Volume 24
December 2014

Statistical COMPUTING & GRAPHICS

[American Statistical Association Sections of Computing and Graphics]

A joint newsletter of the Statistical Computing & Statistical Graphics Sections of the American StatisticalAssocia-

Statistical Computing & 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: Michael Minnotte, Statistical Computing, and Hadley Wickham, Statistical Graphics, provide their insights:

MICHAEL MINNOTTE, CHAIR OF STATISTICAL COMPUTING SECTION

As members of the ASA Sections on Statistical Computing and Statistical Graphics, we recognize the central role that computing plays in the modern practice of statistics. How successfully do we communicate that to our students? In the rush to cover the multitude of topics that make up a modern statistics course, computation may appear as an afterthought or, worse, be left out altogether.

In introductory classes, it is common to supplement the standard homework assignments with additional problems or assignments to analyze full datasets using the software of the instructor's choice. Even in such a commonplace setting, there are many issues at play. Do you select a statistical program on the basis of ease of use, favor in your students' disciplines, personal preference for your own use, or other factors still? Do you provide all datasets in ready-to-use form, or do you make your students collect their own data for at least some assignments? If the latter, how do you ensure that good statistical practice is followed, not to mention ethical research practices? I still remember the group of students who wanted to measure pain tolerance by asking volunteers to place their hands in ice water for as long as they could stand. Somehow, I doubted that the Institutional Review Board would have been thrilled with that proposal.

More advanced classes often provide more opportunities for computation, but if we restrict ourselves to straightforward application of the methods we're teaching, we fail to realize the potential to educate our students in the benefits of statistical computing. Simulation allows us to examine our methods and assumptions in a way little else will. I've had students run simulations computing level and power for various combinations of two-sample test, distribution, and sample size. Use of heavy-tailed distributions such as the t-with 3 degrees of freedom quickly demonstrates that the two-sample t-test is quite robust to its normality assump-

Table of Contents

tion, while the F test of variance equality is very much not. Likewise, actually having students program statistical methods in the language of your (or their!) choice can force a deeper understanding of both computing and those methods in a way that simple mathematical analysis or the use of prebuilt functions may not.

If you have any innovative approaches to including and encouraging the use of statistical computing in your courses, I'd love to hear about them. Send me an email (michael.minnotte@und.edu) or, better yet, put it in a post to the section message board and email list so we can all benefit.

Changing topics, thanks to our Program Chair, Nicholas Lewin-Koh, and everyone else who made the Statistical Computing program at the 2014 Joint Statistical Meetings so successful. And I add my personal thanks to Chair-Elect David Van Dyk, Past Chair Montserrat Fuentes, and the rest of the section officers who stepped up and took over my duties when my wife and I received word shortly before JSM that the baby we were adopting was about to make an early appearance, leading to a sudden change in travel plans. Aaron Michael Minnotte is doing well, and gives his thanks also!

Our 2015 Program Chair, Feng Liang, is busy arranging an equally exciting program for the 2015 JSM in Seattle. The invited program is set, but be sure to get your contributed papers and contributed topic sessions submitted by the February 2 deadline!

Congratulations to the new members of the Section on Statistical Computing Executive Board: 2015 Chair-Elect David Poole, 2015 Program Chair-Elect Wendy Martinez, and 2015-2017 Council of Sections Representative Jonathan Lane. And many thanks for your service to the members who will be leaving the board in 2015: Montse Fuentes, Nicholas Lewin-Koh, and 2012-2014 Council of Sections Representative, Erik Iverson.

You should be receiving a ballot soon with candidates for 2015 ASA elections, including section officers; be sure to return that to make your voice heard on the future of the sections. If you'd like to get more involved in the section, running for office and serving is a great way to do that. David Poole will be looking for 2016 candidates next fall, and I'm sure he'd love to put your name on the ballot.

Finally, thanks to Usha Govindarajulu and Anushka Anand, Publications Officers for Statistical Computing and Statistical Graphics, respectively, and editors of the newsletter you're currently reading. Newsletter editor was a separate, appointed position for many years, but the boards of the two sections decided to ask the officer in the otherwise-underutilized role of Publications Officer to take on this additional responsibility for the future. Usha and Anushka have agreed to step up and I know they'll do a great job for us.

Hadley Wickham, CHAIR OF STATISTICAL GRAPHICS SECTION

It seems like only a few weeks ago that I was in Boston for JSM 2014, but it's already time to start thinking about JSM 2015, to be held in Seattle. There are three main ways that you can contribute to the graphics lineup:

- Submit a contributed paper! We'd love to learn about the exciting research that you've been working on this year.

- Don't ignore the advantages of submitting a poster session: you get to chat one-on-one with people interested in your work in a way that you don't get in a talk session. A great tip from Amelia McNamara is to print your poster on fabric. They look great, they're easy to transport and you can repurpose into nerd couture!

-

- Organise a topic-contributed session. A topic-contributed session can be either paper presentations (at least three presenters and no more than two discussants) or a panel (3–6 speakers). Organising a topic-contributed session is a great way to put the spotlight on a topic that you think is important. See the JSM website for more details.

- Host a round table lunch. This is a great opportunity to chat with a bunch of interested JSM attendees about a topic of your choice. Please get in touch with the program chair Matt Shotwell, matt.shotwell@vanderbilt.edu, if you're interested.

Would you like to have a more significant role in the section? We're always looking for volunteers who'd like to help out. Please shoot an email to me at hadley@rstudio.com if you're interested in learning more.

We're also planning on another data expo. Look for details soon!

Statistical Computing Program Chair Report: (Nicholas Lewin-Koh)

The statistical computing section had another good year at JSM. This year our invited and topic contributed sessions covered a wide breadth of topics including software developments, big data and computational topology. It is important for us as a section to represent the breadth of statistical computing.

This year we received 9 proposals, While the section is allotted 4 invited sessions, we were able to secure a 5th slot through the competition and two additional slots through the graphics session since some of our proposals were more appropriate for the graphics section. The remainder were fielded as topic contributed sessions. In total we had 5 topic contributed sessions. Sessions were well attended and generally got good feedback. We also cosponsored many sessions with other sections.

Our student presentations this year were of very high quality. Two of the students were from Iowa State University, one in the graphics category on a perceptual paradox and one from computing on lossy compression of images. The other two winners were from NC State on sparse regression and TU Berlin on visualizing continuous embeddings.

There is still a lot of ambiguity, as to what is statistical computing? Is it the theory of computing? Is it algorithms? Or is it the implementation? Of course it is all of the above, this year we were able to cater to all the aspects of statistical computing with sessions on mixed model software, hierarchical modeling, and efficient estimation in large data settings. I hope we are able to continue to cater to our diverse membership.

Statistical Graphics Program Chair (Michael Kane)

Over the last few years the technological infrastructure of graphics has become a developed web technology. With this technological change comes a change in how we present visualizations and whom we present them to.
JavaScript and its associated visualization libraries allow us to quickly create graphical representations of data that users can interact with and, with the web as a medium for dissemination, these representations routinely reach large audiences. Data visualization has become mainstream.

This is not to say that visualization is a solved problem." Technologies are still being developed, methodologies for incorporating interactivity are still being conceived, and we will continue to contend with creating elective visualizations from data. However, as these areas continue to mature it is worth considering their affect on our practice and how they foster new avenues for exploration.

One avenue that I have been watching is the growth of data exploration environments, which allow statisticians and domain experts to systematically manage, prioritize, and track data visualization for exploration and hypothesis generation. Much of this work traces back to Bell Labs in the 1980s.
Now, with new more capable technological tools, we have begun to hash-out theses ideas more fully and take these ideas even further with projects like RCloud (AT&T Labs Research, 2014), shiny (RStudio and Inc., 2014), and trelliscope (Hafen, 2014).

I believe this avenue in particular has the potential to define how data-driven science is performed.

Our Graphics Community members continue to develop new methodologies based on the very practical need to understand the structure of data. Our savvy and emphasis on demonstrating new ideas through technology make our talks some of the most exciting. And the pace of development makes me eager to see what comes next. What will the grammar of interactive graphics look like? Which tools presented next year will make us wonder, How did we explore data before this?

Acting as the JSM Program Chair for this past-years Graphics section and helping to facilitate the circulation of methodologies in graphical representations of data was tremendously rewarding. I would like to thank everyone who participated and all those who helped make the sessions such a success.

References

AT&T Labs Research (2014). RCloud: collaboratively develope share *R scripts*. URL https://github.com/att/rcloud.

Hafen R (2014). *trelliscope: Trelliscope*. R package version 0.8.0.2, URL https://github.com/tesseradata/trelliscope.

RStudio, Inc (2014). *shiny: Web Application Framework for R*. R package version 0.10.2.1, URL http://CRAN.R-project.org/package=shiny.

Treasurer's Report of Statistical Computing (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 will increase the award for student paper winners at JSM to $800 each for up to three winners. Our balance is about $20K.

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

The ASA also invites members to submit proposals for initiatives that support the Association's mission; see http://www.amstat.org/about/MemberInitiativesInstructions.cfm.

Treasurer/Awards Chair Report of Statistical Graphics (Kenneth Shirley)

Congratulations to Susan VanderPlas of Iowa State University, who won the 2014 graphics student paper competition with her paper, "The Curse of Three Dimensions: Why Your Brain Is Lying to You". Susan was awarded a $500 travel award (and her registration for JSM 2014 was paid for by the Section on Statistical Graphics). We would also like to thank our judges, John Castelloe, Erik Iverson, and Heike Hofmann, for their hard work in evaluating the submissions.

Our section is currently seeking meetups and other small meetings/conferences to sponsor if they have a statistical graphics component. Please contact the chairperson, Hadley Wickham, or the treasurer, Kenny Shirley, if you are organizing such a meeting and would like to request funding.

We are currently offering a $500 travel award to anyone who organizes a topic-contributed session for JSM 2015, to incentivize more people to get involved in our section's activities!

Awards Chair Report of Statistical Computing and Graphics

The Statistical Computing and Statistical-Graphics Sections of the ASA are co-sponsoring a student paper competition on the topics of Statistical Computing and Statistical Graphics. Students are encouraged to submit a paper in one of these areas, which might be original methodological research, some novel computing or graphical application in statistics, or any other suitable contribution (for example, a software-related project). The selected winners will present their papers in a topic-contributed session at the 2015 Joint Statistical Meetings. The Sections will pay registration fees for the winners as well as a substantial allowance for transportation to the meetings and lodging.

Anyone who is a student (graduate or undergraduate) on or after September 1, 2014 is eligible to participate. An entry must include an abstract, a six page manuscript (including figures, tables and references), blinded versions of the abstract and manuscript (with no authors and no references that easily lead to identifying the authors), a C.V., and a letter from a faculty member familiar with the student's work. The applicant must be the first author of the paper.

The faculty letter must include a verification of the applicant's student status and, in the case of joint authorship, should indicate what fraction of the contribution is attributable to the applicant. We prefer that electronic submissions of papers be in Postscript or PDF. All materials must be in English.

Students may submit papers to no more than two sections and may accept only one section's award. Students must inform both sections applied to when he or she wins and accepts an award, thereby removing the student from the award competition for the second section.

All application materials MUST BE RECEIVED by 5:00 PM EST, Sunday, December 14, 2014 at the address below. They will be reviewed by the Student Paper Competition Award committee of the Statistical Computing and Graphics Sections. The selection criteria used by the committee will include innovation and significance of the contribution as well as the professional quality of the manuscript. Award announcements will be made by January 15th, 2015.

Additional important information on the competition can be accessed on ASA's "Student Paper Competition/ Travel Award to Attend the Joint Statistical Meetings" page at: http://www.amstat.org/sections/studentpaperawards.cfm, or at the website of the Statistical Computing Section, http://www.statcomputing.org.

Inquiries and application materials should be emailed or mailed to:

Student Paper Competition
c/o Aarti Munjal
Colorado School of Public Health
University of Colorado Denver
aarti.munjal@ucdenver.edu

ASA John M. Chambers Statistical Software Award - 2015

The Statistical Computing Section of the American Statistical Association announces the competition for the John M. Chambers Statistical Software Award. In 1998 the Association for Computing Machinery presented its Software System Award to John Chambers for the design and development of S. Dr. Chambers generously donated his award to the Statistical Computing Section to endow an annual prize for statistical software written by, or in collaboration with, an undergraduate or graduate student.

The prize carries with it a cash award of $1000, plus a substantial allowance for travel to the annual Joint Statistical Meetings (JSM) where the award will be presented.

Teams of up to 3 people can participate in the competition, with the cash award being split among team members. The travel allowance will be given to just one individual in the team, who will be presented the award at JSM. To be eligible, the team must have designed and implemented a piece of statistical software. The individual within the team indicated to receive the travel allowance must have begun the development while a student, and must either currently be a student, or have completed all requirements for her/his last degree after January 1, 2014. To apply for the award, teams must provide the following materials:

Current CV's of all team members.

A letter from a faculty mentor at the academic institution of the individual indicated to receive the travel award. The letter should confirm that the individual had substantial participation in the development of the software, certify her/his student status when the software began to be developed (and either the current student status or the date of degree completion), and briefly discuss the importance of the software to statistical practice.

A brief, one to two page description of the software, summarizing what it does, how it does it, and why it is an important contribution. If the team member competing for the travel allowance has continued developing the software after finishing her/his studies, the description should indicate what was developed when the individual was a student and what has been added since.

An installable software package with its source code for use by the award committee. It should be accompanied by enough information to allow the judges to effectively use and evaluate the software (including its design considerations.) This information can be provided in a variety of ways, including but not limited to a user manual (paper or electronic), a paper, a URL, and online help to the system.

All materials must be in English. We prefer that electronic text be submitted in Postscript or PDF. The entries will be judged on a variety of dimensions, including the importance and relevance for statistical practice of the tasks performed by the software, ease of use, clarity of description, elegance and availability for use by the statistical community. Preference will be given to those entries that are grounded in software design rather than calculation. The decision of the award committee is final.

All application materials must be received by 5:00pm EST, Tuesday, February 17, 2015 at the address below. The winner will be announced in May and the award will be given at the 2015 Joint Statistical Meetings.

Inquiries and application materials should be emailed or mailed to:

Student Paper Competition
c/o Aarti Munjal
Colorado School of Public Health
University of Colorado Denver
aarti.munjal@ucdenver.edu

Brief on this year's awards

We had a great competition this year with 22 submissions, 19 in the Computing Section and 3 for the Graphics Section. We would like to acknowledge our judges for their hard work in evaluating these submissions: John Castelloe, Erik Iverson, and Heike Hofmann. We had one winner from the Graphics Section and three winners from the Computing Section, as listed below:

This year's winners are:

Computing: Gina Grünhage (Technische Universität Berlin) - Visualizing the Effects of a Changing Distance Using Continuous Embeddings
Computing: Geoffrey Thompson (Iowa State University) - An Adaptive Method for Lossy Compression of Big Images
Computing: Guan Yu (University of North Carolina at Chapel Hill) - Sparse Regression Incorporating Graphical Structure Among Predictors
Graphics: Susan Vanderplas (Iowa State University) - The Curse of Three Dimensions: Why Your Brain Is Lying to You

Council of Sections Representatives Report

Jane L. Harvill, John Monahan, & Erik Iverson

The Council of Sections (CoS) of the American Statistical Association (ASA) met twice during the Joint Statistical Meetings (JSM) in Boston. This report is a summary of discussion items of interest to the Section on Statistical Computing.
1. The Archive Committee of ASA is requesting sections send any pictures of the section chairs (past and present) to the Archive Committee chairperson, John McKenzie at mckenzie@babson.edu.
2. The ASA has a mentoring program, and encourage sections to participate.
3. A new proposed Section on Genomics and Genetics was proposed, and later approved.
The discussion on the new section spurred a lengthy discussion that entered into other, related issues. The issues that are directly related to the Section on Statistical Com-
puting are listed below.
(a) Some sections are concerned about how the addition of another section will affect the number of invited sessions at JSM. In particular, it was suggested that JSM Program Chairs handle the issue to make sure that the newer sections are not getting too many invited sessions. Related to this issue are the following comments.

Some members felt that since the number of invited sessions is increasing,
there is no reason to not create a Genomics Section.
What is the purpose of invited sessions? The quality of many has declined,
or has become so specialized, that they are not well attended except by the
select few who propose and participate in them.

(b) Some sections expressed the concern that a section on Genomics and Genetics
was too specialized, or that those areas were already represented through other
venues.

(c) Much of the discussion was about making more rigorous requirements for creating
a new section. The following items were suggested, but no motion was made.

Groups interested in becoming a section should first be required to form an
interest group.
Increase the number of signatures required for becoming a section.
In addition to signatures from ASA members, a new section must have signatures from non-ASA members who would be willing to join. (One member suggested 100 ASA member signatures and 100 ASA non-member signatures.)

(d) Rick Peterson, ASA liaison to the CoS, pointed out that fewer than one-half of
ASA members are members of a section.

4. Student paper awards was the next topic on the agenda; in particular, was the issue
of the number of awards to which a student can submit a paper. A vote was taken,
and 17 sections were in favor of allowing students to submit to at most two awards; 5
section were in favor of one.

(a) Since papers are submitted electronically, it would be preferable that ASA monitor the number of submissions by a student so that the rule is not student-enforced.

(b) Rick Peterson said that, under the current system, it would be hard to implement a monitoring system, or to generate a master list of names of students who submit applications. However, he will let the CoS know what information he needs to
get an automated system in place.

(c) ASA does know which students win awards, and should not allow a student to receive more than one award.

NEWS FROM RSTUDIO

R users are doing some of the most innovative and important work in science, education, and industry. In the past year we introduced many new and enhanced products to assist them.

RStudio IDE: Last December we issued v.98 of RStudio and throughout the year have made working with the new capabilities in Shiny, RMarkdown, Packrat easier along with many other new features.

Shiny and Shiny Server: In March we released a new website to facilitate Shiny app development: shiny.rstudio.com, and added many new features; while Shiny Server added support for interactive documents.

NEW dplyr: dplyr provides a set of tools for efficiently manipulating datasets in R.

NEW ggvis: Like ggplot2, ggvis is built on concepts from the grammar of graphics, but adds interactivity, a new data pipeline, and renders in a web browser.

NEW tidyr: tidyr makes it easy to "tidy" your data. Tidy data is easy to work with: it's easy to munge (with dplyr), visualise (with ggplot2 or ggvis) and model.

RMarkdown v2: In RMarkdown v2 we combined the flexible outputs of Pandoc with good default options and an R-friendly interface to make reproducible document generation simpler.

packrat: packrat v0.4 was released to CRAN this Fall. It manages package dependencies in order for R projects to be more isolated and reproducible.

NEW PROFESSIONAL EDITIONS

Last December we introduced **RStudio Server Pro** to enable growing deployments under a commercial license supported directly from RStudio. Features unique to RStudio Server Pro include session load balancing, R user authentication and customized session configuration

.

In February, we announced **Shiny Server Pro** to help R users share their work with colleagues and customers securely while providing IT groups the management and performance tuning features they need to deploy Shiny apps and interactive documents at scale.

These are actually only a partial list. Many of these were topics at our JSM presentations in Boston, webinars, and advanced training courses. We appreciate all the support we heard from JSM participants for the work we're doing.

Roger Oberg (roger@rstudio.com)

Computing News from SAS

Rick Wicklin, SAS, Rick.Wicklin@sas.com

For Section members who are interested in learning SAS or improving their SAS skills, SAS is pleased to provide the SAS University Edition, which is free to anyone (inside or outside higher education) for learning or for academic research. The University Edition contains Base SAS, SAS/STAT, SAS/IML®, and SAS/ACCESS® Interface to PC Files. Download it today onto your Windows, Mac, or Linux system.

Since the previous newsletter, SAS has released two versions of its analytical products:

The 13.1 release of SAS/STAT® software (December 2013) includes new procedures for Bayesian choice models, nonparametric survival models for interval-censored data, and item response theory models. New graphical enhancements include path diagrams for structural equation models and enhanced graphics for visualizing survival functions. For details about the SAS 13.1 analytical products, including links to papers that were written for SAS Global Forum 2014, see "SAS Analytics: Highlights of 13.1 Releases."

The 13.2 release of SAS/STAT software (August 2014) includes new procedures for fitting longitudinal models by using generalized estimating equations, fitting proportional hazards regression models for interval-censored data, and analyzing spatial point patterns. For details, see "SAS Analytics: Highlights of 13.2 Releases."

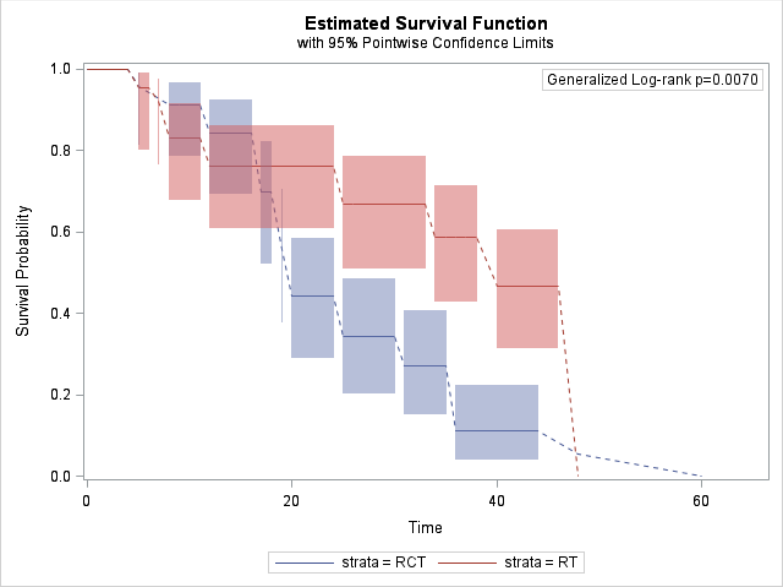
Sometimes it is hard to stay up-to-date on recent advances in SAS software. The following links might help:

The paper "SAS/STAT 13.1 Round-Up" by Bob Rodriguez and Maura Stokes provides an overview of selected features in SAS/STAT 13.1. The paper also provides a peek at SAS/STAT 13.2.

If you like to watch videos, there are dozens of 10-minute videos that each present an overview of a single topic in SAS. Recent video topics include the LASSO method for selecting regression effects, fitting multivariate adaptive regression splines, and visualizing matrices by using heat maps.

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

For a fun SAS-oriented blog that highlights statistical computation, programming, and graphics, subscribe to Rick



Wicklin's blog, The DO Loop, which is published semi-weekly.

Update from Revolution Analytics

David Smith, Chief Community Officer

For the past 7 years, Revolution Analytics has been supporting the R project and promoting the use of R for statistical analysis in business applications. In the past year, we have expanded our Revolution R product line to include two new editions: Revolution R Open and Revolution R Plus. We have also expanded our training and services offerings to introduce new on-line and in-person courses, and to offer certification for Revolution R Enterprise developers. We were also very proud to be named a Visionary in the Gartner Magic Quadrant for Advanced Analytics.

Revolution R Open

Revolution R Open is a new, free and open source distribution of R from Revolution Analytics with additional capabilities focused on performance and reproducibility. It is based on the latest release of R from the R Foundation, making it compatible with all R scripts, packages and applications that connect with R (including RStudio).

Revolution R Open is available to download from MRAN (the Managed R Archive Network). (There you can also find information about R and Revolution R Open, and explore R packages and task views.) Additional features in Revolution R Open include:

Binaries linked with the Intel Math Kernel Libraries. This brings multi-threaded computations to R, which can improve R's performance (especially for matrix operations on multi-CPU systems) without needing to modify any of your code.

The Reproducible R Toolkit, which is designed to make it easy to write and share R scripts with reproducible output, especially in the context of changing R packages. By adding the following two lines to the top of your script:

```
library(checkpoint)

checkpoint("2014-11-20")
```

Revolution R Open will ensure that all the necessary R packages are installed with versions that existed as of the specified date. This ensures that your code will continue to give the same results even after packages are updated, and that colleagues to whom you send R scripts get the same results as well. It works by drawing on an archive of daily snapshots of of all R packages that we store on MRAN.

64-bit binary distributions for Windows, Linux (Red Hat, CentOS, Debian and OpenSUSE) and Mac.

DeployR Open

Another new open-source project from Revolution Analytics is DeployR Open, which is designed to make it easy to integrate R into other applications via Web Services. DeployR Open is a server-based framework where R programmers can publish scripts, which can then be called by application de-

where R programmers can publish scripts, which can then be called by application developers writing in Java, JavaScript or .NET. You can learn more and download DeployR Open .

Open-Source Support with Revolution R Plus

Our goal is to promote the adoption of R in companies, where the biggest barriers are concerns around technical support and the use of open source licenses.

To provide companies with the confidence they need to adopt R, we now offer Revolution R Plus: an annual subscription which provides technical support and open-source assurance for R and other open-source components that works with R. Revolution R Plus covers Revolution R Open, DeployR Open, and other open-source projects described at this site.

The technical support program includes email and phone service from 8AM to 6PM in the userÕs local time zone. And with Open Source Assurance, Revolution Analytics will defend customers in court should anyone else make a copyright, patent, trademark or trade secret claim against the open source software we cover. If you know of any companies that are thinking about adopting R but have concerns about technical support or the use of open source software, they can learn more at this site.

Updates to Revolution R Enterprise

We are continuously updating Revolution R Enterprise to add more capabilities for handling and analyzing Big Data with R and for securely integrating R into business applications.

New statistical models for arbitrarily large data sets include "decision forests" (ensembles of trees), interactive decision trees, stepwise selection for GLMs, and Stochastic Gradient Boosting for boosted classification and regression trees. All of these models can now be run directly on data in Hadoop clusters, in Teradata databases, or in the Cloud via the Amazon Marketplace.

Revolution R Enterprise is sold as an annual subscription to commercial customers, but is always available free of charge to anyone in academia. Supported site-wide licenses are also available to academic institutions and non-profits for just $999 per year.

We now offer certification for Revolution R Enterprise users, with proctored exams available on -site or on-line. In addition to our catalog of AcademyR on-site and on-line courses, we also have a new on-line training option for Revolution R Enterprise, which you can take for free at DataCamp.com.

Supporting the R Community

Revolution Analytics is dedicated to supporting and growing the R community. There are now about 150 local R user groups worldwide, and we would love to see more! If you run an R user group or event or would like to start one, you can join our sponsorship program at http://revolutionanalytics.com/r-community.

Update from Tableau Software

Bora Beran, Program Manager of Statistics and Calculations

Tableau is known for helping users of wide range of skill levels accomplish analysis tasks easily in its interactive visualization environment. Sometimes with a click of a button and sometimes via scripting if more flexibility and expressibility is needed. In October 2014, Leland Wilkinson joined Tableau as VP of Statistics to help further our commitment to make statistics accessible to more people.

In the 8.1 release, Tableau added support for multiplicative models to its time series forecasting feature as well as allowed users to author calculations in Tableau using R which became an instant favorite for advanced users. Tableau users took advantage of these features to gain more insights from their data through exploration as well as sharing their visual analytic dashboards powered by R. Tableau continues to make it easy to analyze data at different levels now empowered with an easy interface to complex R calculations for non-technical consumers of their reports. Learn more at our site:

Customer use cases span from statistical tests to classification (k-means, decision trees), data transformations (multidimensional scaling), sentiment analysis and even optimization. In 2014 customer segmentation analysis and social media dashboards were most common among Tableau users.

Tableau's integration with R not only provides Tableau users access to a large and rapidly growing set of statistical libraries but also offers R users access to easily explore their data at the speed of thought. With Tableau, users manipulate data visually by drilling up/down hierarchies with drag and drop actions, applying filters by making selections in the visualization and partitioning data visually to run R scripts on each slice. Tableau users reap the benefits of specialized database connectors (from Excel and SQL Server to Hadoop and Teradata), query optimizations and caching to analyze big data efficiently. Learn more at this link.

News from TIBCO Software

TIBCO is focused on extending the reach of the R language to the enterprise, by support the growth of the R community, and investing in our enterprise-grade, alternative R interpreter: TIBCO Enterprise Runtime for R (TERR). In the past year, we have release multiple updates to TERR, expanded the usage of TERR within TIBCO Spotfire, and promoted the embedding of TERR into multiple different platforms. We also continued our support of the R Community, including co-sponsoring useR 2014, R Finance 2014 and EARL 2014, as well as multiple R meetups around the world. The TERR engine is available as a free download for the R community from tap.tibco.com, and is compatible with the RStudio IDE.

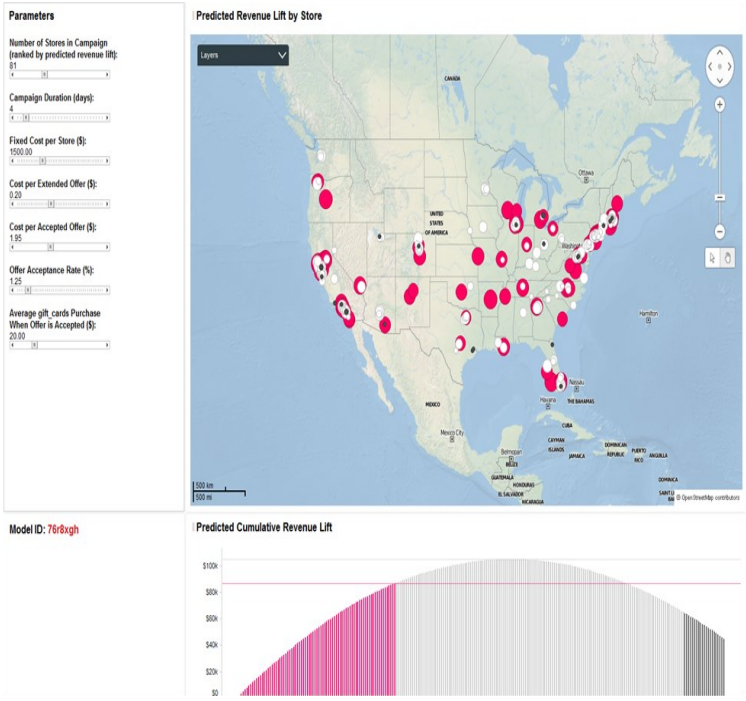
TERR Releases

TERR is faster and more robust than the open source R engine. It was built from the ground up for embedding in other platforms, based on our 25+ years of experience as the developers of S-PLUS. Over the last year, we have expanded the compatibility of TERR with the open source R engine, focusing on parallelization, clustering, nonlinear optimization, and compatibility with key CRAN packages. Approximately 2000 packages from CRAN are now fully compatible with the TERR platform.

We have also made it easier for R users to develop using TERR, by implementing compatibility with the RStudio IDE on Windows and Linux (including RStudio Server), and compatibility with ESS.

Spotfire

TIBCO Spotfire provides a powerful environment for interactive visualization and data discovery for a wide audience. Over the past year, we have expanded the availability of advanced analytics in Spotfire through the R language. TERR is now embedded in every desktop copy of Spotfire, and users can enhance their ad hoc analyses and applications using the R language.



Applying predictive models within Spotfire to do "What if?" analyses

This year, we added a Forecasting Tool that allows Spotfire users to easily do a Holt-Winters forecast on a time series, made TERR accessible to all Spotfire users without additional licensing, streamlined the process of creating and developing Spotfire applications that leverage TERR, and made it easy to call TERR directly from the Spotfire expression language. We also made it possible to reuse Spotfire Data Connections from TERR, so that TERR users can benefit from the wide range of enterprise Big Data sources Spotfire connects to.

Embeddable TERR

This year, we also made great strides in making TERR available across many different platforms and environments, so that TERR users can easily develop in open source R, and then deploy their code in the TERR engine—without having to reimplement and retest their code for the move into production. Highlights of this year include:

- TERR can now be integrated with TIBCO's Complex Event Processing platforms, TIBCO Business Events and TIBCO Streambase, so that predictive models can be scored in real time to drive automated decisions for retail offers, customer churn prevention, equipment maintenance and many other applications.
- TIBCO Cloud Compute Grid was launched, enabling TERR users to run massively parallel computations on the cloud using their Amazon AWS accounts.

Spotfire and TERR applications on Hadoop for Big Data analytic problems, which won the Cloudera 2014 Data Impact Award for Advanced Analytics.

Embedding of TERR by TIBCO partner Lavastorm Analytics, to provide advanced analytics based on the R language in their flagship Lavastorm Analytics Engine.

TERR compatibility with Interactive R Statistics nodes in KNIME, enabling KNIME users to benefit from the power and speed of TERR.

Lou Bajuk-Yorgan
Sr. Dir., Product Management,
TIBCO Analytics
lbajuk@tibco.com, @loubajuk

References:

Download the TERR Developer Edition:

Overview of Spotfire and TERR:

TIBCO Cloud Compute Grid:

Announcement from Tim Hesterberg, Statistical Computing Treasurer

The ASA is in the process of revising the Curriculum Guidelines for Undergraduate Programs in Statistical Science.
The draft guidelines:
http://www.amstat.org/education/curriculumguidelines.cfm
call for better statistical computing skills.

I am writing a supplement to the guidelines, about resampling in the undergraduate curriculum. I describe the bootstrap and permutation tests, discuss their value in teaching, and discuss some issues. Some points may surprise you, e.g.\ you should use 10,000+ samples, and how poor the common bootstrap percentile interval is in small samples.

I also talk about how inaccurate common formula methods are---e.g. a t-test is not reasonably accurate until n ge 5000, for exponentialpopulations---and how resampling can be used for diagnostics and
better inferences. We can use computing power to do better statistics.

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 Presidents and Officers of both sections for their contribution to this newsletter via officer reports, news, and announcements.

The Statistical Computing & Statistical Graphics Newsletter is a publication of the Statistical Computing and Statistical Graphics Sections of the ASA.

All communications regarding ASA membership and the Statistical Computing and Statistical Graphics Section,
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
email{asainfo@amstat.org}