Data Visualisation with R

Thomas Rahlf

Data Visualisation with R

111 Examples

Second Edition

Springer

Thomas Rahlf
Rheinische Friedrich-Wilhelms-Universität
Bonn
Bonn, Germany

ISBN 978-3-030-28443-5 ISBN 978-3-030-28444-2 (electronic)
https://doi.org/10.1007/978-3-030-28444-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface to the Second English Edition

Due to the continued interest in the book, Springer Verlag has decided to publish an English translation of the second edition. Again, many thanks to Ralf Gerstner from Springer, and also to Tracey Duffy for translating all of the additions.

In 2017, Ulrike Grömpling reviewed the English translation of the first edition in the *Journal of Statistical Software*. Inspired by the book, and following the release of the second edition, she has published an R package “prepplot” which greatly simplifies the configuration of figure regions for base R graphics. I highly recommend trying this one out.

Meanwhile, the third edition of Paul Murrell’s standard work *R Graphics* has been published. When I read in his foreword that my book—among others—was one inspiration for restructuring part of his book, I felt very honored.

Bonn, Germany Thomas Rahlf

Preface to the Second German Edition

The feedback on the first edition of this book (published by Open Source Press in 2014 and now out of print), and on the English edition that has been published in the meantime, was so pleasing that I was happy to accept the offer made by Springer Verlag to publish a new, updated edition of the book. The concept of explaining complete examples and the restriction to Base Graphics have been retained. Compared to the first edition of the book, this new, updated edition contains 11 additional, new examples, hence the new subtitle "111 Examples". There are two main additions: firstly, a section on visualising network relationships has been added to the chapter on categorical data. In addition to examples of classic network diagrams, an adapted heat map, and a multiple bar chart, this section also contains a chord diagram and a riverplot. Although the chord diagram and riverplot may appear unusual at first glance, they can now be found in publications of respected scientific journals such as Science, Nature or Cell. Examples on the use of georeferenced grid formats and on cartograms have been added to the chapter on maps. Three examples for the integration of data created with R in interactive JavaScript illustrations have also been added to the book. R now provides multiple concepts and packages that can be used to create JavaScript visualisations more or less directly. Ultimately, such packages form a type of container in R. In each case, a specific syntax developed by the authors of these packages translates the scripts written in this form into the notation required for the underlying JavaScript library. This means that we have to rely on the scope of the language of the R package and on the quality and flexibility of the translation routines. I am not sure if this is the right path. In this book, I have taken a different path. In three examples in Chap. 12, the data are prepared with R such that they are integrated in existing, only slightly adapted JavaScript code. This is done using Highcharts and Mapael, two JavaScript libraries that can be used to create very aesthetic illustrations "out of the box" with minimal change effort.

The primary objective of this book is still to explain how to create presentation graphics. For the exploratory visualisation within the scope of data analysis, I refer to the book Graphical Data Analysis with R (CRC Press, 2015) by Antony Unwin. The first people I would like to thank are Agnes Herrmann and Iris Ruhmann at

Springer, who have enabled me to create this updated edition of the book. For helpful comments, suggestions, and exchange of ideas for this edition, I would also like to thank Alberto Cairo, Martin S. Fischer, Sebastian Jeworutzki, Nikola Sander, Antony Unwin, January Weiner, and Stefan Fichtel. January Weiner has kindly included a comment in his riverplot package that is helpful for example 6.4.4.

Bonn, Germany Thomas Rahlf

Preface to the English Edition

This book is a translation of the German book “Datendesign mit R” that was published 2014 by Open Source Press. Due to the encouraging strong interest in the German edition Springer Verlag offered to publish an English translation. First of all I would like to thank Ralf Gerstner from Springer for this and for his helpful suggestions for improvement, as well as Annika Brun for translating most of the text, Colin Marsh for copy editing, and Katja Diederichs for converting all scripts from German to English. Last year I benefited a lot from a communication with Antony Unwin. His book “Graphical Data Analysis with R” can be seen as complementary to my own: while this one focusses on presentation of graphics, you will benefit from his book if you are interested in exploring data graphically.

Bonn, Germany	Thomas Rahlf

Preface to the German Edition

Some 20 years ago, when I reviewed a score of books on statistical graphics and graphic-based data analysis, things were completely different: there were proprietary formats and operating systems, their character sets were incompatible, and graphic and statistical software was expensive. Since the turn of the century, the situation has changed fundamentally: the Internet has come of age, open-source projects have attracted more and more followers, and a handful of enthusiasts provided version 1.0 of the free statistical programming language R. Many developers were inspired to collaborate on this project. R reached version 3 in 2013, and in addition to the basic software, more than 7000 freely available extension packs are currently available. Companies and organisations such as Google, Facebook or the CIA are using R for their data analysis. Its graphic capabilities are again and again emphasised as its strong point. Pretty much all technologies relevant for data visualisation are quickly integrated into R. Through numerous functions, detailed designs of every imaginable figure, creation of maps and much more are made possible. All it takes is to know how—and that is where this book wants to contribute.

What This Book Wants to Be—and What It Doesn't Want to Be

This book is not an introduction that systematically explains all the graphic tools R has to offer. Rather, its aim is to use 100 complete script examples to introduce the reader to the basics of designing presentation graphics, and to show how bar and column charts, population pyramids, Lorenz curves, box plots, scatter plots, time series, radial polygons, Gantt charts, heat maps, bump charts, mosaic and balloon charts, and a series of different thematic map types can be created using

R's Base Graphics System. Every example uses real data and includes step-by-step explanations of the figures and their programming. The selection is based on my personal experiences—it is likely that readers will find one or another illustration lacking, and consider some too detailed. However, a large scope should be covered. This book is aimed at:

- R experts: You can most likely skip Part I. For you, the examples are particularly useful, especially the code.
- Readers that have heard of R and maybe even tried R before and are not daunted by programming; you will profit from both parts.
- Beginners: for you, the finished graphics pictured here will be most helpful. You will see what R can do. Or, in other words: you will realise that there is such a tool as R, and that it can be used to create graphics you have wanted to create for a long time, but merely never knew how. The code will be too complicated for you, but you may be able to commission others to do your graphics programming in R.

Windows, Mac, and Linux

All of the scripts and working steps will yield identical results when executed in Windows, Mac OS X or Linux. All of the examples were created in Mac OS X and then tested in Ubuntu 12.04 and an evaluation copy of Windows 8.1.

Acknowledgements

The following people deserve my thanks for hints, comments, feedback, data, discussions or help: Gregor Aisch, Insa Bechert, Evelyn Brislinger, Giuseppe Casalicchio, Arnulf Christl, Katja Diederichs, Günter Faes, Mira Hassan, Mark Heckmann, Daniel Hienert, Bruno Hopp, Uwe Ligges, Lorenz Matzat, Meinhard Moschner, Stefan Müller, Paul Murrell, David Phillips, Duncan Temple Lang, Martijn Tennekes, Patrick R. Schmid, Thomas Schraitle, Valentin Schröder, Torsten Steiner, Michael Terwey, Katrin Weller, Bernd Weiss, Nils Windisch, Benjamin Zapilko, and Lisa Zhang. This manuscript particularly benefited from discussions with an infographic designer and a data journalist. Stefan Fichtel looked over every figure and provided critical feedback. For selected figures, he designed his own suggestions; this has been an invaluable help. We did not always agree, and I have disregarded his advice here or there. Therefore, any remaining errors and shortcomings are mine. Björn Schwentker went to the trouble of proof reading large parts of the manuscript. I am very grateful for his valuable notes which have surely

made some parts of the text clearer and more readable. Finally, I want to thank Markus Wirtz for tackling the experiment of ultimately printing everything into a book.

On the Internet

The figures are conceived for different final output options. The format of the book implies that some details have become very small, e.g. in maps and radial column charts. Particularly for such cases, please refer to the book's website, on which all figures are available in high resolution or as vector graphics in PDF format:
http://www.datavisualisation-r.com

Bonn, Germany
Thomas Rahlf

Contents

Part I Basics and Techniques

Part II Examples