- a graphics system, for interactively working with models. For example using **R-Tcl/Tk**, **Rggobi** or the R-interface to Graphviz.

- an analyzing unit that combines data and model with the possibility of using different inference algorithms in the analyzing step.

A minimal version of **grbase** is planned for January 2004.

An invited session concerned with the gR developments is being planned for the Joint Statistical Meeting in Toronto, 8-12 August 2004.

See `http://www.math.auc.dk/gr/gr2003/` for more information about the workshop and related links, including links to the aforementioned software.

Claus Dethlefsen
Aalborg University, Denmark
`dethlef@math.auc.dk`

Book Reviews

John Maindonald and John Braun: Data Analysis and Graphics Using R — An Example-based Approach

The aim of the book is to describe the ideas and concepts of many statistical methodologies, that are widely used in applications, by demonstrating the use of R on a number of examples. Most examples in the book use genuine data collected by the authors in their combined several decades of statistical consulting experience. The authors see the book as a companion to other books that include more mathematical treatments of the relevant theory, and they avoid mathematical notation and mathematical description of statistical methods. The book is aimed at both scientists and students interested in practical data analysis. Data and new R functions used in the book are included in the DAAG package available from the authors' web sites and through the Comprehensive R Archive Network (CRAN).

The book begins with a nice summary of the contents of the twelve chapters of the book. Chapter 1, *A Brief Introduction to R*, provides enough information on using R to get the reader started. Chapter 2, *Style of Data Analysis*, demonstrates with many examples the use of R to carry out basic exploratory data analysis involving both graphical and numerical summaries of data. The authors not only describe how to create graphs and plots but also show the reader what to look for in the data summaries and how to interpret the summaries in the context of each particular example. Chapter 3, *Statistical Models*, describes the authors' view on the importance of models as a framework for statistical analysis. Chapter 4, *An Introduction to Formal Inference*, introduces the basic ideas of random sampling and sampling distributions of statistics necessary to understand confidence intervals and hypothesis testing. It also includes chi-square tests for contingency tables and one-way ANOVA.

The next several chapters demonstrate the use of R to analyze data using linear models. Chapter 5, *Regression with a Single Predictor*, Chapter 6, *Multiple Linear Regression*, Chapter 7, *Exploiting the Linear Model Framework*, and Chapter 8, *Logistic Regression and Other Generalized Linear Models*, use increasingly complex models to lead the reader through several examples of practical data analysis.

The next three chapters discuss more specialized topics that arise frequently in practice. Chapter 9, *Multi-level Models, Time Series, and Repeated Measures*, goes through examples that use more complicated error structures than examples found in previous chapters. Chapter 10, *Tree-based Classification and Regression Trees*, provides an introduction to tree-based regression and classification modeling. Chapter 11, *Multivariate Data Exploration and Discrimination*, describes both principle components analysis and discriminant analysis.

The final chapter, Chapter 12, *The R System — Additional Topics*, is a far more detailed introduction to R than that contained in the initial chapters. It is also intended as a reference to the earlier chapters.

The book is a primer on the nuts-and-bolts use of R for the types of statistical analysis that arise commonly in statistical practice, and it also teaches the reader to think statistically when interpreting the results of an analysis. The strength of the book is in the extensive examples of practical data analysis with complete examples of the R code necessary to carry out the analyses. Short R commands appear on nearly every page of the book and longer R code examples appear frequently as footnotes.

I would strongly recommend the book to scientists who have already had a regression or a linear models course and who wish to learn to use R. However, my recommendation has a few caveats. The first chapter of the book takes the reader through an introduction to R that has the potential to be a little frustrating for a reader with no prior R experience. For example, the first plotting command given is

```
plot(ACT ~ Year, data=ACTpop, pch=16)
```

The meaning of the `pch=16` option is described and the option `data=ACTpop` is self evident, but the syntax `ACT ~ Year` is not explained and is potentially confusing to an R beginner who does not automatically translate ~ into "is modeled by". Page 5 gives the advice to create a new workspace before experimenting with R functions, but provides no details on how one actually does this. Most examples of R code in the book do contain adequate descriptions, but there are a number of exceptions.

A second caveat is that the descriptions of statistical methods are an adequate refresher, but are inadequate as a primary source of information. The authors indicate clearly that the book is meant to complement other books in the presentation of, and the mathematical description of, statistical methods. I agree that the book would not work well as a stand-alone text book for a course on statistical modeling. However, it is also not short and I would hesitate to require students to buy it in addition to another comprehensive textbook. The scope of the book is greater than simply serving as a companion book for teaching R.

Despite my hesitation to use this book in teaching, I give it a strong recommendation to the scientist or data analyst who wishes an easy-to-read and an understandable reference on the use of R for practical data analysis.

Bret Larget
University of Wisconsin—Madison
`larget@stat.wisc.edu`

Changes in R 1.8.0

by the R Core Team

MacOS changes

- As from this release there is only one R port for the Macintosh, which runs only on MacOS X. (The 'Carbon' port has been discontinued, and the 'Darwin' port is part of the new version.) The current version can be run either as a command-line application or as an 'Aqua' console. There is a 'Quartz' device `quartz()`, and the download and installation of both source and binary packages is supported from the Aqua console. Those CRAN and BioC packages which build under MacOS X have binary versions updated daily.

User-visible changes

- The defaults for `glm.control(epsilon=1e-8, maxit=25)` have been tightened: this will produce more accurate results, slightly slower.

- sub, gsub, grep, regexpr, chartr, tolower, toupper, substr, substring, abbreviate and strsplit now handle missing values differently from "NA".

- Saving data containing name space references no longer warns about name spaces possibly being unavailable on load.

- On Unix-like systems interrupt signals now set a flag that is checked periodically rather than calling longjmp from the signal handler. This is analogous to the behavior on Windows. This reduces responsiveness to interrupts but prevents bugs caused by interrupting computations in a way that leaves the system in an inconsistent state. It also reduces the number of system calls, which can speed up computations on some platforms and make R more usable with systems like Mosix.

Changes to the language

- Error and warning handling has been modified to incorporate a flexible condition handling mechanism. See the online documentation of `tryCatch()` and `signalCondition()`. Code that does not use these new facilities should remain unaffected.

- A triple colon operator can be used to access values of internal variables in a name space (i.e. `a:::b` is the value of the internal variable `b` in name space `a`).

- Non-syntactic variable names can now be specified by inclusion between backticks `` `Like This` ``. The `deparse()` code has been changed to output non-syntactical names with this convention, when they occur as operands in expressions. This is controlled by a `backtick`