

Hastie, T., Tibshirani, R., Eisen, M., Brown, P., Scherf, U., Weinstein, J., Alizadeh, A., Staudt, L., and Botstein, D. (2000b). Gene shaving: a new class of clustering methods for expression arrays. Technical report, Stanford University. 21

Li, M. N. and Rossini, A. (2001). RPVM: Cluster statistical computing in R. *R News*, 1(3):4–7. 21, 23

Message Passing Interface Forum (1997). MPI-2: Extensions to the Message-Passing Interface. <http://www.mpi-forum.org/docs/docs.html>. 21

Rossini, A., Tierney, L., and Li, N. (2003). Simple parallel statistical computing in r. Technical Report Working Paper 193, UW Biostatistics Working Paper Series. <http://www.bepress.com/uwbiostat/paper193>. 21, 23

Scyld Computing Corporation (1998). Beowulf introduction & overview. <http://www.beowulf.org/intro.html>. 21

Tierney, L. (2002). Simple Network of Workstations for R. Technical report, School of Statistics, University of Minnesota. <http://www.stat.uiowa.edu/luke/R/cluster/cluster.html>. 21

Yu, H. (2002). Rmpi: Parallel statistical computing in R. *R News*, 2(2):10–14. 21

Brett Carson, Robert Murison, Ian A. Mason
The University of New England, Australia
bcarson@turing.une.edu.au
rmurison@turing.une.edu.au
iam@turing.une.edu.au

R Help Desk

Getting Help – R’s Help Facilities and Manuals

Uwe Ligges

Introduction

This issue of the *R Help Desk* deals with its probably most fundamental topic: Getting help!

Different amounts of knowledge about R and R related experiences require different levels of help. Therefore a description of the available manuals, help functions within R, and mailing lists will be given, as well as an “instruction” how to use these facilities.

A secondary objective of the article is to point out that manuals, help functions, and mailing list archives are commonly much more comprehensive than answers on a mailing list. In other words, there are good reasons to read manuals, use help functions, and search for information in mailing list archives.

Manuals

The manuals described in this Section are shipped with R (in directory ‘.../doc/manual’ of a regular R installation). Nevertheless, when problems occur before R is installed or the available R version is outdated, recent versions of the manuals are also available from CRAN¹. Let me cite the “R Installation and Administration” manual by the *R Development Core*

Team (R Core, for short; 2003b) as a reference for information on installation and updates not only for R itself, but also for contributed packages.

I think it is not possible to use adequately a programming language and environment like R without reading any introductory material. Fundamentals for using R are described in “An Introduction to R” (Venables et al., 2003), and the “R Data Import/Export” manual (R Core, 2003a). Both manuals are the first references for any programming tasks in R.

Frequently asked questions (and corresponding answers) are collected in the “R FAQ” (Hornik, 2003), an important resource not only for beginners. It is a good idea to look into the “R FAQ” when a question arises that cannot be answered by reading the other manuals cited above. Specific FAQs for Macintosh (by Stefano M. Iacus) and Windows (by Brian D. Ripley) are available as well².

Users who gained some experiences might want to write their own functions making use of the language in an efficient manner. The manual “R Language Definition” (R Core, 2003c, still labelled as “draft”) is a comprehensive reference manual for a couple of important language aspects, e.g.: objects, working with expressions, working with the language (objects), description of the parser, and *debugging*. I recommend this manual to all users who plan to work regularly with R, since it provides really illuminating insights.

For those rather experienced users who plan to create their own packages, writing help pages and optimize their code, “Writing R Extensions” (R Core,

¹<http://CRAN.R-project.org/>

²<http://CRAN.R-project.org/faqs.html>

2003d) is the appropriate reference. It includes also descriptions on interfacing to C, C++, or Fortran code, and compiling and linking this code into R.

Occasionally, `library(help=PackageName)` indicates that package specific manuals (or papers corresponding to packages), so-called vignettes, are available in directory `'.../library/PackageName/doc'`.

Beside the manuals, several books dealing with R are available these days, e.g.: introductory textbooks, books specialized on certain applications such as regression analysis, and books focussed on programming and the language. Here I do not want to advertise any of those books, because the "optimal" choice of a book depends on the tasks a user is going to perform in R. Nevertheless, I would like to mention [Venables and Ripley \(2000\)](#) as a reference for programming and the language, and [Venables and Ripley \(2002\)](#) as a comprehensive book dealing with several statistical applications and how to apply those in R (and S-PLUS).

Help functions

Access to the documentation on the topic *help* (in this case a function name) is provided by `?help`. The `"?"` is the probably most frequently used way for getting help, whereas the corresponding function `help()` is rather flexible. As the default, help is shown as formatted text. This can be changed (by arguments to `help()`, or for a whole session with `options()`) to, e.g., HTML (`htmlhelp=TRUE`), if available. The HTML format provides convenient features like links to other related help topics. On Windows, I prefer to get help in Compiled HTML format by setting `options(chmhelp=TRUE)` in file `'Rprofile'`.

When exact names of functions or other topics are unknown, the user can search for relevant documentation (matching a given character string in, e.g. name, title, or keyword entries) with `help.search()`. This function allows quite flexible searching using either regular expression or fuzzy matching, see `?help.search` for details. If it is required to find an object by its partial name, `apropos()` is the right choice – it also accepts regular expressions.

Documentation and help in HTML format can also be accessed by `help.start()`. This function starts a Web browser with an index that links to HTML versions of the manuals mentioned above, the FAQs, a list of available packages of the current installation, several other information, and a page containing a search engine and keywords. On the latter page, the search engine can be used to "search for keywords, function and data names and text in help page titles" of installed packages. The list of

keywords is useful when the search using other facilities fails: the user gets a complete list of functions and data sets corresponding to a particular keyword.

Of course, it is not possible to search for functionality of (unknown) contributed packages or documentation packages (and functions or data sets defined therein) that are not available in the current installation of R. A listing of contributed packages and their titles is given on CRAN, as well as their function indices and reference manuals. Nevertheless, it might be still hard to find the function or package of interest. For that purpose, an R site search³ provided by Jonathan Baron is linked on the search page of CRAN. Help files, manuals, and mailing list archives (see below) can be searched.

Mailing lists

Sometimes manuals and help functions do not answer a question, or there is need for a discussion. For these purposes there are the mailing lists⁴ *r-announce* for important announcements, *r-help* for questions and answers about problems and solutions using R, and *r-devel* for the development of R.

Obviously, *r-help* is the appropriate list for asking questions that are not covered in the manuals (including the "R FAQ") or the help pages, and regularly those questions are answered within a few hours (or even minutes).

Before asking, it is a good idea to look into the archives of the mailing list.⁵ These archives contain a huge amount of information and may be much more helpful than a direct answer on a question, because most questions already have been asked – and answered – several times (focussing on slightly different aspects, hence these are illuminating a bigger context).

It is really easy to ask questions on *r-help*, but I would like to point out the advantage of reading manuals *before* asking: Of course, reading (a section of) some manual takes much more time than asking, and answers from voluntary help providers of *r-help* will probably solve the recent problem of the asker, but quite similar problems will certainly arise pretty soon. Instead, the time one spends reading the manuals will be saved in subsequent programming tasks, because many side aspects (e.g. convenient and powerful functions, programming tricks, etc.) will be learned from which one benefits in other applications.

³<http://finzi.psych.upenn.edu/search.html>, see also <http://CRAN.R-project.org/search.html>

⁴accessible via <http://www.R-project.org/mail.html>

⁵archives are searchable, see <http://CRAN.R-project.org/search.html>

Bibliography

- Hornik, K. (2003): *The R FAQ*, Version 1.7-3. <http://www.ci.tuwien.ac.at/~hornik/R/> ISBN 3-901167-51-X. 26
- R Development Core Team (2003a): *R Data Import/Export*. URL <http://CRAN.R-project.org/manuals.html>. ISBN 3-901167-53-6. 26
- R Development Core Team (2003b): *R Installation and Administration*. URL <http://CRAN.R-project.org/manuals.html>. ISBN 3-901167-52-8. 26
- R Development Core Team (2003c): *R Language Definition*. URL <http://CRAN.R-project.org/manuals.html>. ISBN 3-901167-56-0. 26
- R Development Core Team (2003d): *Writing R Extensions*. URL <http://CRAN.R-project.org/manuals.html>. ISBN 3-901167-54-4. 26

Venables, W. N. and Ripley, B. D. (2000): *S Programming*. New York: Springer-Verlag. 27

Venables, W. N. and Ripley, B. D. (2002): *Modern Applied Statistics with S*. New York: Springer-Verlag, 4th edition. 27

Venables, W. N., Smith, D. M., and the R Development Core Team (2003): *An Introduction to R*. URL <http://CRAN.R-project.org/manuals.html>. ISBN 3-901167-55-2. 26

Uwe Ligges
 Fachbereich Statistik, Universität Dortmund, Germany
ligges@statistik.uni-dortmund.de

Book Reviews

Michael Crawley: *Statistical Computing: An Introduction to Data Analysis Using S-Plus*

John Wiley and Sons, New York, USA, 2002
 770 pages, ISBN 0-471-56040-5
<http://www.bio.ic.ac.uk/research/mjcraw/statcomp/>

The number of monographs related to the S programming language is sufficiently small (a few dozen, at most) that there are still great opportunities for new books to fill some voids in the literature. *Statistical Computing* (hereafter referred to as SC) is one such text, presenting an array of modern statistical methods in S-Plus at an introductory level.

The approach taken by author is to emphasize "graphical data inspection, parameter estimation and model criticism rather than classical hypothesis testing" (p. ix). Within this framework, the text covers a vast range of applied data analysis topics. Background material on S-Plus, probability, and distributions is included in the text. The traditional topics such as linear models, regression, and analysis of variance are covered, sometimes to a greater extent than is common (for example, a split-split-split-split plot experiment and a whole chapter on ANCOVA). More advanced modeling techniques such as bootstrap, generalized linear models, mixed effects models and spatial statistics are also presented. For each topic, some elementary theory is presented and usually an example is worked by hand. Crawley also discusses more philosophical issues such as randomization, replication, model parsimony, appropriate transformations, etc. The book contains a very thor-

ough index of more than 25 pages. In the index, all S-Plus language words appear in bold type.

A supporting web site contains all the data files, scripts with all the code from each chapter of the book, and a few additional (rough) chapters. The scripts pre-suppose that the data sets have already been imported into S-Plus by some mechanism—no explanation of how to do this is given—presenting the novice S-Plus user with a minor obstacle. Once the data is loaded, the scripts can be used to re-do the analyses in the book. Nearly every worked example begins by attaching a data frame, and then working with the variable names of the data frame. No detach command is present in the scripts, so the work environment can become cluttered and even cause errors if multiple data frames have common column names. Avoid this problem by quitting the S-Plus (or R) session after working through the script for each chapter.

SC contains a very broad use of S-Plus commands to the point the author says about `subplot`, "Having gone to all this trouble, I can't actually see why you would ever want to do this in earnest" (p. 158). Chapter 2 presents a number of functions (`grep`, `regexpr`, `solve`, `sapply`) that are not used within the text, but are usefully included for reference.

Readers of this newsletter will undoubtedly want to know how well the book can be used with R (version 1.6.2 base with recommended packages). According to the preface of SC, "The computing is presented in S-Plus, but all the examples will also work in the freeware program called R." Unfortunately, this claim is too strong. The book's web site does contain (as of March, 2003) three modifications for using R with the book. Many more modifications