

Western Kentucky University

From the SelectedWorks of Matt Bogard

January 2010

Using R

Contact
Author

Start Your Own
SelectedWorks

Notify Me
of New Work



Available at: http://works.bepress.com/matt_bogard/10

Using R

Many people that are involved in data analysis may have one software package that they prefer to use, but may never have heard of [R](#), or even thought about trying something new. R is a statistical programming language with a command line interface that is becoming more and more popular every day. If you are used to a point and click interface (like SPSS or Excel for example) the learning curve for R may be quite steep but well worth the effort. If you are already familiar with a programming language, like SAS for instance, (SPSS also has a command line interface) the switch to R may not be that difficult but could be very rewarding. In fact there is a free version of a reference available entitled '[R for SAS and SPSS Users](#)' that I found very helpful getting started.

How I use R

I often use SAS for data management, recoding, merging and sub-setting data sets, and cleaning and formatting data. I'm getting better at doing some of these things in R but I just have more experience in SAS. I first became interested in using R for spatial lag models and geographically weighted regression, which requires the construction of spatial weights matrices. (These techniques aren't as straight forward if at all possible in SAS). I have also used R for data visualization, data mining/machine learning, as well as social network analysis. I give my students the option to use R in the statistics course I teach and I have had several pursue it. And most recently, I have become acquainted with SAS IML Studio, which is a product that will better enable me to integrate some of my SAS related projects with some of the powerful tools R provides.

How can you use R?

Google's Chief Economist Hal Varian uses R . His paper [Predicting the Present with Google Trends](#) is a great example of how R can be used for prediction, and is loaded with R code and graphics. Facebook uses R as well:

"Facebook's Data Team used R in 2007 to answer two questions about new users: (i) which data points predict whether a user will stay? and (ii) if they stay, which data points predict how active they'll be after three months?" ([Link](#)- how Google and Facebook are using r by Michael E. Driscoll at Dataspora.com)

The New York Times recently did a feature on R entitled '[Data Analysts Captivated by R](#)' where industry leaders explain why and how they use R.

"Companies like Google and Pfizer say they use the software for just about anything they can. Google, for example, taps R for help understanding trends in ad pricing and for illuminating patterns in the search data it collects. Pfizer has created customized packages for R... Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it."

The article goes on to discuss the proliferate and diverse packages offered by R:

“One package, called BiodiversityR, offers a graphical interface aimed at making calculations of environmental trends easier. Another package, called Emu, analyzes speech patterns, while GenABEL is used to study the human genome. The financial services community has demonstrated a particular affinity for R; dozens of packages exist for derivatives analysis alone.”

In a recent issue of Forbes ([Power in the Numbers](#) by Quentin Hardy May 2010) R is discussed as a powerful analysis tool as well as its potential to revolutionize business, markets, and government:

‘Using an R package originally for ecological science, a human rights group called Benetech was able to establish a pattern of genocide in Guatemala. A baseball fan in West Virginia used another R package to predict when pitchers would get tired, winning himself a job with the Tampa Bay Rays. An R promoter in San Francisco, Michael Driscoll, used it to prove that you are seven times as likely to change cell phone providers the month after a friend does. Now he uses it for the pricing and placement of Internet ads, looking at 100,000 variables a second’

“Nie posits that statisticians can act as watchdogs for the common man, helping people find new ways to unite and escape top-down manipulation from governments, media or big business...It's a great power equalizer, a Magna Carta for the devolution of analytic rights...”

The proprietary adaptation of R, Revolution R was recently discussed on Fox Business; you can see the interview [here](#).

A good way to get a look at what researchers are doing with R would be to review presentation topics at a past UseR conference, [from 2008](#):

Andrew Gelman, **Bayesian generalized linear models and an appropriate default prior**

Michael Höhle, **Modelling and surveillance of infectious diseases - or why there is an R in SARS**

E. James Harner, Dajie Luo, Jun Tan, **JavaStat: a Java-based R Front-end**

Robert Ferstl, Josef Hayden, **Hedging interest rate risk with the dynamic Nelson/Siegel model**

Jacob Michaelson, Andreas Beyer, **Random Forests for eQTL Analysis: A Performance Comparison**

Graham J. WilliamsDeploying , **Data Mining in Government - Experiences With R/Rattle**

Jing Hua Zhao, Qihua Tan, Shengxu Li, Jian'an LuanSome, **Perspectives of Graphical Methods for Genetic Data**

Who else uses R?

A review of past participant lists from UseR also gives you a pretty good idea about companies that are using R. Past participants includes:

Merril Lynch, Merck Research Laboratories , Siemens AG,, BASF SE, Astra Zeneca, Bayer CropScience, Novartis Pharma, AT&T Labs, Bell Labs, Bank of Canada, Credit Suisse, Max Planck Institute, US Naval Academy, Harvard, Vanderbilt University + many universities.

UCLA has great [online resources](#) for R and Stanford offers a course called Elements of Statistical Learning which utilizes a [textbook](#) with the same title that heavily references R code.

Downloading and using R

For more information about R and to download it to try it yourself, the place to start is the R Project website: <http://www.r-project.org/>

A few other great references for R can be found here:

[Quick R \(Statistics Guide\)](#)- <http://www.statmethods.net/index.html>

[Revolution Analytics](#)- (Blog) - <http://blog.revolutionanalytics.com/>

<http://www.r-bloggers.com/>

[Facebook](#)- R Bloggers Group

[Statistical Computing with R: A Tutorial](#) (Illinois State University) –

<http://math.illinoisstate.edu/dhkim/rstuff/rtutor.html>

[\(UCLA R Class Notes\)](#) -*great example code for getting started with R* –
<http://www.ats.ucla.edu/stat/R/notes/>

[Using R for Introductory Statistics](#) (John Verzani) - <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

References:

Changing the Face of Analytics. Fox Business Network via YouTube:

http://www.youtube.com/watch?v=4nOMhTBuXl8&feature=youtube_gdata

Data Analysts Captivated by R's Power. New York Times. January 6, 2009.

Driscoll, Michael E. How Google and Facebook Are Using R. Dataspora Blog. February 19, 2009. <http://dataspora.com/blog/predictive-analytics-using-r/>

Hardy, Quentin. Power in Numbers. Forbes Magazine. May 24, 2010.

Hastie, Tibshirani, and Friedman. The Elements of Statistical Learning. Second Edition. February 2009. <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

Hyunyoung Choi and Hal Varian. Predicting the Present with Google Trends. Google Inc. Draft Date April 10, 2009.
http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/googleblogs/pdfs/google_predicting_the_present.pdf

R 4stats.com: <http://rforsasandspssusers.com/>

The R Project: <http://www.r-project.org/>

UCLA Resources to Help You Learn and Use R: <http://www.ats.ucla.edu/stat/R/>