



Journal of Statistical Software

November 2015, Volume 68, Book Review 5.

doi: 10.18637/jss.v068.b05

Reviewer: David W. Zeitler
Grand Valley State University

Graphical Data Analysis with R

Antony Unwin
CRC Press, Boca Raton, 2015.
ISBN 978-1-4987-1523-2. 310 pp. USD 59.46.
<http://www.crcpress.com/9781498715232>

This text has the potential of bringing sophisticated visualization to a broad audience without resorting to mathematical formalizations or the skills of a graphics artist. It engages the reader with interesting graphics right from the start and overall is clear and unimimidating. Code for all examples is provided in the text and is available on a supporting website. What's more, the code works as is, rather unusual and refreshing.

Graphics code in the book is based on the **ggplot2** package providing sophisticated and appealing graphics without resorting to special purpose code. Use of the **dplyr** package for data manipulation makes the necessary task of data manipulation much simpler.

The structure of the text provides a logical straightforward introduction to graphical data analysis starting with single continuous and categorical variables progressing to bivariate and on to multivariate data. This is followed up with big picture (overview) graphics, time series, data quality (missing values and outliers) and comparison graphics (simple dashboards). Chapters three through eleven provide short sections on modeling and testing for the types of structure illuminated by the graphics of the chapter, an overview of main points of the chapter and exercises based on readily available datasets to help the reader explore the concepts. Each exercise set ends with an 'intermission' exercise inviting the reader to consider a question or three about a well known work of art.

Not everything is perfect. The time series chapter could have been longer in my opinion. It's more important than the brief treatment indicates. More could have been done with ensemble graphics (dashboards) and a separate collection of case studies would have been a useful addition. Case studies are provided but are mixed in with the ensemble graphics chapter. There also could have been a discussion of graphics and big data. Efficiently handling large to huge data sets is becoming a critical skill for the data analyst and often requires specific approaches to graphical analysis.

One last note that should probably go with any text using R. The nature of open source software is that it's always changing. This constant change is both a blessing and a curse. One could with little doubt run SAS code written back in the mid 1970's on today's SAS

system. The base R language and packages are quite stable, however newer R packages can and do change within months. Most of the time backward compatibility is maintained, but the potential for unintended compatibility issues is greater in open source software than it is in relatively static software. Hence the recent introduction of **packrat** from the team at RStudio and **checkpoint** from Revolution Analytics which help to manage change within a users R environment. That being said, the text code uses well-written and stable packages that are not likely to suffer from backwards compatibility issues, but it still needs to be diligently maintained either in reprints and editions or on the website.

Reviewer:

David Zeitler
Grand Valley State University
Statistics Department
1 Campus Drive
Allendale, MI 49401, United States of America
E-mail: zeitlerd@gvsu.edu
URL: <http://faculty.gvsu.edu/zeitlerd/>