'From Statistics to Data Mining'
Computer Lab Session
General Introduction to Data Analysis with 

Master 1 COSI / CPS^2
Saint-Étienne, France

Fabrice Muhlenbach

Laboratoire Hubert Curien, UMR CNRS 5516
Université Jean Monnet de Saint-Étienne
18 rue du Professeur Benoît Lauras
42000 SAINT-ÉTIENNE, FRANCE
http://perso.univ-st-etienne.fr/muhlfabr/

1 Outline

1. Introduction to R (1/2)
2. Introduction to R (2/2)
3. Probability, Random Variables and Probability Distributions
4. Linear Algebra (1/2)
5. Linear Algebra (2/2)
6. Principal Component Analysis
7. Linear Regression (1/2)
8. Linear Regression (2/2)
9. Clustering (1/2)
10. Clustering (2/2)

2 Recommended Lectures for the Lab Sessions

2.1 R basics

- R Core Team (2017), "The R language definition", Version 3.4.1 (2017-06-30): This draft version presents the basics for language.

  Available online on the R Core Team website:
  `https://cran.r-project.org/doc/manuals/R-lang.pdf`
  Table of Contents:

1. Introduction
2. Objects
3. Evaluation of expressions
4. Functions
5. Object-oriented programming
6. Computing on the language
7. System and foreign language interfaces
8. Exception handling
9. Debugging
10. Parser

- Adler (2010), "R in a Nutshell – a Desktop Quick Reference" or Adler (2011) « R, L'essentiel » (French translation): Reference book for learning R.
  Package: **nutshell**: Data for "R in a Nutshell"
  Table of Contents:
  - Part I. R Basics
    - 1. Getting and Installing R
    - 2. The R User Interface
    - 3. A Short R Tutorial
    - 4. R Packages
  - Part II. The R Language
    - 5. An Overview of the R Language
    - 6. R Syntax
    - 7. R Objects
    - 8. Symbols and Environments
    - 9. Functions
    - 10. Object-Oriented Programming
    - 11. High-Performance R
  - Part III. Working with Data
    - 12. Saving, Loading, and Editing Data
    - 13. Preparing Data
    - 14. Graphics
    - 15. Lattice Graphics
  - Part IV. Statistics with R
    - 16. Analyzing Data
    - 17. Probability Distributions
    - 18. Statistical Tests
    - 19. Power Tests
    - 20. Regression Models
    - 21. Classification Models
    - 22. Machine Learning
    - 23. Time Series Analysis
    - 24. Bioconductor
- Allerhand (2011), "A Tiny Handbook of R": This small textbook provides a roadmap for the R language and programming environment with signposts to further resources and documentation.
  Table of Contents:
  1. Introduction to R
  2. Data Structures

3. Tables and Graphs
4. Hypothesis Tests
5. Linear Models

- Kabacoff (2011), "R in Action": This book presents both the R system and the use cases that make it such a compelling package for business developers. The book begins by introducing the R language, including the development environment. Focusing on practical solutions, the book also offers a crash course in practical statistics and covers elegant methods for dealing with messy and incomplete data using features of R.

  Table of Contents:

  - Part I Getting Started
    - 1 Introduction to R
    - 2 Creating a dataset
    - 3 Getting started with graphs
    - 4 Basic data management
    - 5 Advanced data management
  - Part II Basic Methods
    - 6 Basic graphs
    - 7 Basic statistics
  - Part III Intermediate Methods
    - 8 Regression
    - 9 Analysis of variance
    - 10 Power analysis
    - 11 Intermediate graphs
    - 12 Resampling statistics and bootstrapping
  - Part IV Advanced Methods
    - 13 Generalized linear models
    - 14 Principal components and factor analysis
    - 15 Advanced methods for missing data
    - 16 Advanced graphics

- Meys and de Vries (2012), "R For Dummies": This book is an easy-to-follow guide that explains how to use R for data processing and statistical analysis. It also gets you started with presenting your data with compelling and informative graphics. You will gain practical experience of using R in a variety of settings and get familiar with some powerful R data analysis tools.

  Table of Contents:

  - Part I: R You Ready?
    - Chapter 1: Introducing R: The Big Picture
    - Chapter 2: Exploring R
    - Chapter 3: The Fundamentals of R
  - Part II: Getting Down to Work in R
    - Chapter 4: Getting Started with Arithmetic
    - Chapter 5: Getting Started with Reading and Writing
    - Chapter 6: Going on a Date with R
    - Chapter 7: Working in More Dimensions
  - Part III: Coding in R
    - Chapter 8: Putting the Fun in Functions
    - Chapter 9: Controlling the Logical Flow

    - Chapter 10: Debugging Your Code
    - Chapter 11: Getting Help
  - Part IV: Making the Data Talk
    - Chapter 12: Getting Data into and out of R
    - Chapter 13: Manipulating and Processing Data
    - Chapter 14: Summarizing Data
    - Chapter 15: Testing Differences and Relations
  - Part V: Working with Graphics
    - Chapter 16: Using Base Graphics
    - Chapter 17: Creating Faceted Graphics with Lattice
    - Chapter 18: Looking At ggplot2 Graphics
  - Part VI: The Part of Tens
    - Chapter 19: Ten Things You Can Do in R That You Would've Done in Microsoft Excel
    - Chapter 20: Ten Tips on Working with Packages

2.2 Programming with R

- Braun and Murdoch (2007), "A First Course in Statistical Programming with R": This book introduces students to statistical programming, using R as a basis. Unlike other introductory books on the R system, this book emphasizes programming, including the principles that apply to most computing languages, and techniques used to develop more complex projects.

  Table of Contents:

  1. Getting started
  2. Introduction to the R language
  3. Programming statistical graphics
  4. Programming with R
  5. Simulation
  6. Computational linear algebra
  7. Numerical optimization

- Chambers (2008), "Software for Data Analysis: Programming with R": This book –written by one of the authors who invented **S** language (ancestor of R)– guides the reader in programming with R, from interactive use and writing simple functions to the design of R packages and intersystem interfaces.

  Table of Contents:

  1. Introduction: Principles and Concepts
  2. Using R
  3. Programming with R: The Basics
  4. R Packages
  5. Objects
  6. Basic Data and Computations
  7. Data Visualization and Graphics
  8. Computing with Text
  9. New Classes
  10. Methods and Generic Functions
  11. Interfaces I: C and Fortran
  12. Interfaces II: Other Systems
  13. How R Works

- Matloff (2011), "The Art of R Programming: A Tour of Statistical Software Design": This book is for those who wish to develop *software* in R, more than using it in an ad hoc way –to plot a histogram here, perform a regression analysis there, and carry out other discrete tasks involving statistical operations.

  Table of Contents:

  1. Getting Started
  2. Vectors
  3. Matrices and Arrays
  4. Lists
  5. Data Frames
  6. Factors and Tables
  7. R Programming Structures
  8. Doing Math and Simulations in R
  9. Object-Oriented Programming
  10. Input/Output
  11. String Manipulation
  12. Graphics
  13. Debugging
  14. Performance Enhancement: Speed and Memory
  15. Interfacing R to Other Languages
  16. Parallel R

- Grolemund (2014), "Hands-On Programming with R –Write Your Own Functions and Simulations".

  Learn how to program by diving into the R language, and then use your newfound skills to solve practical data science problems. With this book, you'll learn how to load data, assemble and disassemble data objects, navigate R's environment system, write your own functions, and use all of R's programming tools.

  RStudio Master Instructor Garrett Grolemund not only teaches you how to program, but also shows you how to get more from R than just visualizing and modeling data. You'll gain valuable programming skills and support your work as a data scientist at the same time.

  Work hands-on with three practical data analysis projects based on casino games Store, retrieve, and change data values in your computer's memory Write programs and simulations that outperform those written by typical R users *Use R* programming tools such as if else statements, for loops, and S3 classes Learn how to write lightning-fast vectorized R code Take advantage of R's package system and debugging tools Practice and apply R programming concepts as you learn them.

  Table of Contents:

  – Part I. Project 1: Weighted Dice
  – 1. The Very Basics
  – 2. Packages and Help Pages
  – Part II. Project 2: Playing Cards
  – 3. R Objects
  – 4. R Notation
  – 5. Modifying Values
  – 6. Environments
  – Part III. Project 3: Slot Machine
  – 7. Programs
  – 8. S3
  – 9. Loops
  – 10. Speed

2.3 Data Analysis — Data Mining — Data Science

- Cichosz (2015), "Data Mining Algorithms: Explained Using R": This book is a practical, technically-oriented guide to data mining algorithms that covers the most important algorithms for building classification, regression, and clustering models, as well as techniques used for attribute selection and transformation, model quality evaluation, and creating model ensembles. The author presents many of the important topics and methodologies widely used in data mining, whilst demonstrating the internal operation and usage of data mining algorithms using examples in R.

  Table of Contents:

  - Part I. Preliminaries
  - Part II. Classification
  - Part III. Regression
  - Part IV. Clustering
  - Part V. Getting Better Models

- Torgo (2011), "Data Mining with R: Learning with Case Studies" (aka **DMwR**): The main goal of this book is to introduce the reader to the use of R as a tool for data mining. One of the key issues in data mining is size. A typical data mining problem involves a large database from which one seeks to extract useful knowledge, and in this book we will use MySQL as the core database management system.
  Package **DMwR**: Data Mining with R.
  Table of Contents:

  1. Introduction
  2. Predicting Algae Blooms
  3. Predicting Stock Market Returns
  4. Detecting Fraudulent Transactions
  5. Classifying Microarray Samples

- Mailund (2017), "Beginning Data Science in R: Data Analysis, Visualization, and Modelling for the Data Scientist".

  Discover best practices for data analysis and software development in R and start on the path to becoming a fully-fledged data scientist. This book teaches you techniques for both data manipulation and visualization and shows you the best way for developing new software packages for R. Beginning Data Science in R details how data science is a combination of statistics, computational science, and machine learning. You'll see how to efficiently structure and mine data to extract useful patterns and build mathematical models. This requires computational methods and programming, and R is an ideal programming language for this. This book is based on a number of lecture notes for classes the author has taught on data science and statistical programming using the R programming language. Modern data analysis requires computational skills and usually a minimum of programming.

  Table of Contents:

  - Introduction to R Programming
  - Reproducible Analysis
  - Data Manipulation
  - Visualizing Data
  - Working with Large Datasets
  - Supervised Learning

  - Unsupervised Learning
  - More R Programming
  - Advanced R Programming
  - Object Oriented Programming
  - Building an R Package
  - Testing and Package Checking
  - Version Control
  - Profiling and Optimizing

- Beysolow II (2017), "Introduction to Deep Learning Using R: A Step-by-Step Guide to Learning and Implementing Deep Learning Models Using R".

  Understand deep learning, the nuances of its different models, and where these models can be applied.

  The abundance of data and demand for superior products/services have driven the development of advanced computer science techniques, among them image and speech recognition. *Introduction to Deep Learning Using R* provides a theoretical and practical understanding of the models that perform these tasks by building upon the fundamentals of data science through machine learning and deep learning. This step-by-step guide will help you understand the disciplines so that you can apply the methodology in a variety of contexts. All examples are taught in the R statistical language, allowing students and professionals to implement these techniques using open source tools.

  Table of Contents:

  - Introduction to Deep Learning
  - Mathematical Review
  - A Review of Optimization and Machine Learning
  - Single and Multilayer Perceptron Models
  - Convolutional Neural Networks (CNNs)
  - Recurrent Neural Networks (RNNs)
  - Autoencoders, Restricted Boltzmann Machines, and Deep Belief Networks
  - Experimental Design and Heuristics
  - Hardware and Software Suggestions
  - Machine Learning Example Problems
  - Deep Learning and Other Example Problems
  - Closing Statements

- Williams (2011), "Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery": This book will guide the reader through the various options that Rattle –the free and open source software built on top of the R statistical software package– provides and serves to guide the new data miner through the use of Rattle. Many excursions into using R itself are presented, with the aim of encouraging readers to use R directly as a scripting language. Through scripting comes the necessary integrity and repeatability required for professional data mining.

  Table of Contents:

  - I. Explorations
    * 1. Introduction
    * 2. Getting Started
    * 3. Working with Data
    * 4. Loading Data
    * 5. Exploring Data
    * 6. Interactive Graphics

    - 7. Transforming Data
  - II. Building Models
    - 8. Descriptive and Predictive Analytics
    - 9. Cluster Analysis
    - 10. Association Analysis
    - 11. Decision Trees
    - 12. Random Forests
    - 13. Boosting
    - 14. Support Vector Machines
  - III. Delivering Performance
    - 15. Model Performance Evaluation
    - 16. Deployment
  - IV. Appendices
    - A. Installing Rattle
    - B. Sample Datasets

- Abedin (2017), "Modern R Programming Cookbook –Recipes for emerging developers in R programming and data scientists to simplify their R programming capabilities".

  R is a powerful tool for statistics, graphics, and statistical programming. It is used by tens of thousands of people daily to perform serious statistical analyses. It is a free, open source system whose implementation is the collective accomplishment of many intelligent, hard-working people. There are more than 2,000 available add-ons, and R is a serious rival to all commercial statistical packages. The objective of this book is to show how to work with different programming aspects of R. The emerging R developers and data science could have very good programming knowledge but might have limited understanding about R syntax and semantics. Our book will be a platform develop practical solution out of real world problem in scalable fashion and with very good understanding. You will work with various versions of R libraries that are essential for scalable data science solutions. You will learn to work with Input / Output issues when working with relatively larger dataset. At the end of this book readers will also learn how to work with databases from within R and also what and how meta programming helps in developing applications.

  Table of Contents:

  1. Installing and Configuring R and its Libraries
  2. Data Structures In R
  3. Writing Customized Functions
  4. Conditional and Iterative Operations
  5. R Objects and Classes
  6. Querying, Filtering, and Summarizing
  7. R For Text Processing
  8. R and Databases
  9. Parallel Processing In R

- Hodnett and Wiley (2018), "R Deep Learning Essentials, 2nd Edition –A step-by-step guide to building learning models using TensorFlow, Keras, and MXNet".

  Deep learning is probably the hottest technology in data science right now, and of R is one of the most popular data science languages However, of R is not considered as an option for deep learning by many people, which is a shame, as of R is a wonderful language for data science. This book shows that R is a viable option for deep learning, because it supports libraries such as MXNet and Keras. When I decided to write this book, I had numerous goals. First, I wanted to show how to apply deep

learning to various tasks, and not just to computer vision and natural language processing. This book covers those topics, but it also shows how to use deep learning for prediction, regression, anomaly detection, and recommendation systems. The second goal was to look at topics in deep learning that are not covered well elsewhere; for example, interpretability with LIME, deploying models, and using the cloud for deep learning. The last goal was to give an overall view of deep learning and not just provide machine learning code. I think I achieved this by discussing topics such as how to create datasets from raw data, how to benchmark models against each other, how to manage data when model building, and how to deploy your models. My hope is that by the end of this book, you will also be convinced that of R is a valid choice for use in deep learning.

Table of Contents:

3 Using R

3.1 Why R?

R is an open source software package, licensed under the GNU General Public License (GPL). Moreover R is *better* than its commercial counterparts in many respects: capability, community and performance.

3.2 What is R?

R is an open-source software environment for statistical computing and graphics. R compiles and runs on Windows, Mac OS X, and numerous UNIX platforms (such as Linux). For most platforms, R is distributed in binary format for ease of installation. The R software project was first started by Robert Gentleman and Ross Ihaka. The language was very much influenced by the **S** language, which was originally developed at Bell Laboratories by John Chambers and colleagues. Since then, with the direction and talents of R*'s core development team*, R has evolved into the *lingua franca* for statistical computations in many disciplines of academia and various industries.

R is much more than just its core language. It has a worldwide repository system, the *Comprehensive R Archive Network* (CRAN) for user-contributed add-on packages to supplement the base distribution. A core set of packages is included with the installation of R, and they are 5,300 additional packages (as of April 2012). In total, R currently has functionality to address an enormous range of problems and still has room to grow.

R is designed around its core scripting language but also allows integration with compiled code written in C, C++, Fortran, Java, etc., for computationally intensive tasks or for leveraging tools provided for other languages.

The R system is a software environment for statistical computing and graphics. It includes many different components. The term R to refer to a few different things:

- A computer language
- The interpreter that executes code written in R
- A system for plotting computer graphics described using the R language
- The Windows, Mac OS, or Linux application that includes the interpreter, graphics system, standard packages, and user interface.

3.3 What for?

R can be used:

- for data mining (Torgo, 2011; Tufféry, 2010; Williams, 2011; Ledolter, 2013; Zhao, 2013; Zhao and Cen, 2013; Cichosz, 2015)
- to produce professionnal graphics (Chang, 2012b; Mittal, 2011; Murrell, 2006; Wickham, 2009)
- to work with large datasets and manipulating data from files and databases (Spector, 2008; Torgo, 2011)
- for data mashups, e.g. plotting foreclosure data on a map image after downloading and manipulating the underlying data (Leipzig and Li, 2011)
- for the development and application of Bayesian inference in statistics (Albert, 2009)
- for multivariate data analysis (Wehrens, 2011; Oja, 2010; Husson et al., 2010)and visualization (Sarkar, 2008)
- for biology (Seefeld and Linder, 2007): bioinformatics and computational biology (Gentleman, 2008; Mathur, 2010), biostatistical design and analysis (Logan, 2010), applied statistical genetics (Foulkes, 2009), ecology (Stevens, 2009; Soetaert and Herman, 2009), analysis of phylogenetics and evolution (Paradis, 2006), biostatistical analyses of clinical trial data (Chen and Peace, 2010)
- as a tool for dealing with data in behavioral science (Li and Baron, 2011)
- to analyze large datasets in parallel (McCallum and Weston, 2011)
- for programming graphical user interfaces (GUIs) (Lawrence and Verzani, 2012)
- for business analytics (Ohri, 2012; Jank, 2011)
- for computational finance and financial engineering (Würtz et al., 2009)
- in econometrics (Kleiber and Zeileis, 2008)
- to handle, visualise, and analyse spatial data (Bivand et al., 2013)
- as the software tool in the development of "Six Sigma projects", the breakthrough quality management methodology (Cano et al., 2012)
- to do statistical analysis for linguists and psycholinguists (Baayen, 2008)
- with Ruby for helping you find real answers to everyday problems (Chang, 2012a)
- to apply statistical analysis of competing risks and multistate data (Beyersmann et al., 2012)

- to perform morphometrics –morphometric analysis is the study of shape and size variations and co-variations and their covariations with other variables (Claude, 2008)
- to deal with time series –variables measured sequentially in time (Cowpertwait and Metcalfe, 2009; Cryer and Chan, 2008; Pfaff, 2010; Shumway and Stoffer, 2011)
- for doing interactive and dynamic plots on a computer screen as part of data exploration and modeling (Cook and Swayne, 2007)
- with JavaScript for data visualizations used as communication tools in the field of web development (Barker, 2013)
- for social science research (Vinod, 2010).

4 R Installation and Upgrading

4.1 Getting and Installing Interactive R Binaries

R has been ported to every major desktop computing platform. Because R is open source, developers have ported R to many different platforms. Additionally, R is available with no license fee.

1. Visit the official R website. On the site, you should see a link to "Download."
2. The download link actually takes you to a list of mirror sites. The list is organized by country. You'll probably want to pick a site that is geographically close, because it's likely to also be close on the Internet, and thus fast, e.g., you can use the link for the Université Lyon 1.
3. Find the right binary for your platform and run the installer.

4.2 Upgrading R

Upgrading R on Windows is not so easy. To make upgrading and updating easier, it is possible to run the command `updateR()` from the *installr* package.

Another way is to reinstall a new version of R, copy the downloaded R packages of the previous version and paste them in the directory of the new version.

5 R Basics

5.1 Beginning with R

The prompt '>' indicates that R is waiting for commands.

Several ways of sending commands to R:

- Type commands in the console (not recommended)
- Using the script editor.

Using scripts allows to:

- Store, re-use or later modify your statistical analysis
- Share code with others.

R scripts are text files containing a series of R commands. The usual extension for such files is '.R'

Comments: everything between '#' and the end of line. It is strongly recommended to document scripts with comments.

5.2 Font Conventions

This document describes how to do computations in R. This requires that the user types input, and R responds with text or graphs as output.

R code will be presented surrounded by a frame, keywords from the R language will be written in a bold typewriter font, and comments will be typed in a slanted font (all text after the pound sign '#' within the same line is considered a comment).

After R is started, there is a console awaiting for input. At the prompt (>), the user can enter numbers and perform calculations. The result number from the R computation will appear between square brackets.

```
> This was typed by the user          # This is a comment
[1] This is a response from R
```

```
> 1 + 2        # Simple addition
[1] 3
```

Other commands return multiple values, and each line of results will be labeled to aid the user in deciphering the output. For example, the sequence of integers from 1 to 30 may be displayed as follows:

```
> 1:30
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
[25] 25 26 27 28 29 30
```

The numbers in the square brackets give the index of the element immediately to the right.

Notice that if the user's instructions will take more than one line, a plus sign (+) will appear at the beginning of a new line for indicating that the line is incomplete. In this case, R is waiting for the end of the code before running it.

To quit your R session, type:

```
> q()
```

An R session consists of starting R, working with it and then quitting.

You quit R by typing q() or (in Windows) by selecting "File | Exit". R displays the prompt at the beginning of the line as a cue for you. This is where you interact with the system (you do not type the prompt). The vast majority of our work with R will consist of executing functions. For now, we will say that a function is a small, single-purpose executable program that R can be instructed to run. A function is identified by a name followed by parentheses. Often the parentheses are separated by words. These are called arguments. We say that we execute a function when we type the function's name, with the parentheses (and possibly arguments) and then hit the Enter key.

Let us go through a simple session. This will give you a feel for R. At any point, if you wish to quit, type q() to end the session. To quit, you can also use the "File | Exit" menu. At the prompt, type (recall that you do not type the prompt):

```
> help.start()
```

This will display the introductory help page in your browser. You can access help.start() from the "Help | Html" help menu. In the "Manuals" section of the HTML page, click on the title "An Introduction to R". This will bring up a well-written tutorial that introduces R and its capabilities.

5.3 Getting Help and Built-in Examples

R provides extensive documentation. For example, entering ?c or help(c) at the prompt gives documentation of the function c in R. Please give it a try.

```
> help(c)
```

If you are not sure about the name of the function you are looking for, you can perform a fuzzy search with the apropos function.

```
> apropos("nova")
[1] "anova"                    "anova.glm"
    ...
```

Finally, there is an R specific Internet search engine at `http://www.rseek.org` for more assistance.

It is often convenient to use help.start(). This brings up an Internet browser that will show a menu of several options, including a listing of installed packages. The base package contains many of the routinely used functions. Another function that is often used is help.search().

For example, to see if there are any functions that do optimization (finding minima or maxima), type:

```
> help.search("optimization")
```

And the browser window will display the following information:

The search string was "optimization"

Code demonstrations:

nws::pportfolio	parallel portfolio optimization	(Run demo in console)

Help pages:

igraph::fastgreedy.community	Community structure via greedy optimization of modularity
igraph::multilevel.community	Finding community structure by multi-level optimization of modularity
igraph0::fastgreedy.community	Community structure via greedy optimization of modularity
vcd::RepVict	Repeat Victimization Data
...	...

A useful alternative to help() is the example() function:

```
> example(mean)
```

```
mean> x <- c(0:10, 50)

mean> xm <- mean(x)

mean> c(xm, mean(x, trim = 0.10))
[1] 8.75 5.50
```

This example shows simple use of the mean() function as well as how to use the trim argument.

References

Abedin, J. (2017). *Modern R Programming Cookbook – Recipes for emerging developers in R programming and data scientists to simplify their R programming capabilities*. Packt Publishing.

Adler, J. (2010). *R in a Nutshell – a Desktop Quick Reference*. O'Reilly.

Adler, J. (2011). *R, L'essentiel*. Paris: Pearson Education France. [Book available at the library of the Faculty of Science and Technology].

Albert, J. (2009). *Bayesian Computation with R* (2^{nd} ed.). Use R! Springer.

Allerhand, M. (2011). *A Tiny Handbook of R*. Springer.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.

Barker, T. (2013). *Pro Data Visualization using R and JavaScript*. Apress.

Beyersmann, J., A. Allignol, and M. Schumacher (2012). *Competing Risks and Multistate Models with R*. Use R! Springer.

Beysolow II, T. (2017). *Introduction to Deep Learning Using R: A Step-by-Step Guide to Learning and Implementing Deep Learning Models Using R*. Apress.

Bivand, R. S., E. Pebesma, and V. Gómez-Rubio (2013). *Applied Spatial Data Analysis with R*. Use R! Springer.

Braun, W. J. and D. J. Murdoch (2007). *A First Course in Statistical Programming with R*. Cambridge University Press.

Cano, E. L., J. M. Moguerza, and A. Redchuk (2012). *Six Sigma with R – Statistical Engineering for Process Improvement*. Use R! Springer.

Chambers, J. M. (2008). *Software for Data Analysis: Programming with R*. Springer.

Chang, S. S. (2012a). *Exploring Everyday Things with R and Ruby: Learning About Everyday Things*. O'Reilly.

Chang, W. (2012b). *R Graphics Cookbook – Practical Recipes for Visualizing Data*. O'Reilly.

Chen, D.-G. and K. E. Peace (2010). *Clinical Trial Data Analysis Using R*. Chapman & Hall / CRC.

Cichosz, P. (2015). *Data Mining Algorithms: Explained Using R*. John Wiley & Sons, Ltd.

Claude, J. (2008). *Morphometrics with R*. Use R! Springer.

Cook, D. and D. F. Swayne (2007). *Interactive and Dynamic Graphics for Data Analysis: With R and GGobi*. Use R! Springer.

Cowpertwait, P. S. P. and A. V. Metcalfe (2009). *Introductory Time Series with R*. Use R! Springer.

Cryer, J. D. and K.-S. Chan (2008). *Time Series Analysis: With Applications in R* (2^{nd} ed.). Use R! Springer.

Foulkes, A. S. (2009). *Applied Statistical Genetics with R: For Population-based Association Studies*. Use R! Springer.

Gentleman, R. (2008). *R Programming for Bioinformatics*. Chapman & Hall / CRC.

Grolemund, G. (2014). *Hands-On Programming with R –Write Your Own Functions and Simulations*. O'Reilly Media.

Hodnett, M. and J. F. Wiley (2018). *R Deep Learning Essentials, Second Edition –A step-by-step guide to building learning models using TensorFlow, Keras, and MXNet*. Packt Publishing.

Husson, F., S. Lê, and J. Pagès (2010). *Exploratory Multivariate Analysis by Example Using R*. Computer Science & Data Analysis. Chapman & Hall / CRC.

Jank, W. (2011). *Business Analytics for Managers*. Use R! Springer.

Kabacoff, R. (2011). *R in Action*. Manning Publications.

Kleiber, C. and A. Zeileis (2008). *Applied Econometrics with R*. Use R! Springer.

Lawrence, M. F. and J. Verzani (2012). *Programming Graphical User Interfaces in R*. The R Series. Chapman & Hall / CRC.

Ledolter, J. (2013). *Data Mining and Business Analytics with R*. John Wiley & Sons, Ltd.

Leipzig, J. and X.-Y. Li (2011). *Data Mashups in R*. O'Reilly.

Li, Y. and J. Baron (2011). *Behavioral Research Data Analysis with R*. Use R! Springer.

Logan, M. (2010). *Biostatistical Design and Analysis Using R: A Practical Guide*. John Wiley & Sons, Ltd.

Mailund, T. (2017). *Beginning Data Science in R: Data Analysis, Visualization, and Modelling for the Data Scientist*. Apress.

Mathur, S. K. (2010). *Statistical Bioinformatics with R*. Academic Press.

Matloff, N. S. (2011). *The Art of R Programming: A Tour of Statistical Software Design*. San Francisco, CA: No Starch Press, Inc.

McCallum, Q. E. and S. Weston (2011). *Parallel R*. O'Reilly.

Meys, J. and A. de Vries (2012). *R For Dummies*. John Wiley & Sons, Ltd.

Mittal, H. V. (2011). *R Graphs Cookbook*. Packt Publishing.

Murrell, P. (2006). *R Graphics* (1^{st} ed.). Computer Science & Data Analysis. Chapman & Hall / CRC.

Ohri, A. (2012). *R for Business Analytics*. Springer.

Oja, H. (2010). *Multivariate Nonparametric Methods with R: An approach based on spatial signs and ranks*. Lecture Notes in Statistics. Springer.

Paradis, E. (2006). *Analysis of Phylogenetics and Evolution with R* (1^{st} ed.). Use R! Springer.

Pfaff, B. (2010). *Analysis of Integrated and Cointegrated Time Series with R* (2^{nd} ed.). Use R! Springer.

R Core Team (2017). The R language definition.
URL `https://cran.r-project.org/doc/manuals/R-lang.pdf`.

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Use R! Springer.

Seefeld, K. and E. Linder (2007). Statistics Using R with Biological Examples.
URL `http://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf`.

Shumway, R. H. and D. S. Stoffer (2011). *Time Series Analysis and Its Applications: With R Examples* (3^{rd} ed.). Springer Texts in Statistics. Springer.

Soetaert, K. and P. M. J. Herman (2009). *A Practical Guide to Ecological Modelling: Using R as a Simulation Platform*. Springer.

Spector, P. (2008). *Data Manipulation with R*. Use R! Springer.

Stevens, M. H. H. (2009). *A Primer of Ecology with R*. Use R! Springer.

Torgo, L. (2011). *Data Mining with R: Learning with Case Studies*. Chapman & Hall / CRC.

Tufféry, S. (2010). *Data mining et statistique décisionnelle: L'intelligence des données* (3^e ed.). Paris: Editions Technip.

Vinod, H. D. (Ed.) (2010). *Advances in Social Science Research Using R*. Lecture Notes in Statistics. Springer.

Wehrens, R. (2011). *Chemometrics with R: Multivariate Data Analysis in the Natural Sciences*. Use R! Springer.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Use R! Springer.

Williams, G. (2011). *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. Use R! Springer.

Würtz, D., Y. Chalabi, W. Chen, and A. Ellis (2009). *Portfolio Optimization with R/Rmetrics*. Zurich: Rmetrics Association and Finance Online Publishing.

Zhao, Y. (2013). *R and Data Mining: Examples and Case Studies*. Academic Press.

Zhao, Y. and Y. Cen (2013). *Data Mining Applications with R*. Elsevier.