

R Programming as a Part of Bigdata Course. The R Environment.

John Aa. Sørensen, lektor
Section for Information Technology, DTU Diplom
Center for Continuing Education

2 February 2017

The R Environment.

```
#
#           "R_Intro"  TABLE OF CONTENTS
#
# 1.0: Install R and RStudio.
#      Install R on Windows/Mac.
#      Install RStudio on Windows/Mac.
#      The four RStudio windows.
#      Setting workspace and folders for R Packages.
#      Get the version numbers of installed R and RStudio.
#
# 2.0: Enter the main R Language documentation.
#
# 2.1: Initial example applying R for generation of uniformly
#      distributed numbers and plotting the histogram.
#
# 2.2: Overview demos immediately available in the R base system,
#      and running a selection of demos.
#
# 2.3: Example creating, plotting and saving plot of normalized
#      distributed data.
```

The R Environment.

```
# 2.4: An overview of packages and libraries in R.  
#  
# 2.5: An overview of datasets available in R package 'datasets'.  
#       Check if a package is installed.  
#  
# 3.0: Calling an R script file from R code.  
# 3.1: Reusing results of an analysis - using output as input.  
#  
# 4.0: A few R console and other commands.  
#  
# 5.0: Getting and setting working directory and path to packet libray.  
#       cf. [Kabacoff, 2015] page 12 Table 1.3.  
#
```

The R Environment.

```
#                               REFERENCES
#
# [Adler, 2012] Joseph Adler, "R In a Nutshell", O'Reilly, 2012.
# [Kabacoff, 2015] Robert I. Kabacoff, "R in Action",
#   2'Ed, Manning Publications, 2015.
# [R Core, 2015] R Core Team and contributors worldwide,
#   "The R Language Manual System", from help.start()
# [short, 2004] Tom Short, "Short Reference Card"
#   cran.r-project.org/doc/contrib/Short-refcard.pdf
# [Teetor, 2011] Paul Teetor, "R Cookbook", O'Reilley, 2011.
# [Torfs, 2014] Paul Torfs, Caludia Brauer, "A (very) Short ..."
#   cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf
# [Zhao, 2013] Yanchang Zhao, "R and Data Mining", Elsevier, 2013.
# [Zhao, 2013] Yanchang Zhao, "R Reference Card for Data Mining",
#   www.rdatamining.com/docs/r-reference-card-for-data-mining
#
```

The R Environment.

```
#####  
# 1.0: Install R and Rstudio  
#  
# Initially use link to the 32/64bit R installation package  
# to Windows: https://cran.r-project.org/bin/windows/base/  
# to Mac: https://cran.r-project.org/bin/macosx/  
#  
# Then use this link to the RStudio installation package to  
# Windows and Mac:  
# https://www.rstudio.com/products/rstudio/download/  
#  
# The four RStudio windows and their functions.  
#  
# Upper left Window: R source file editing and R code  
# execution.  
# Lower left window: R Console. Input/output directly to/from R.  
# Upper right window: The R program variables (environment) and  
# console history.  
# Lower right window: Help, plots, packages.  
#  
# Now open the file 1_Intro.R in the RStudio Editor and  
# work with the examples below starting from 2.0.
```

The R Environment.

```
#  
# Setting workspace for session for R code files (script files),  
# from RStudio top menu -> Session -> Set Working Directory ->  
#     Choose your Directory -> select directory.  
#  
# Setting the path to folder for installing R packages.  
# It's suggested to create a folder at the top C: level for installing  
# the R packages, e.g. c:/R_packages/ .  
# Then the R packages can be installed using the following:  
#     Package from CRAN repository to this laptop's disk:  
#         install.packages("cluster",lib="C:/R_packages/")  
#     Package from this laptop's disk to this laptops RAM:  
#         library("cluster",lib="C:/R_packages/")  
#  
# The version of the current R installation is obtained by  
# version CR (CR: Return key)
```

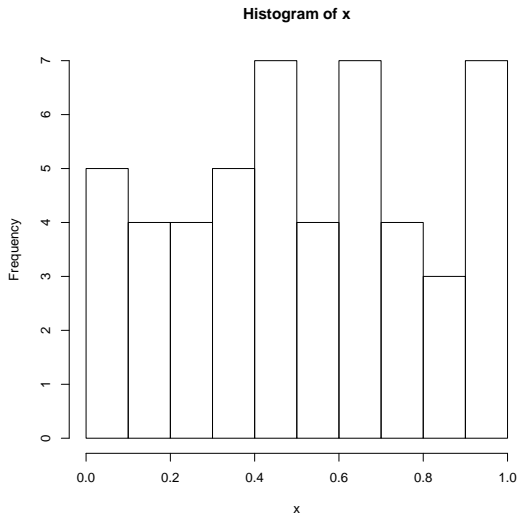
The R Environment.

```
#####  
# 2.0: The main R Language documentation system.  
#  
help.start() # This is the main entry to the R project documentation.  
# Check briefly the main enties for:  
# -Manuals  
# -Reference  
# -Miscellaneous Material  
# -Material Specific to the Windows/Mac Port  
#  
# Store the original graphical parameters for later restore.  
opar.org <- par(no.readonly=TRUE)
```

The R Environment.

```
#####  
# 2.1: Initial example applying R for generation of uniformly  
#       distributed numbers and plotting their histogram.  
#  
x <- runif(50) # Get manual runif(), by console input ?runif() .  
# If min and max not specified, default 0 and 1 are used.  
summary(x)      # Get the manual by console input ?summary() .  
# Check output from summary().  
quantile(x) # Compare this to summary.  
hist(x) # Generate histogram of 50 uniformly distributed numbers.  
#  
# Generate pdf of histogram.  
pdf(file="fig_1_2_1_hist.pdf")  
hist(x) # Generate histogram of 50 uniformly distributed numbers.  
dev.off()  
par(opar.org)
```


The R Environment



The R Environment.

```
#####  
# 2.2: Overview of demos available in base R installation,  
#       and running a demo.  
#  
# Try a selection of demo packages, from the R project.  
#  
# List all possible demos in packages available.  
demo(package = .packages(all.available = TRUE))  
demo()  
#  
# Run example demos  
demo(graphics) # Run demo graphics from package 'graphics'  
library(help = "stats") # Manual on the package 'stats'  
demo(lm.glm)    # Run demo lm.glm from package 'stats'
```

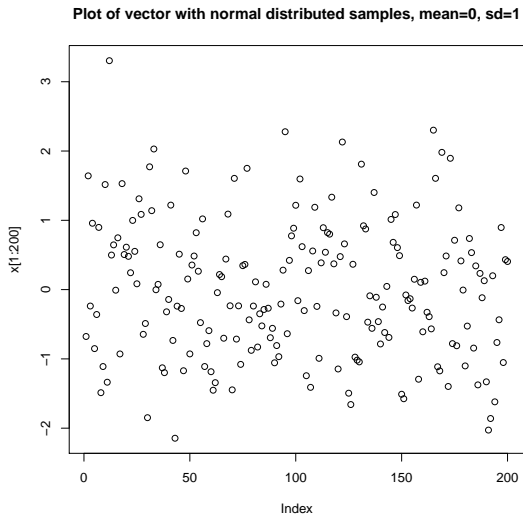
The R Environment.

```
#####  
# 2.3: Example on creating, plotting and saving plot of  
#       normalized distributed data.  
#  
# Exemplify plots of normal distributed data.  
# Each of the plots can be exported as e.g. pdf to further  
# application in documentation.  
#  
set.seed(1023) # Setting seed for normal distributed generator.  
N <- 10000     # Use N samples from Normal distribution  
x <- rnorm(N, mean=0, sd=1.0) # Normal distributed sequence.  
plot(x[1:200]) # Plot part of the dataset.  
title(main="Plot of vector of normal distrib. samples, mean=0, sd=1")  
hist(x)  
title(sub="Normally distributed samples, mean=0, sd=1")  
boxplot(x)  
title(main="Box-plot of normal sequence.", xlab="x", ylab="y")  
#
```

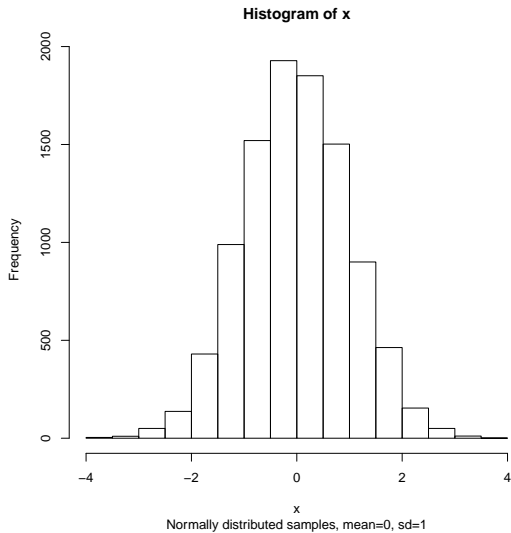
The R Environment.

```
# Generate pdf plots.
pdf(file="fig_1_2_dataset.pdf")
plot(x[1:200]) # Plot part of the dataset.
title(main="Plot of vector of normal distrib. samples, mean=0, sd=1")
dev.off()
pdf(file="fig_1_2_hist.pdf")
hist(x)
title(sub="Normally distributed samples, mean=0, sd=1")
dev.off()
pdf(file="fig_1_2_boxplot.pdf")
boxplot(x)
title(main="Box-plot of normal sequence.", xlab="x", ylab="y")
dev.off()
par(opar.org)
```

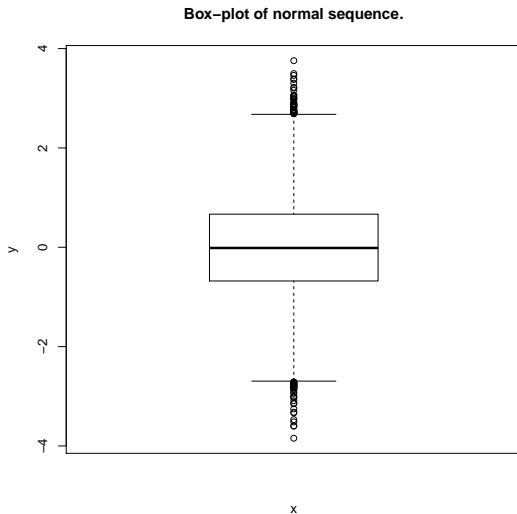
The R Environment



The R Environment



The R Environment



The R Environment.

```
#####  
# 2.4: An overview of packages and libraries in R.  
#  
# Packages and libraries in R (CRAN)  
#  
help.start()    # Main entry to the R project documentation.  
#  
#Link to CRAN packages http://cran.r-project.org/web/packages/  
#Continue to the name list of packages  
#http://cran.r-project.org/web/packages/available\_packages\_by\_name.html  
#  
# List the path to library of packages on present installation.  
.libPaths()  
#  
# List all packages in library on the present installation.  
library()  
#  
loc='C:/R_packages/' # Create folder for packages.
```


The R Environment.

```
# Install a package in the package library at laptop.
# General package on statistics etc.
install.packages("e1071",lib=loc)
#
# ... then load the package e1071 into laptop RAM.
#
library("e1071", lib=loc)
#
# ... and finally check the help function for the package.
# Help function on the modules in package e1071.
# Displayed in right lower RStudio window.
help(package="e1071", lib=loc)
#
# Check if a package is installed
# any(grepl("<name of package>", installed.packages()))
any(grepl("e1071", installed.packages(lib=loc))) # ->[1] TRUE
```

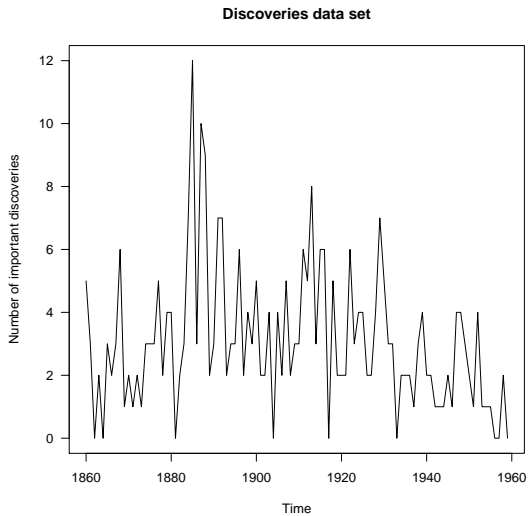
The R Environment.

```
#####  
# 2.5: An overview of datasets in the R package 'datasets'.  
#  
help(package="datasets") # Help on package with datasets  
#  
# Example edataset euro  
?euro()  
# This code is from the ?euro() manual example.  
# Run this example  
require(graphics)  
dotchart(euro,  
          main = "euro data: 1 Euro in currency unit")  
dotchart(1/euro,  
          main = "euro data: 1 currency unit in Euros")  
dotchart(log(euro, 10),  
          main = "euro data: log10(1 Euro in currency unit)")
```

The R Environment.

```
# Example dataset discoveries
??discoveries()
# Run the example
require(graphics)
plot(discoveries, ylab = "Number of important discoveries",
      las = 1)
title(main = "Discoveries data set")
#
# Generate a pdf plot of discoveries
pdf(file="fig_1_2_5_discoveries.pdf")
plot(discoveries, ylab = "Number of important discoveries",
      las = 1)
title(main = "Discoveries data set")
dev.off()
par(opar.org) # Restore the original graphics parameters.
```

The R Environment



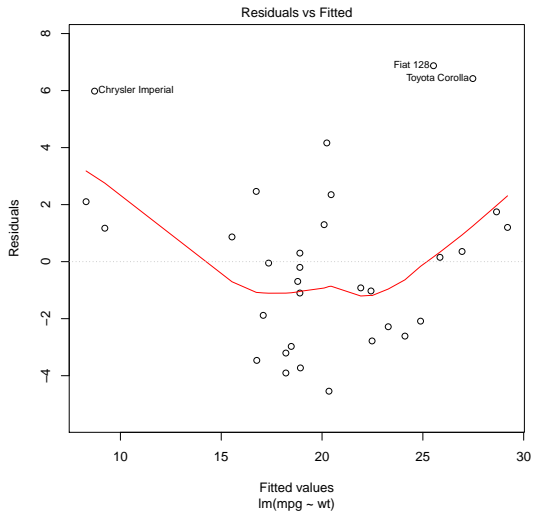
The R Environment.

```
#####  
# 3.0: Calling an R script file from R code.  
#  
# Initially copy this code to filename in workspace using  
# the RStudio editor. Remember to remove the leftmost comments.  
#  
# # Filename: Script_1.R  
#  
# # Script for generation of a vector with 200 Gaussian  
# # distributed samples with mean=1 and spread=2  
#  
# N <- 200; mean <- 1; sd <-2;  
# x <- rnorm(N, mean, sd) # Generate normal distributed  
#                          # sequence.  
#  
# Then run this R code to verify the function of a script call.  
#  
source("Script_1.R") # Call script for normal distributed data.  
plot(x)  
title(main="Plot of vector with normal distributed samples")  
hist(x)  
title(sub="Normally distributed samples")
```

The R Environment.

```
#####  
# 3.1: Reusing the result of an analysis  
#       Using output as input.  
#       [Kabacoff, 2015] page 17  
#  
# Linear regression predicting miles pr. gallon (mpg)  
# from car weight (wt)  
lm(mpg~wt, data=mtcars) # Display result on console.  
lmfit <- lm(mpg~wt, data=mtcars) # Create new object  
summary(lmfit)  
pdf(file="fig_1_3_1_lmfit.pdf")  
plot(lmfit)  
dev.off()  
par(opar.org) # Restore the original graphics parameters.
```

The R Environment



The R Environment.

```
#####  
# 4.0: A few R console commands etc.  
#  
# Ctrl l: Clear console.  
# ls(): List all objects in workspace.  
# rm(list=ls()): Remove all objects from workspace.  
# rm(a): Remove the object a.  
# q(): quit the R program.  
# version: List present version of installed R.  
# Sys.sleep(): Suspend execution in given time without busy wait.  
# Esc: Stop running code in RStudio.
```


The R Environment.

```
#####  
# 5.0: Controlling workspace,  
#       cf. [Kabacoff, 2015] page 12 table 1.3  
#  
wd_ini <- getwd() # Get the initial working directory.  
wd_ini          # Display the initial working directory.  
#  
# Set working directory to a new folder: */Session_2  
# Alternatively use manually the RStudio  
# "Session" -> "Set working directory".  
# Note: These operations do not create the actual  
#       folder */Session_2.  
#  
wd_1="/Session_2" # String for new working directory.  
?paste()          # Check manual for paste().  
# Concatenate strings to new working directory.  
wd_new=paste(wd_ini,wd_1, sep="")  
# Concatenated without space separator, rep. by ""  
#
```

The R Environment.

```
wd_new          # Display the new working directory path.
setwd(wd_new)   # Finally set new working directory */Session_2.
getwd()         # Check that new working directory is as deired.
setwd(wd_ini)   # Restore the initial working directory.
getwd()         # Check that the initial working directory is back.
#
# Path to the library
.libPaths()
##### End of script #####
```

References I



Joseph Adler (2012)

R in a Nutshell

O'Reilly



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