



Journal of Statistical Software

February 2018, Volume 83, Book Review 2.

doi: 10.18637/jss.v083.b02

Reviewer: James E. Helmreich
Marist College

The R Primer (2nd ed.)

Claus Thorn Ekstrøm
Chapman & Hall/CRC, Boca Raton, 2017.
ISBN 978-1-138-63197-7. 408 pp. USD 59.95 (P).
<http://www.rprimer.dk/>

Introduction

R help pages can be quite dense for the novice (and more experienced users as well). Vignettes can be long and are written by package, which may or may not coincide with a specific topic. Here, with *The R Primer*, we have a how-to text presenting brief introductions to a broad array of topics. The text is written primarily for those with an idea of what they need to accomplish, and who wish to discover how to write the necessary R code and understand the output provided. I believe it will also be quite useful as a reference for those with a beginning to intermediate level of familiarity with R. While some of the methods described in *The R Primer* are introductory, others go well beyond that level into fairly sophisticated modeling, graphical, and data manipulation applications. But in the sense of a basic introduction to how various methods are implemented in R, the book is an excellent primer indeed.

Broadly, the text covers four areas: importation and manipulation of data; statistical analyses; graphics; and basic information on help, packages, and the workspace. The writing is clear and easy to follow. Each section begins with a brief statement of a problem, and continues with enough discussion, code, and analysis of R output (as appropriate to the topic) so as to give the reader a starting point for their own analyses.

Data importation and manipulation

The first two chapters concern data import and export, and manipulation. Examples are given of importing many different formats, including **Excel** spreadsheets, **SAS**, **SPSS**, **Stata** and **SYSTAT** files. Others include formats frequently found on the web and from various databases: **HTML**, **XML**, **JSON**, **SQL**. The chapter on data manipulation covers the basics of vector operations in R, the **apply** family of functions, dealing with missing values, and even such prosaic problems as a French one and a half (1,5) to a US one and a half (1.5), as well as calculating the area under a curve. There are brief introductions to matrix and

array operations, how R works with dates and times, as well as very basic structures such as factors, dataframes and lists.

Statistical analyses

This is the longest by far of the chapters, running nearly 200 pages. Though in a short text it is impossible to cover every type of analysis, Ekstrøm has chosen a quite comprehensive list to address. One can quibble about the topics covered (e.g. instrumental variables regression and propensity score analysis are omitted). But to my mind a more serious omission is any mention of Bayesian techniques and interfacing with **JAGS** or **Stan**.

That said, topics do cover a wide spectrum. They are grouped into a number of categories, including: descriptive statistics; linear models; generalized linear models; analysis of repeated measurements; model validation; contingency tables; multivariate methods; resampling and bootstrapping; nonparametric methods; survival analysis; and quite a few miscellaneous topics as well. Each such category has a handful of specific methods outlined via a succinct case study. So for instance Repeated Measurements contains discussions of fitting a linear mixed-effects model; fitting the same with serial correlation; fitting a generalized linear model; a generalized estimating equation model; analyzing a time series with an ARIMA model; and decomposing a time series into trend, seasonal, and residual components. The author assumes some familiarity with the method on the part of the reader; his intent is to describe how to do something in R, and how to interpret the output.

For any given topic, the discussion tends to run between one to five pages, and typically features a short description and analysis of a dataset using the method of interest. Frequently Ekstrøm uses functions from his own package **MESS**, but he also brings to bear functions from a broad array of packages. Alerting the reader to the existence of appropriate functions and packages is one of the strong points of the text. He also mentions packages that may be useful even if he does not use them explicitly in the analysis presented. Code and output is given, with descriptions of syntax and interpretation of the output.

Graphics

Ekstrøm chooses to treat base R graphics in the *Primer*. Though Ekstrøm does mention the packages **lattice** for trellis graphics and **ggplot2** for an implementation of the grammar of graphics, these are clearly beyond the scope of this text. He begins with the **par** settings available, colors and color palettes, as well writing mathematical formulas. Along with standard higher level graphics (histograms, bar plots, box plots, scatter plots, plots for **lm** objects) more esoteric plots are suggested: violin plots, bubble plots, hanging rootograms, pyramid plots and others. Surface plots, heat maps and interactive 3D scatterplots (among others) are demonstrated as well. Readers may find smaller items helpful as well: adding error bars to a bar plot; multiple line plots; fiddling with layout, legends; multiple axes or broken ones; even Venn diagrams. Ekstrøm must pick and choose what to present, and he does so in a way that gives the reader a real sense of how flexible R graphics can be. He discusses the export of graphics in detail (including as native \LaTeX code), as well as embedding fonts in Postscript or PDF files.

The author occasionally will editorialize. In the section on creating pie charts, Ekstrøm begins by saying more or less: don't (to be fair in a less harsh manner), and includes an appropriate

reference to [Tufté \(2001\)](#). He also points the reader to what he considers a better alternative, the `waffle` function from the package of the same name.

R information

The short final chapter is about getting information. Thus finding help pages, vignettes, and seeing the source code for functions are covered. He omits any discussion of parsing a R help page, which seems a significant oversight. There is a nice treatment of obtaining the source code for hidden functions using the `getAnywhere` function. Package manipulation (installing and updating – including from repositories such as GitHub or Bioconductor – loading, unloading, viewing the contents of, etc.) is thoroughly covered. Here we also find sections on manipulation of the workspace: setting the working directory; loading packages upon startup; using session histories; locating files interactively; and working with the operating system. Many of these are the little nitty-gritty details of working with R that are easy for a beginner to ignore at first but are needed as one’s facility with R improves.

Conclusion

The information in the primer is, perhaps, available in various places online and in multiple help pages. There is real value though in having short, clear, well written presentations of important topics all in one place. But one of its more important benefits as I see it is what might be termed the serendipity of discovery: how do you find what you do not know to look for? Just as perusing a brick-and-mortar bookstore allows one to happen across an interesting book by accident that Amazon would never suggest, looking through Ekstrøm’s *Primer* allows the reader to stumble upon functions and techniques that he or she might not have known to look for at all, but nevertheless will find quite useful. For instance, though I have been working in an educational setting with S-PLUS and R for nearly two decades, I had not appreciated (or had forgotten) that plotting characters 21 through 25 could have different colored borders and centers (p. 290), or knew the specific command to load data from the clipboard (p. 8; different for a PC and a Mac). Minor points surely, but the kind of small thing that could make life a bit easier. I recall the early struggle I had with the syntax for mathematical expressions in graphics: the *Primer* discussion is brief, but gives nearly everything one needs to know (to get started at least) in about a page of examples that stikes me as far more accessible than the standard R help page. This is a useful handbook that serves well as a quick resource. It also strikes me as the best second text to use – after *Learning Base R* ([Leemis 2015](#)) – for those completely new to R) as an ancillary text for students in courses using R. I will be certainly be recommending *The R Primer* to my students in the future.

References

Leemis LM (2015). *Learning Base R*. Ascended Ideas.

Tufté ER (2001). *The Visual Display of Quantitative Information*. 2nd edition. Graphics Press.

Reviewer:

James E. Helmreich

Marist College

Department of Mathematics

3399 North Road

Poughkeepsie, NY 12514, United States of America

E-mail: James.Helmreich@Marist.edu

URL: <http://foxweb.marist.edu/users/james.helmreich/>