Sessions Overview

John Aa. Sørensen

DTU Diplom

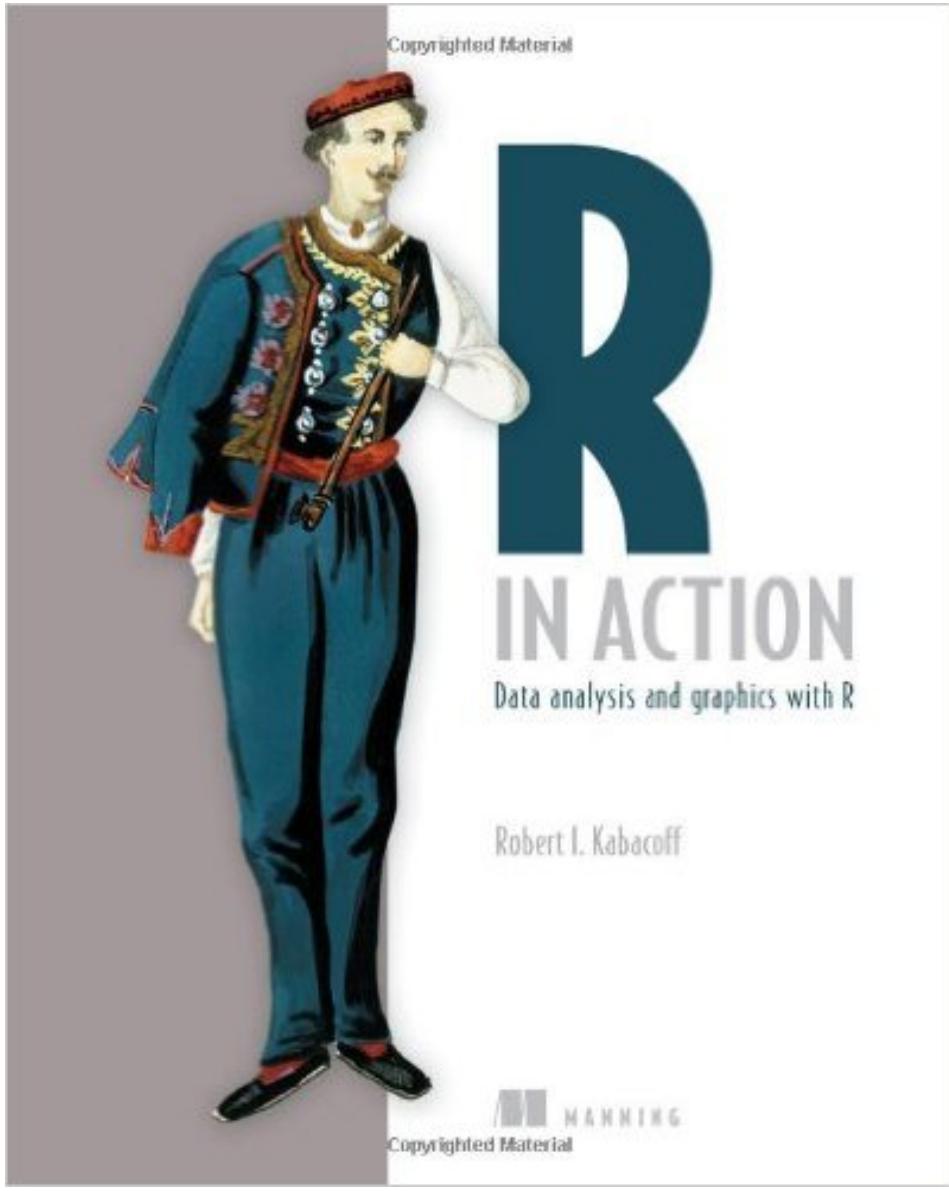
jaas@dtu.dk

2. February 2017

Session 1 Install R and RStudio

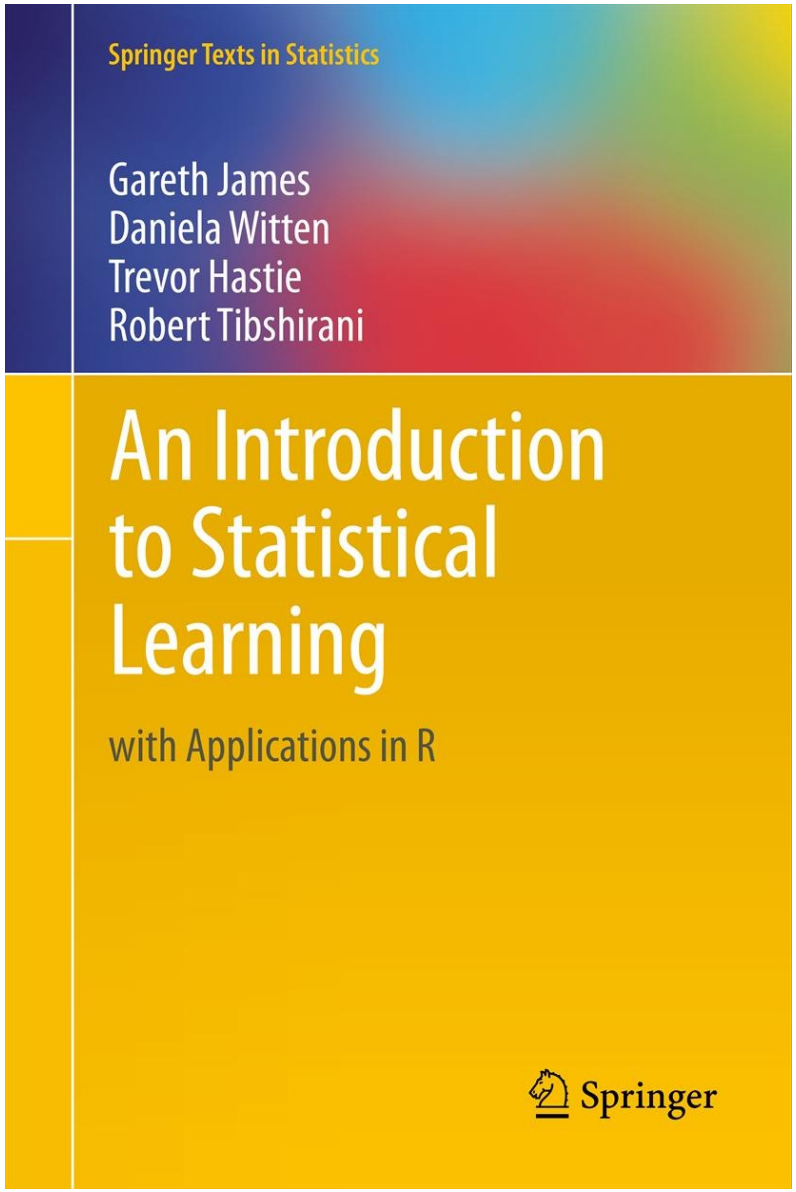
- Install R and RStudio on Windows/Mac.
- Enter the main R Language documentation.
- Initial example on applying R for generation of uniformly distributed numbers and plotting the histogram.
- Overview of demos running demos available in R base system.
- Example on creating, plotting and saving a plot.
- An overview of packages and libraries in R.
- An overview of datasets available in the R package 'datasets'.
- A few R console commands.
- Course material.

Course Text Book



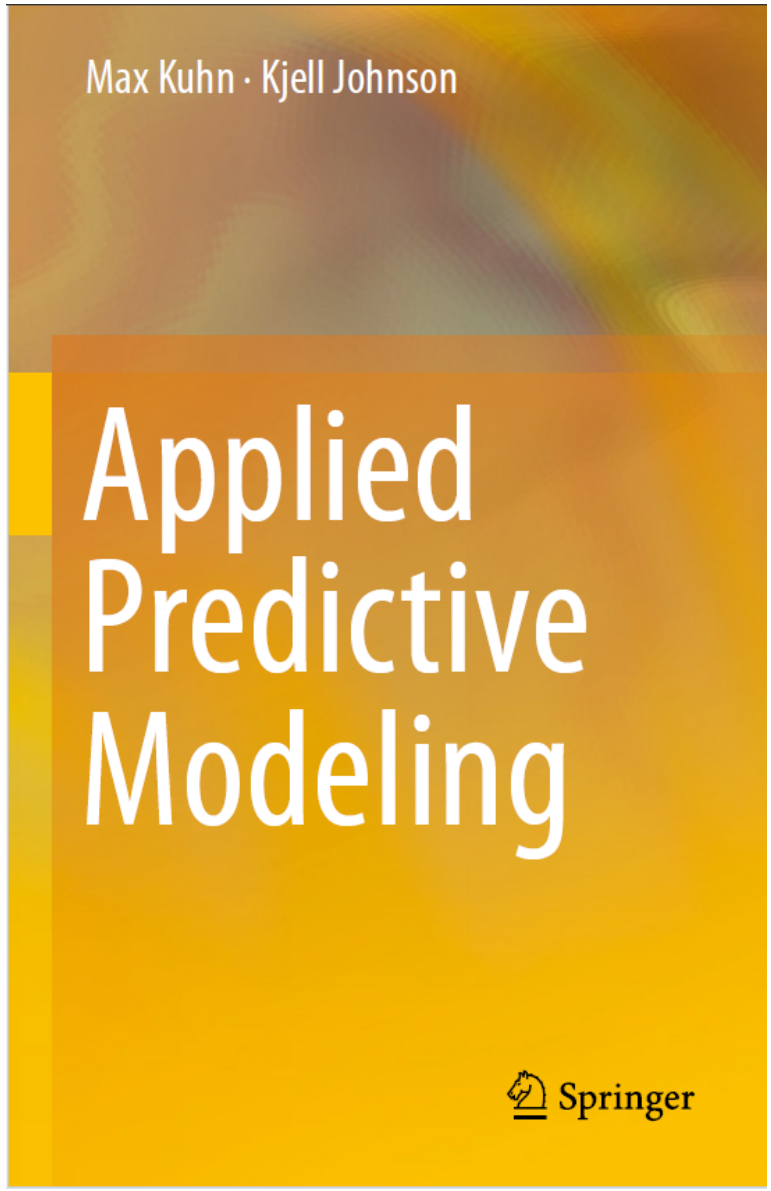
Kabacoff, "R in Action", 2'Ed, Manning Publications, 2015.

Excellent Supplementary Text Book I



Is available in pdf from the authors websites as pdf.

Excellent Supplementary Text Book II



Is available in pdf by logging into the DTU Library.

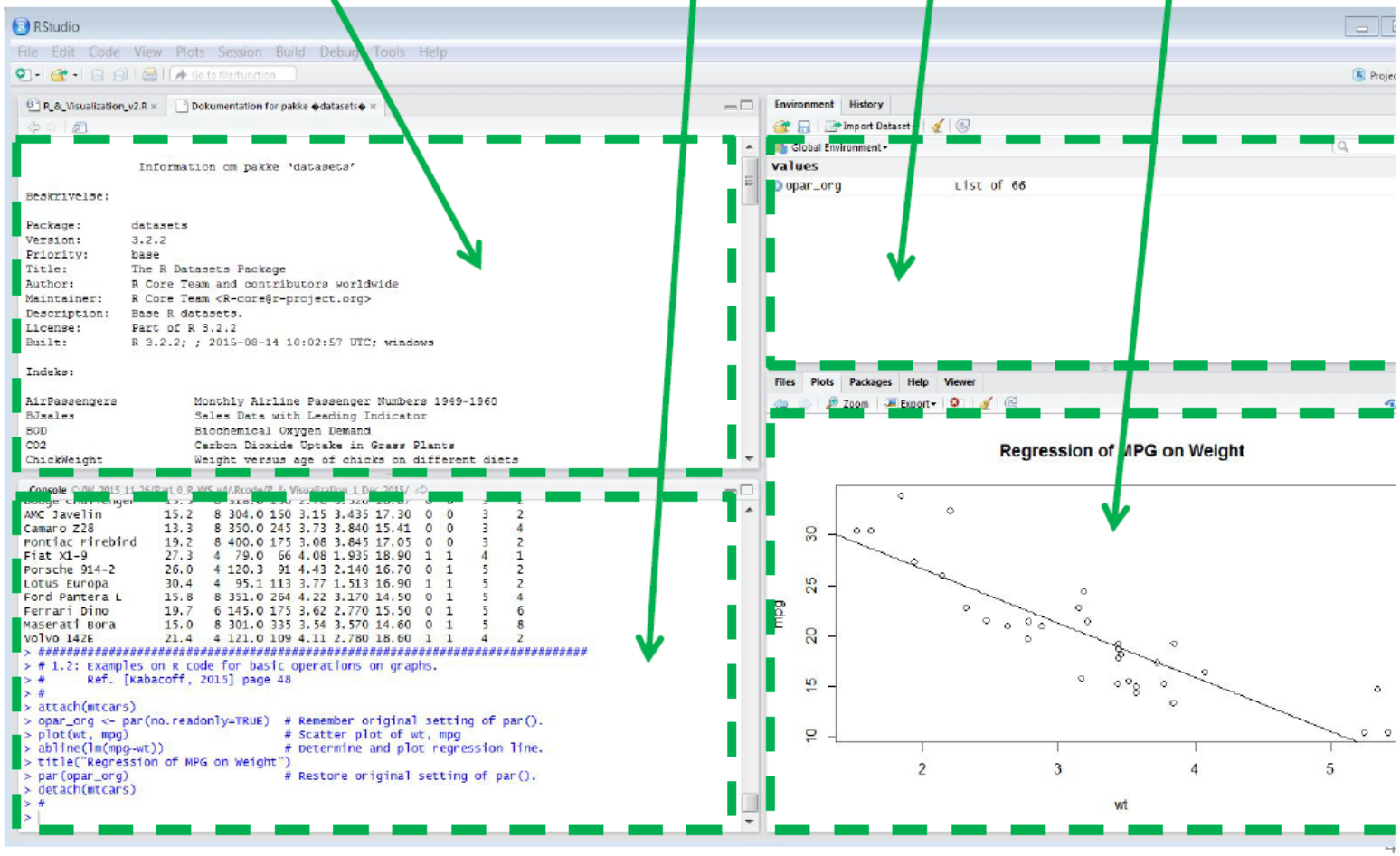
R, RStudio IDE, Code and Supplementing References

- R
- RStudio IDE (Integrated Development Environment).
- Code, which can run directly on a Windows or Mac platform.
- Further supplementary key references and short notes typically in pdf.

RStudio IDE

The Four Rstudio Windows:

R Scripts (code), Console, Workspace, Plots



Session 2 Introduction to R Language Elements

- Examples on scalar operations and examples on functions.
- Examples on vector definitions and operations.
- Examples on matrix definitions and vector, matrix operations.
- Examples on defining and applying data frames.
  - Create a data frame with customer assessments of two products.
  - Insert a new variable in a data frame. (Insert a new column in a data frame).
  - Merge two sets of observations for the same set of variables (Insert new rows in a data frame).
  - Identify missing values in a data frame.
  - Excluding missing values from a data frame preparing for analysis.
  - Grouping observations.
  - Sorting observations according to one variable.
  - Plot selected parts of data frame.
  - Deleting (removing) rows or columns in dataframes.
- Example on defining and applying factors (ordinal variables).
- Example on defining and applying lists.

Session 3 R Data Import Export

- Importing data from a comma separated values (csv) text file.
- Importing data from a csv file exported from Excel.
- Merging two data frames.
- Writing a data frame to csv file in working directory.
- Hands-on visualization of own data or data from the R package 'data'.

Session 4 R Intro Visualization

- Plot of regression line to pdf or jpg file.
- Applying graphic parameters for fonts, colors, axes, labels.
- RColorBrewer package, for color palettes.
- Gray-levels, text objects, margin controle.
- Comparing 2 plots and example on scatter plots.
- One figure with multiple plots.
- Combining a scatter plot with two box plots.
- Comparing groups by using parallel box plots.
- Plotting a large number of labeled values on a simple horizon scale, sorted and colored.
- Bubble plot with point size area proportional to a variable, using the mtcar dataset.
- 3d Scatter plot and mouse controlled spinning of a 3D plot.

Session 5 R Intro Clustering

- Install packages needed for clustering.
- Step 1: Choose appropriate attributes.
- Step 2: Scale the data.
- Step 3: Screen for outliers.
- Step 4: Calculate distances.
- Step 5: Select a clustering algorithm.
- Step 6: Obtain one or more cluster solutions.
- Step 7: Determine the number of clusters present.
- Step 8: Obtain final clustering solution.
- Step 9: Visualize the results.
- Step 10: Interpret the clusters.
- Step 11. Validate the results. Analyze stability.
- Step 11: The CCC (Cubic Cluster Criteria) referenced by NbClust as indicator for unimodality.
- Examples using the partitioning clustering k-means.
- Examples using the partitioning around medoids.

Session 6 R Intro Classification

- Install packages needed for classification.
- Preparing data for classification examples.
- Logistic regression example.
- Creating a decision tree.
- Conditional inference trees.
- Random forest.
- Support vector machines.
- Distance and dissimilarity in heterogeneous data, mixing continuous, categorical and percentage data.
- Dynamic Time Warp (DTW) between time sequences,

Session 7 Text Mining & Visualization of Graphs.

- Install packages for textmining and graphs visualization.
- Textmining: Generate a textcorpus, wordclouds.
- Graph visualization: Representation of graphs.
- Graph visualization: Plotting datastructures as graphs.

References I

Joseph Adler (2012)

R in a Nutshell

OReilly

Robert I. Kabacoff (2015)

R in Action

Manning Publications 2'Ed.

R Core Team and contributors worldwide (2015)

The R Language Manual System

CRAN e.g. via RStudio

Tom Short, (2004)

Short Reference Card

CRAN cran.r-project.org/doc/contrib/Short-refcard.pdf

Paul Teetor

R Cookbook

O'Reilley

References II

Paul Torfs, Caludia Brauer

A (very) Short Introduction to R.

CRAN cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf

Yanchang Zhao

R and Data Mining.

Elsevier 2013.

Yanchang Zhao

R Reference Card for Data Mining.

www.rdatamining.com

www.rdatamining.com/docs/r-reference-card-for-data-mining.pdf