

## Table of Contents

I. Web resources for R – Part 1.....	3
II. Web resources for R – Part 2.....	6
III. More on packages.....	9
IV. Popularity of R.....	12
V. Books.....	13
VI. Timeline for the development of R.....	14
VII. Companies that “sell” R.....	17

# Extra information on R

Christopher R. Bilder  
University of Nebraska-Lincoln  
Department of Statistics

September 16, 2011

[www.chrisbilder.com/workshop](http://www.chrisbilder.com/workshop)

© 2011 Christopher R. Bilder

© 2011 Christopher R. Bilder

Extra.3

Extra.4

## I. Web resources for R – Part 1

Please see my general resources web page at [www.chrisbilder.com/resources](http://www.chrisbilder.com/resources).

The R Listserv is the primary way that people obtain answers to questions when the Help is not enough. Questions are posted to the listserv and other R users respond to them. Responses are frequently given by R Core group members (<http://www.r-project.org/contributors.html>). Below are important links for the listserv:

1) Subscribe: <http://www.r-project.org/mail.html>

There are a number of R listservs with R-help being the main listserv. This listserv is active with usually more than 100 posts a day, so I recommend subscribing to it in digest mode.

While I still subscribe to R-help, I no longer consistently read it.

2) Archives: <http://finzi.psych.upenn.edu/search.html>

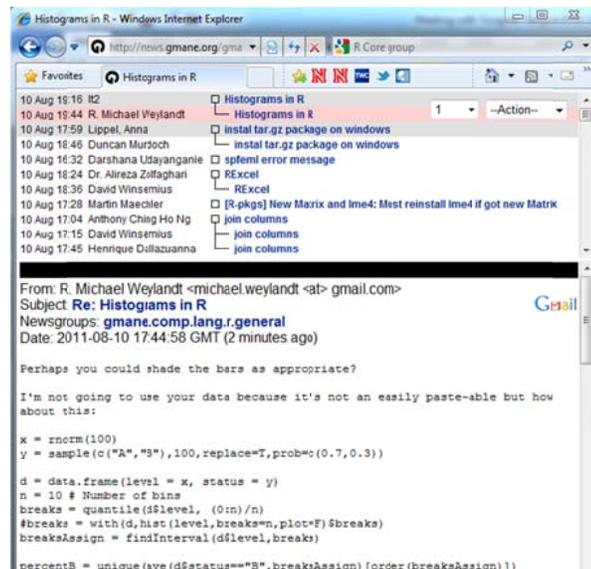
Before posting a question on the listserv, search here first! Also, be careful with the “Target” used in the

search. I recommend searching the R-help listserv archives back to 2002.

Try a search with my name!

3) Reading listserv posts on the web: <http://news.gmane.org/gmane.comp.lang.r.general>

This is a more convenient way to read R-help due to its threaded content:



© 2011 Christopher R. Bilder

© 2011 Christopher R. Bilder

A more recent way to ask questions and/or find answers is at stackoverflow (<http://stackoverflow.com/questions/tagged/r>). This is a general website of questions and answers for programming languages. The interface is somewhat better than the older listserv, because content is given in a more organized manner.

## II. Web resources for R – Part 2

### 1) R Journal: <http://journal.r-project.org>

This is the official publication for the R project. The journal was created in 2009 to provide a peer-reviewed paper environment for contributions to R. For example, I have a paper in the December 2010 issue regarding my binGroup package. The journal also gives development notes on R.

The journal replaced the “R News” newsletter that was first published in 2001. The beginning of the first issue for R News (<http://journal.r-project.org/archive.html>) is a nice read on “What is R?”

Another useful journal on R is the *Journal of Statistical Software* (<http://www.jstatsoft.org/>). This journal was created in 1996. While it publishes peer-reviewed articles for any statistical software package, the vast majority of its articles are on R now.

### 2) Blogs: <http://www.r-bloggers.com>

There are quite a few blogs on R, and the above link provides a central location for their feeds. My favorite R blog is Revolutions at <http://blog.revolutionanalytics.com>, which is written

primarily by David Smith of Revolution Analytics. I have learned a lot from reading this blog! You can even see a posting featuring some of my work at <http://blog.revolutionanalytics.com/2009/12/because-its-friday-detecting-cylons.html>.

### 3) Twitter: <http://twitter.com/#!/RLangTip>

This provides the R language tip of the day, which is now given by Revolution Analytics. It gives some nice basic tips on using R in 140 characters or less.

There are also many other people tweeting on R. By searching the #Rstat hashtag in Twitter, you can see recent tweets.

You can even tweet directly from R with the TwitterR package.

### 4) R Wiki: <http://rwiki.sciviews.org/doku.php>

Wiki for R

### 5) R Seek: <http://www.rseek.org>

Google search that filters out non-R websites

### 6) R by example: <http://www.mayin.org/ajayshah/KB/R/index.html>

### 7) My teaching websites: [www.chrisbilder.com](http://www.chrisbilder.com)

Courses that exclusively use R are:

- STAT 870 – Multiple regression analysis
- STAT 875 – Categorical data analysis
- STAT 892 – Applied time series analysis
- STAT 950 – Bootstrap methods and their application

Each of these websites contains my lecture notes and R programs. The websites also contain in-class videos of the lectures.

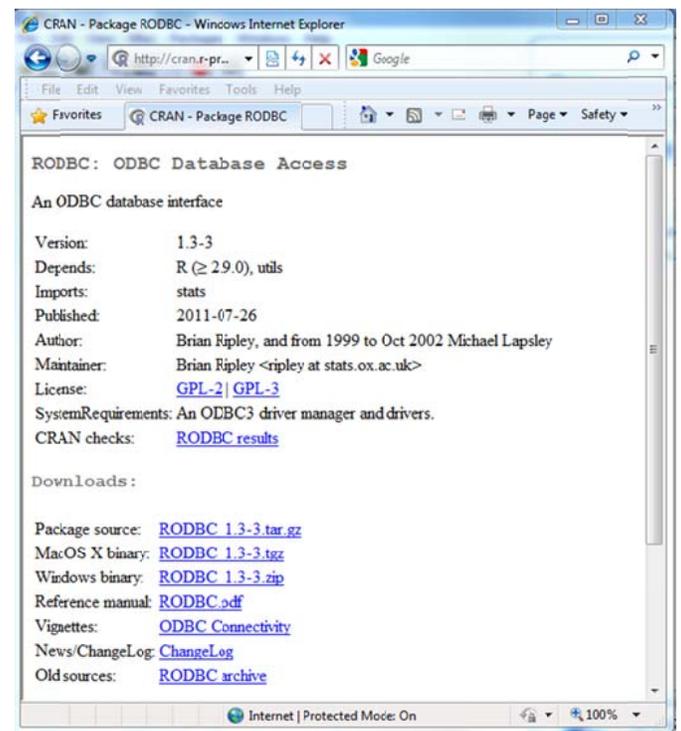
### III. More on packages

A list of all packages is available at <http://cran.r-project.org/web/packages>. Because there are over 3,000 packages, it may be difficult a specific one that fits your needs. This is acknowledged as a problem for R, and there have been some recent changes to help users.

Tips to find a package:

- Page down and look through the packages
- Search the web page at [http://cran.r-project.org/web/packages/available\\_packages\\_by\\_name.html](http://cran.r-project.org/web/packages/available_packages_by_name.html) for keywords. Try a search for “group testing”
- Search the listserv archive

Once you find a package of interest, its corresponding web page contains useful information. All packages have a manual that duplicates its help web pages. Many packages also have vignettes that give paper like introductions. For example, below is the web page for RODBC that mentions its reference manual and vignette:



Can you trust the functions in packages?

Here are my levels of trust:

- Packages available in the default installation of R: Yes

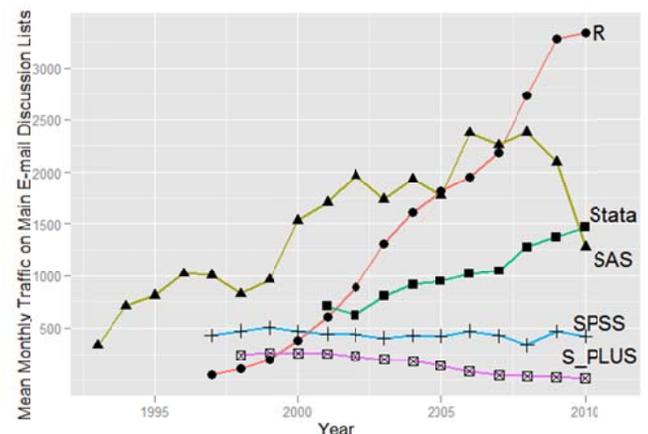
- Packages written by leaders in the area of interest: Most likely, yes
- Packages written by people you trust: Most likely, yes
- Packages that have been peer-reviewed for the *R Journal* or the *Journal of Statistical Software*: Most likely, yes
- Packages from unknown authors: Hopefully
- Packages with version numbers beginning with a 0: Hopefully
- Packages created for a student's dissertation: Hopefully
- Packages just recently created: Hopefully

A higher level of caution should be used with packages falling in the “Hopefully” group. My comments here are not meant to alarm you about the correctness of R. Rather, there are a vast number of R packages contributed by users, and many of them can perform calculations that no other software can; however, contributed packages need to be used in a judicious manner.

If there is any doubt about a package, please remember all code is available for examination.

### IV. Popularity of R

Mean monthly traffic from R and other listservs (Muenchen, 2011, <http://sites.google.com/site/r4statistics/popularity>):



The author says the plot may be a little misleading due to SAS discussion migrating to other forums. The author also cites work by Arthur Tabachnek who found that R was the most discussed topic on the SAS listserv in 2009.

## V. Books

The trend among statistics books today is to incorporate “R” into them. In fact, I am writing a book named “Categorical Data Analysis with R” that should come out late in 2012 from CRC Press.

Recent books on R that you may find useful:

- *An Introduction to R* by Venables, Smith, and the R Development Core Team – This comes with the installation of R (Select HELP > MANUALS > AN INTRODUCTION TO R).
- *Software for Data Analysis: Programming with R* by Chambers
- *An R Companion to Applied Regression* by Fox and Weisberg
- *R Graphics* (2<sup>nd</sup> edition) by Paul Murrell

Older books on R that are still good:

- *Modern Applied Statistics with S* (4<sup>th</sup> edition) by Venables and Ripley
- *S Programming* by Venables and Ripley

## VI. Timeline for the development of R

### 1970s

- 1976: A preliminary version of S was created by Bell Labs as a programming language like C, but for statistics. John Chambers is one of the primary inventors.

### 1980s

- 1988: Statistical Sciences Inc. creates the S-Plus software package and sells the software to users.

### 1990s

- 1993: Statistical Sciences merges with Mathsoft and buys the exclusive license to market S.
- 1996: A paper by Ross Ihaka and Robert Gentleman is published in the *Journal of Computational and Graphical Statistics* on an early version of R.
- 1999: John Chambers receives the Association for Computing Machinery (ACM) Award for S.

### 2000s

- 2000: A first full release of R is made. It is based on the S syntax, but it is not exactly S. R is often referred to as “not unlike S”.
- 2001: I first hear about R early this year!

- 2004: The first UseR! conference is held in Austria and the R Foundation is formed.
- 2004: During a Joint Statistical Meetings (JSM) session that I attended, a SPSS executive says his company and other statistical software companies have felt R’s impact and they are changing their business model.
- 2007: The first USA-based UseR! conference is held at Iowa State University.
- 2008: The editor for the *Journal of the American Statistical Association* says during a JSM presentation that R has become the de facto statistical software package for researchers.
- 2009: A New York Times article [“Data Analysts Captivated by R’s Power”](#) is published on 1-6-09. The article contains the now famous quote by Anne Milley of SAS:

I think it addresses a niche market for high-end data analysts that want free, readily available code. We have customers who build engines for aircraft. I am happy they are not using freeware when I get on a jet.

Milley later says that she should not have made the airplane comment. SAS starts promoting its ability to call R programs from PROC IML later this same year.

- 2009: The first issue of the R Journal (<http://journal.r-project.org>) is published in May 2009, and it replaces “R News”.

### 2010s

- 2010: The second USA-based UseR! Conference is held at the National Institutes of Standards.
- 2011: R 2.14.0 will be released in October – The new release will have some significant speed improvements (see <http://blog.revolutionanalytics.com/2011/08/with-byte-compiler-r-214-will-be-even-faster.html>).
- 2012: The third USA-based UseR! Conference will be held at Vanderbilt University on June 12-15.

Chambers (2010) and Venables and Ripley (2000) provided much of the pre-2000 history.

## VII. Companies that “sell” R

- Revolution Analytics
- Tibco (S-Plus)