Introduction to Data Visualization in

Below:

Data Entry and Manipulation

Basic Graphics

Useful Links

Why R?

- programming language and software environment for stats and graphics

- free (<https://www.r-project.org/>) and easy to use

- widely used and well supported

Data Entry and Manipulation

#1 Set a working directory to where your data will be

setwd('/Users/davorka/Desktop/DATA')

getwd()

# alternatively via: Misc/Change Working directory

#2 Reading in the data

#standard txt file

data<-read.table('data.txt', header = T, sep = ',')

head(data)

#previous saved RData

load('data2.RData')

head(data2)

str(data2)

table(data2$gender)

#manipulating the data

newdata<-data2[1:100, -3] # syntax data[rows, columns]

newdata<-newdata[newdata$gender == 1,]

str(newdata)

newdata$weight<-newdata$weight/2.205

colnames(newdata)[3]<-"weightKG"

str(newdata)

# saving newdata

save(newdata,file='newdata.RData') #or

write.table(newdata, "newdata.txt", sep=",")

rm(newdata) #remove newdata, rm(list=ls()) to remove all

Basic Graphics in R

#R allows a range of graphical parameters (<https://www.statmethods.net/advgraphs/parameters.html>) and has a large palate of predefined colors (http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf), but you can also custom define colors

help(rgb)

myred <- rgb(186, 12, 10, max = 255, alpha = 50)

mygreen <- rgb(0, 140, 140, max = 255, alpha = 50)

attach(data2) # vs detach()

plot(height)

data2<-data2[height>140,]

detach()

attach(data2)

plot(height)

# Basic histogram

hist(height,

breaks = 10, #number of cells, or breaks = c(140,160,170,180,200)

main = 'Histogram of heights from data set 2', #title

xlab = 'height', #x axis

ylab = 'frequency', #y axis

col = myred)

#Scatter Plot

plot(weight~height, main = 'A plot of weight vs. height', pch = 16, col = mygreen, las = 1)

#by gender

plot(weight~height, main = 'A plot of weight vs. height', pch = 16, col = gender, las = 1)

legend("topleft", inset=.02, legend=c( "gender 1", "gender 2"), col=c("red", "black"), pch = 16)

# Adding smooth lines

plot(weight~height, main = 'A plot of weight vs. height', pch = 16, col = mygreen, las = 1)

data2<-data2[order(data2$height),]

dethach()

attach(data2)

loefit<-loess( weight~height, span=1) #span = smooting parameter 0-1 , with 1 smoothest lines

lines(loefit$fitted~loefit$x, type = 'l')

#by gender

plot(weight~height, main = 'A plot of weight vs. height', pch = 16, col = gender, las = 1)

data21<-data2[gender==1,]

loefit<-loess(data21$weight~data21$height, span=0.5)

lines(loefit$fitted~loefit$x, type = 'l', lwd = 3)

data22<-data2[gender==2,]

loefit<-loess(data22$weight~data22$height, span=0.5)

lines(loefit$fitted~loefit$x, type = 'l', lwd = 3, col = 'red')

pairs(data2) # see col = gender <-useful matrix plot

#Boxplot

boxplot(weight~gender, main = 'A boxplot of weight', xlab = 'gender', pch = 16, col = mygreen, las = 1) # see with notch = T

#multiple plots

par(mfrow=c(1,3))

hist(height, breaks = 10, main = 'Histogram of heights from data set 2', xlab = 'height', ylab = 'frequency', col = myred)

plot(weight~height, main = 'A plot of weight vs. height', pch = 16, col = gender, las = 1)

data21<-data2[gender==1,]

loefit<-loess(data21$weight~data21$height, span=0.5)

lines(loefit$fitted~loefit$x, type = 'l', lwd = 3)

data22<-data2[gender==2,]

loefit<-loess(data22$weight~data22$height, span=0.5)

lines(loefit$fitted~loefit$x, type = 'l', lwd = 3, col = 'red')

legend("bottomright", inset=.02, legend=c( "gender 1", "gender 2"), col=c("red", "black"), pch = 16)

boxplot(weight~gender, main = 'A boxplot of weight', xlab = 'gender', pch = 16, col = mygreen, las = 1) # see with notch = T

#Lattice is great for multiple panels

library(lattice)

xyplot(weight~height|gender, pch = 16, col = mygreen)

#add a grid and a smoother

xyplot(weight~height|gender, panel = function(x, y) {

+ panel.grid(h = -1, v = 2)

+ panel.xyplot(x, y, pch = 16, col = mygreen)

+ panel.loess(x, y, span=0.5, lwd = 2, col = myred)

+ })

xyplot(weight~height|age, panel = function(x, y) {

+ panel.grid(h = -1, v = 2)

+ panel.xyplot(x, y, type = 'l', col = mygreen, lwd = 3)

+ })

Useful Links

Google any question you have about R

Coursera ( <https://www.coursera.org/> ) – audit R classes for free

For a nice tutorial also see:

<https://rpubs.com/drsong/9575>

<http://shinyapps.org/apps/RGraphCompendium/index.php?utm_content=buffer631a3&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer>