Journal of Statistical Software

July 2022, Volume 103, Book Review 1. doi: 10.18637/jss.v103.b01

Reviewer: James E. Helmreich
Marist College

Learning Base R (2nd Edition)

Lawrence M. Leemis
Lawrence M. Leemis, Department of Mathematics, The College of William and Mary, Williamsburg, Virginia, 2022.
ISBN 978-0-9829174-5-9. 356 pp. USD 40.00 (P).
http://www.math.wm.edu/~leemis/

I am very glad to see that a second edition of *Learning Base R* (LBR) has been published. I liked the first edition and used it many times in an introductory data science course. This course is taken primarily by mathematics, data science, finance and science students. The text serves well as a lab manual for the course. The chapters are short and I typically use two or three chapters for a weekly lab. It would be easy to incorporate material into daily lectures as well. I have students input and experiment with the LBR code as well as work exercises. Chapters cover diverse basic topics, including: vectors, arrays, matrices, using and writing functions, logical operations, coercion, data frames, iteration, recursion and simulations. There are longer chapters covering probability, statistics and linear algebra as well as diverse other topics. When the first edition came out, I wrote a more detailed positive review in this journal, see Helmreich (2016). The second edition has been expanded by nearly a hundred pages and a similar number of new exercises. Many new commands are introduced as well.

The most important and best improvements are the new exercises. In the first edition, some of the short chapters had only two or three or, in the case of the chapter on packages, only one question. Every chapter in the second edition has new exercises, on average over nine new exercises per chapter (the packages chapter now has fifteen). They are quite good, Leemis has made them both creative and engaging.

Leemis is frequently able to write questions that build upon and extend or introduce new material or ways of thinking to the student. He emphasizes how operations can be vectorized in R. For instance, my students who have had some coding experience will look at a problem such as calculating

$$\sum_{i=1}^{15} \left(\frac{2^i}{i!} - \frac{\cos(3i)}{i^4} \right),$$

and will want to set up a loop for the calculation. The charge however is to find the result using two commands, or something like:

```
R> i <- 1:15
R> sum(2^i/cumprod(i) - cos(3 * i)/i^4)
```

Using vectors this way is an important and useful aspect of R programming, and Leemis puts considerable emphasis upon it.

As another example he asks students to calculate

$$\sum_{i=1}^{9} \sum_{j=i+1}^{9} \frac{2^i 3^j}{(i+j)!}.$$

Leemis provides a hint, suggesting use of the `outer` function. This command is new to students, so they will need to read and interpret the help file in order to use the new command. This sort of thing is excellent: no one person knows how to use more than a small percentage of what is available in R. One of the most important things I try to instill in my students is the ability to discover and learn to use new functions on their own. It is nice to find an educator of a similar mind in Leemis.

No new chapters have been added in the second edition, though all have been expanded and had light revisions and editing improvements. Chapters are generally structured in a similar manner to the first edition, however a few have had more significant revisions. The *Coercion* chapter has separate sections on the `is` and `as` families of functions, and works at length with dates. The *Iteration* chapter includes a new section on algorithm development, as well as separate sections on debugging and optimization. The *Relational Operators* chapter contains a new section on the `switch` function. The longer *Statistics* chapter has seen serious revision and addition of new material. New to this chapter are the a discussion of ANOVA; regression including non-linear and logistic; as well as time series analysis. Leemis's focus is on how to perform these analyses in R and so assumes the reader has the appropriate statistics background. Finally the *Packages* chapter has been restructured and lengthened considerably, with new material on **ggplot2** and **lubridate**.

Overall, Leemis has not changed how he presents the methods and commands. These were already carefully done and are clear. What he has succeeded in doing is adding more explanatory and motivational material. For instance, in the *Arrays* chapter, he motivates why arrays might be useful with new applications in meteorology and photography. The new *Algorithm Development* section of the *Iteration* chapter similarly looks holistically at the process of creating an algorithm, using two well known examples.

Leemis's text continues to be my go-to suggestion when asked about an accessible introduction to R. The second edition improves upon an already nice treatment; I am glad to have it on my bookshelf.

References

Helmreich JE (2016). "Learning Base R." *Journal of Statistical Software, Book Reviews*, **69**(4), 1–3. doi:10.18637/jss.v069.b04.

Reviewer:

James E. Helmreich
Marist College
Department of Mathematics
3399 North Road
Poughkeepsie, NY 12601, United States of America
E-mail: James.Helmreich@Marist.edu
URL: http://foxweb.marist.edu/users/james.helmreich/

Journal of Statistical Software https://www.jstatsoft.org/
published by the Foundation for Open Access Statistics https://www.foastat.org/
July 2022, Volume 103, Book Review 1 *Published:* 2022-07-13
doi:10.18637/jss.v103.b01