Data Visualization with R

Rob Kabacoff

2018-09-03

Contents

Welcome

R is an amazing platform for data analysis, capable of creating almost any type of graph. This book helps you create the most popular visualizations - from quick and dirty plots to publication-ready graphs. The text relies heavily on the ggplot2 package for graphics, but other approaches are covered as well.

My goal is make this book as helpful and user-friendly as possible. Any feedback is both welcome and appreciated.

Preface

How to use this book

You don't need to read this book from start to finish in order to start building effective graphs. Feel free to jump to the section that you need and then explore others that you find interesting.

Graphs are organized by

- the number of variables to be plotted
- the type of variables to be plotted
- the purpose of the visualization

Chapter	Description
Ch 1	provides a quick overview of how to get your data into R and how to prepare it for analysis.
Ch 2	provides an overview of the `ggplot2` package.
Ch 3	describes graphs for visualizing the distribution of a single categorical (e.g. race) or quantitative (e.g. income) variable.
Ch 4	describes graphs that display the relationship between two variables.
Ch 5	describes graphs that display the relationships among 3 or more variables. It is helpful to read chapters 3 and 4 before this chapter.
Ch 6	provides a brief introduction to displaying data geographically.
Ch 7	describes graphs that display change over time.
Ch 8	describes graphs that can help you interpret the results of statistical models.
Ch 9	covers graphs that do not fit neatly elsewhere (every book needs a miscellaneous chapter).
Ch 10	describes how to customize the look and feel of your graphs. If you are going to share your graphs with others, be sure to skim this chapter.
Ch 11	covers how to save your graphs. Different formats are optimized for different purposes.
Ch 12	provides an introduction to interactive graphics.
Ch 13	gives advice on creating effective graphs and where to go to learn more. It's worth a look.
The Appendices	describe each of the datasets used in this book, and provides a short blurb about the author and the Wesleyan Quantitative Analysis Center.

There is **no one right graph** for displaying data. Check out the examples, and see which type best fits your needs.

Prequisites

It's assumed that you have some experience with the R language and that you have already installed R and RStudio. If not, here are some resources for getting started:

- A (very) short introduction to R
- DataCamp - Introduction to R with Jonathon Cornelissen
- Quick-R
- Getting up to speed with R

Setup

In order to create the graphs in this guide, you'll need to install some optional R packages. To install **all** of the necessary packages, run the following code in the RStudio console window.

```
pkgs <- c("ggplot2", "dplyr", "tidyr",
          "mosaicData", "carData",
          "VIM", "scales", "treemapify",
          "gapminder", "ggmap", "choroplethr",
          "choroplethrMaps", "CGPfunctions",
          "ggcorrplot", "visreg",
          "gcookbook", "forcats",
          "survival", "survminer",
          "ggalluvial", "ggridges",
          "GGally", "superheat",
          "waterfalls", "factoextra",
          "networkD3", "ggthemes",
          "hrbrthemes", "ggpol",
          "ggbeeswarm")
install.packages(pkgs)
```

Alternatively, you can install a given package the first time it is needed.

For example, if you execute

`library(gapminder)`

and get the message

`Error in library(gapminder) : there is no package called ‘gapminder’`

you know that the package has never been installed. Simply execute

`install.packages("gapminder")`

once and

`library(gapminder)`

will work from that point on.