



A joint newsletter of the Statistical Computing & Statistical Graphics Sections of the American Statistical Association

Statistical Computing & Statistical Graphics newsletter

[To keep members of our community informed of our section news.]

Welcome to Statistical Computing and Statistical Graphics newsletter

News from our Chairs: David van Dyk, Statistical Computing, and Naomi Robbins, Statistical Graphics, both provide their insights:

The rise of data science, including big data and data analytics, has recently attracted enormous attention in the popular press and has significantly affected research methodology in a wide-range of scholarly disciplines and commercial endeavors. Massive data streams are now the norm in every sector of society. Marketing and sales, online activity, transportation, climate, medical records, and myriad other human activities and physical processes are under constant surveillance. We hope this information will be beneficial, be it through increased commercial profitability, tracking and fighting contagions, improving security, or understanding and curtailing threats such as global warming. Accomplishing this requires new statistical frameworks, computational methods, and tools for interacting with massive databases.

For statisticians today's quickly evolving data-centric environment

is a call to arms for transforming our interactions with data and with computing. New collaborative efforts with communities of researchers specializing in database management and in distributed and parallel systems are needed to tackle emerging challenges. These challenges and opportunities are especially poignant for statisticians specializing in computational and graphical methods.

The American Statistical Association is well positioned to help formulate discussion around the role of statistics in data science, to navigate the way forward in this quickly evolving environment, and to provide forums for communication and collaboration among data scientists, including statisticians and non-statisticians alike. The Board of Directors of the ASA released [a statement on the role of statistics in data science](#) in September 2015 and is facilitating a number of activities that may be of interest to the membership of the Sections on



**DAVID VAN DYK,
CHAIR OF
STATISTICAL
COMPUTING
SECTION**

Table of Contents

Chair News.....	1-3
Program Chair news.....	4-5
Treasurer reports & article.....	6
Computing awards news.....	7
Council of Sections news.....	8
News from R Studio & article.....	9
Computing news from SAS.....	10-11
Announcement & Notes.....	12
List of officers.....	13

Computational Statistics and the Section of Graphical Statistics. These include conferences that aim to transcend traditional professional boundaries between communities of data scientists (the [Conference on Statistical Practice](#) and the [Women in Statistics and Data Science Conference](#)); collaboration with other professional societies to bring high-profile data scientists who would not ordinarily come to the JSM and to bring statisticians to other meetings; and to facilitate the development of statistically rich data science curricula for secondary, two-year college, and undergraduate education. A pair of recent *Amstat News* articles ([July](#), [August](#)) highlight progress in curricular development in a number of universities. In the November issue, Tommy Jones (Director of Data Science at Impact Research) explores [the identity of statistics in data science](#).

If you have ideas about how the Computing and Graphics Sections (or the ASA) can better serve our membership as we navigate data science and our role in it, don't hesitate to contact me!

Although JSM 2016 is still seven months away, planning for the worlds largest gathering of statisticians is well underway! The invited programs of both Sections are already set. You can read about the nine sessions sponsored or co-sponsored by the Computing Section in Wendy Martinez's re-

port in this newsletter. The *Journal of Computational and Graphical Statistics* will be honored for the first time with a special invited session at JSM 2016. Finally, the awards committees are now busily evaluating nominations for the *Student Paper Competition*, for the *John Chambers Statistical Software Award*, and the prestigious *ASA Computing and Graphics Award*. The winners will be honored in our annual joint section mixer in Chicago. Watch for announcements about the winners, and participants, keep your fingers crossed!

If you would like to present your work at JSM 2016, it is not too late to submit a proposal for a [topic-contributed session](#) or an abstract for a [contributed talk or poster](#). Also consider participating in a contributed "speed sessions"; you will be given a five minute slot for an oral presentation with a follow up LCD poster later in the day. The deadline for topic-contributed session proposals is January 14, 2016 and for contributed submissions is February 1, 2016. We are also looking for volunteers to chair technical sessions. If you are interested, please email Wendy Martinez by February 12, 2016 at mar-tinez.wendy@bls.gov.

David van Dyk
Section on Statistical Computing
Section Chair (2015)



NAOMI ROBBINS, CHAIR OF STATISTICAL GRAPHICS SECTION

It's hard to believe that 2015 is almost over. The Section on Statistical Graphics had a successful year. Our invited sessions at JSM were well attended and well received. Thanks to our Program Chair, speakers, organizers, and chairs of all our sessions. The invited papers for the 2016 JSM are already submitted but now is the time to submit proposals for topic contributed sessions, contributed sessions, and poster sessions.

JSM is not the only opportunity to present your work and network. The [Conference on Statistical Practice \(CSP\)](#) is a smaller conference that emphasizes applications. The 2016 conference to be held in San Diego in February has a number of short courses and sessions scheduled on statistical graphics. The subject of both full-day short courses is graphics. Garrett Golemund of R Studio will present *Interactive Graphics and Reports with RMarkdown and Shiny* and I will be teaching a short course entitled *Communicating Data Clearly*. Peihua Qiu of the University of Florida will teach a half-day course on *Modern*

Statistical Process Control Charts and Their Use as a Tool for Analyzing Big Data. Wendy Martinez and James Wisnowski will speak at a Text Analytics session; Xan Gregg and I will speak at a Data Visualization session; Garrett Golemund, Brandon Le Beau, Emily Beylerian, and Christopher Franck will speak at sessions on Interactivity with R Shiny; and Richard Heiberger and William Cecere will speak at a Graphical Design and Software session. In addition, Till Bergmann and Rhonda Crate have Visualization in the title of their talks in sessions that are not exclusively on graphics. Ed Mulrow will be showing a poster on visualization.

We welcome volunteers who want to become more involved with the chapter. One way to volunteer is to organize sessions for JSM. Another is to help with guidance on avoiding misleading graphs ([see separate article](#)). Still another is to let officers know of positions you'd like to hold. Thanks to Di Cook for her work as secretary and treasurer, for always being available to give advice, answer questions, or do whatever needed to be done. Thanks to Anushka Anand for her excellent work as newsletter editor. Thanks also to Hadley Wickham, our past Chair; and to Regina Nugent and Mario Morales, our section reps. Best wishes to Michael Kane, our incoming chair and Xie Yihue, our incoming program chair.

Statistical Computing Program Chair Report (*Wendy Martinez*):

Now is the time to get ready for the Joint Statistical Meetings 2016 to be held in Chicago from July 30 – August 4, 2016. The theme of JSM 2016 will be “The Extraordinary Power of Statistics.” We have an exciting line-up of invited sessions scheduled, as you can see below.

Sunday	Making the Most of R Tools
Monday	Tricks and Treats in Classification and Regression Trees
Tuesday	Robust Variable Selection Approaches in a Large Model Space
Wednesday	Uncertainty Estimation for Massive Data Sets Approaches for Text Analysis
Thursday	Julia for Modern Statistical Computing

We also have three more invited sessions on the schedule that were submitted to our section. Two other sections graciously agreed to co-sponsor them. These sessions are listed here.

Wednesday

Reproducibility in Statistics and Data Science
(*Section on Defense and National Security*)

Computing for Non-Linear Methods with Large Datasets (*Government Section*)

Thursday

Analysis, Storage, and Privacy for Big Data
(*Section on Defense and National Security*)

As always, please check the JSM online program for any updates and details on the sessions, as schedules can change. Now, we need to work on the contributed program.

CALL FOR VOLUNTEERS TO CHAIR CONTRIBUTED SESSIONS

For those going to the 2016 JSM, I am inviting you to serve as a contributed session chair. It is a great way to network and to support your section. To volunteer, please email me by February 12, 2016 at martinez.wendy@bls.gov.

CALL TO SUBMIT CONTRIBUTED SESSION ABSTRACTS

We encourage you to organize a topic-contributed session or to submit a contributed abstract for the 2016 Joint Statistical Meetings (JSM). There are two types of contributed sessions: *regular* and *topic-contributed*. The submission period of **proposals for topic-contributed sessions** (paper or panel) opened on December 1, 2015 and will close on January 14, 2016. For more information on topic-contributed sessions, refer to: <http://www.amstat.org/meetings/jsm/2016/topiccontributed.cfm>. Note that **proposals** for topic-contributed sessions must be submitted online by the deadline.

TOPIC-CONTRIBUTED PAPER SESSIONS

These sessions consist of a collection of contributed talks and discussions (if desired) that share a common topic. There are three (3) format options for a Topic-Contributed paper session:

1. Five (5) papers
2. Four (4) papers and one (1) discussant
3. Three (3) papers and two (2) discussants

Once the session is officially accepted, all speakers in a Topic-Contributed paper session **MUST** submit an abstract to the ASA office by **February 1, 2016**. On the abstract form, speakers should indicate that they are speaking in a Topic-Contributed session and provide the name of the session organizer.

Statistical Computing Program Chair Report (*Wendy Martinez*) - continued :

Statistical Graphics Program Chair Report (*Matthew S. Shotwell*):

TOPIC-CONTRIBUTED PANEL SESSIONS

The traditional panel discussion format is used, wherein a minimum of three and a maximum of five panelists provide commentary on a particular topic. Panel sessions each require **ONE ABSTRACT** submitted by the organizer and must be received by **February 1, 2016**. The Topic-Contributed panel organizer must coordinate efforts through the appropriate JSM Program Committee member.

Key Dates for JSM 2016 (check JSM 2016 website at www.amstat.org for current information)

December 1, 2015 – February 1, 2016

Online submission of abstracts, invited posters, introductory overview lectures, topic-contributed, and contributed abstracts open.

January 14, 2016

Online submission of topic-contributed session proposals deadline

March 31, 2016 – April 18, 2016

Online abstract editing open

May 2, 2016

Registration and housing open

May 17, 2016

Draft manuscript deadline

June 1, 2016

Early registration deadline

June 2, 2016 – June 30, 2016

Regular registration

The Section on Statistical Graphics (SSG) sponsored several invited and contributed sessions at the Joint Statistical Meetings in Seattle, which focused on interactive graphics technology, graphical software frameworks, and visualizing complex neuroimaging data.

In addition to the SSG program, statistical graphics played a prominent role in other parts of the program, including the Jackie Dietz Best *Journal of Statistics Education* paper award, which was presented to Jennifer Kaplan (University of Georgia) and colleagues for their paper titled "Investigating Student Understanding of Histograms".

As the activities associated with JSM 2015 conclude, I would like to welcome the incoming SSG program committee representative, Yihui Xie (RStudio), whose contributions include [knitr](#) and the '[animation](#)' R package, among others.

We look forward to Yihui's work on the JSM 2016 program, which is themed "The Extraordinary Power of Statistics."

**Statistical Computing Secretary/
Treasurer Report (Tim Hesterberg):**

The Statistical Computing budget is stable; our big expenses are food at the mixer at the Joint Statistical Meetings (JSM), and student paper awards. We increased the award for student paper winners at JSM to \$800 each for up to three winners. Our balance is about \$21K.

If you have ideas for ways the section could help members (using funds or otherwise), please contact a section officer.

**Statistical Graphics Secretary/
Treasurer Report (Dianne Cook):**

The graphics section finances are very healthy. We sponsored two student paper awards this past year, each being \$800 to support travel to the JSM.

We are interested in spending the reserves down some. Ways that it is possible to get help with graphics work are:

- (1) Color printing costs for accepted articles in a major journal, to the amount of \$2500
- (2) Help for travel expenses for invited speakers external to the ASA, and also by request for organizing a topic contributed session at JSM
- (3) Support for organizing a local graphics meetup group, on the order of \$200, to cover some food and beverage related costs.
- (4) Other requests will be considered

ASA Statement on Misuses of Statistical Graphics (Naomi Robbins):

The ASA Scientific and Public Affairs Committee (SPA) is concerned about the recent misuses of statistical graphics (specifically, the Planned Parenthood graphic and the VA graphic from last year). These figures are available from a recent blog post by [Dr. George Cobb](#).

The SPA sent a draft statement to the ASA Board of Directors, suggesting that the Board issue the statement strongly endorsing best practices around communicating information through graphics and visualizations. The statement is on hold right now since the Board would like it to include specific guidance.

Please send an email to Naomi Robbins, naomi@nbr-graphs.com if you are interested in joining a committee to co-operate with the SPA to assist policymakers and the public with understanding how to create displays that communicate statistical information most effectively.

**Statistical Computing Awards Chair
(Patrick Breheny):**

The Statistical Computing and Statistical Graphics Sections of the ASA are co-sponsoring three separate competitions this year: [the student paper competition](#), [the John M. Chambers Statistical Software Award](#), and [the Statistical Computing and Graphics Award](#). The first two competitions are annual, while the third award is, in principle, bi-annual although has not actually been awarded since 2010.

The [student paper competition](#), which rewards original work in the area of methodological research or some other novel computing or graphical application in statistics, carries with it a cash award of \$1,000. The competition is open to anyone who is a student (graduate or undergraduate) on or after September 1, 2015. This year we anticipate awarding four prizes, with winners coming from both the computing and graphics sections (last year, each section awarded two prizes).

The [Chambers Award](#), in contrast to the student paper competition, is an award for the development of statistical software -- nominations must include the software, and judges try out the software to determine how useful it will prove to the statistical and scientific communities. Previous winners include Deepayan Sarkar (lattice), Had-

ley Wickham (ggplot), and Michael Kane (bigmemory). The award carries with it a cash prize of \$1,000, and is open to anyone who is either currently a student or completed her/his last degree after January 1, 2015.

Finally, the most prestigious of the awards, the [Statistical Computing and Graphics Award](#), is intended to recognize an individual or team for innovation in computing, software, or graphics that has had a great impact on statistical practice or research. So far, this award has only been given once, in 2010, to Robert Gentleman and Ross Ihaka for their role in the development of R. The award is open to anyone except the members of the Awards Committee.

The deadline for applications for all three competitions was December 15, 2015, with judging to be completed and winners chosen by January 15, 2016. Winners for all three competitions will be given their awards at the 2016 Joint Statistical Meetings.

Statistical Computing Council of Sections representatives' report

(Jane Harvill, Jonathan Lane, John Monahan):

Over the past several years there has been an ongoing debate about the proliferation of Sections with the American Statistical Association. On the one hand the many new sections reflect the diversity of the association, but on the other sections are sometimes perceived to be in competition with each other. [The Council of Sections \(CoS\)](#) discussed these issues at their annual business meetings at the Joint Statistics Meetings in August. This required action on a proposal for a new section that was before the council, but also involved rethinking the general procedures for forming and maintaining a section.

The [Lifetime Data Analysis \(LIDA\)](#) Interest Group of ASA petitioned the [Council of Sections](#) Governing Board to become a Section of the ASA. This proposal was discussed at the JSM, but the majority of the CoS membership voted in September not to approve the new section. LIDA will continue as an Interest Group within the ASA.

During the CoS Business Meetings, the proposal for the LIDA section prompted a discussion on the general requirements for creating a new section and for remaining a section within ASA. The council decided to increase the size threshold for Sections. Given that the ASA has grown overall as an association, it was decided that it is sensible for the minimum Section

membership to be larger as well. CoS also decided that any any proposed section must first function as an Interest Group for at least three years before becoming a Section. In the past Sections could be proposed without first being an Interest Group. Specifically, the CoS voted to amend the CoS Charter to the following.

1. Regarding the formation of new Sections, Article IV, paragraph 2 to include, "A Section may be chartered by a majority vote of the Council of Sections based upon evidence of 3 years of experience as an interest group with a membership of at least 50 full members each year, a summary of accomplishments during the Interest Group period, a petition of at least 200 full members, a Proposed Section Charter, and a list of officers."
2. Regarding the dissolution of existing Sections, Article IV, paragraph 3 to include, "Failure to maintain at least 200 dues-paying members for two consecutive years."

R Consortium News from RStudio (Hadley Wickham)

In the last twelve months, probably the biggest news for the R community is the creation of the [R Consortium](#). The goal of R Consortium is to provide a way for companies that have built their businesses on top of R to contribute back to the community. Some of the current members are RStudio, Microsoft, TIBCO Analytics, Google and Oracle. You can see a complete membership on the [R Consortium homepage](#).

One of the ways that the R Consortium gives back to the R community is through the [Infrastructure Steering Committee](#). I'm the chair of the infrastructure steering committee, and our responsibility is to find projects that widely benefit the R community and help support their development. One of the first ways we're doing this is through an open call for [funding](#). Proposals need to outline how they will broadly benefit the R community, and are due on January 10. If you think you have a project that would be a good fit, please reach out to me if you have any questions.

Hadley Wickham
Chief Scientist, Rstudio

Conferences and Discount Rates (Naomi Robbins)

Once again, [O'Reilly](#) is offering us discounts for their conferences. [Strata + Hadoop World](#) is the leading event on how big data and ubiquitous, real-time computing is shaping the course of business and society. It brings together the world's best data scientists and business leaders to share hard-won knowledge and innovations in technology and strategy. Check out the impressive program and make plans to join [Strata + Hadoop World](#) in San Jose March 28-31, 2016. Save 20% on most passes with discount code UGDVZNY20.

Mary Treseler, director of strategic content for Design at O'Reilly, explains that in today's fast-changing economy, designers increasingly shape businesses, products, and services. To work at their fullest potential, designers need to know more about more, working across teams and disciplines.

She developed the O'Reilly Design Conference to help designers gain the skills they need to make the best impact on the world, in business and society. It will be held in San Francisco January 19 -22, 2016. Registration is now open; see pricing and packages.

You can use **discount code PCDVZSF20** to save 20% on registration for the [O'Reilly Design Conference](#).

Computing News from SAS (Rick Wicklin):

In July 2015, SAS introduced the 14.1 release of analytical products such as SAS/STAT® software. SAS/STAT 14.1 includes two new procedures and many enhancements, including the following:

- The new SURVEYIMPUTE procedure provides missing value imputation for survey data.
- The new GAMPL procedure fits generalized additive models via penalized likelihood estimation. Figure 1 shows an example of graphical output from the GAMPL procedure.
- The HPGENSELECT procedure supports LASSO selection for building generalized linear models. This extends the model-building capability of the GLMSELECT procedure, which has supported the LASSO, adaptive LASSO, and elastic net methods for linear models for several releases.
- The HPSPLIT procedure provides classification and regression trees via a standard modeling syntax.
- The MCMC procedure for Monte Carlo Markov chain analysis now supports the Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS) sampling algorithms for continuous parameters. It also supports leading and lagging variables, an ordinary differential equation solver, and a general integration function.
- The IRT procedure (for item response theory models) fits generalized partial-credit models.
- The POWER procedure supports power and sample size analyses for

Cox proportional hazards regression models for survival data.

For more information, see the video [“Highlights of SAS/STAT 14.1,”](#) which is one of [100 videos about how to use SAS analytical software.](#)

If you are a student or a teacher, or you just want to increase your knowledge of SAS programming, SAS provides the free [SAS® University Edition](#), which contains Base SAS®, SAS/STAT, SAS/IML®, several time series forecasting procedures, and SAS/ACCESS® Interface to PC Files. SAS University Edition is available for Windows, Mac, and Linux systems.

To stay up to date with SAS activities throughout the year, subscribe to the bi-monthly [SAS Statistics and Operations Research Newsletter.](#)

Rick Wicklin, SAS
Rick.Wicklin@sas.com

Computing News from SAS (continued):

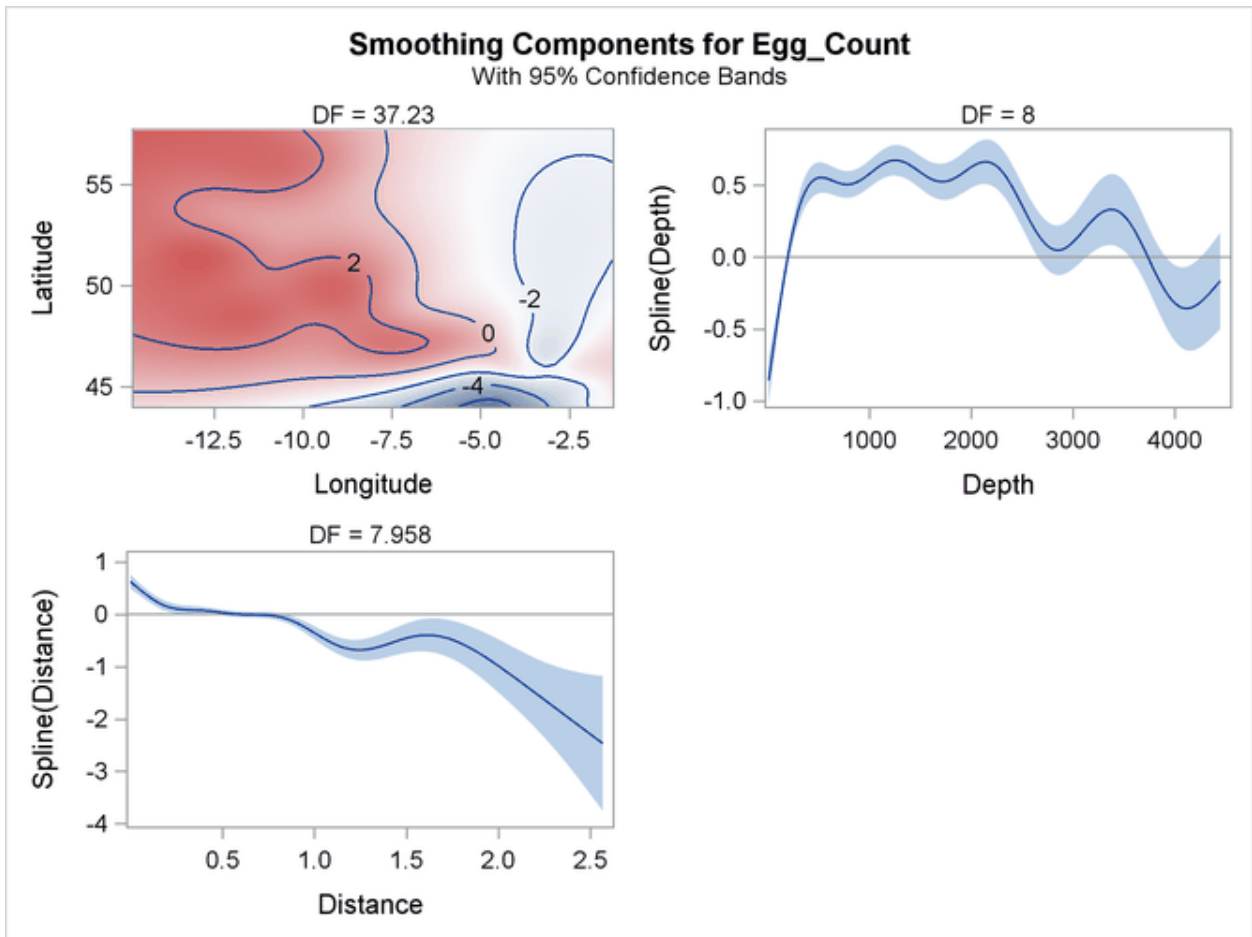


Figure 1: Nonparametric Negative Binomial Model for Count Data Fit by Using PROC GAMPL

ANNOUNCEMENTS:

The [Government Statistics Section](#) is sponsoring Data Challenge 2016 to take place at the JSM 2016 meetings. The contest is open to anyone who is interested in participating, including college students and professionals from the private or public sector. This contest challenges participants to analyze a government data set using statistical and visualization tools and methods. There will be two award categories – Professional and Student. One award in the amount of \$500 will be given for entries in the Professional category. Three awards will be given for entries in the Student category. These are \$500 for first place, \$300 second place, and \$200 for third place.

The data set for the [GSS Data Challenge 2016](#) will be the Department of Transportation's General Estimates System (GES). Contestants are free to use any or all of the data in their analysis. Merging data from the GES with other data sets is also encouraged. A link to the website with the data and other useful information (data dictionaries, manuals, etc.) is given [here](#). Contestants must submit an email to Wendy Martinez at: (martinez.wendy@bls.gov) by January 12, 2016 to enter as the date has been extended. The information required in the email and further details can be found on the [GSS website](#).

Notes and Acknowledgements:

The newsletter was put together by Section on Statistical Computing Publications Officer, Usha Govindarajulu, with assistance and contribution from Section on Statistical Graphics Publications Officer, Anushka Anand. Thank you to all the Chairs and Officers of both sections for their contribution to this newsletter via officer reports, news, and announcements.

If you would like to make future contributions to our newsletter or have any questions, please free to contact the publication officers or any other section officer listed on the last page of this newsletter. Thank you.

Statistical Computing Section Officers 2015

David A Van Dyk, Chair
d.van-dyk@imperial.ac.uk
 +44 20 7594 8574

David Poole, Chair-Elect
poole@research.att.net
 (973) 360-7337

Tim Hesterberg, Secretary / Treasurer
timhesterberg@gmail.com
 (206) 876-1790

Jane Harvill, COS Representative 2013-2015
Jane_Harvill@Baylor.Edu
 (254) 710-1517

Jonathan Wesley Lane, COS Representative 2015-2017
iono.lane@gmail.com
 (715) 252-2103

John Francis Monahan, COS Representative 2014-2016
monahan@ncsu.edu
 (919) 515-0622

Feng Liang, Program Chair 2015-2015
iangf@illinois.edu
 (217) 333-6017

Martinez, Wendy L., Program Chair –Elect 2016-2016
martinez.wendy@bls.gov
 (202) 691-7400

Usha Govindarajulu, Publications Officer 2014-2016
usha@alum.bu.edu
 (718) 221-6599

Patrick Breheny, Awards Chair
patrick-breheny@uiowa.edu
 (319) 384-1584

Stacy Lindborg, COS Gov Board 2014-2016
 (see right)

Rick Peterson, ASA Staff Liaison (see right)

The Statistical Computing & Statistical Graphics News-
 letter is a publication of the Statistical Computing and
 Statistical Graphics Sections of the ASA.

Statistical Graphics Section Officers 2015

Naomi Robbins, Chair
naomi@nbr-graphs.com
 (973) 694-6009

Michael Kane, Chair-Elect
michael.kane@yale.edu
 (203) 737-4768

Dianne Cook, Secretary / Treasurer
dicook@monash.edu
 +61 (0) 3990 52608

Matthew Shotwell, Program Chair 2015-2015
matt.shotwell@vanderbilt.edu
 (615) 875-3397

Yihui Xie, Program Chair-Elect 2015-2015
yihui@rstudio.com
 (859) 230-4326

Rebecca Nugent, Section Rep 2014-2016
rnugent@stat.cmu.edu
 (412) 268-7830

Mario Morales, Section Rep 2013-2015
emetricz@hotmail.com
 (917) 420-3433

Anushka Anand, Publications Officer 2014-2015
aanand@tableausoftware.com
 (206) 633-3400

Patrick Breheny, Awards Chair (see left)

Stacy Lindborg, COS Gov Board 2014-2016
stacy.lindborg@biogenidec.com
 (617) 914-1979

Rick Peterson, ASA Staff Liaison
[email {rick@amstat.org}](mailto:rick@amstat.org)
 (703) 684-1221

All communications regarding ASA membership and the Sta-
 tistical Computing and Statistical Graphics Sections,
 including change of address, should be sent to

American Statistical Association
 1429 Duke Street
 Alexandria, VA 22314-3402 USA
 TEL (703) 684-1221
 FAX (703) 684-2036
asainfo@amstat.org