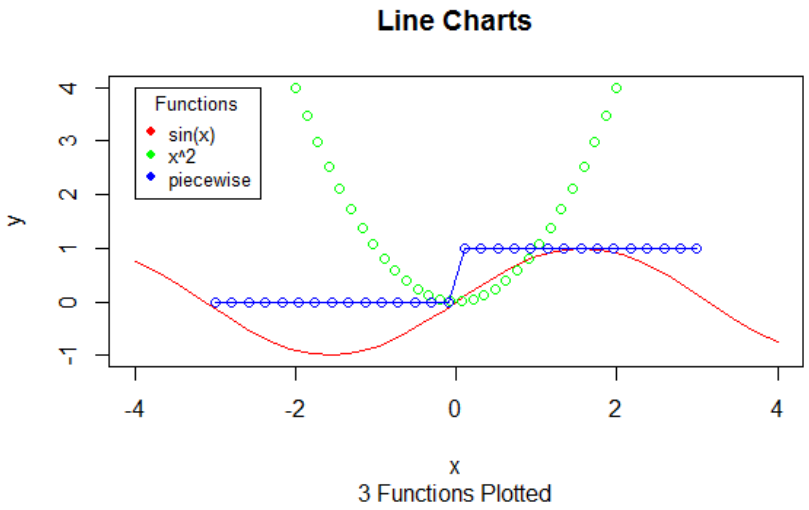
Lab 1

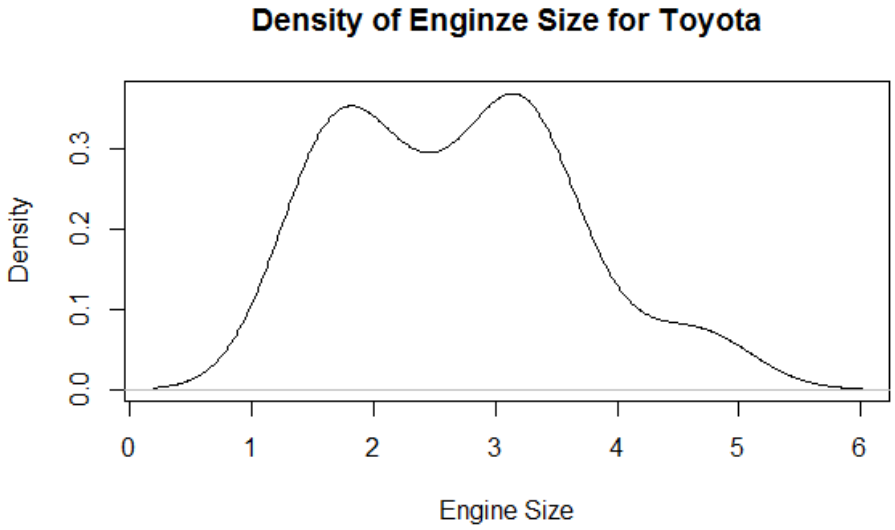
Steven Lin

Exercise 1

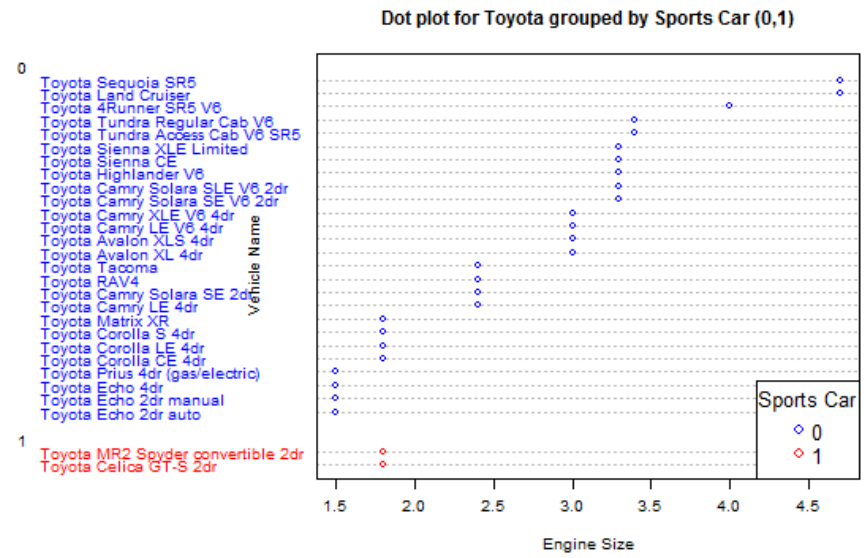


Exercise 2

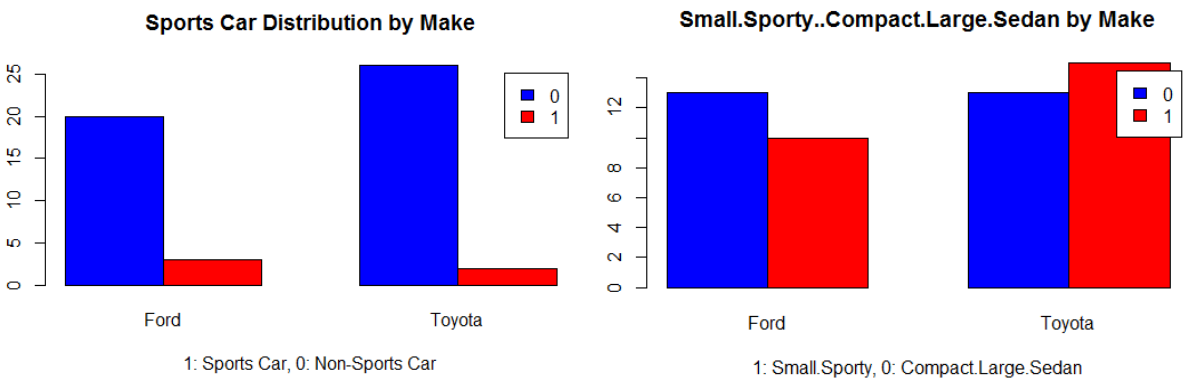
a)



b)


Dot plot for Toyota grouped by Sports Car (0,1)
0
Toyota Sequoia SR5
Toyota Land Cruiser
Toyota 4Runner SR5 V6
Toyota Tundra Regular Cab V6
Toyota Tundra Access Cab V6 SR5
Toyota Sienna XLE Limited
Toyota Sienna CE
Toyota Highlander V6
Toyota Camry Solara SLE V6 2dr
Toyota Camry Solara SE V6 2dr
Toyota Camry XLE V6 4dr
Toyota Camry LE V6 4dr
Toyota Avalon XLS 4dr
Toyota Avalon XL 4dr
Toyota Tacoma
Toyota RAV4
Toyota Camry Solara SE 2dr
Toyota Camry LE 4dr
Toyota Matrix XR
Toyota Corolla S 4dr
Toyota Corolla LE 4dr
Toyota Corolla CE 4dr
Toyota Prius 4dr (gas/electric)
Toyota Echo 4dr
Toyota Echo 2dr manual
Toyota Echo 2dr auto
1
Toyota MR2 Spyder convertible 2dr
Toyota Celica GT-S 2dr
Vehicle Name
Sports Car
0
1
1.5
2.0
2.5
3.0
3.5
4.0
4.5
Engine Size

c)


Sports Car Distribution by Make
0
5
10
15
20
25
0
1
Ford
Toyota
1: Sports Car, 0: Non-Sports Car
Small.Sporty..Compact.Large.Sedan by Make
0
2
4
6
8
12
0
1
Ford
Toyota
1: Small.Sporty, 0: Compact.Large.Sedan

Rcode

```
# Lab session 1 exercise

##### Load data ######################################################################

# My PC
main = "C:/Users/Steven/Documents/Academics/3_Graduate School/2014-2015 ~ NU/"

# Aginity
#main = "\\\\nas1/labuser169"

course = "MSIA_411_Data_Visualization"
datafolder = "/Lab/Data"
setwd(file.path(main,course, datafolder))


##### Ex1  ###########################################################################

#### 1-a ####
# http://www.statmethods.net/graphs/line.html

x1 = seq(-4,4,len=30)
y1 = sin(x1)

x2 = seq(-2,2,len=30)
y2 = x2^2

x3 = seq(-3,3,len=30)
y3 = 1*(x3>0)

# get the range for the x and y axis
xrange = range(c(x1,x2,x3))
yrange = range(c(y1,y2,y3))

# set up the plot
plot(xrange, yrange, type="n", xlab="x",ylab="y")
colors = c("red","green","blue")
linetype = c("l","p","o")

# add lines
lines(x1, y1, type=linetype[1],col=colors[1])
lines(x2, y2, type=linetype[2], col=colors[2])
lines(x3, y3, type=linetype[3], col=colors[3])

# add a title and subtitle
title("Line Charts", "3 Functions Plotted")
```

```
fnames = c("sin(x)","x^2","piecewise")

# add a legend
legend("topright", legend = fnames, cex=0.8, col=colors,
     pch=c(16,16,16),title="Functions")

##### Ex2  ################################################################

carsdata =read.csv("04cars data.csv",header=TRUE,
               na.strings=c("","*","NA"))

head(carsdata)
str(carsdata)

#### 2-a ####
index_Toyota = grep("Toyota",(carsdata$Vehicle.Name))
carsdata_Toyota = carsdata[index_Toyota,]

d=density(carsdata_Toyota$Engine.Size..l.)
plot(d, main="Density of Enginze Size for Toyota", xlab="Engine Size")

#### 2-b ####

# sort by engine size
carsdata_Toyota = carsdata_Toyota[order(carsdata_Toyota$Engine.Size..l.),]

#group them based on Sports.car
#to group things, must first turn them into factors
carsdata_Toyota$Sports.Car = factor(carsdata_Toyota$Sports.Car)


levels(carsdata_Toyota$Sports.Car)

#now create color groups:
carsdata_Toyota$color[carsdata_Toyota$Sports.Car=="0"] = "blue"
carsdata_Toyota$color[carsdata_Toyota$Sports.Car=="1"] = "red"

dotchart(carsdata_Toyota$Engine.Size..l.,
       labels=carsdata_Toyota$Vehicle.Name,
       cex=.6,
       groups=carsdata_Toyota$Sports.Car,
       color = carsdata_Toyota$color,
       xlab = "Engine Size",
       ylab = "Vehicle Name",
       main = "Dot plot for Toyota grouped by Sports Car (0,1)")

legend("bottomright" ,legend = c("0","1"),
     cex=0.8, col=c("blue","red"),
```

```
      pch=c(1,1), title="Sports Car")

#### 2-c ####

# Create a grouped bar chart for Toyota and Ford cars in
# order to compare the distribution of
# "Small.Sporty..Compact.Large.Sedan" and
# "Sports.Car" between the two car makes.

#distribution of cylinders in sports and non-sports cars
index_Toyota = grep("Toyota",(carsdata$Vehicle.Name))
index_Ford = grep("Ford",(carsdata$Vehicle.Name))

dataT = carsdata[c(index_Toyota),]
dataF = carsdata[c(index_Ford),]
dataT$make = "Toyota"
dataF$make = "Ford"

data2 = rbind(dataT,dataF)

table1 =table(data2$Sports.Ca,data2$make) #watch out for the order

table2 =table(data2$Small.Sporty..Compact.Large.Sedan
            ,data2$make) #watch out for the order

par(mfrow=c(1,2))

barplot(table1,
      main="Sports Car Distribution by Make",
      col=c("blue","red"),
      xlab="1: Sports Car, 0: Non-Sports Car",
      legend=rownames(table1),beside=TRUE)


barplot(table2,
      main="Small.Sporty..Compact.Large.Sedan by Make",
      col=c("blue","red"),
      xlab="1: Small.Sporty, 0: Compact.Large.Sedan",
      legend=rownames(table2),beside=TRUE)

par(mfrow=c(1,1))
```