i=1}^{m} a_{ij}$ = the number of coauthors that coauthored paper j, and

c_{ij} = the tie-strength between authors i and j.

Finally if $D_{m\times m} = C^2_{m\times m}$, then d_{ii} = the vertex degree of author i.

Suppose the coauthor-by-paper adjacency matrix A is given by

	paper1	paper2	paper3
coauthor1	1	0	1
coauthor2	0	1	1
coauthor3	1	1	0
coauthor4	1	0	0

[WEG2009, p.8]

$$\Rightarrow A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}_{4\times 3}$$

There is a one-to-one correspondence between the adjacency matrix representation of a network and directed graphs.

[SHA2008, p.11]

Consider a bipartite "coauthor-by-paper" social network. Let A be the adjacency matrix of size m x n representing the graph of the network, with m = number of coauthors, and n = number of papers. Then,

$C_{m\times m} = A_{m\times n} \cdot A^T_{n\times m}$ = coauthorship proximity matrix, and

$P_{n\times n} = A^T_{n\times m} \cdot A_{m\times n}$ = paper-by-paper proximity matrix.

where,

$c_{ii} = \sum_{j=1}^{n} a_{ij}$ = number of papers author i published,

$p_{jj} = \sum_{i=1}^{m} a_{ij}$ = number of coauthors coauthored paper j, and

c_{ij} = tie-strength between coauthors i and j.

Finally, if $D_{m\times m} = C^2_{m\times m}$ then

d_{ii} = vertex degree of coauthor i.

[SHA2008, p.11]

~~Example:~~

Suppose the coauthor-by-paper adjacency matrix A is given by

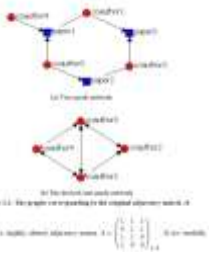
[htpb]

	paper1	paper2	paper3
coauthor1	1	0	1
coauthor2	0	1	1
coauthor3	1	1	0
coauthor4	1	0	0

$$\Rightarrow A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}_{4\times 3}$$

There is a one-to-one correspondence between the matrix representation of a ~~social~~ network and directed graphs.

[WEG2009, p.8]
2.1.4.1 EXAMPLE



If we carefully examine the networks in Figures 2.1 and 2.2, we observe that these

different 2-mode networks in fact have the same 1-mode graphical network representation. This is due to the fact that when converting to a 1-mode network, some network features are lost; much.

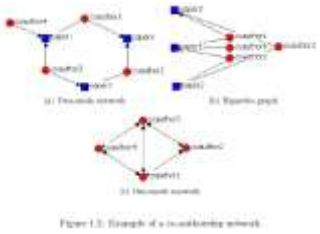
[WEG2009, p.9]
the same effect when one projects from 3-D to 2-D. This is an example of how the one-mode network does not provide sufficient information on how peer-ties are formed. As a result, the analysis of two-mode networks, indeed, multi-mode networks and one-mode networks should be performed concurrently.

The blockmodel does not show how cliques were formed. The ultimate solution to this problem is to consider the weighted adjacency matrix as opposed to the binary adjacency and then construct the distributions of dyads

and higher order interactions.

[WEG2009, p.9]
2.2 RESEARCH ISSUES FOR SOCIAL NETWORKS
Having now laid out the basics of social networks, we would like to raise some research issues. The calculation given above of the two one-mode networks from a two-mode network is dependent on the fact that the networks in question are binary, i.e. the adjacency matrices contain only 0s and 1s. Clearly it would be desirable to allow non
[WEG2009, p.10]
binary entries. Moreover, it seems clear that allowing for the possibility of multi-mode networks is desirable. Another desirable extension is the possibility of allowing the networks to grow or change over time. Two other extensions involve making inferences about missing edges and/or nodes. We address these in turn.

[SHA2008, p.12 cont



If we carefully examine the networks in Figures 1.2(a) and 1.3(a) we observe that these
[SHA2008, p.13]
different 2-mode networks in fact have the same 1-mode graphical network representation, ~~see Figures 1.2(c) and 1.3(c).~~ This is due to the fact that when converting to 1-mode some network features are lost;

the same effect when someone projects from 3-D to 2-D. This is an example of how the 1-mode network does not provide sufficient answer of how peer-ties are formed. As a result, the analysis of two-mode networks and one-mode networks should
be performed concurrently. …
~~This is important when cliques are present and one needs to determine which members formed which clique. The entrepreneurial and laboratory style of coauthorship networks, which will be discussed in Chapter 5 are different styles, yet the blockmodel of the 1-mode network identifies both as one style.~~
The blockmodel does not show how cliques were formed. The ultimate solution to this problem is to consider the weighted adjacency matrix as opposed to the binary adjacency and then construct the distributions of dyads,
[SHA2008, p.14]
triads, tetrads, pentads, hexads, heptads, and octads.

[WEG2009, p.10 cont]

2.2.1 MULTI-MODE NON-BINARY NETWORKS

The Iraq War gives a perfect example of a situation in which there are obviously more than two classes of actors. At the very least, there are the allied military, the insurgents, and the civilians. Indeed, the allied military could be broken down into U.S. military and other forces. The insurgents could easily be broken into the Shiite militias, the Al-Quaeda insurgents, and the Iranian-sponsored insurgents. The civilians could easily be broken into Shiites, Sunnis, and Christian civilian populations. In the two-mode network case, the two-mode network can be broken into 2! = 2 one-mode networks. This corresponds to the two one-dimensional faces of the two-mode adjacency matrix. Because we have only one unique way of forming the transpose with a two-dimensional matrix, we have only two one-mode networks resulting. Supposing for a moment that we have a three-mode network. Suppose *n, m, k* are respectively the number of actors in each of the three groups. Then these are the dimensions of the sides of the 3D cuboid adjacency matrix. A cuboid matrix is in fact a tensor of rank 3; however, for the purposes of this proposed research discussion, we will use the term cuboid instead.

The 3D cuboid has six faces leading to six different ways to view the block in terms of size, namely, *n* x *m* x *k, n* x *k* x *m, m* x *n* x *k, m* x *k* x *n, k* x*n* x *m* and *k* x *m* x *n*. As a result, the transpose of the 3D cuboid matrix can be done in 3! = 6 different ways. Because the transpose can be done in different ways, the two-mode and one-mode adjacency matrices can be formed in much more complex ways than the simple two-mode social networks. *We propose to explore the three-mode and, more generally, the N-mode social networks through their adjacency tensors of rank 3 and rank N respectively.*

Turning back for a moment from considerations of multimode networks, we focus on non-binary matrices. The reason that the computations for Cmxm = A_{mxn} * A^T_{nxm} and P_{nxn} = A^T_{nxm}* A_{mxn} work in Example 2.1.4.1 is that the matrices are binary. In particular, 1 x 1 = 1 and 0 x 0 = 0 . Of course, if the elements, a_i, are frequency counts or probabilities, this idempotent multiplication will not be the case. It is clear that the multimode case is even more complex. We believe we have an approach based on tensor

[WEG2009, p.11]

decomposition. *We propose to explore methods for addressing conversion of multimode non-binary adjacency tensors and matrices into lower mode networks.*

2.2.2 EVOLVING NETWORKS

Networks evolve over time or at least our knowledge of the actors and their connections evolve over time. Figure 2.3 illustrates the preliminary assessment of connections among the 9/11 hijackers. Figure 2.4 illustrates the assessments of the expanded social network based on additional intelligence information. Of course, it is expected that social networks will evolve as new actors enter the milieu. In addition, of course, while the set of actors in the network itself could be static, the relationships may shift and evolve so that old connections are broken and new one established. The difficulty from a mathematical perspective is that as new actors come into or leave the network, the size of the adjacency matrix or tensor changes.

[SHA2008, p.55]

A cuboid matrix is in fact a tensor of rank 3; however, for the purposes of this research I will use the term cuboid instead.

[SHA2008, p.56]

I would like to discuss how a cuboid is being transposed in 3D. Unlike the 2D rectangular matrix, which only has two faces, the 3D cuboid has six faces leading to six different ways to view the block in terms of size, namely, *n* x *m* x *p, n* x *p* x *m, m* x *n* x *p, m* x *p* x *n, p* x*n* x *m* and *p* x *m* x *n*. As a result, the transpose can be done in six different ways.

[WEG2009, p.11 cont]
Thus the underlying mathematical framework is different. If we view the one-mode adjacency matrix as an operator on a finite dimensional vector space, then as the network evolves, the dimension of the relevant vector space also changes. The implication is that there is no common mathematical framework for the network. The solution would appear to be to assume that the social network has an infinite number of nodes, all but a finite number of the are inactive at any given time. However, they may be activated with null links also being activated. This perspective allows a common infinite-dimensional framework to be in place at all times. Of considerable importance, is the fact that the strength of ties may be time-dependent. This perspective is motivated by our alcohol-modeling in Fairfax County, VA. Not only are new residents coming into the community as well as old ones leaving, but the strength of connections in their multimode network (individuals and alcohol outlets) as measured by conditional probabilities is changing on multiple time scales. The size of this social network is on the scale of 1,000,000 actors. *We propose to develop common mathematical framework for evolving social networks including multimodal networks.*

2.2.3 ESTIMATING MISSING LINKS AND MISSING NODES

Because edges determine connectivity between nodes, they are crucial to the structure of networks and knowing whether or not there is a missing edge in an incompletely observed network is of great importance. In many sampled networks, edges are imperfectly observed because of under-coverage or because actors are intentionally suppressing their roles and linkages to serve different purposes. A clear example of the

[WEG2009, p.12]
latter are networks of terrorists as in Figures 2.3 and 2.4. But criminal networks, networks of spies, even networks of corporations and of countries may want to suppress connections in order to gain strategic advantage. We suggest

a mathematical model to predict unobserved edges and vertices in a network based on covariate information on vertices and edges. The covariates are the exogenous attributes of actors. There are two types of attributes a set of nodes or edges can have, quantitative attributes, which are numerical summaries associated with entities and qualitative attributes, which are categorical summaries associated with entities. Our proposed model consists of two similarity measures calculated simultaneously using both the quantitative and the qualitative attributes derived

exogenously. We note that Marchette and Priebe (2008) develop a method for predicting edges based on a constrained random dot product graph which use endogenous properties of the graph itself.

Our idea is that if two vertices have a high similarity measure, then there is a high probability the vertices have edge connecting them or there is a high potential for forming an edge. Nodes and edges do not necessarily have the same set of attributes. Depending on the network setup and the properties of the entities, different networks may have completely different set of node attributes. Therefore, before applying the proposed method of estimating missing linkages, covariate information needs to be carefully defined.

[SHA2008, p.70]
Estimating Missing Edges And Vertices Using Covariate Information

Because edges determine connectivity between nodes, they are crucial to the structure of networks and knowing whether or not there is a missing edge in an incompletely observed network is of great importance. In many sampled networks, edges are imperfectly observed because of under-coverage or because actors are intentionally suppressing their roles and linkages to serve different purposes.

In this chapter, I present

a mathematical model to predict unobserved edges and vertices in a network based on covariate information on vertices and edges. The covariates are the exogenous attributes of entities. There are two types of attributes a set of nodes or edges can have, quantitative attributes, which are numerical summaries associated with entities and qualitative attributes, which are categorical summaries associated with entities. The model consists of two similarity measures calculated simultaneously using both the quantitative and the qualitative attributes derived

[SHA2008, p.71]
externally as opposed to endogenous approach. In the process of computing the similarity measure between nodes using the quantitative information I use the inner (dot) product technique to obtain an estimate. On the other hand, I use contingency tables and the χ^2 test to obtain another estimate to compute the similarity using qualitative information. The probability of having an edge between two given vertices is then a weighted sum of the two estimates.

If two pairwise vertices wind up having a high similarity measure then there is a high probability the vertices have edge connecting them or there is a high potential for forming an edge. Nodes and edges do not necessarily have the same set of attributes. Depending on the network setup and the properties of the entities, different networks may have completely different set of nodes attributes. Therefore, before applying the method of estimating missing linkages, covariate information need to be carefully defined.

[WEG2009, p.12 cont]
For example, in the author-coauthor social networks, possible attributes on authors and coauthors are *age, education, gender, spoken languages, discipline, number of publications.* However, possible attributes related to papers include *field, topic, keywords, year of publication, publisher, single/multiple author(s)*. In the alcoholconsumer settings, *age, ethnicity, smoker, drug-user, alcoholic, income, job-class* are possible consumers attributes, whereas *zip-code, location, hours-of-day, days-of-week* are some possible attributes associated with alcohol outlets. *We propose to develop methods of inferring the probability of possible missing edges based on inner product and related similarity measures using exogenous attributes.*

Vertices are not less important than edges. In fact, actors are the main element of a network; without actors a network is meaningless. Actors play a significant role in determining the dynamics of a network. We propose
a technique to estimate missing vertices (nodes) in a network. The method is again based on covariate information for vertices (actors) rather than edges, and utilizes the line space of edges which becomes the space of vertices as discussed
above.

In the line space of graphs, vertices become edges and edges become vertices. Consequently, to estimate a missing vertex in the space of

[WEG2009, p.13]
graphs, it suffices to estimate the missing edge corresponding to that vertex in the line space of graphs. In this

regards, we propose to use a mapping to transform from the space of graphs to the line space. Because graphs and matrices are isomorphic (one-to-one and onto), there is a function (transformation) that takes the graph and transforms it from the original space onto the line space and vice versa using matrices. In this sense, the matrix is the operator.

[SHA2008, p.71 cont]
For example, in the author-coauthor social networks, possible attributes on authors and coauthors are *age, education, gender, spoken languages, discipline, number of publications*. However, possible attributes related to papers include *field, topic, keywords, year of publication, publisher, single/multiple author(s)*. In the alcoholconsumer settings, *age, ethnicity, smoker, drug-user, alcoholic, income, job-class* are possible consumers attributes, whereas *zip-code, location, hours-of-day, days-of-week* are some possible attributes associated with ABC stores.

[SHA2008, p.76]
Vertices are not less important than edges. In fact, actors are the main element of a network; without actors a network is meaningless. Actors play a significant role in determining the dynamics of a network. In this section, I will introduce
a technique to estimate missing vertices (nodes) in a network. The method is again based on covariate information for vertices (actors) rather than edges, and utilizes the line space of edges which becomes the space of vertices as discussed
in section (2.9).

~~In optimization theory, maximizing a problem in the dual space is equivalent to; and sometimes tends to be more feasible than, minimizing it in the original space.~~

In the line space of graphs, vertices become edges and edges become vertices. Consequently, to estimate a missing vertex in the space of

graphs, it suffices to estimate the missing edge corresponding to that vertex in the line space of graphs. In this

[SHA2008, p.76]
regards, I use a mapping to transform from the space of graphs to the line space. Because graphs and matrices are isomorphic (one-to-one and onto), there is a function (transformation) that takes the graph and transforms it from the original space onto the line space and vice versa using matrices. In this sense, the matrix is the operator.

[WEG2009, p.14]

Definition 2.10. The *line graph* of $G(V, E)$ also known the *edge graph*, denoted G^l, is a graph satisfying the following criteria:

1. Each vertex of $G^l(V^l, E^l)$ represents an edge of $G(V, E)$.

[WEG2009, p.15]

2. Two vertices v_i^l and v_j^l of $G^l(V^l, E^l)$ are adjacent; i.e. $(v_i^l, v_j^l) \in E^l$ if and only if their corresponding edges are adjacent in $G(V, E)$.

The line graph is intersection graph of the edges of $G(V, E)$, it represents the adjacencies between edges of $G(V, E)$.

van Rooij and Wilf (1965) showed that if G is connected, the sequence $G, G^l, (G^l)^l, \left((G^l)^l\right)^l, \cdots$ of line graphs have four possible behaviors:

1. If G is a cycle graph, C_n, then $G^l$and each subsequent line graph is isomorphic to G itself. Cyclic graphs are the only connected graphs for which $G^l = G$.
2. If G is a claw $K_{1,3}$, then G^l and all subsequent line graphs are C_3.
3. If G is a path graph P_n, then each subsequent line graph is a shorter path P_{n-1} until eventually P_0 terminates with an empty graph.
4. In all remaining cases, the sizes of the line graphs increase without bound.

The idea then is to convert G(V, E) into $G^l(V^l, E^l)$, and then apply the method for inferring edges to $G^l(V^l, E^l)$. *We propose to investigate the feasibility of inferring missing nodes by using the duality principle of line graphs and the procedures for inferring missing edges described above.*

[SHA2008, p.67]

2.9 Line Graphs

Definition 2.9.1. The line graph of $G(V, E)$ also known the edge graph, denoted G^l, is a graph satisfying the following criteria:

1. Each vertex of $G^l\left(V^l, E^l\right)$ represents an edge of $G(V, E)$.
2. Two vertices v_i^l and v_j^l of $G^l\left(V^l, E^l\right)$ are adjacent; i.e. $\left(v_i^l, v_j^l\right) \in E^l$ if and only if their corresponding edges are adjacent in $G(V, E)$.

The line graph is intersection graph of the edges of $G(V, E)$, it represents the adjacencies between edges of $G(V, E)$.

van Rooij and Wilf (1965) showed that if G is connected the sequence $G, G^l, \left(G^l\right)^l, \left(\left(G^l\right)^l\right)^l$ of line graphs have four possible behaviors:

1. If G is a cycle graph C_n then G^l and each subsequent line graph is isomorphic to G itself. Cyclic graphs are the only connected graphs for which $G^l \cong G$.
2. If G is a claw $K_{1,3}$, then G^l and all subsequent line graphs are C_3.
3. If G is a path graph P_n, then each subsequent line graph is a shorter path P_{n-1} until eventually P_0 terminates with an empty graph.
4. In all remaining cases, the sizes of the line graphs increase without bound.

[WEG2009, p.15 cont]

3 COMPUTER NETWORKS

The concepts that are applied to social networks can also be applied to computer networks. If the computers (or routers) are nodes (actors), then their edges correspond to direct connections. Thus the node degree of a computer (or router) is just the number of other computers directly connected to it. Computer networks are generally designed (or perhaps not designed) on an evolving basis, with new connections being made on a greedy basis. Like human social networks computer networks tend to be scale free

A scale-free network is a network whose degree distribution follows a power law. The fraction *P(k)* of nodes in the network having *k* connections to other nodes is, for large values of *k*, P(k) ~$k^{-\gamma}$ where γ is a constant whose value is typically in the range $2 < \gamma < 3$This means a few computers (nodes) will have a high node degree, most likely servers, while many computers will have a low node degree, most likely clients. It is our premise that neither situation is particularly good from a network security perspective. A very high degree node, a node having high *centrality* and especially

[WEG2009, p.16]

betweenness centrality, is vulnerable because it is likely a target of a *denial-of-service attack*. On the other hand, a node with low degree is unimportant in the network could easily be the target of a *trojan* and could be made to function as a *zombie*. A low degree node is unlikely to have much attention from system administrators, and consequently could be spewing out spam email or worse yet private and classified information. Finally an unimportant node could be used by a threatening insider.

3.2 NETWORK ASSESSMENT METRICS

We describe and use the following metrics in assessing and improving network robustness and efficiency:

3.2.1 Average Shortest Path Length (ASPL) - ASPL can be used as an indicator of the efficiency of the network and is the average of all of the shortest paths from every node to the other connected nodes. This means that the smaller the ASPL, the more efficient the network may be.

3.2.2 Network Diameter (ND) - Network diameter can be used as an indicator of the efficiency factors of a network and is the longest of all of its shortest paths. This could be defined as the "maximum of all of the minimum paths" (MaxMin).

Generally, the size of a network's diameter is indicative of how spread out the network may be, and the more spread out the network is, potentially, the less efficient it may be.

3.2.3 Average Node Degree (AND) and Average Node Degree Squared (ANDS) - ANDS represents the average of the squared values of the degrees of all of the nodes on the network. Typically, one would use the average of the node degrees as an indicator. This value, however, will never vary if the total number of links on the network (Network Link Budget) remains constant. As this will be the case in some of our proposed research and evaluations, we opt to use the squared values of the node degrees as opposed to the actual node degrees.

[REZ2009, p46]

A scale-free network is a network whose degree distribution follows a power law. The fraction *P(k)* of nodes in the network having *k* connections to other nodes is, for large values of *k*, P(k) ~$k^{-\gamma}$ where γ is a constant whose value is typically in the range $2 < \gamma < 3$.

[REZ2009, p35]

3.4. Network Assessment Metrics

Based on these definitions, I describe and use the following metrics in assessing and improving network robustness and efficiency:

• Average Shortest Path Length (ASPL) – ASPL can be used as an indicator of the efficiency of the network and is the average of all of the shortest paths from every node to the other connected nodes. This means that the smaller the ASPL, the more efficient the network may be.

…(text and equation)…

• Network Diameter (ND) – Network diameter can be used as an indicator of the efficiency factors of a network and is the longest of all of its shortest paths. This can be defined as the "maximum of all of the minimum paths" (MaxMin):

ND = Max [Min (all SPLs)]

Generally, the size of a network's diameter is indicative of how spread out the network may be, and the more spread out the network is, potentially, the less efficient it may be.

• ~~Degree Standard Deviation (DSD~~) *(text+equation) this just was reordered elsewhere*

• Average Node Degree (AND) and Average Node Degree Squared (ANDS) – ANDS represents the average of the squared values of the degrees of all of the nodes in the network. Typically, one would use the Average Node Degree (AND) as ~~a measurement or~~ an indicator. This value, however, will never vary if the total number of links and nodes on the network (Network Link Budget) remains constant. As this is the case in most of the research in this dissertation, the AND will not be a useful indicator. As a result, I opted to use the squared values of the node degrees as opposed to the actual node degrees in order to be able to capture differences in the node degree centralizations between a given network and a revised instance of such a network: ANDS = SSND / *N* where SSND is the Sum of the Squares of the Node Degrees and *N* is the total number of nodes on the network

[WEG2009, p.16 cont]
Large values of ANDS may be indicative that the network is vulnerable, since the network may take a significant hit if the nodes with high connectivity were to be removed or Lost. Large values of ANDS could also be an indication that the network is "over-connected" and there are excess links, and therefore it is not a very efficient network. As a result, ANDS may be used as a criterion for robustness, as well as one for efficiency of networks.

[WEG2009, p.17]
3.2.4 Degree Standard Deviation (DSD) - DSD represents the standard deviation of all the degrees of all the nodes in the network and is an indicative of variability of the degree densities among the nodes.

3.2.5 Average Node Betweenness (ANB) - ANB represents the average (mean) of the betweenness values for all of the nodes on the network. Node Betweenness represents the number of times a node appears in the set of all the shortest paths that connect all of the nodes of the network.

3.2.6 Highest Link Betweenness (HLB) - HLB represents the highest value of the betweenness from among the betweenness values for all of the links on the network.

3.2.7 Network Degree Centralization (NDC) - Degree centrality measures how 'concentrated' the degree centralities of the actors are in the network.

3.2.8 Network Closeness Centrality Mean (NCCM) - NCCM represents the mean of the closeness values for all of the nodes on the network.

3.2.9 Average Node-Based Disconnection Ratio (ANBDR) -Node-Based Disconnection Ratio measures the ratio of the network disconnection caused by the removal of a node on the network. ANBDR is the average of all of the disconnection ratios caused by removing a node on the network
This is a measure of robustness of the network.

3.2.10 Average Link-Based Disconnection Ratio (ALBDR) - Link-Based Disconnection Ratio measures the ratio of the network disconnection caused by the removal of a link on the network. ALBDR is the average of all of the disconnection ratios caused by removing a link on the network.
Similarly, this is also a measure of robustness.

3.3 NETWORK OPTIMIZATION
As indicated above, our premise is that nodes that are too central are vulnerable and thus make the network vulnerable. Likewise, nodes that are of low centrality are also

[WEG2009, p.18, 19, 20, most of 21 omitted: no antecedents found, via quick searches in Epstein&Axtell(1996) and North&Macal(2007). The writing is choppy, but Wegman should have known this material, and the proposal was written fairly quickly.]

[REZ2009, p37]
Large values of ANDS may be indicative that the network is vulnerable, since the network may take a significant hit if the nodes with high ~~volume of~~ connectivity were to be removed or lost. Large values of ANDS could also be an indication that the network is "over-connected" and there are excess links, and therefore it is not a very efficient network. As a result, ANDS may be used as a criterion for robustness, as well as one for efficiency of networks.

[REZ2009, p35] *(this was just re-ordered)*
• Degree Standard Deviation (DSD) – DSD represents the standard deviation of all the degrees of all the nodes in the network and is an indicative of variability of the degree densities among the nodes: *(equation)*
[REZ2009, p37]
• Average Node Betweenness (ANB) – ANB represents the average of the betweenness values for all of the nodes in the network

• Highest Link Betweenness (HLB) – HLB is the largest of the betweenness values for all of the links in the network.

• Network Degree Centralization (NDC) – Degree centrality measures how 'concentrated the degree centralities of the actors are in the network.

• Network Closeness Centrality Mean (NCCM) – NCCM represents the mean of the closeness values for all of the nodes in the network
[REZ2009, p38]
• Average Node-Based Disconnection Ratio (ANBDR) – Node-Based Disconnection Ratio measures the ratio of the network disconnection caused by the removal of a node on the network. ANBDR is the average of all of the disconnection ratios caused by removing a node on the network.

• Average Link-Based Disconnection Ratio (ALBDR) –Link-Based Disconnection Ratio measures the ratio of the network disconnection caused by the removal of a link on the network. ALBDR is the average of all of the disconnection ratios caused by removing a link on the network.

*The use of "**actors**" is odd here, although Wegman and students have elsewhere sometimes applied SNA terms to computer networks in place of standard network terms, such as nodes.*

[WEG2009, p.21]

An agent-based model (ABM) is a computational model for simulating the actions and interactions of autonomous individuals in a network, with a view to assessing their
[WEG2009, p.22]
effects on the system as a whole. It combines elements of game theory, complex systems, computational sociology, multi-agent systems, and evolutionary programming. Monte Carlo Methods are used to introduce randomness.
Two informative books in the area are North and Macal (2007) and Epstein and Axtell (1996). The latter book was among the very first large-scale attempt to create agent-based models.

The next section likely comes originally from Brush(1967) or Binder(2001), or somewhere else, but I did find a copy that precedes [WEG2009], so it is unlikely that Wegman wrote this. The inclusion of unquoted text is certainly consistent. Axtell was on Sharabati's committee, so that may be a connection.

[WEG2009, p.22 (most), 23, 24 (most) skipped, no antecedents seen]

[WEG2009, p.24] ***Very likely original(s) were Brush or Binder.***
5.4 SUMMARY
Agent-based modeling does not have a strong mathematical and statistical foundation. One interesting connection to agent-based modeling is the Ising model, which has a fairly well-developed mathematical/statistical underpinning.
The Ising model, named after the physicist Ernst Ising, is a mathematical model in statistical mechanics. See Brush (1967) and Binder (2001).
It has since been used to model diverse phenomena in which bits of information, interacting in pairs, produce collective effects.

The Ising model is defined
[WEG2009, p.25]
on a discrete collection of elements (agents) with variables called spins, which can take on the value 1 or -1. The spins, s_i, interact in pairs, with energy that has one value when the two spins are the same, and a second value when the two spins are different.

While the Ising model is a simplified microscopic description of ferromagnetism, it is still extremely important because of the universality of the continuum limit. Universality means that the fluctuations near the phase transition are described by a continuum field with a free energy or Lagrangian which is a function of the field values. Just as there are many ways to discretize a differential equation, all of which give the same answer when the lattice spacing is small, there are many different discrete models that have the exact same critical behavior, because they have the same continuum limit.

Wikipedia, Agent-based model, 10/31/07 (first appearance of text)
en.wikipedia.org/w/index.php?title=Agent-based_model&oldid=168343713
26 September 2008
en.wikipedia.org/w/index.php?title=Agent-based_model&oldid=241095322 09/26/08
An agent-based model (ABM) is a computational model for simulating the actions and interactions of autonomous individuals in a network, with a view to assessing their

effects on the system as a whole. It combines elements of game theory, complex systems, emergence, computational sociology, multi agent systems, and evolutionary programming. Monte Carlo Methods are used to introduce randomness. …
Joshua M. Epstein and Robert Axtell developed the first large-scale ABM, the Sugarscape, to simulate and explore the role of social phenomenon such as seasonal migrations, pollution, sexual reproduction, combat, and transmission of disease and even culture.

It is possible that the beginning words were picked up from something Wegman and/or Said wrote, but more likely, they were derived from someone well-published in the ABM field. By late 2008, Said and Wegman had written some ABM papers, so history of antecedents in unclear. This is minor, but odd.

English text at Chinese website, 12/30/08
zh-yue.wikipedia.org/w/index.php?title=User:Hillgentleman/i_sing&oldid=234927

The Ising model, named after the physicist Ernst Ising, is a mathematical model in statistical mechanics.
It has since been used to model diverse phenomena in which bits of information, interacting in pairs, produce collective effects.

en.wikipedia.org/w/index.php?title=Ernst_Ising&oldid=273634420 02/27/09[706]
The Ising model is defined

on a discrete collection of elements (agents) with variables called spins, which can take on the value 1 or -1. The spins, s_i, interact in pairs, with energy that has one value when the two spins are the same, and a second value when the two spins are different.
English text at Chinese website, 12/30/08
While the Ising model is a simplified microscopic description of ferromagnetism, it is still extremely important because of the *universality* of the continuum limit. Universality means that the fluctuations near the phase transition are described by a continuum field with a free energy or Lagrangian which is a function of the field values. Just as there are many ways to discretize a differential equation, all of which give the same answer when the lattice spacing is small, there are many different discrete models that have the exact same critical behavior, because they have the same continuum limit.

[706] Of course, this followed [WEG2009], but *likely* was copied from elsewhere.

[WEG2009, p.25 cont]

The Ising model undergoes a phase transition between an ordered and a disordered phase in 2 dimensions or more. In 2 dimensions, the Ising model has a strong/weak duality (between high temperatures and low ones) called the Kramers-Wannier duality. The fixed point of this duality is at the second-order phase transition temperature.

The experimentally observed critical fluctuations of ferromagnets near the Curie point and of fluids at the vapor/liquid critical point are described exactly by the critical fluctuations of the Ising model. The same is true for the simplest statistical models in three dimensions whose fluctuations can be described by a single scalar field, the local magnetization in a near-critical magnet or the local density in a near-critical fluid. All these systems have fluctuating clusters whose fractal scaling laws and long distance correlation functions are quantitatively predicted by the model. Apart from the continuum limit, many discrete systems can be mapped exactly or approximately to the Ising system. The grand canonical ensemble formulation of the lattice gas model, for example, can be mapped exactly to the canonical ensemble formulation of the Ising model. The mapping allows one to exploit simulation and analytical results of the Ising model to answer questions about the related models.
The Ising model can be thought of as a Markov random field on a square grid, where the maximal graph cliques are edges (i.e. pairs of neighboring vertices).

While a very simplified version of an ABM, the Ising model clearly has elements that have local interactions based on rules for the spins and which have emergent behavior on a macroscale. The Ising model is generally constructed on a lattice. So it is an extremely simplified version of what might be thought of as an ABM. Nonetheless, it is suggestive
[WEG2009, p.26]
that a more sophisticated mathematical/statistical framework could be developed for agent-based models.
We propose to develop the mathematical/statistical framework of agent-based models operating in a social network environment. Agent-based models do not have a strong mathematical foundation. However, in analogy to the Ising model, we propose to investigate the limiting behavior of agent-based models on a lattice as the lattice size shrinks.

[WEG2009, p.26 (rest), p.27 skipped –summary of proposed tasks]

The entire section on Ising models seems very strange, out of place with the rest of the proposal, which was related to work that had been done by Wegman's students. This text seems very unlikely to have been written originally by Wegman or his students.

English text at Chinese website, 12/30/08

The Ising model undergoes a phase transition between an ordered and a disordered phase in 2 dimensions or more. In 2 dimensions, the Ising model has a strong/weak duality (between high temperatures and low ones) called the Kramers-Wannier duality. The fixed point of this duality is at the second-order phase transition temperature.

The experimentally observed critical fluctuations of ferromagnets near the Curie point and of fluids at the vapor/liquid critical point are described exactly by the critical fluctuations of the Ising model. The same is true for the simplest statistical models in three dimensions whose fluctuations can be described by a single scalar field, the local magnetization in a near-critical magnet or the local density in a near-critical fluid. All these systems have fluctuating clusters whose fractal scaling laws and long distance correlation functions are quantitatively predicted by the model. Apart from the continuum limit, many discrete systems can be mapped exactly or approximately to the Ising system. The grand canonical ensemble formulation of the lattice gas model, for example, can be mapped exactly to the canonical ensemble formulation of the Ising model. The mapping allows one to exploit simulation and analytical results of the Ising model to answer questions about the related models.
The Ising model can be thought of as a Markov random field on a square grid, where the maximal graph cliques are edges (*i.e.* pairs of neighboring vertices).

The End