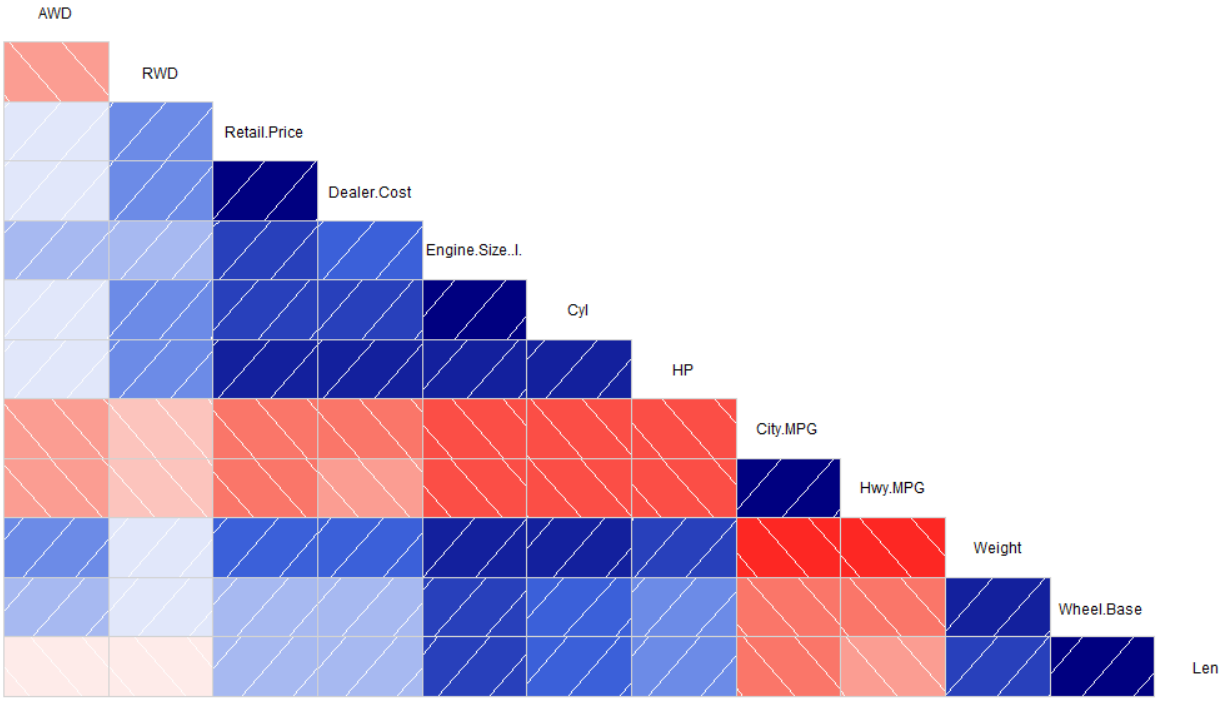
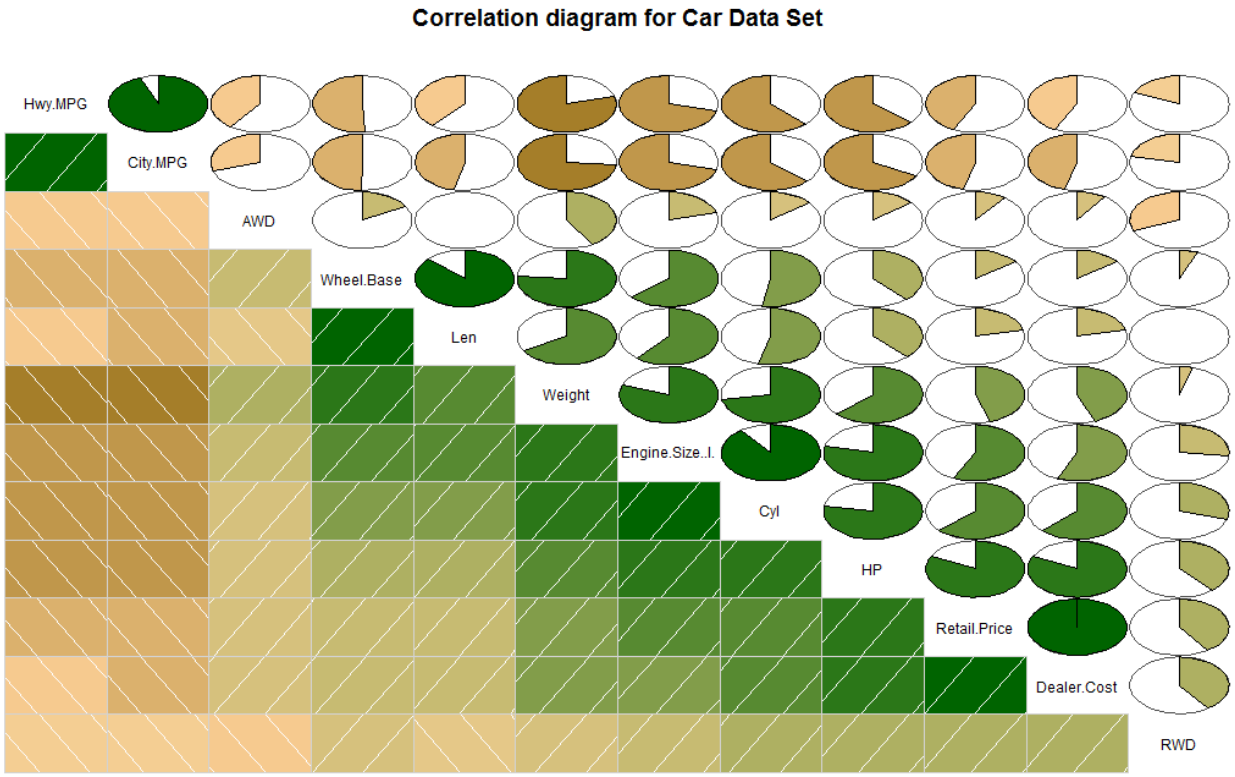
Lab 2
Steven Lin

Exercise 1

a)

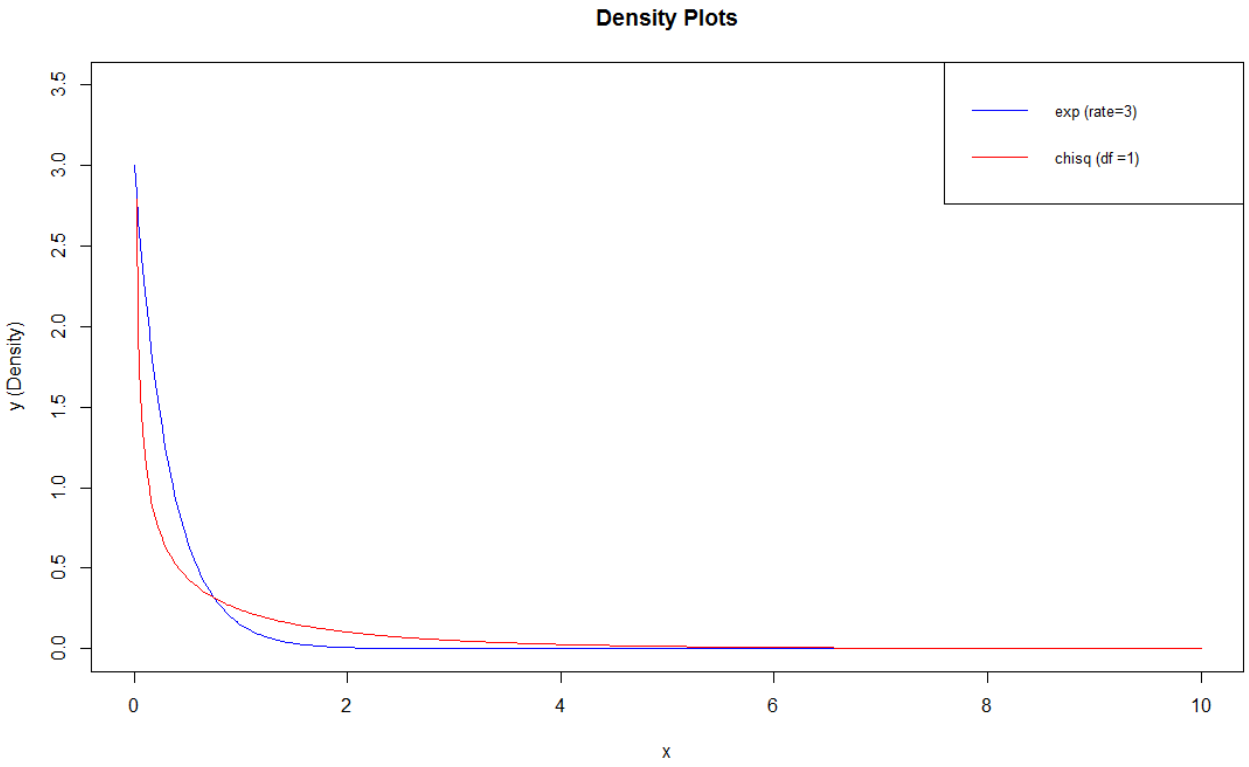


b)

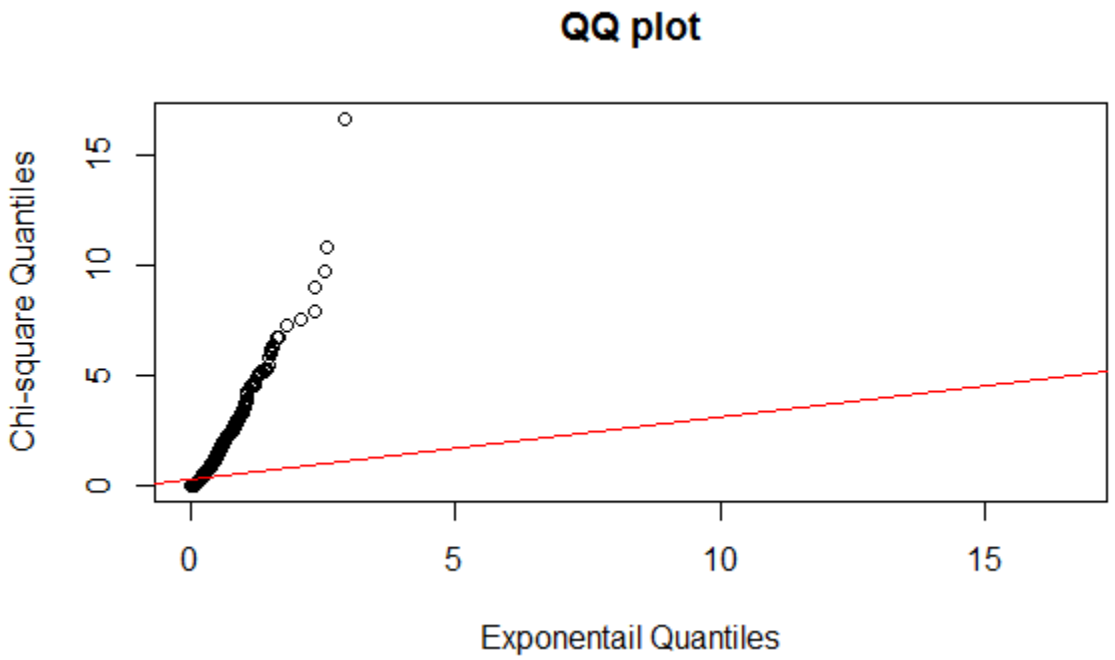

Correlation diagram for Car Data Set
Hwy.MPG
City.MPG
AWD
Wheel.Base
Len
Weight
Engine.Size..l.
Cyl
HP
Retail.Price
Dealer.Cost
RWD

Exercise 2

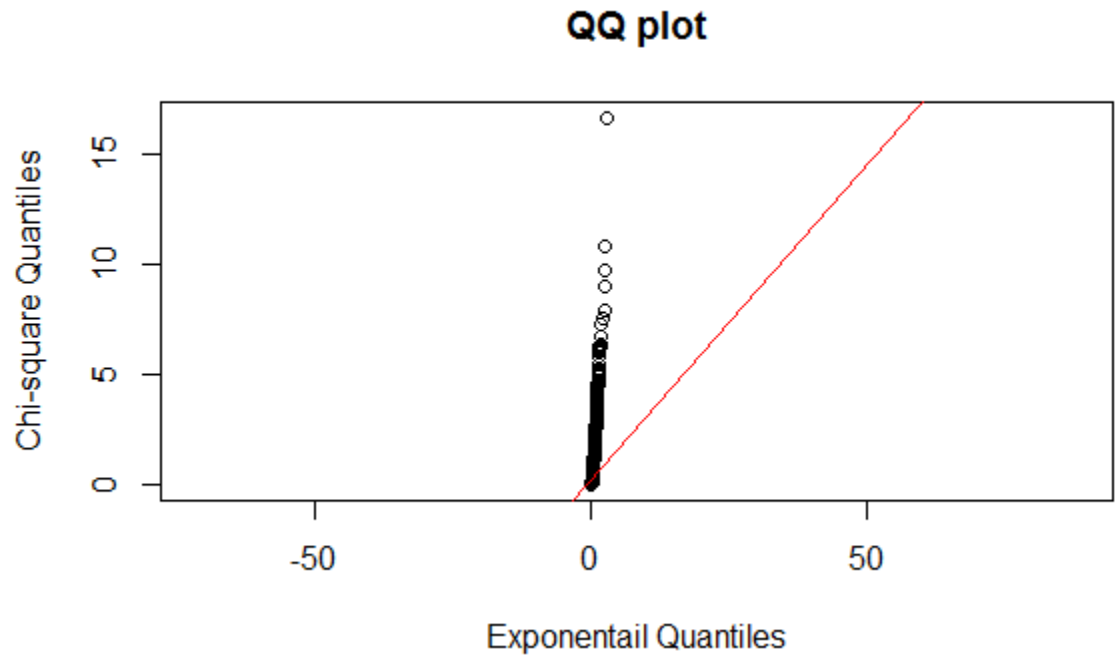
a)


Density Plots
exp (rate=3)
chisq (df =1)
y (Density)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
2
4
6
8
10
x

b)

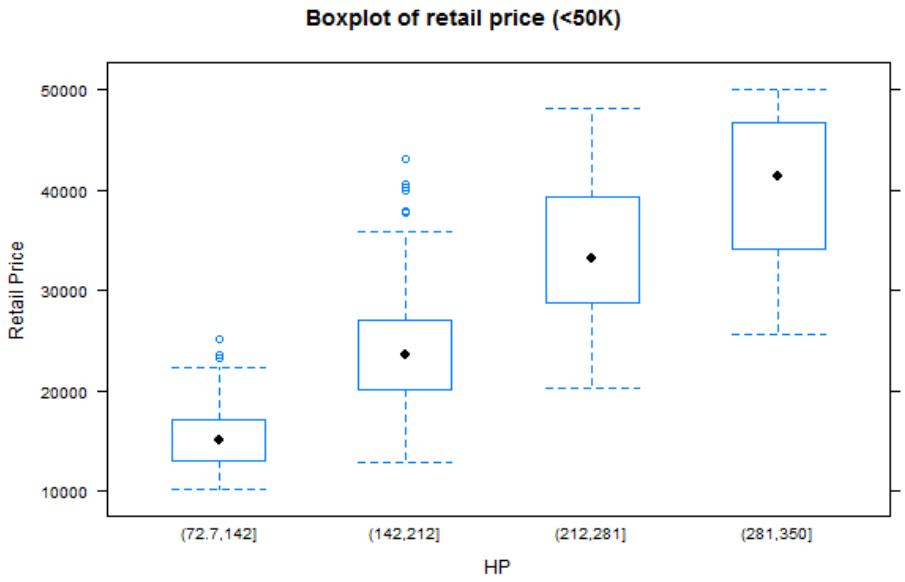

QQ plot
Chi-square Quantiles
0
5
10
15
0
5
10
15
Exponentail Quantiles

c)

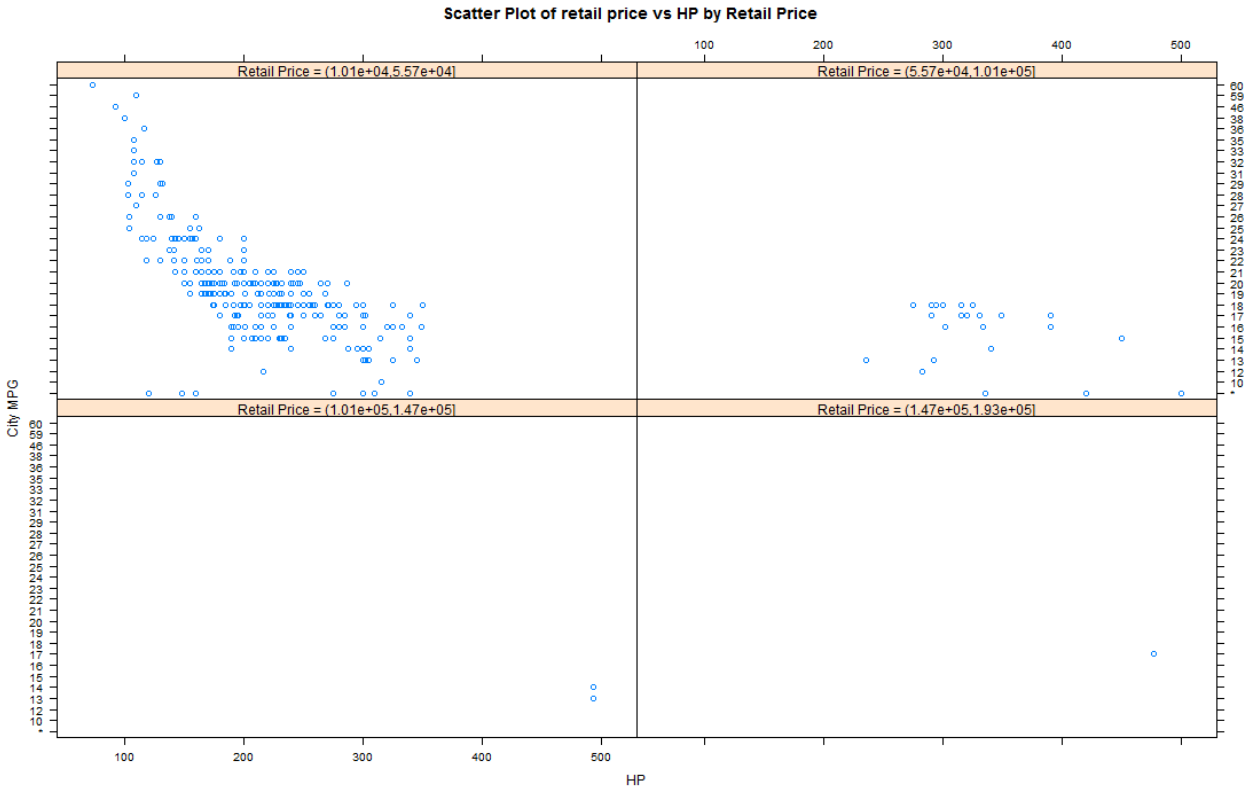

QQ plot
Chi-square Quantiles
0
5
10
15
-50
0
50
Exponentail Quantiles

Exercise 3

a)



b)



Rcode

```
# Lab session 2 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 ####

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

# install.packages('corrgram')
library(corrgram)

# Create a "corrgram" for columns 8 through 19 with the following
# features: no PCA order, empty upper panel, and shades for the
# lower panel. Make sure to name the variables on the diagonal.

corrgram(carsdata[,8:19], order = F, panel=panel.shade,
       lower.panel = panel.shade, upper.panel = NULL,
       text.panel = panel.txt,cex.labels= 1.1,
       main = "Correlation diagram for Car Data Set")


#### 1-b ####

#For the same columns as in part (a), create a "corrgram" with the #following features: PCA order,
shades for the lower panel and
# pie charts for the upper panel. #Correlation diagrams are made
# up of four colors; your task is to use the following four colors: #"darkgoldenrod4",
"burlywood1","darkkhaki" and "darkgreen".

corrgram(carsdata[,8:19], order = T, panel=panel.shade,
       lower.panel = panel.shade, upper.panel = panel.pie,
       text.panel = panel.txt,cex.labels= 1.1,
       main = "Correlation diagram for Car Data Set",
       col.regions = colorRampPalette(c("darkgoldenrod4",
```

```
                    "burlywood1",
                    "darkkhaki","darkgreen")))



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

#### 2-a ####

# plot the density function of a exp distribution rate =1
# plot the density function of a chisq distribution 1 df

x = seq(0,10,length=500)
dens_exp = dexp(x,rate = 3)
dens_chi = dchisq(x,df = 1)

# get the range for the x and y axis
xrange = c(0,10)
yrange = c(0,3.5)

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

# add lines
lines(x, dens_exp, type=linetype[1],col=colors[1])
lines(x, dens_chi, type=linetype[2], col=colors[2])

# add a title and subtitle
title("Density Plots")

fnames = c("exp (rate=3)","chisq (df =1)")

# add a legend
legend("topright", legend = fnames, cex=0.8, col=colors,
    lty=c(1,1))

#### 2-b ####


# Generate 1000 exponential random variables with rate equal to 3,
# and 1000 chi-squared random variables with 1 degree of freedom.
x_exp = rexp(1000, rate=3)
x_chi  = rchisq(1000,df = 1)

#Create a QQ plot and plot the line that goes through the first
```

```
#and third quantiles (that is qqline).

qqplot(x_exp,x_chi,
     ylim=range(c(x_exp,x_chi)),
     xlim=range(c(x_exp,x_chi)),
     main="QQ plot",
     xlab="Exponentail Quantiles",
     ylab="Chi-square Quantiles")

qqline(x_exp,col='red')

#### 2-c ####

# change the asp (the y/x aspect ratio) to line looks like 45 degree
qqplot(x_exp,x_chi,
     ylim=range(c(x_exp,x_chi)),
     xlim=range(c(x_exp,x_chi)),
     main="QQ plot",
     xlab="Exponentail Quantiles",
     ylab="Chi-square Quantiles",
     asp=4)

qqline(x_exp,col='red')




##### Ex3  ###################################################################

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

# Select all cars with Retail.Price less than $50,000.
data3 = carsdata[carsdata$Retail.Price < 50000,]

# Split the HP (horsepower) of these cars into 4 categories
data3$HP = cut(data3$HP,4)

# add HP to level ranges
#levels(data3$HP) = paste("HP = ",levels(data3$HP),sep="")


# create boxplots for their retail price by conditioning on the categories

bwplot(data3$Retail.Price~data3$HP,data=data3,
      main="Boxplot of retail price (<50K) ",
      xlab="HP", ylab="Retail Price")
```

```
#### 3-b ####

data4  =read.csv("04cars data.csv",header=TRUE,
                 na.strings=c("","*","NA"))

# Condition on Retail.Price by creating four ranges

data4$Retail.Price = cut(data4$Retail.Price,4)
levels(data4$Retail.Price)

# add Retail Price to level ranges
levels(data4$Retail.Price) = paste("Retail Price = ",
                       levels(data4$Retail.Price),sep="")


# For each of these categories, create
# scatterplots for City.MPG versus HP (horsepower).

#index.cond provides the order of the panels (ideally should order panels # by the order of the factor,
but it doesn't seem to do this R)
xyplot(data4$City.MPG ~ data4$HP | data4$Retail.Price,
    main="Scatter Plot of retail price vs HP by Retail Price",
    xlab="HP", ylab="City MPG",
    index.cond=list(c(3,4,1,2)))
```