



---

# Journal of Statistical Software

January 2016, Volume 69, Book Review 4.

doi: 10.18637/jss.v069.b04

---

Reviewer: James E. Helmreich  
Marist College

---

## Learning Base R

Lawrence M. Leemis

Lightning Source, LaVergne, 2016.

ISBN 978-0-9829174-8-0. 263 pp. USD 29.99 (P).

<http://www.math.wm.edu/~leemis/>

---

Many institutions require some level of technology instruction for their students, the nature of which will vary of course by subject area. For students in STEM fields this likely means an introduction to some high-level programming language and certain basic coding techniques (if, for, while loops, writing functions, etc.). For most STEM fields, an exposure to statistics is also a major requirement. Combining the two in an introductory data analysis or statistics course using R is a reasonable approach. *Learning Base R* would serve as an excellent ancillary text for such a course; indeed it should be considered for the same purpose with a linear algebra course as well. Leemis' text is a very straightforward tutorial that is designed for use in (computer) laboratory settings. The level is such that even (first year) students with no computing experience beyond gaming and social media should find it accessible. It is easily the most accessible text in this sense of which I am aware. Its use as an introduction for STEM field students to scientific programming with R is well worth considering.

Leemis acknowledges the origins of R as a statistics platform, but presents R for the most part as a general scientific programming language. The book is divided into 28 short chapters in black and white. They take typically 15 minutes to work through, though some of the later chapters are quite a bit longer. The intention is that the reader reproduce the commands shown in the chapter and then play around a bit with them to get a handle on what they are actually doing. Exercises at the end of each chapter are straightforward questions involving use of the chapter commands and techniques. The exposition in each chapter ends with a list of the new commands used, as well as other related commands (e.g., having discussed `t.test` in Chapter 26, he lists `wilcox.test`, `chisq.test` and several others).

The book starts out with the very basics, downloading R and elementary calculator operations. He follows with chapters on the idea and basic manipulation of objects, vectors, matrices and arrays. Each of these chapters is quite brief at between 5 to 8 pages. A full period lab should be able to cover two or more of these chapters. Alternatively, these short chapters could be incorporated as a small part of daily presentations – the book is quite flexible in how it could be used in a course. Curiously a discussion of the principal data structure in R – *Data Frames* – is put off until Chapter 16.

Simple *User-Written Functions* are introduced in Chapter 8. There is an exceptionally nice group of exercises here, e.g., writing functions to calculate harmonic, geometric and quadratic means. Chapter 9 – *Utilities* – is an odd collection of topics: use of the history command, changing the prompt, a discussion of floating point presentations, and how to call a help page. To my mind a discussion of the workspace and setting the working directory fit here but are put off until a later chapter. The explication of a help page is somewhat perfunctory, especially given how dense they can be to a new student – the entire chapter could have been devoted to the topic. Chapter 10 introduces *Complex Numbers*.

The next several chapters start to get into topics that provide a strong flavor of the language. We have chapters on *Character Strings* (including names and factors), *Logical Elements*, *Relational Operators*, very short chapters on *Coercion* and *Lists*, and finally *Data Frames*. These chapters provide the basics of creating, working with, manipulation and extraction of elements from these objects.

Chapters 17 and 18 discuss datasets that arrive with R, as well as the importation of external data files. This latter is a complicated subject that is surveyed quite (perhaps too) briefly. But so much of this is syntax- as well as platform-dependent that it may well be best left to the individual instructor. Here we get introduced to setting the working directory and commands to save output as files. Leemis neglects to mention that it is easy to copy and paste from the GUI into an editor, a minor quibble.

Chapter 19 introduces the four-prefix families of probability functions. Brief mathematical discussions of the binomial, Poisson, uniform and normal distributions are provided, followed by easy calculations to show R ‘gets it right.’ He lists over a dozen other families as well for informational purposes.

Two chapters cover *High-Level Graphics* and *Custom Graphics*. The former includes histograms, boxplots, normal QQ plots, scatterplots, as well as contour, image, perspective and several other lesser used plots for quantitative data; pie, dotchart, bar for categorical data, and a time series plot. None are gone into in great detail, but showing the ease that fairly complex graphics can be created is the main point, and is well done. The *Custom Graphics* chapter might be subtitled ‘fiddling with the details’ (customizing axes, adding text and a legend, etc.). It would be impossible to cover everything that can be done, but Leemis does an admirable job of giving an idea of the possibilities. The chapter finishes with four nice examples (without code) that endeavor to ‘tell the story’ without extensive introduction. These succeed admirably in providing a *raison d’être* for the whole high-level graphics enterprise. Not given, but in this vein, it would have been nice to show what is possible with a reference to the R Graphics Gallery <http://www.r-graph-gallery.com/all-graphs/>. Also, while the text is in black and white, a few comments about using color (as for instance depicting a group factor with several levels on a scatterplot) would seem to be in order but are omitted.

From here, Leemis presents four chapters on scientific programming topics: conditional statements (`if`); looping (`while`, `for` and `repeat`); recursively defined functions; and a chapter on simulation. The middle two chapters provide code for several well known algorithms; these include: the sieve of Eratosthenes, Bubble Sort, the Babylonian square root algorithm, Binary Search and Tower of Hanoi. Basic debugging techniques are also discussed. The presentation is quite readable and clear.

The next two chapters are subject oriented: *Statistics* and *Linear Algebra*. The chapter on statistics discusses confidence intervals ‘by hand’ as well as `t.test`; the Kolmogorov-Smirnov

test (with a listing of several other standard test commands); an introduction to linear modeling using `lm`; and a very nice and comprehensive discussion of formula syntax. The linear algebra chapter includes everything necessary for the first undergraduate course in the subject. Recall that a much earlier chapter introduced matrices; here we find R methods for elementary operations, inverses, determinants, eigenvalues and vectors, and matrix decompositions.

The final chapter, as befitting a text on *base* R, introduces the reader to packages. Both base and the installation of contributed packages are discussed, as well as navigation of the help files associated with each package. Oddly (to me at least) he picks a few commands from here and there and spends several pages using them. I would have preferred a few sentences on each of (what he considers to be) the important packages and what they contain. He does refer the reader to CRAN task views, an important resource, but some information on **MASS**, **hmisc**, **mosaic**, **lattice**, **ggplot2** and others would seem to be in order for those new to R.

As the author points out, there is no natural progression through topics presented here. He has made reasonable choices, but others might differ. To a large extent chapters are standalone, and could easily be omitted or rearranged. Certainly R finds its main use in the field of statistics, but Leemis succeeds in showing its applicability in other fields (especially linear algebra), at least at introductory levels. The text is well written, clear and concise. The editing is extraordinary (I found but one typo on a very close read).

Leemis' text provides an excellent introduction to R, especially suitable for those learning their first high-level programming language. The book provides not just an introduction to R, but to general scientific programming. For those interested in getting to know R on their own it is a very gentle introduction. As a laboratory manual, short course, or ancillary course text it would excel.

### 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](mailto:James.Helmreich@Marist.edu)  
URL: <http://foxweb.marist.edu/users/james.helmreich/>