Reviewer: Dirk Eddelbuettel
Debian Project

R in Action

The book *R in Action* by Robert Kabacoff is a very good introduction to R. This short review will highlight several strong points of the book, and also mention a number of shortcomings which unfortunately detract from what is overall a very commendable introduction.

The book is divided into four parts. The first, titled "Getting Started", provides a gentle introduction into getting and installing R. This is followed by a chapter on creating datasets, and another with an initial introduction to graphs. Two chapters on basic and advanced data management – which also cover data structures, control flow and functions as a first introduction to programming – round out this part.

The second part is the shortest of the four. It contains a simple overview of basic statistical methodologies. It focuses on descriptive statistics and plots, as well as basic inference. This is followed by the third part "Intermediate Methods" which is the longest of the four parts. Comprised of five chapters, it begins with regression analysis and covers a wealth of material including regression diagnostics and post-regression analysis. A chapter on analysis of variance follows, after which power analysis comes back to inference before a chapter on intermediate graphs completes this part.

The final part focuses on advances methods and contains four chapters. Generalized linear models extend the previous coverage on regression. Principal component and factor analysis cover multivariate statistics. The third chapter discusses missing data problems in some depth before a last chapter comes back to graphics with a discussion spanning **lattice**, **ggplot2**, and interactive graphics.

Besides these four core parts, the book also contains eight short appendices. These provide a (useful but somewhat mixed) collection of tips and discussions ranging from graphical user interfaces, customization of the R startup to data and graph export, a reminder on matrix algebra, an enumeration of all the many packages used in the book and two more sections of large data as well as on updating an R installation. This all is good material, but one wonders if it could not have been folded into the main section.

Throughout the book, Kabacoff displays a fine mastery of R idioms which reveal a genuine

familiarity with the material, and the underlying statistics. This is indeed a book one can learn from on a number of issues.

I found only a few minor issues: The discussion of annotating data in Chapter 2 could have benefitted from mentioning the `comment` or `attributes` functions. Similarly, the recoding of the age variable in Section 4.3 could be done more compactly using `cut`. In Table 5.5, the chart displaying a normal distribution uses a fixed set of points from `pretty` leading a to piecewise linear curve even though `curve(dnorm)` could have been used for a smoother display. Chapter 5 calls looping inefficient and suggests `apply` functions though this is no longer as clear-cut as it once was given the improvements to the R parser.

Similarly, a few issues reveal some shortcomings in the editing of the book. Plots appear to be using different default fonts even within the same chapter. Some graver errors remain: The derivation of the ordinary least squares esimator in Section 8.2 misses a minus sign in the last and major equation. And maybe worst of all, in a book filled with countless citations of other works, there is no bibliography or reference section to actually provide details of the quoted works.

These shortcomings notwithstanding, Kabacoff has provided a very good book which introduces R in a hands-on manner covering an admirable breadth of material. *R in Action* should find broad use in many introductory and even intermediate courses on R as well as for self-study by researchers and practioners alike.

Reviewer:

Dirk Eddelbuettel
Debian Project
Chicago, IL, United States of America
E-mail: edd@debian.org
URL: http://dirk.eddelbuettel.com/