

Journal of Behavioral Data Science, 2021, 1 (1), 170–172.
DOI: <https://doi.org/10.35566/jbds/v1n1/p7>

Book Review: Mastering Software Development in R

Kévin Allan Sales Rodrigues^[0000–0003–4925–5883]

University of São Paulo, São Paulo, Brazil
kevin@usp.br

Book review of **Mastering Software Development in R** by Roger D. Peng, Sean Kross and Brooke Anderson (2017). Victoria, British Columbia, Canada: Leanpub. 472 pages. Price \$0.00 to \$50.00 (e-book).
<https://leanpub.com/msdr>

The book **Mastering Software Development in R** is an excellent introduction to the use of R (R Core Team, 2020) software, and its focus is on teaching how to develop packages for R and how to create complex graphs with *ggplot2*. The package *ggplot2* is one of the most popular and downloaded R packages and was created in 2005 by Hadley Wickham as a data visualization package for the statistical programming language R based on the grammar of graphics—a general scheme for data visualization which breaks up graphs into semantic components such as scales and layers. In particular, *ggplot2* can serve as a replacement for the base graphics in R and contains a number of defaults for web and print display of data. The book covers a wide variety of other packages including: *choroplethr*, *choroplethrMaps*, *data.table*, *datasets*, *devtools*, *dlnm*, *dplyr*, *faraway*, *forcats*, *GGally*, *ggmap*, *ggthemes*, *ghit*, *GISTools*, *grid*, *gridExtra*, *httr*, *knitr*, *leaflet*, *lubridate*, *magrittr*, *methods*, *microbenchmark*, *package*, *pander*, *plotly*, *profvis*, *pryr*, *purrr*, *rappdirs*, *raster*, *RColorBrewer*, *readr*, *rmarkdown*, *scales*, *sp*, *stats*, *stringr*, *testthat*, *tidyr*, *tidyverse*, *tigris*, *titanic*, and *viridis*.

Developing R packages and making specialized statistical graphics are very relevant skills today because as new models and statistical methodologies emerge, there must be software available to apply the cutting edge theory to real problems. In addition, publishing packages on CRAN (comprehensive R archive network) is a way of scientific dissemination that can increase the impact of a scientific research because the packages allow for quick applications of the methodology developed in the research. The importance of data visualization is obvious and a well-thought-out graph can synthesize a lot of information (descriptive or inferential) clearly and intuitively.

The book has at least three main advantages: it is affordable (can even be acquired for free), it serves both as an introductory book for R and as a “bridge”

for more advanced books such as: Wickham (2015), Wickham (2016), Wickham (2019) and Xie, Allaire, and Grolemund (2018), and it goes straight to the point, allowing for faster and more fluid learning. It is an ideal book for anyone who wants to learn advanced R topics without investing a lot of time. I believe that this book is important for anyone who is starting to develop their own packages (including experienced researchers who are not familiar with programming or software development).

Now, I want to compare the book with four other books on R. Wickham (2015) is the reference book for anyone who wants to learn how to create their own R packages. It covers all the steps of creating a package, from organizing function codes to disseminating the package. Wickham (2016) is a book on *ggplot2* written by the main author of this package and consequently is a reference book on *ggplot2*. Wickham (2019) is a book that addresses more advanced R topics, such as metaprogramming and techniques to improve the performance of R codes. Xie et al. (2018) is the first official book authored by the core R Markdown developers that provides a comprehensive and accurate reference to the R Markdown ecosystem. The four books together cover the same content as the reviewed book and each was written by the package developers themselves or at least by people who have contributed a lot to the area corresponding to its content. The reviewed book is able to provide a broad overview of several important R topics but clearly does not offer the depth of a reference book. The reviewed book is great for beginning learners of the topics that are not covered in a first R course and the books mentioned here are useful if deeper understanding of any of the topics covered in the reviewed book is needed.

The book contains a brief introduction and 4 chapters with well-defined scopes. The introduction states the R packages that will be used in this book. Chapter 1 covers the introduction to R and how to clean and to tidy data. Chapter 2 covers introductory programming topics, such as if, else and object-oriented programming and other more advanced topics like profiling and benchmarking, robust error handling and debugging. Chapter 3 deals with building packages for R and covers R package development, writing good documentation and vignettes using *knitr* and *R Markdown*, writing tests for an R package using the *testthat* package, continuous integration¹ tools such as Travis and Appveyor, and distributing packages via CRAN and GitHub. Finally, Chapter 4 covers building graphics with the *ggplot2* package, creation of simple and dynamic maps, creation of new *ggplot2* theme by modifying an existing theme, creation of new geom function to implement a new feature or simplify a workflow, and other related topics.

Each of the book's 4 chapters begins with a description of what will be learned in a short paragraph and follows by a list of topics covered in the chapter. At the beginning of each section, there is also a list of topics covered, except for the sections of Chapter 4 and Sections 2.8, 3.1 and 3.10. These sections without a list

¹ The topic of continuous integration is a little known topic in the statistical community and has the role of ensuring that the package continues to function properly after successive updates.

of topics covered are either conceptual in nature or the title is self-explanatory. The list of topics covered helps navigate the book and learn exactly what you want without having to read the entire book. This makes the book to be a quick reference to find information in short time. Although the book does not include exercises, it constantly encourages the readers to experiment with variations of the codes presented and for that it is enough to copy, paste and edit the codes of the book itself in R. For readers who have already taken an introductory R course or acquired basic knowledge of R through practice, Chapters 3 and 4 will certainly be most interesting. Chapter 3 provides technical details for the development and publication of packages on CRAN and GitHub. Publishing packages on CRAN and GitHub are great ways to share R code with the scientific community, but it's also possible to simply share the package in a standardized way at one's company, university, or research institute. Chapter 4 teaches how to develop customized visualization tools through packages such as *ggplot2* and *ggmap*.

In general, the book achieved the proposed objectives of what? Some changes can be made to improve a reader's experience. For example, it can list the R packages required for each of the 4 chapters so the reader can prepare the computational environment in advance for a specific chapter, such as Chapter 3 or 4. It also benefits by adding an index to show the pages on which each package it is mentioned since the book has more than 400 pages. By doing so, the reader interested in using a specific package can be directed more quickly to the examples of the desired package.

Acknowledgements

The author gratefully acknowledge CNPq (Brazilian National Council for Scientific and Technological Development) for the financial support.

References

- R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org>
- Wickham, H. (2015). *R packages: Organize, test, document, and share your code*. O'Reilly Media, Inc.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wickham, H. (2019). *Advanced r (2nd ed.)*. CRC press.
- Xie, Y., Allaire, J. J., & Golemund, G. (2018). *R markdown: The definitive guide*. CRC Press.