

Five Tutorials for Data Visualization and Analysis with R

Ruba Aljafari

Joseph W. Clark

University of Nebraska at Omaha

Abstract:

There is a growing interest among students and faculty in learning the tools of “data science”, among them data visualization. Unfortunately from a course-design standpoint, the best textbooks (like the best data scientists) do not settle on a single tool but jump to a different software system in almost every chapter. This package of five hands-on tutorials is designed to help instructors teach a wide variety of data science activities using a single tool: the R programming language.

R is a jack-of-all-trades technology for data analysis and visualization that will allow students to do a variety of analyses without having to learn a different system or language for each one. Much in demand by employers of data scientists, it is a fine choice of tool on which to base an introductory course. We propose that instructors use existing textbooks for the “why” and use our tutorials for the “how to” either as homework or as in-class labs.

Contents of ZIP file:

- Tutorials
 - #1: Working with a single variable
 - #2: Working with two variables
 - #3: Time series analysis
 - #4: Working with three or more variables
 - #5: Acquiring data from the web
- Data sets
- Data set descriptions
- Scripts: all R code used in the tutorials

Teaching Note:

- (a) Learning Objectives:** This is a set of step-by-step tutorials for visual data analysis with R. They will help students of data science learn how to produce a variety of data visualizations in the R language: histograms, box plots, scatterplots, time series, interaction plots, heat maps, and more. They are intended to augment textbooks and lectures on data visualization that either do not give the “how to”, or provide instruction using different software systems.
- (b) Requirements:** R 2.15.1 or later version on any personal computer.
<http://www.r-project.org/>
- (c) Target audience:** Undergraduate or graduate students in a data analysis or data visualization course. Some programming experience is recommended. These tutorials alone do not cover the basics like how to install R, write and run scripts, so some additional materials on R may be useful as well.
- (d) Recommended accompaniment:** Two excellent textbooks on visual data analysis are *Data Analysis with Open Source Tools* (O’Reilly) by Philipp K. Janert, and *Visualize This* (Wiley) by Nathan Yau. Both texts have the same fatal flaw (from the perspective of an instructor): the authors are experts who switch to a different software tool in almost every chapter. We recommend using their texts for the “why” and our tutorials for the “how to”. In fact the tutorials line up well with chapters 2, 3, 4, and 5 of the Janert book.
- (e) Usage:** Tutorials may be used as in-class labs or assigned as homework. Since each tutorial includes a few different types of problems, instructors may instead break them up and use the separate tasks as examples during lectures. Tutorial #5 on acquiring data from the web may be used at any time as it does not require the others as prerequisites.