

## Introduction to R

All statisticians should be proficient in [C](#) (for speed), [perl](#) (for data manipulation), and [R](#) (for interactive analyses and graphics). Think "CPR".

As described on the [R project web page](#):

"R is a system for statistical computation and graphics. It consists of a language plus a run-time environment with graphics, a debugger, access to certain system functions, and the ability to run programs stored in script files.

"The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. Most of the user-visible functions in R are written in R. It is possible for the user to interface to procedures written in the C, C++, or FORTRAN languages for efficiency. The R distribution contains functionality for a large number of statistical procedures. Among these are: linear and generalized linear models, nonlinear regression models, time series analysis, classical parametric and nonparametric tests, clustering and smoothing. There is also a large set of functions which provide a flexible graphical environment for creating various kinds of data presentations. Additional modules are available for a variety of specific purposes."

I use R for all interactive statistical analyses and graphics. Virtually all of the software I produce is now written as add-on packages for R. The computationally intensive portions of such software are written in C, but writing such software as a package for R makes the data input/output to the C code extremely easy, and makes it easy to create documentation and provide graphical facilities. Moreover, R has a very extensive mathematics library, so I don't need to re-write things that have already been well coded.

While I intended this page to provide an introduction similar to my [Introduction to Perl](#) page, instead, I am leaving this as a list of links and then a few tips at the bottom.

---

## Topics

- [Resources](#)
- [Notes on R for Windows](#)
- [Notes on R for Mac OSX](#)
- [Various hints](#)
- [Building R packages for Windows](#)

---

## Resources

### Web

- [My notes on R for Mac OSX](#)
- [My notes on R for Windows](#)
- [R Archive](#)
- [R FAQ](#)
- [R wiki](#)
- [R for Windows FAQ](#)
- [R for Macintosh FAQ](#)
- [R for Macintosh developer page](#)
- [Building R from source on Windows](#)
- [R Journal](#)
- [R mailing lists](#)
- [R-help mailing list archive](#)
- [R-SIG-Mac mailing list archive](#)
- [Contributed R documentation](#)
- [R tips](#)
- [R primer for introductory statistics](#) [Bret Larget]

### PDF documents

- [SimpleR](#) (John Verzani)

- Using R for data analysis and graphics [[pdf \(693k; 112 pgs\)](#)] (JH Maindonald)
- R reference card [[pdf \(59k, 1 pg\)](#)] (J Baron)
- **The R manuals:**
  - An introduction to R [[pdf](#)] (The official introduction)
  - Writing R extensions [[pdf](#)] (Explanations of writing packages and calling C from R)
  - The R reference index [[pdf](#)] (The R help pages printed out)

## Books

- [A comprehensive list of books related to R](#)
- The most important book:
  - WN Venables and BD Ripley (2002) [Modern applied statistics with S](#), 4th edition. Springer [[Online complements](#)]
- Beginning books:
  - P Dalgaard (2008) [Introductory statistics with R](#), 2nd edition Springer
  - AF Zuur, EN Ieno, EHWG Meesters (2009) [A beginner's guide to R](#), Springer
  - MJ Crawley (2007) [The R book](#), Wiley
  - BS Everitt, T Hothorn (2006) [A handbook of statistical analyses using R](#), Chapman & Hall/CRC
  - J Maindonald, J Braun (2006) [Data analysis and graphics using R](#), 2nd edition, Cambridge University Press
  - J Verzani (2004) [Using R for introductory statistics](#), Chapman & Hall/CRC
- For more advanced programming:
  - WN Venables and BD Ripley (2000) [S programming](#). Springer
  - JM Chambers (2008) [Software for Data Analysis: Programming with R](#), Springer
- Special topics:
  - JJ Faraway (2004) [Linear models with R](#), Chapman & Hall/CRC
  - P Spector (2008) [Data manipulation with R](#), Springer
  - PSP Cowpertwait and A Metcalfe (2009) [Introductory Time Series with R](#). Springer
  - C Ritz, JC Streibig (2009) [Nonlinear regression with R](#), Springer
  - RS Bivand, EJ Pebesma, V Gomez-Rubio (2008) [Applied Spatial Data Analysis with R](#), Springer
  - R Gentleman (2008) [R programming for bioinformatics](#), Chapman & Hall/CRC

## Various hints

### *Math expressions in plots*

One of the nice additions to R (relative to Splus) is the easy inclusion of mathematical expressions in plots, using the function `expression()`. Take a look at `help(plotmath)` to see a big list of what you can do; also look at the examples in the help file for the function `legend`, and consider the following:

```
plot(rnorm(100),rnorm(100),xlab=expression(hat(mu)[0]),
     ylab=expression(alpha^beta),
     main=expression(paste("Plot of ", alpha^beta, " versus ", hat(mu)[0])))
```

### *Emacs and ESS*

For those running R within unix or Linux, I highly recommend running R from within emacs. I like to have a really big emacs window, and so I put the following line in my `~/.bashrc` file. (I use the `bash` shell; if you use `tcsh` or `csh`, put the analogous line in your `~/.tcshrc` or `~/.cshrc` file. After doing this, you can type `be` to get a big emacs window.

```
alias be='emacs -bg black -fg white -geometry 99x50+560+122 &'
```

Place the following lines in your `~/.emacs` file in order to get fancy highlighting and to get access to [ESS](#) (Emacs Speaks Statistics).

```
;; load ESS
(load "/sw/share/emacs/site-lisp/ess-5.3.0/lisp/ess-site")

;; automatic Font Lock mode in TeX mode
(add-hook 'tex-mode-hook 'turn-on-font-lock)
;; if Font Lock necessary for other than .tex files, uncomment following
(global-font-lock-mode t)
```

```

;; modes for other files
auto-mode-alist (append (list ('("\\.c$" . c-mode)
                              ('("\\.tex$" . latex-mode)
                              ('("\\.S$" . S-mode)
                              ('("\\.s$" . S-mode)
                              ('("\\.html$" . html-mode)
                              ('("\\.emacs" . emacs-lisp-mode)
                              )
                  auto-mode-alist)

;; html helper mode from http://www.farne.uklinux.net/emacs-primer.html
(auto-load 'html-helper-mode "html-helper-mode" "Yay HTML" t)
(setq auto-mode-alist (cons ('("\\.html$" . html-helper-mode) auto-mode-alist))
(setq html-helper-do-write-file-hooks t)

;; ESS for Sweave files
(defun Rnw-mode ()
  (require 'ess-noweb)
  (noweb-mode)
  (if (fboundp 'R-mode)
      (setq noweb-default-code-mode 'R-mode)))
(add-to-list 'auto-mode-alist ('("\\.Rnw\\'" . Rnw-mode))
(add-to-list 'auto-mode-alist ('("\\.Snw\\'" . Rnw-mode))
(setq reftex-file-extensions
      ('("Snw" "Rnw" "nw" "tex" ".tex" ".ltx") ("bib" ".bib")))
(setq TeX-file-extensions
      ('("Snw" "Rnw" "nw" "tex" "sty" "cls" "ltx" "texi" "texinfo")))

```

Having done the above, open an emacs window and type `M-x R` and then enter the subdirectory in which you wish to run R. (`M-x` means press the Esc key and then press x. down the diamond button while pressing x.)

### *.Renviron file*

You likely will want to create an `~/.Renviron` file in order to let R know some general parameters. My `~/.Renviron` file contains the following:

```

R_PAPERSIZE=letter
R_LIBS=/Users/kbroman/Rlibs
EDITOR=emacs

```

### *.Rprofile file*

You might also want to create an `~/.Rprofile` file. R code in this file will be run anytime you start R, no matter what subdirectory you start in (unless there's an `.Rprofile` within that subdirectory, in which case that is read instead. Examples of what you may wish to put there include commands to modify the options and/or load certain packages (libraries).

### *Locally installed packages*

To install an R package (e.g., `qtl_0.76.tar.gz`) locally (e.g., in the directory `/users/student/auser/Rlib`), type

```

R INSTALL --library=/users/student/auser/Rlib qtl_0.76.tar.gz

```

In your `~/.Renviron` file, include the line

```

R_LIBS=/users/student/auser/Rlib

```

Within R, when you type `library()`, you should see separate listings of the packages installed locally and those on the main system.