R Programming as a Part of Bigdata Course Visualization

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

16 Februar 2017

Visualization of a Selection of Datasets

```
# filename 4_R_Intro_Visualization.R
#####################################################
# Objective:
# Introduction of selection of visualization methods using
# - introduction to graphics in R
# - exemplifying a selection of graphics function and
#   combinations of functions for visualization of datasets.
#
#       "R Intro Visualization"  TABLE OF CONTENTS
#
# 4.0: Enter main R Language documentation and path to library.
# 4.1: List all data sets in the standard R installation
#      package "datasets".
# 4.2: Examples on R code for basic operations on graphs.
#      Ref. [Kabacoff, 2015] pages 48- . This section includes:
#      - Plot of regression line.
#      - Plot regression line to pdf or jpg file.
# 4.3: Examples on applying graphic parameters for fonts, colors,
#      axes, labels. Ref. [kabacoff, 2015] pages 49 - .
# 4.4: Examples from the RColorBrewer package, for generating
#      color palettes.
# 4.5: Examples on gray-levels, from the R base installation.
```

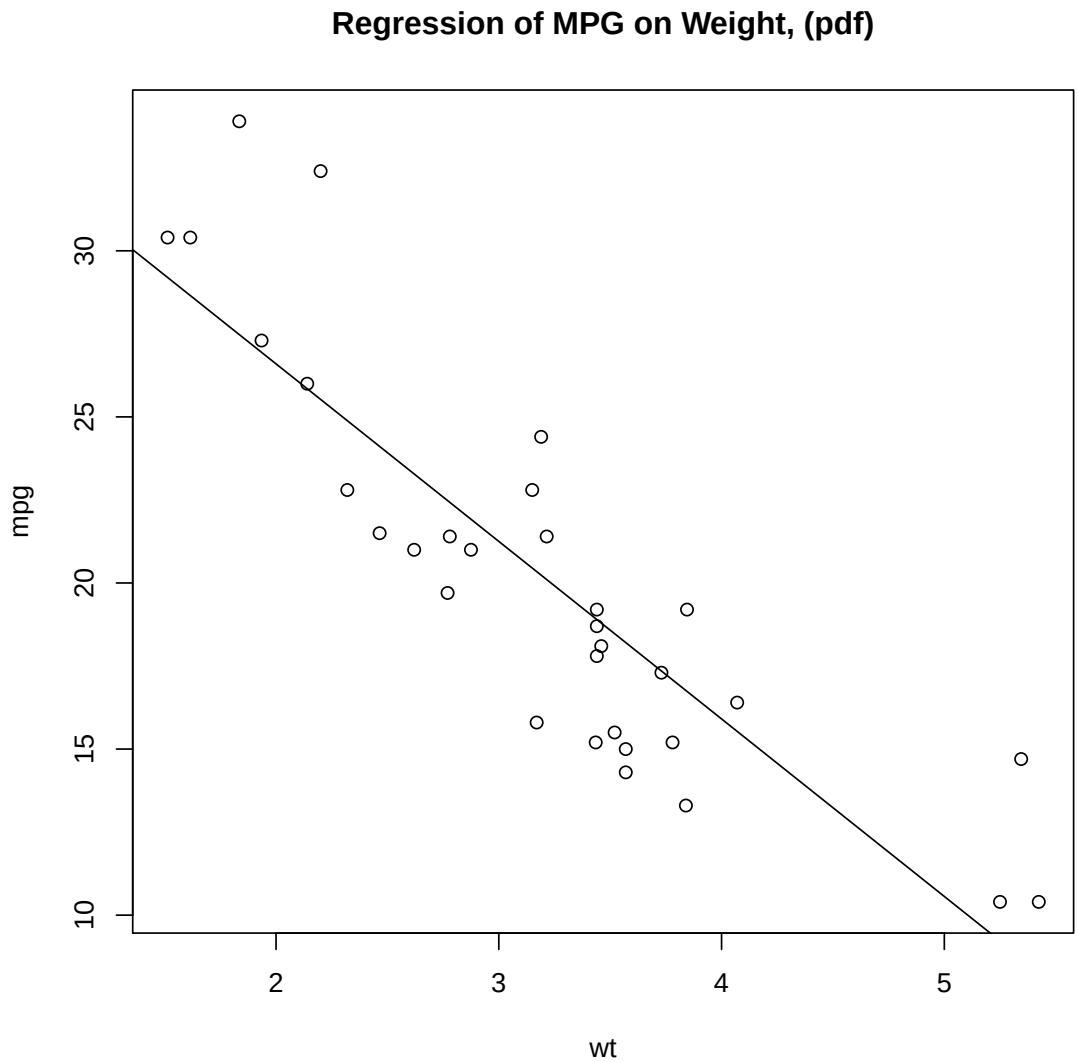
Visualization of a Selection of Datasets

```
# 4.6: Examples on text objects in graphics.
# 4.7: Example on margin control of plots, [Kabacoff, 2015] p. 55
# 4.8: Example on comparing 2 plots, [kabacoff, 2015] list 3.3. p. 61
# 4.9: Example text annotations on graph, [Kabacoff, 2015], p. 61, 62
#              Scatterplot with labeled points.
#              This example generates also a pdf plot
# 4.10: Example: One figure with 4 plots organized 2 x 2.
#       [Kabacoff, 2015] page 65.
# 4.11: Example: One figure with one column, 3 rows, org: 3 x 1.
#       [Kabacoff, 2015] page 65.
# 4.12: Example: One figure with 3 plots, organized: 1 fig, 2 figs.
#       [Kabacoff, 2015] page 67.
# 4.13: Example on combining a scatter plot with two box plots.
#       Ref. [Kabacoff, 2015] p. 69.
# 4.14: Example on Comparing groups by using parallel Box plots.
#       Ref. [Kabacoff, 2015] p. 129.
```

Visualization of a Selection of Datasets

```
# 4.15: Example on plotting a large number of labeled values on a
#       simple horizon scale, sorted and colored.
#       Ref. [Kabacoff, 2015] p. 134.
# 4.16: Example Bubble plot with point size area prop. to a
#       variable, using the mtcar dataset.
#       Ref. [Kabacoff, 2015] p. 267.
# 4.17: Example on 3d Scatter plot, Ref. [Kabacoff, 2015] p. 263.
#       Basic 3D plot
#       3D plot with vertical lines
#       Mouse controlled spinning of a 3D plot.
# 4.18: Creating and visualizing a selection of 3D clusters of
#       Gaussian and Uniform distributions using the function
#       plot3d().
#
```

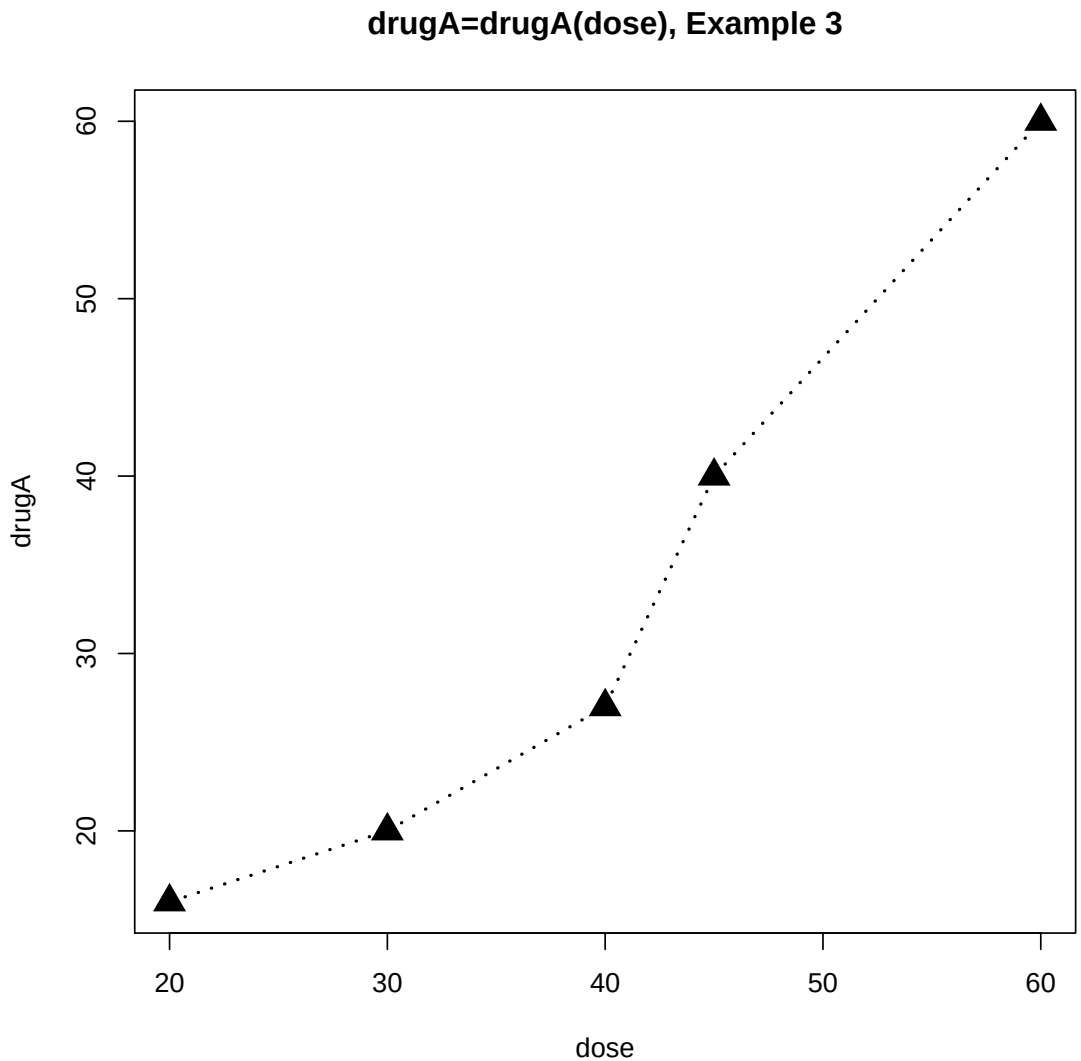
Section 4.2: Regression of mpg (miles pr Gallon) on wt (weight)



Regression of mpg (miles pr Gallon) on wt (weight)

```
#
opar.org <- par(no.readonly=TRUE) # Org. graphical parameters.
##################################################
# 4.2: Examples on R code for basic operations on graphs.
#      Ref. [Kabacoff, 2015] page 48
#
#                        Output graphics in pdf file.
pdf("mpg_regression.pdf")
attach(mtcars)
plot(wt, mpg)
abline(lm(mpg~wt))
title("Regression of MPG on Weight, (pdf)")
detach(mtcars)
dev.off()
par(opar.org)   # Restore original graphics parameters.
#
```

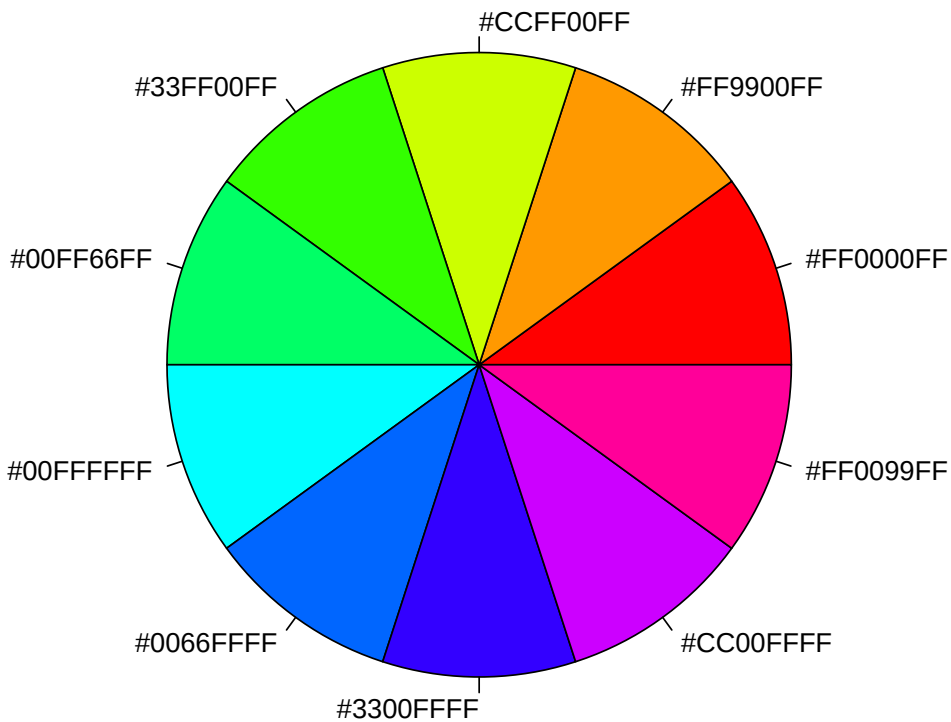
Section 4.3: Plot(Dose,DrugA), Param: b, lty, lwd, pch, cex



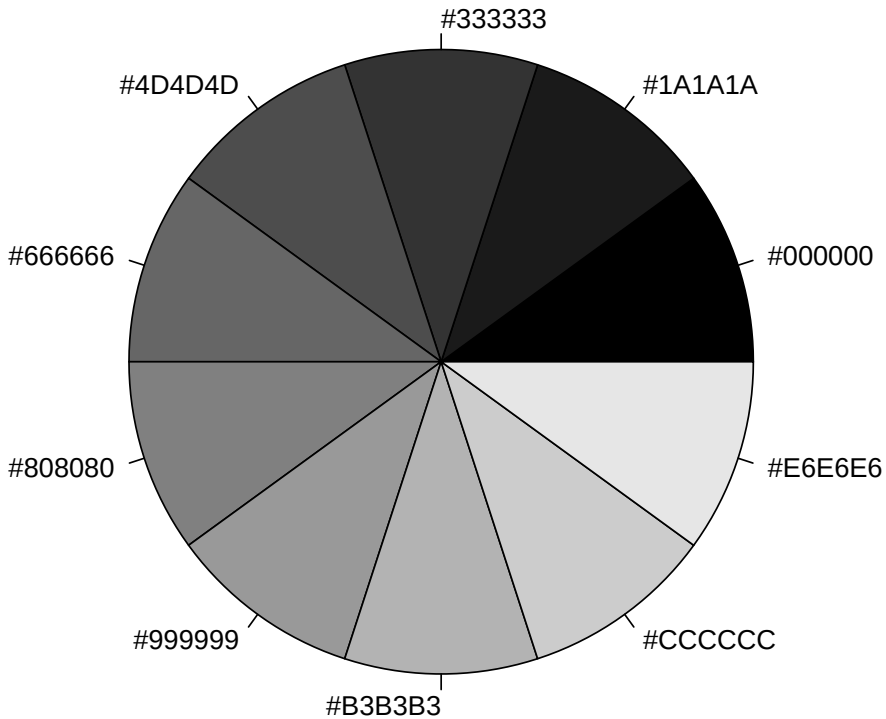
Section 4.3: Graphics param. for text, lines, colors

```
# Combine the plotting parameters, incl. symbol and line width
# scaling in one plot command
plot(dose, drugA, type="b", lty=3, lwd=2, pch=17, cex=2)
title("drugA=drugA(dose), Example 3")
   # b: Plot both points and lines.
   # lty: Line type 3, Fig. 3.5, p. 51 [Kabacoff, 2015].
   # lwd: Scale line width 2 times.
   # pch: use symbol 17 (triangle),
   #      Fig. 3.4, p. 51 [Kabacoff, 2015].
   # cex: scale symbol size 2 times, relative to default.
par(opar.org)    # Restore original parameters.
```

Section 4.5: Color pie



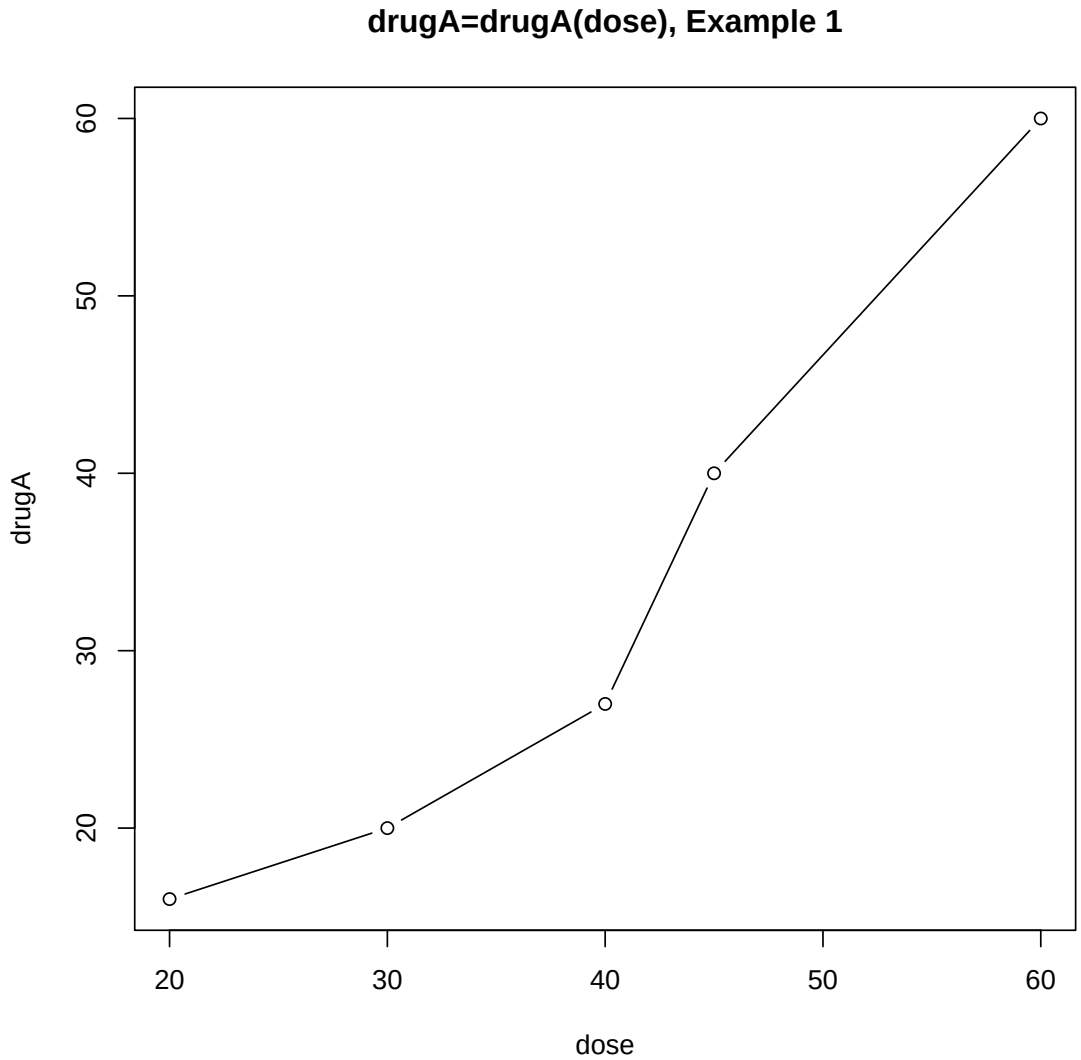
Section 4.5: Gray pie



Section 4.5: Color and Gray Levels.

```
#########################################################
# 4.5: Examples on gray-levels, from the R base installation.
#      Ref. [Kabacoff, 2015] page 53
#
?gray()  # Manual for the gray() function
n <- 10
# Create vector of n contiguous colors.
example_colors <- rainbow(n)
pie(rep(1,n), labels=example_colors, col=example_colors)
# Create vector of colors from a vector of gray levels.
example_grays <- gray(0:n/n)
pie(rep(1,n), labels=example_grays, col=example_grays)
```

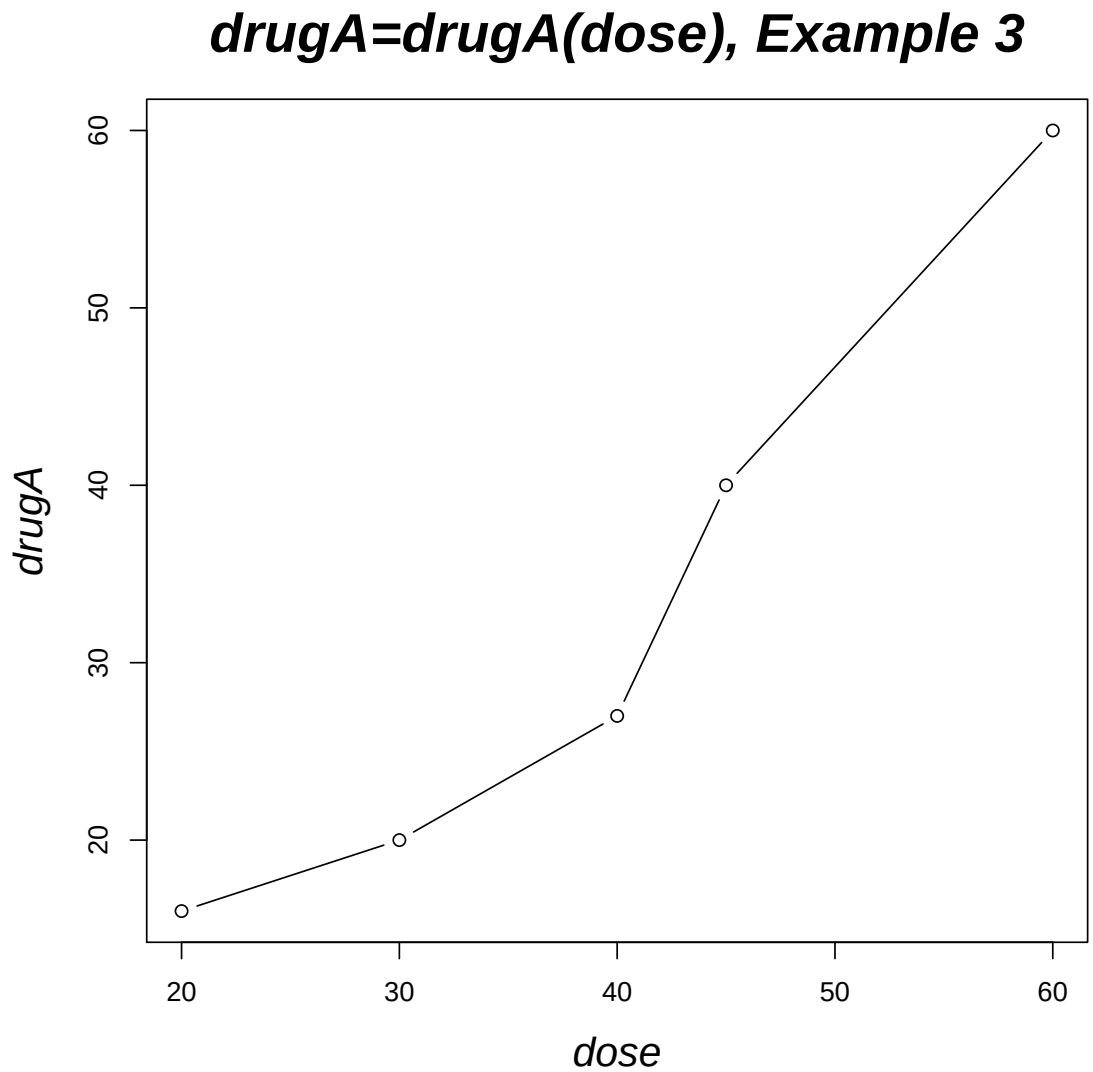
Section 4.6 Text Objects in Graphics.



Section 4.6 Text Objects in Graphics.

```
#########################################################
# 4.6: Examples on text objects in graphics.
#  [Kabacoff, 2015] page 53, Table 3.4 (font size) &
#                            Table 3.5 (font families)
dose <- c(20, 30, 40, 45, 60)
drugA <- c(16, 20, 27, 40, 60)
drugB <- c(15, 18, 25, 31, 40)
#
plot(dose, drugA, type="b")  # b: "Plot lines and points".
title("drugA=drugA(dose), Example 1")
par(opar.org)              # Restore original parameters.
#
# Generate pdf plot
pdf("fig_4_6_txt1.pdf")
plot(dose, drugA, type="b")  # b: "Plot both lines and points".
title("drugA=drugA(dose), Example 1")
dev.off()
par(opar.org)
```

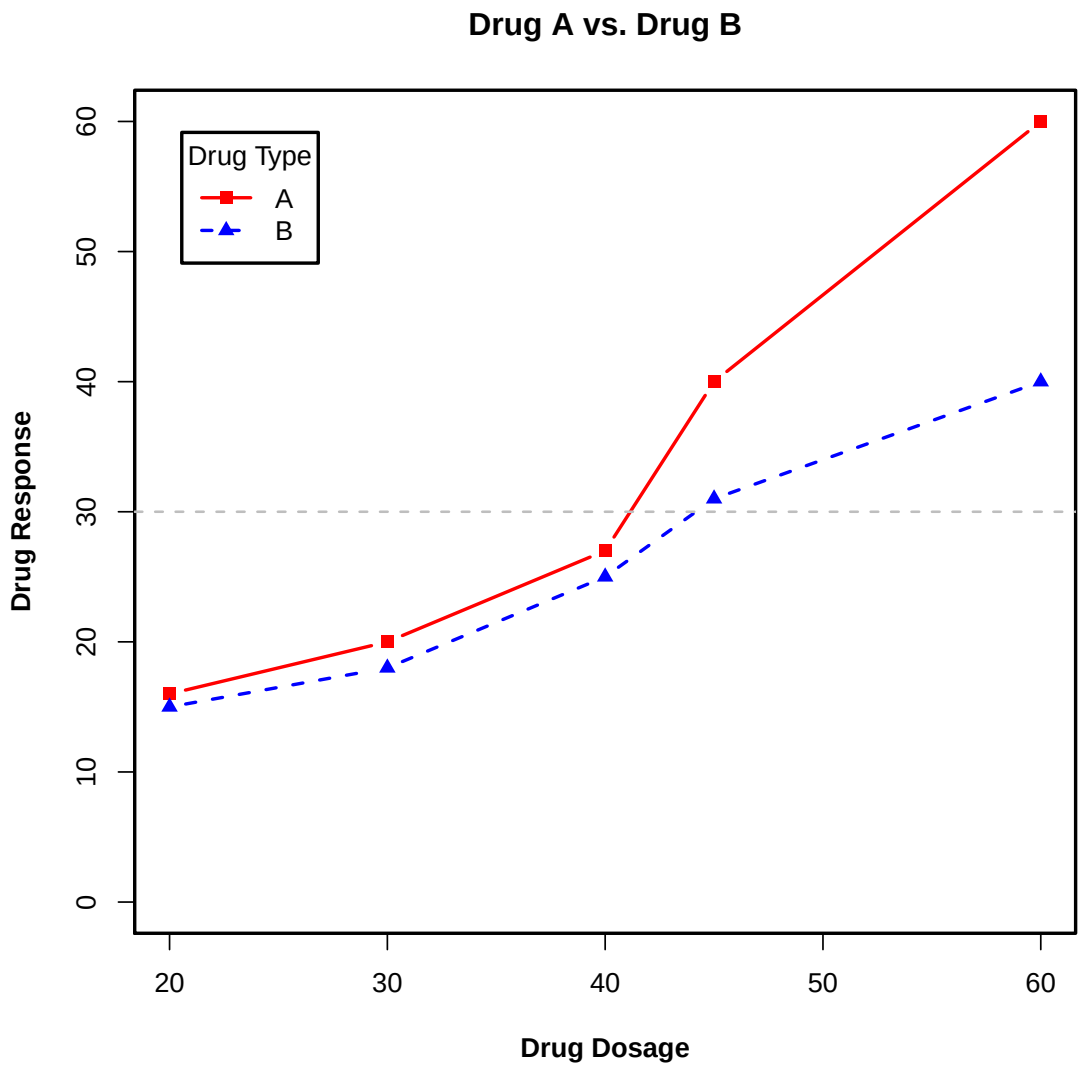
Section 4.6 Text Objects in Graphics.



Section 4.6 Text Objects in Graphics.

```
# Set new parameters for plot
par(font.lab=3, cex.lab=1.5, font.main=4, cex.main=2)
plot(dose, drugA, type="b")  # b: "Plot lines and points".
title("drugA=drugA(dose), Example 3")
par(opar.org)   # Restore original parameters.
#
# Generate pdf plot
pdf("fig_4_6_txt2.pdf")
par(font.lab=3, cex.lab=1.5, font.main=4, cex.main=2)
plot(dose, drugA, type="b")  # b: "Plot lines and points".
title("drugA=drugA(dose), Example 3")
dev.off()
par(opar.org)
```

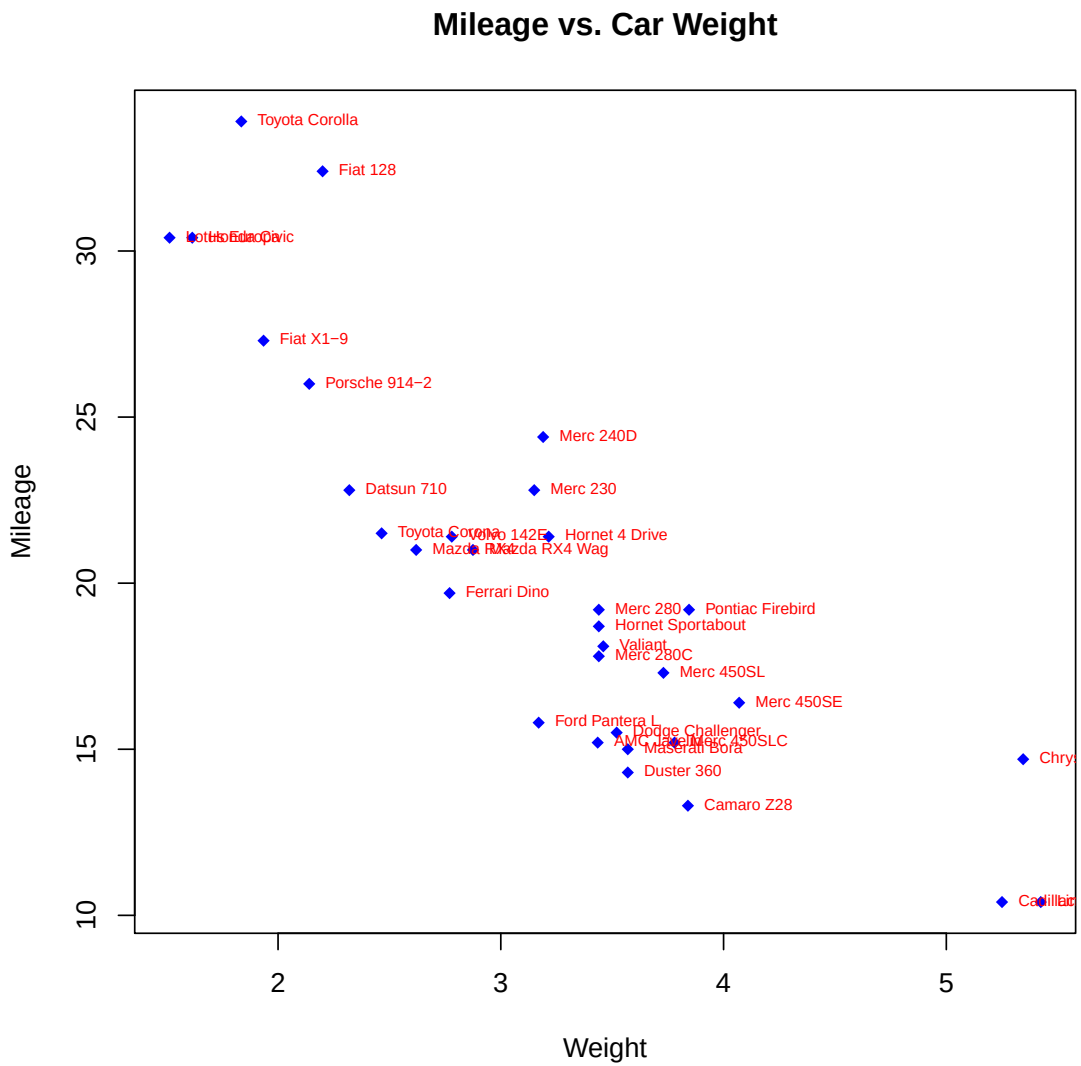
Section 4.8: Plot with legend



Section 4.8: Plot with legend.

```
dose <- c(20, 30, 40, 45, 60)
drugA <- c(16, 20, 27, 40, 60)
drugB <- c(15, 18, 25, 31, 40)
par(opar.org)
par(lwd=2, cex=1, font.lab=2)
plot(dose, drugA, type="b",
     pch=15, lty=1, col="red", ylim=c(0, 60),
     main="Drug A vs. Drug B",
     xlab="Drug Dosage", ylab="Drug Response")
     lines(dose, drugB, type="b",
     pch=17, lty=2, col="blue")
abline(h=c(30), lwd=1.5, lty=2, col="gray")
#
minor.tick(nx=3, ny=3, tick.ratio=0.5)
legend("topleft", inset=0.05, title="Drug Type", c("A","B"),
        lty=c(1, 2), pch=c(15, 17), col=c("red", "blue"))
par(opar.org)   # Restore original parameters.
```

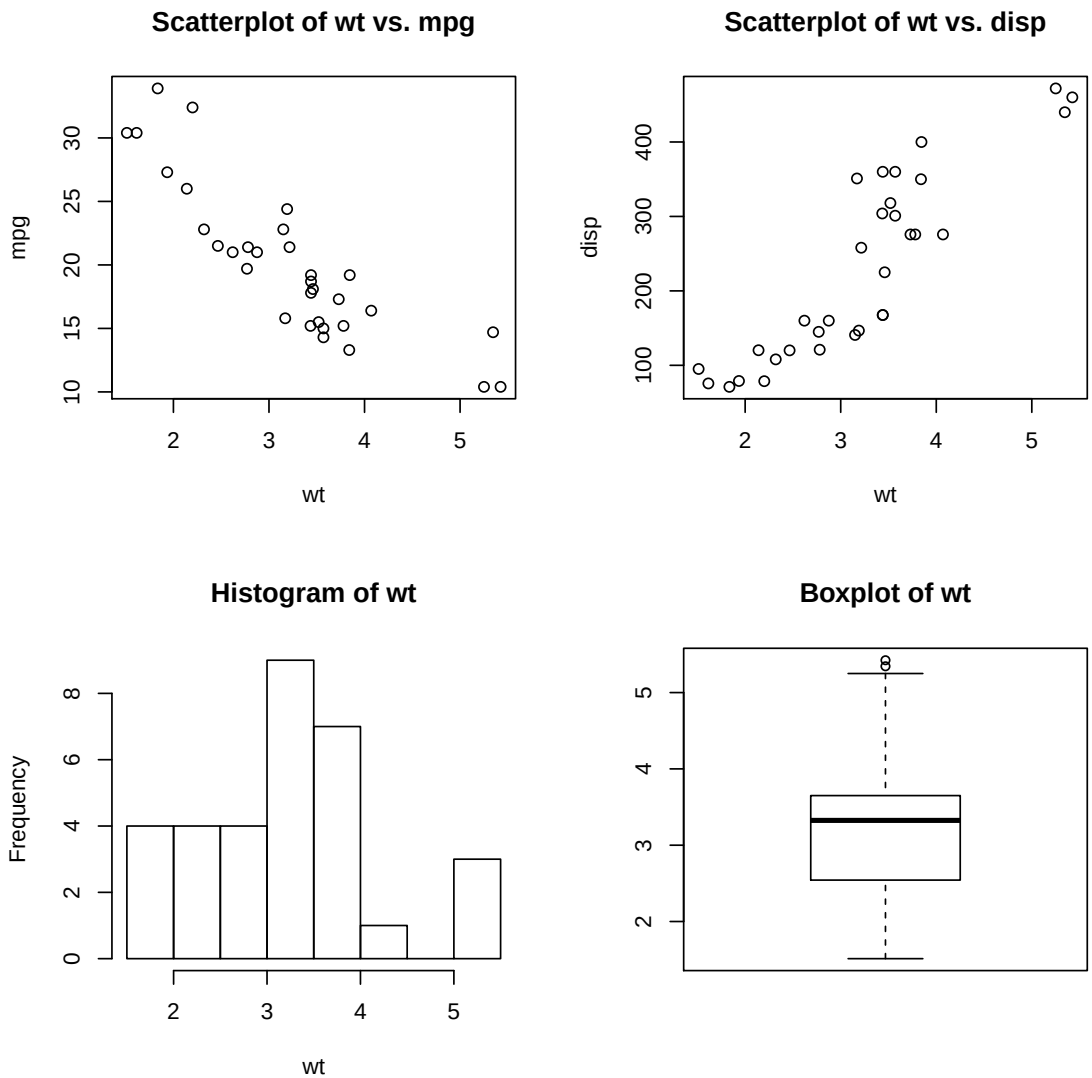
Section 4.9: Milage versus car weight



Section 4.9: Annotation on Scatterplot

```
##########################################
# 4.9: Example on text annotations on graph,
#      ref. [Kabacoff, 2015], p. 61, 62.
#      Scatterplot with labeled points.
par(opar.org)
attach(mtcars)
plot(wt, mpg,
     main="Mileage vs. Car Weight",
     xlab="Weight", ylab="Mileage",
     pch=18, col="blue")
text(wt, mpg,
     row.names(mtcars),
     cex=0.6, pos=4, col="red")
# cex, cf. Table 3.4, p. 53.
# pos, cf. Table 3.9, p. 62.
detach(mtcars)
par(opar.org)
```

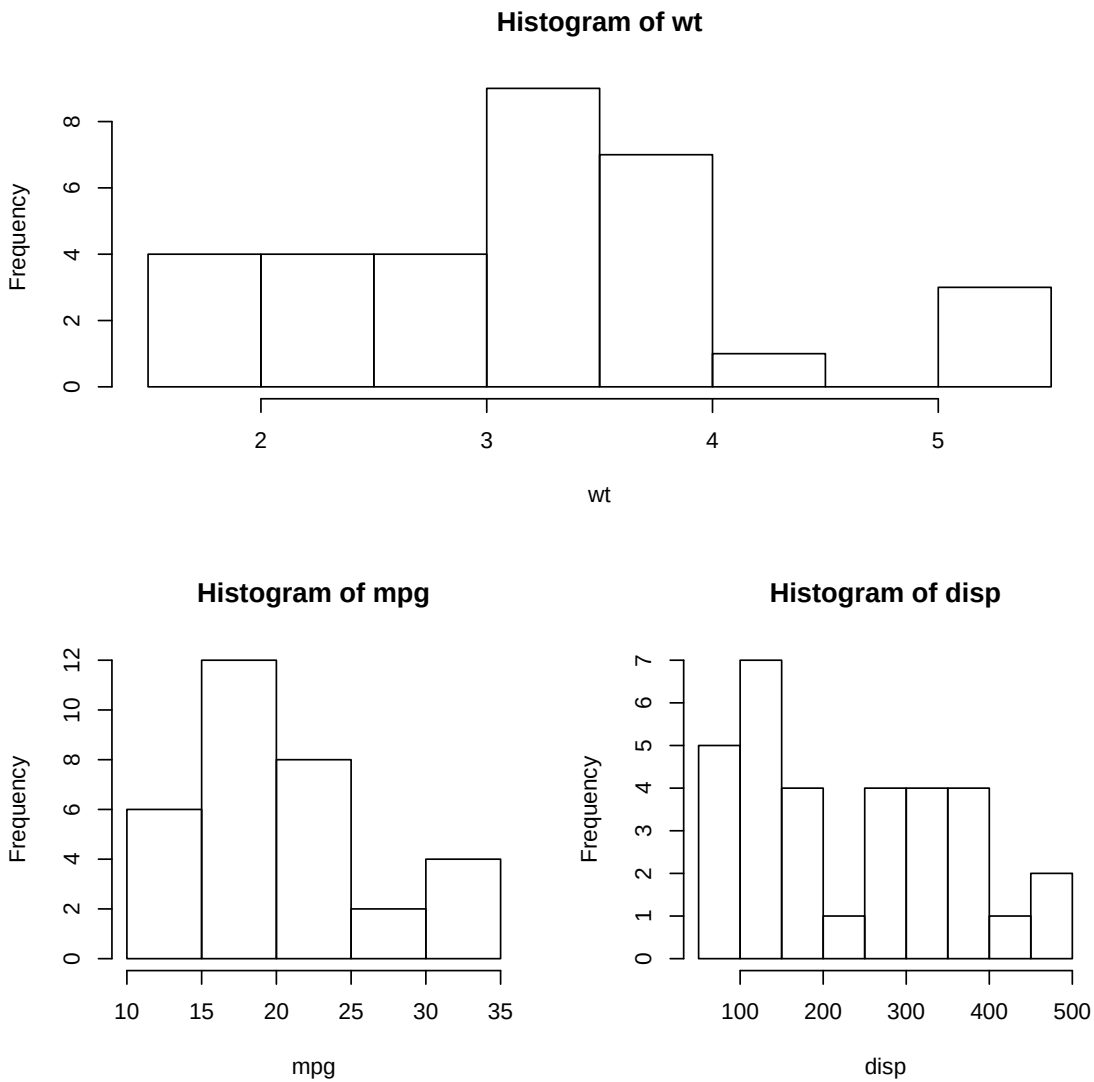
Section 4.10: Figure with 2 x 2 plots



Section 4.10: Figure with 2 x 2 plots

```
#################################################
# 4.10: Example:
#        One figure with 4 plots organized 2 x 2.
#       [Kabacoff, 2015] page 65.
par(opar.org)
attach(mtcars)
par(mfrow=c(2,2))
plot(wt,mpg, main="Scatterplot of wt vs. mpg")
plot(wt,disp, main="Scatterplot of wt vs. disp")
hist(wt, main="Histogram of wt")
boxplot(wt, main="Boxplot of wt")
detach(mtcars)
par(opar.org)
```

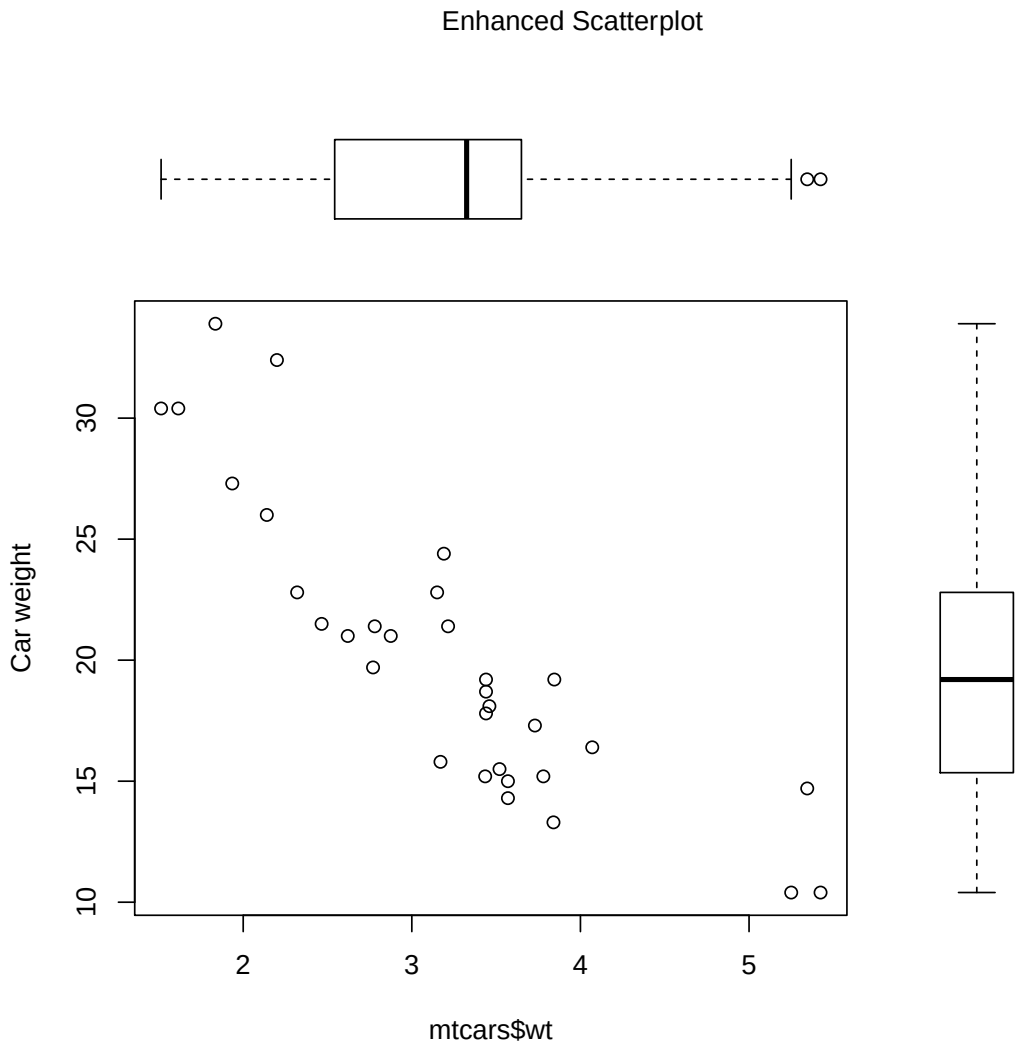
Section 4.12: One figure with 3 plots



Section 4.12: One figure with 3 plots

```
#########################################################
# 4.12: Example:
#   One figure with 3 plots, organized: 1 fig, 2 figs.
#   [Kabacoff, 2015] page 67.
#
attach(mtcars)
?layout()  # Manual for complex plot layout function.
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
hist(wt)
hist(mpg)
hist(disp)
detach(mtcars)
par(opar.org)
```

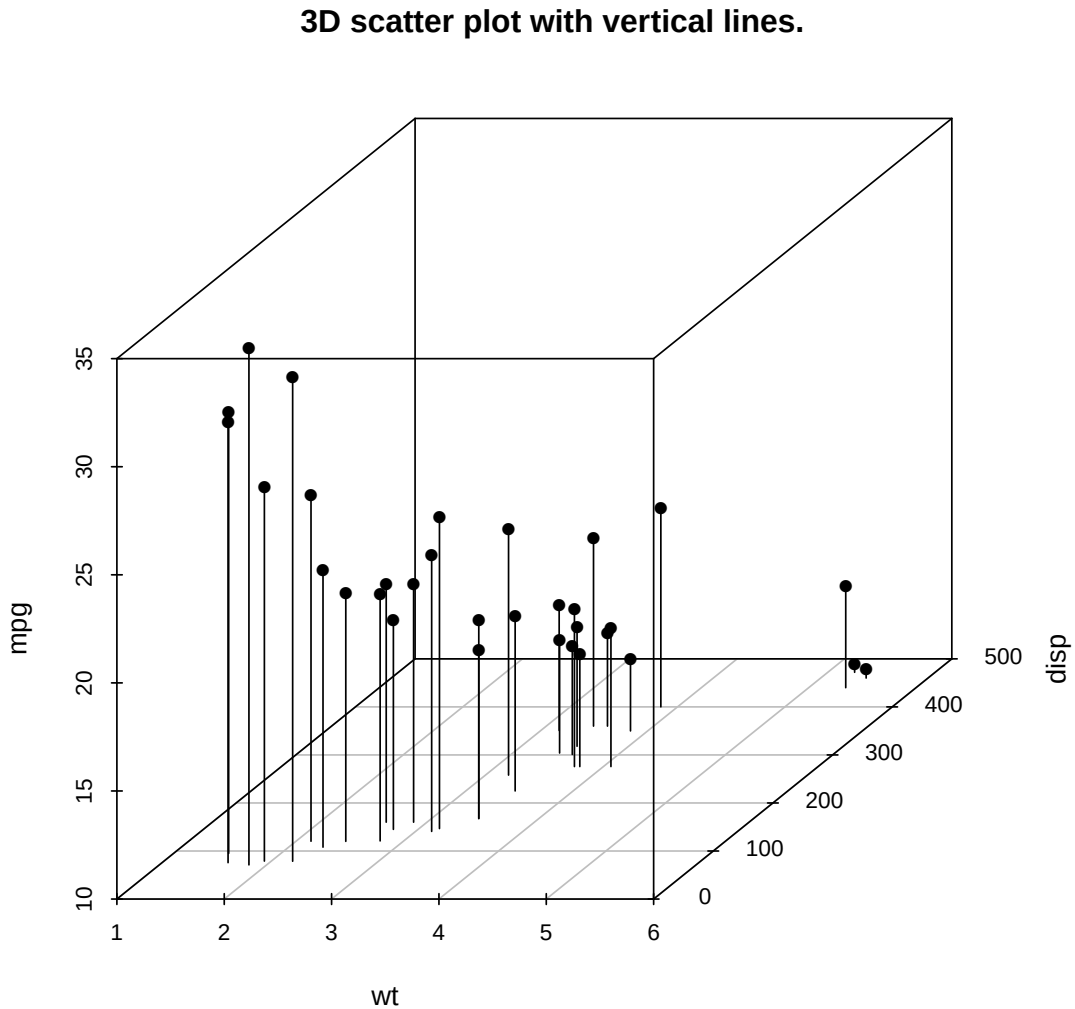
Section 4.13: Scatter plot with two box plots.



Section 4.13: Scatter plot and two box plots

```
#############################################################
# 4.13:
# Example combining a scatter plot with two box plots.
#       Ref. [Kabacoff, 2015] p. 69.
par(opar.org)
par(fig=c(0, 0.8, 0, 0.8)) # Set up scatter plot
 # Center part of fig. from (0,0) to 80% of x and y axis.
plot(mtcars$wt, mtcars$mpg,
     xlabel="Miles per Gallon",
     ylab="Car weight")
par(fig=c(0, 0.8, 0.55, 1), new=TRUE) # Set up box plot.
boxplot(mtcars$wt, horizontal=TRUE, axes=FALSE)
par(fig=c(0.65, 1, 0, 0.8), new=TRUE) # Set up box plot
boxplot(mtcars$mpg, horizontal=FALSE, axes=FALSE)
mtext("Enhanced Scatterplot", side=3, outer=TRUE, line=-3)
    # Print text on one of 4 margins.
    # side=3: top margin.
    # outer=TRUE, use outer margin.
    # line=-3, start on margin line -3.
```

Section 4.17: 3D Scatter plot.



Section 4.17: 3D Scatter Plot

```
#####################################################
# 4.17: Example 3D Scatter plot, [Kabacoff, 2015] p. 263.
#
install.packages("scatterplot3d",lib="C:/R_packages/")
install.packages("rgl",lib="C:/R_packages/")
install.packages("plot3D",lib="C:/R_packages/")
#
library(scatterplot3d,lib="C:/R_packages/")
library(rgl,lib="C:/R_packages/")
library("plot3D",lib="C:/R_packages/")
#
# Basic scatter3d plot
# Use the mtcars data set from package "datasets"
attach(mtcars)
```

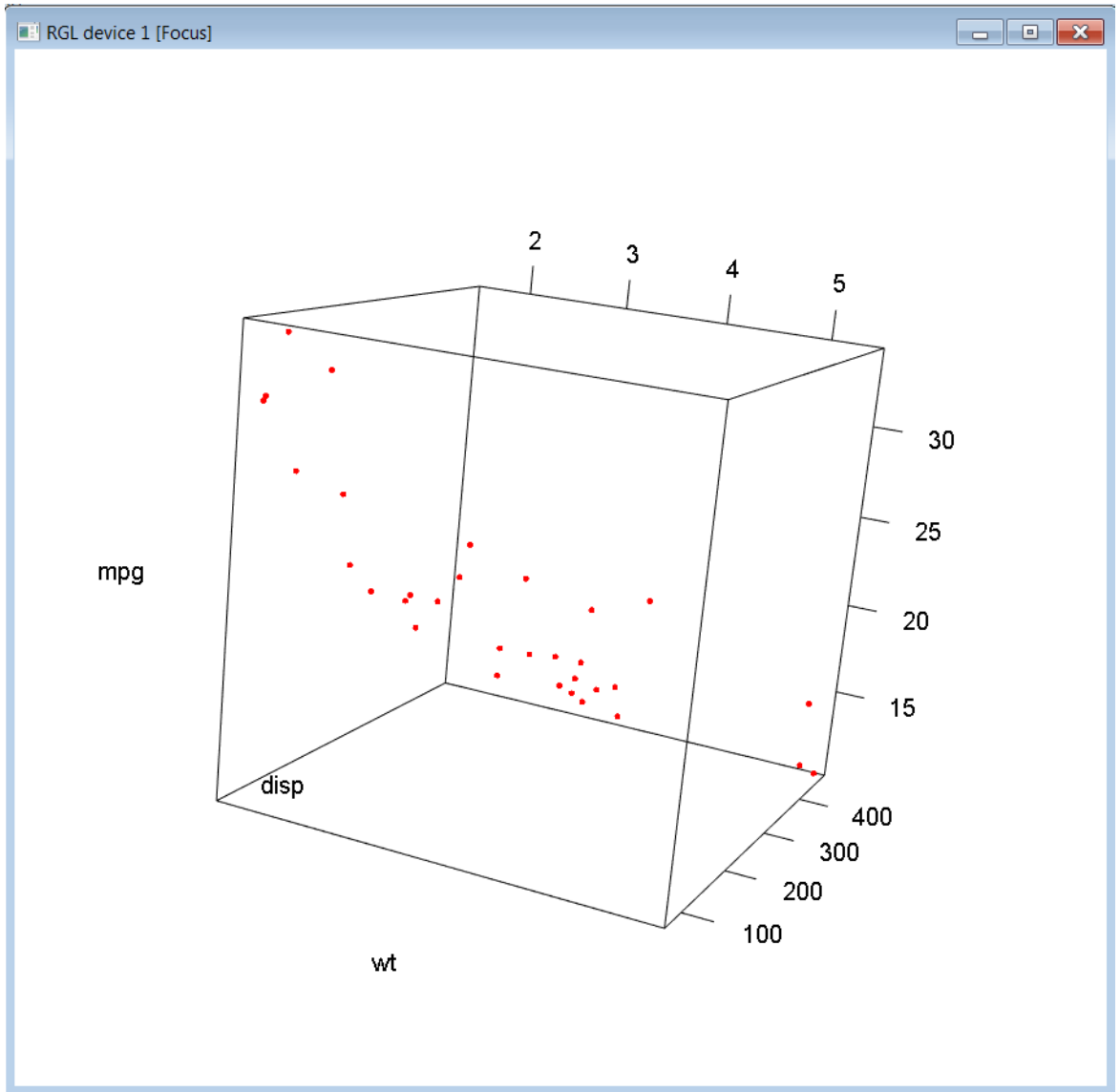
Section 4.17: 3D Scatter Plot

```
scatterplot3d(wt, disp, mpg,
              main="Basic 3D Scatter Plot")
#
# scatterplot3d with vertical lines,
scatterplot3d(wt, disp, mpg,
       pch=16,
       highlight.sd=TRUE,
       type="h",
       main="3D scatter plot with vertical lines.")
```

Section 4.17: 3D Scatter Plot

```
# Generate pdf plot of the above
par(opar.org)
pdf("fig_4_17_scatterplot3d.pdf")
scatterplot3d(wt, disp, mpg,
       pch=16,
       highlight.sd=TRUE,
       type="h",
       main="3D scatter plot with vertical lines.")
dev.off()
par(opar.org)
```

Section 4.17: 3D Plot with Mouse Control of View Direction



Section 4.17: 3D Plot with Mouse Control of View Direction

```
#
# 3D plot with mouse controlled view direction
#
attach(mtcars)
plot3d(wt, disp, mpg, col="red", size=5 )
```

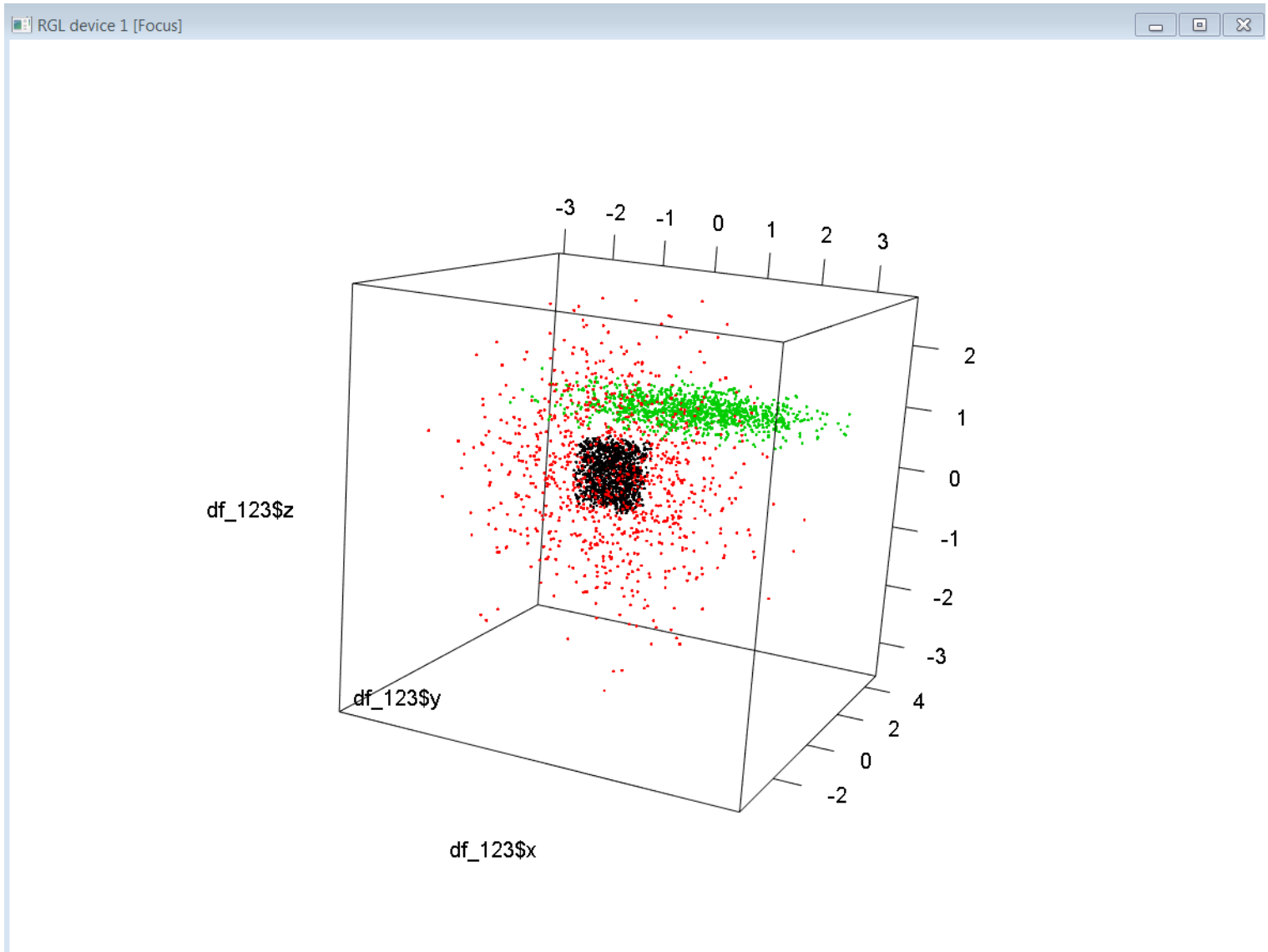
Section 4.18: Examples on Constructing 3D plots.

```
#######################################################
# This example needs the packages from the above Example 4.17.
# 4.18: Example on 3D plot of Gaussian and uniform distributions
#
# Create 3D cluster, uniform distributed, augmented with color v
?runif()
N1 <- 1000  # Use N1 samples per cluster.
x1 <- runif(N1, min=-0.50, max=0.50)   # Uniform from -1.0 to +1
y1 <- runif(N1, min=-0.50, max=0.50)
z1 <- runif(N1, min=-0.50, max=0.50)
col_1 <- rep(1,N1)   # Color no. 1
#
# Create 3D cluster, Gaussian distributed, augmented with color
?rnorm()
N2 <- 1000  # Use N2 samples per cluster.
x2 <- rnorm(N2, mean=0, sd=1)
y2 <- rnorm(N2, mean=0, sd=1)
z2 <- rnorm(N2, mean=0, sd=1)
col_2 <- rep(2,N2)
#
```

Section 4.18: Examples on Constructing 3D plots.

```
# Concatenate the two datasets into one dataframe with colors
# associated each dataset.
#
x <-c(x1,x2); y <-c(y1,y2); z <-c(z1,z2);col_12<-c(col_1,col_2);
df_12 <- data.frame(x,y,z,col_12)
plot3d(df_12$x, df_12$y, df_12$z, col=df_12$col_12, size=3, type
?plot3d()              # Check the manual for plot3d()
# type= s => sphere in each datapoint-
# type= p => points in each datapoint.
#
# Now extend with one more 3D Gaussian cluster with
# standard deviations sdx=1, sdy=1, sdz=0.1 in center 1,1,1.
N3 <- 1000  # Use N2 samples per cluster.
x3 <- rnorm(N3, mean=1.0, sd=1.0)
y3 <- rnorm(N3, mean=1.0, sd=1.0)
z3 <- rnorm(N3, mean=1.0, sd=0.1)
col_3 <- rep(3,N3)
xx <-c(x1,x2,x3); yy <-c(y1,y2,y3); zz <-c(z1,z2,z3);col_123<-c(
df_123 <- data.frame(xx,yy,zz,col_123)
plot3d(df_123$x, df_123$y, df_123$z, col=df_123$col_123, size=3,
##################################################################
```

Section 4.18: 3D Plot with Three Clusters.



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


R Data import/Export (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