



---

# Journal of Statistical Software

August 2019, Volume 90, Book Review 2.

doi: 10.18637/jss.v090.b02

---

Reviewer: Ulrike Grömping  
Beuth University of Applied Sciences Berlin

---

## Practical R for Mass Communication and Journalism

Sharon Machlis

Chapman & CRC, Boca Raton, 2019.

ISBN 978-1-138-72691-8. 246 pp. USD 47.96 (P), USD 140.00 (H), USD 53.96 (e).

<http://www.machlis.com/R4Journalists/>

---

*Practical R for Mass Communication and Journalism* lists three target audiences: “spreadsheet users who want to ‘graduate’ to learning their first programming language, people who know another programming language and now want to learn R and communications professionals who already know some R but want to get some new tips and ideas for using R *in a newsroom or similar setting*” (quoting and emphasis in the original). I think that journalists and other communication professionals from all these audiences will indeed benefit from this book; absolute beginners from other application areas may also find the book a useful resource, since the examples are of general interest and the book does not assume its readers to have any programming knowledge.

The book is structured into 18 chapters of different lengths (from 3 to 28 pages): the brief “Introduction” chapter takes the role of a preface, the second chapter explains how to install R and **RStudio** under Windows and MacOS operation systems and how to obtain the book’s accompanying data and online code. In line with the intended audience, Linux is not covered. Subsequent chapters give appetizers (“See How Much You Can Do in a Few Lines of Code”), cover importing data (Chapter 4), basic data exploration with calculation (Chapter 5), basic visualization of a single numeric variable, possibly for several categories (snow heights in several US cities, Chapter 6), combination of data sets (Chapter 7), grouped analyses in numbers (Chapter 8) and graphs (Chapter 9), writing R functions (Chapter 10), creating maps (Chapter 11), combining previous insights on election data (Chapter 12), date calculations (Chapter 13) and reshaping of data (Chapter 14), R markdown (Chapter 15), web scraping (Chapter 16), a project from start to finish (Chapter 17) and a brief final chapter with hints to additional resources, including further functions, packages and tools, example stories, tutorials and social media resources. Chapter titles try to make dry material sound interesting, e.g. “Help! My data’s in the wrong format!”. Appendices list tasks (How do I . . .), functions and packages; these are printed in the book, but are also available as searchable files online.

The book’s approach is example-based, which should keep the audience interested. Examples include amounts of snowfall in Boston, Chicago and New York City, airline and airport flight reliability data, US house prices or election results. The initial chapters concentrate on general

introductions to basic R concepts, while later sections highlight specific capabilities, always with a view to what a journalist might be interested in. An example-based approach implies that the structure of the exposition follows what's needed for the examples which may not always be what one would teach first from an R perspective. The benefit of always knowing why some R functionality is needed, presumably outweighs the disadvantages. However, in some places the section titles could have been adjusted to better reflect section content: for example, Section 6.5.1 is named “Base R Graphics”, but solely discusses data preparation with R package **dplyr**.

Each chapter initially states the R packages that will be used in the chapter, and closes with a “Wrap-Up” and an “Additional Resources” section; these are quite useful. The initial section might become even more useful by including brief explanations of the purposes of the packages, as I repeatedly found myself googling for the purpose of a listed package.

In terms of packages used for data handling, the R universe is meanwhile quite rich and diverse. The book mainly relies on the **tidyverse** and related packages, though it also introduces (inelegant variants of) base R solutions. Packages heavily used by the book are **rio** for data importing and exporting data, **janitor** for cleaning data, **dplyr** for manipulating data, **pacman** for loading packages, **ggplot2** and relatives for graphics. The book touches upon many special topics without really covering them, in order to give readers an impression and ideas what they might be willing to learn later; for example, it very briefly covers code testing (with package **testthat**) or alludes to text manipulation with regular expressions.

It is not easy for me to embody an absolute R novice; I suspect that novice readers must possess a solid willingness to embark on programming, in order to digest the book; the book's example-based approach will certainly support them in keeping up motivation. I assumed the role of a learner with regard to **ggplot2**, which I had so far largely ignored. Some of the book's explanations made it quite easy to get an intuitive grasp of the **ggplot** approach, and I am now more willing than before to use it, because it has some real benefits for creating modern visualizations. For example, I learnt about making graphs interactive with the help of package **ggiraph** (the **sendplot** solution for base R graphics is much more complicated).

Data visualization is among the best-covered topics in the book, which of course makes a lot of sense for communication applications. The two longest chapters are “Beginning data visualization” and “Maps in R”. From a teaching perspective, I appreciate the book author's decision to provide non-satisfactory interim results, e.g., in Figures 6.17 to 6.21, where she develops a bar chart into its final form of Figure 6.22. This approach should make it easy for readers to follow the learning steps and feel good about having achieved a result, even if it is not yet perfect. While Figure 6.22 shows the final result of the process (except for the axis title “reorder(Winter, -TotalSnow)”, which is not yet production quality; Figure 15.5 eventually shows a production quality graph for these data), many visualizations in the book show states that one would not want to publish, be it because of annotation sizes, annotation content, or axis scaling. I would have preferred to always see a finished result at the end of the process, if only for acting as a role model.

The chapter on maps has a wealth of useful information, not only on the creation of maps but also on adding geographical information to existing data or handling map data in various formats. It is unavoidable that some of the information ages fast, because access to real-life data subject to change. For example, I wanted to try out the book's geocoding hints with some German addresses and found myself stuck because one now has to register with

Google for obtaining an API key before the geocoding functionality can be used. This is also mentioned on the book's accompanying website, where the author thankfully warns that Google will request credit card information even from users who are happy to only use free geocoding services. This piece of information saved me from investing effort in finding out how to register for an API.

Overall, this book can be recommended for its target audiences. It contains a lot of useful information, which is presented with interesting examples. The accompanying website is kept up-to-date. For users who anticipate to use the book with some intensity, it may be worthwhile to invest in a hardback copy, since my paperback copy has already got some loose pages.

**Reviewer:**

Ulrike Grömping

Beuth University of Applied Sciences Berlin

Department II

D-13353 Berlin, Germany

E-mail: [groemping@bht-berlin.de](mailto:groemping@bht-berlin.de)

URL: <http://prof.beuth-hochschule.de/groemping/>