**Question 1.**

#using the required package

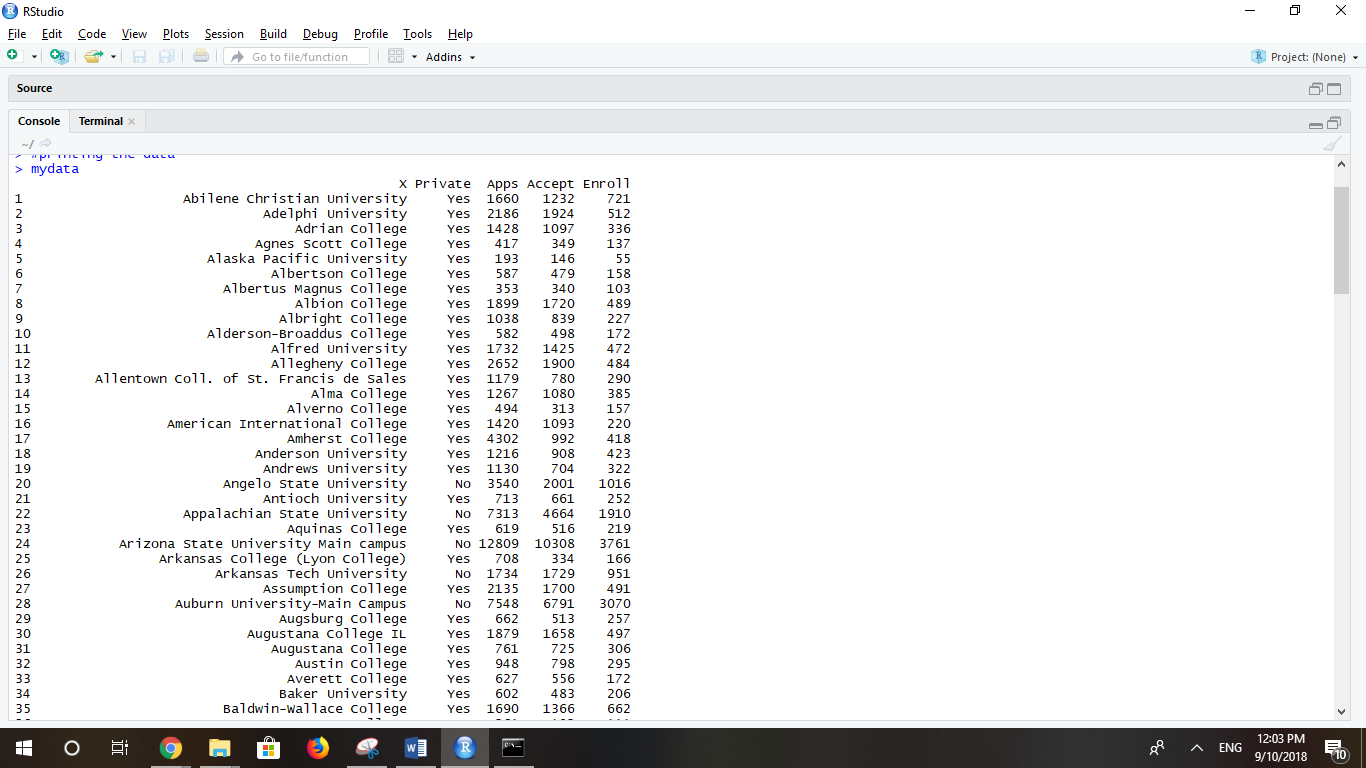
library(gdata)

# read csv file

mydata <- read.csv("C:\\Users\\Yuvesh\\Desktop\\College.csv")

#printing the data

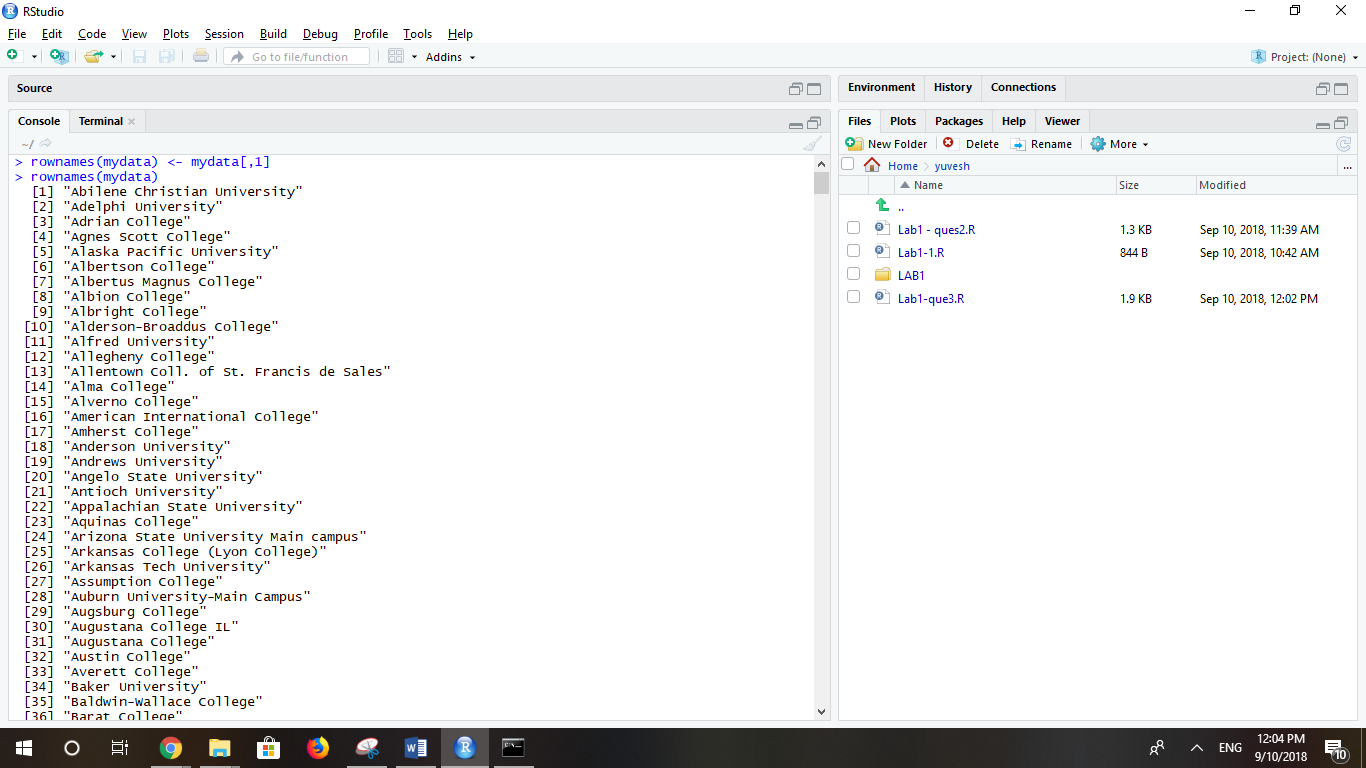
Mydata



#get the row name and data to it

rownames(mydata) <- mydata[,1]

rownames(mydata)

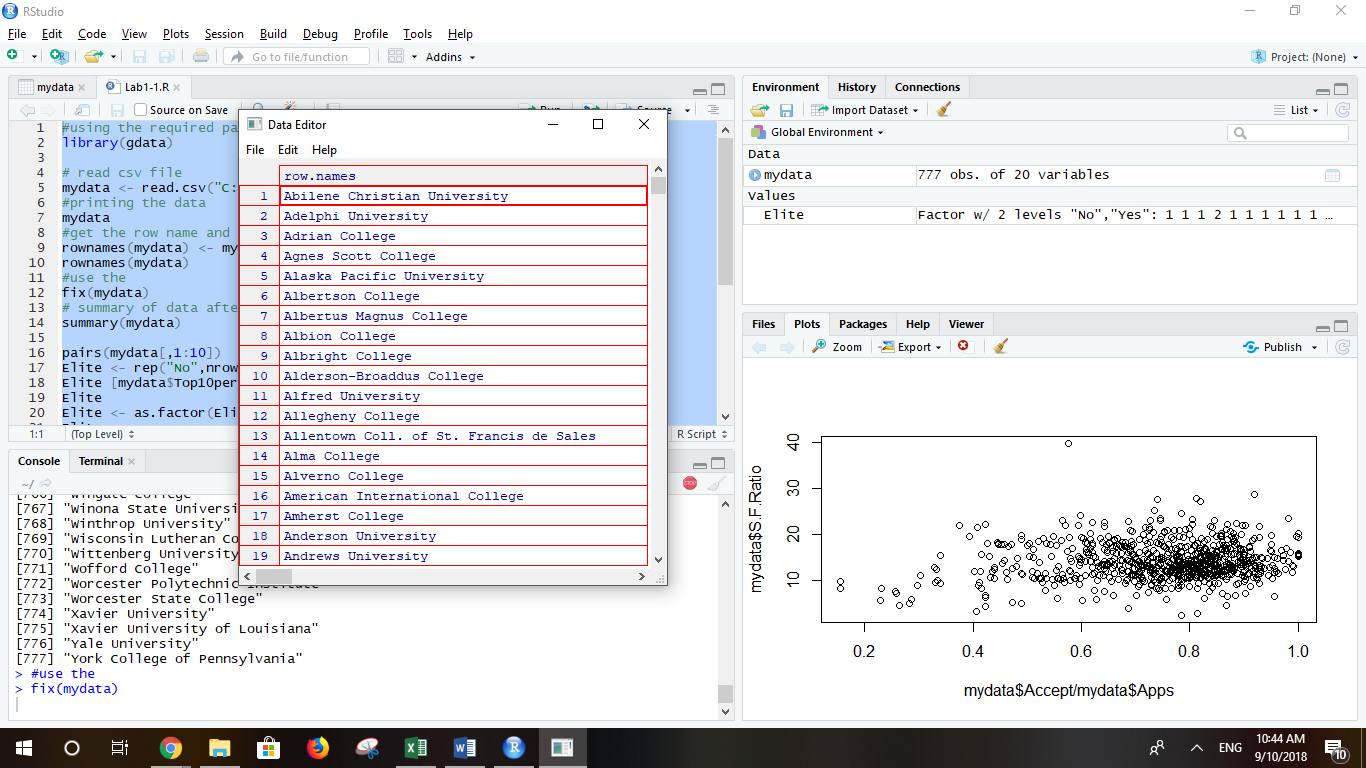


#use the

fix(mydata)

# summary of data after filling row

summary(mydata)

#pairing the data 

pairs(mydata[,1:10])

Elite <- rep("No",nrow(mydata))

Elite [mydata$Top10perc>50] <- "Yes"

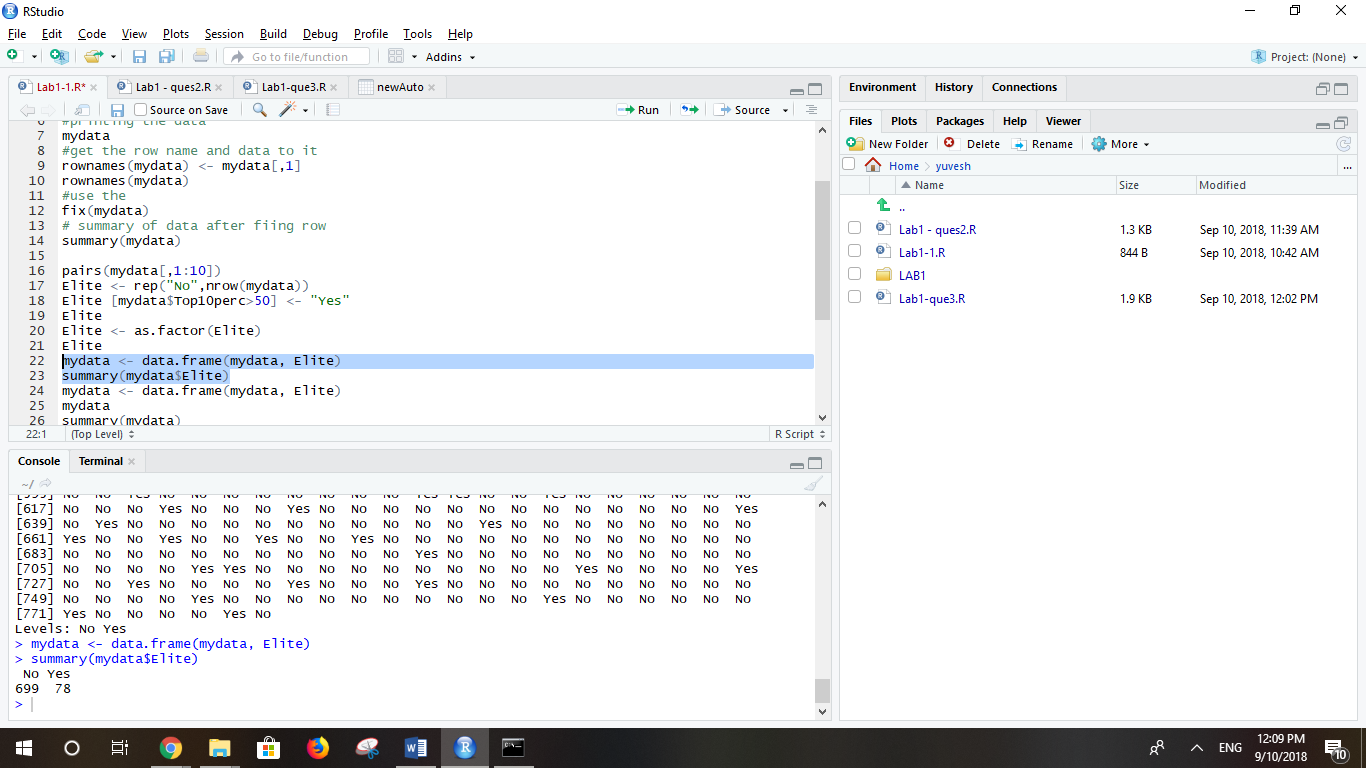
Elite

Elite <- as.factor(Elite)

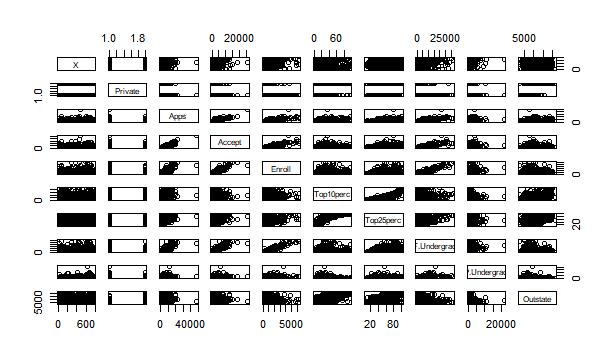
Elite

mydata <- data.frame(mydata, Elite)

summary(mydata$Elite)



Summary(mydata)



# histograms

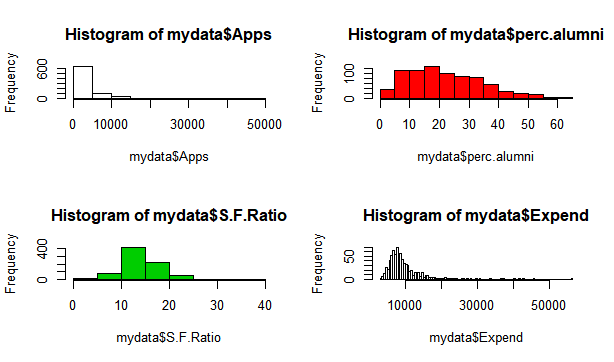
par(mfrow=c(2,2))

hist(mydata$Apps)

hist(mydata$perc.alumni, col=2)

hist(mydata$S.F.Ratio, col=3, breaks=10)

hist(mydata$Expend, breaks=100)

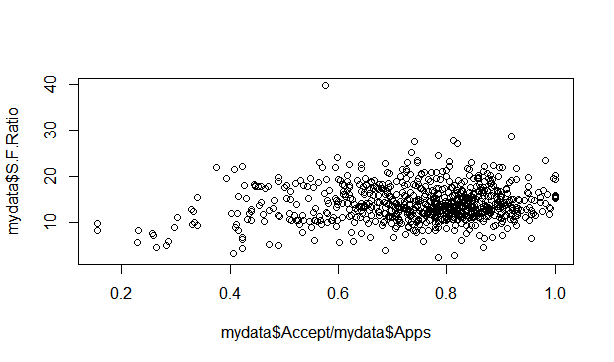


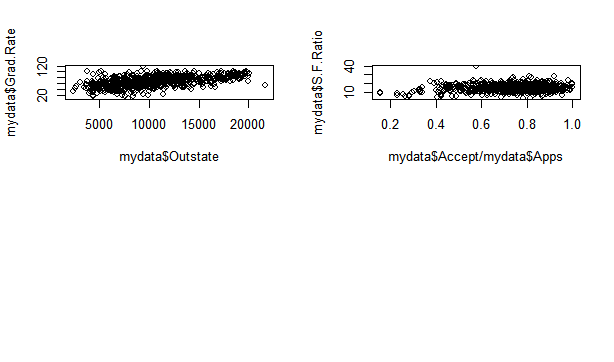
# eploring the data

plot(mydata$Outstate, mydata$Grad.Rate)

# High tuition correlates to high graduation rate.

plot(mydata$Accept / mydata$Apps, mydata$S.F.Ratio)





**Question 2:**

Auto <- read.csv("C:\\Users\\Yuvesh\\Desktop\\Auto.csv", header=T, na.strings="?")

# to elimate the missing data

Auto <- na.omit(Auto)

dim(Auto)

#summayrising the data at the end

summary(Auto)

# (a)

# quantitative: mpg, cylinders, displacement, horsepower, weight,acceleration, year

# qualitative: name, origin

#(b)

# apply the range function to the first eight columns of Auto

sapply(Auto[, 1:8], range)

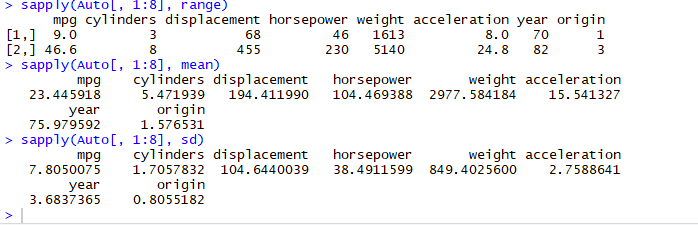
# (c)

# apply the mean function to the first eight columns of Auto

sapply(Auto[, 1:8], mean)

## apply the standard deviation function to the first eight columns of Auto

sapply(Auto[, 1:8], sd)



# (d)

# to delete 10th coloumn

newAuto <- Auto[-10,]

# to delete 85th coloumn

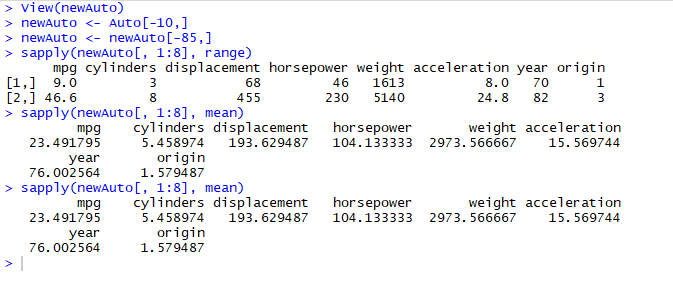
newAuto <- newAuto[-85,]

# apply the range function to the first eight columns of new Auto after deleting 10th row

sapply(newAuto[, 1:8], range)

# apply the mean function to the first eight columns of new Auto after deleting 10th row

sapply(newAuto[, 1:8], mean)

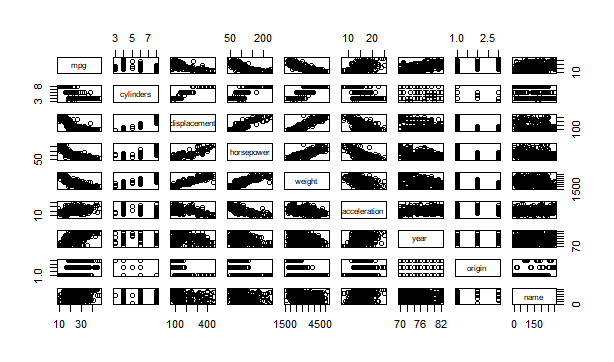


## apply the function to the first eight columns of new Auto after deleting 10th row

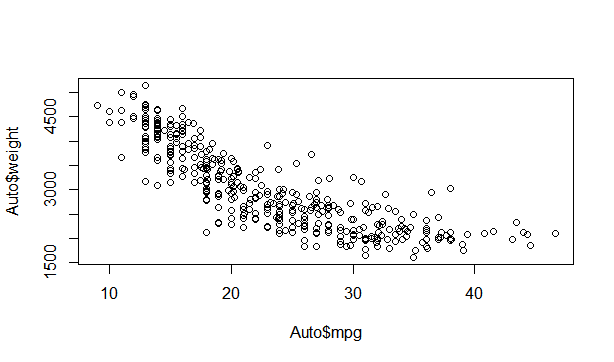
sapply(newAuto[, 1:8], sd)

# (e)

pairs(Auto)

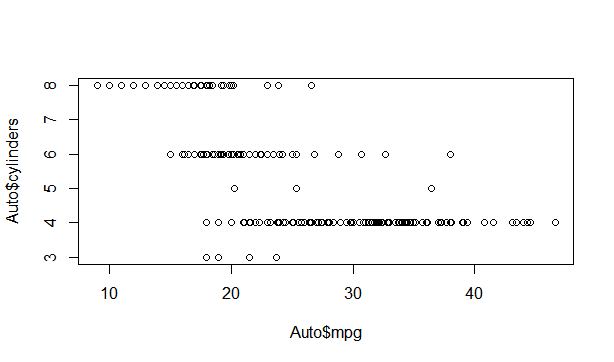


plot(Auto$mpg, Auto$weight)

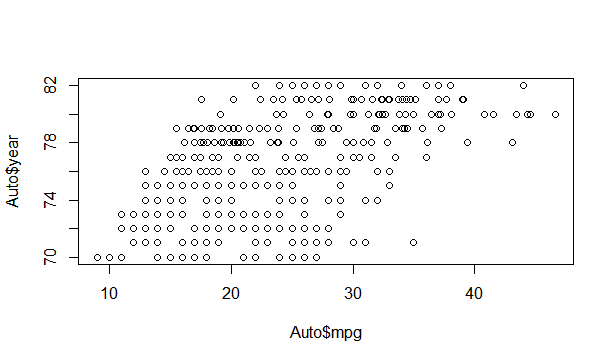
# Heavier weight correlates with lower mpg.

plot(Auto$mpg, Auto$cylinders)

# More cylinders, less mpg.

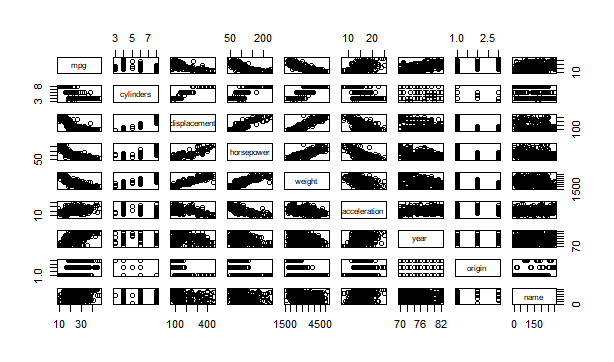


plot(Auto$mpg, Auto$year)

# Cars become more efficient over time.

# (f)

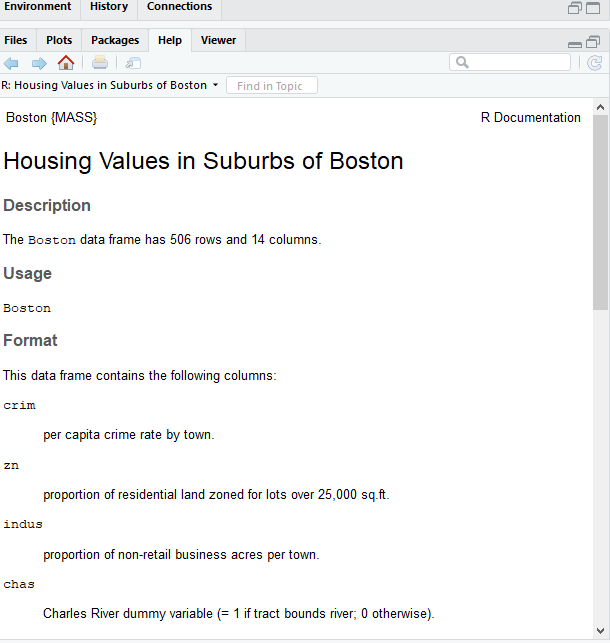
pairs(Auto)



Question 3:

>library(MASS)

>?Boston

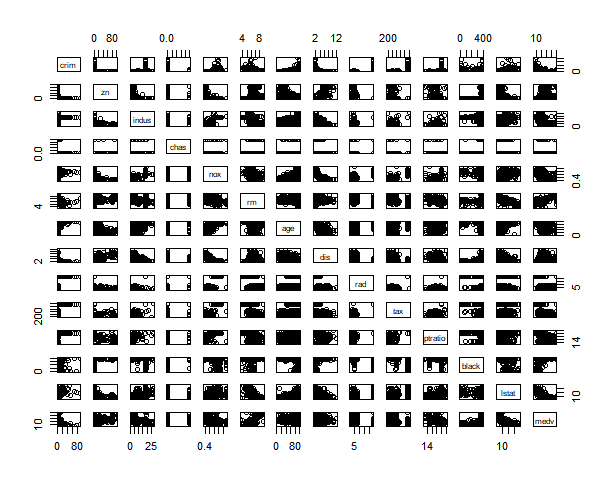


>dim(Boston)



# (b)

pairs(Boston)



# (c)

par(mfrow=c(3,2))

plot(Boston$age, Boston$crim)

# Older homes, more crime

plot(Boston$dis, Boston$crim)

# Closer to work-area, more crime

plot(Boston$rad, Boston$crim)

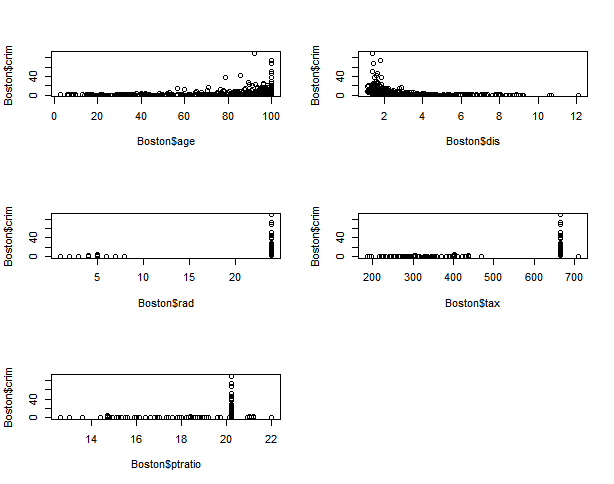
# Higher index of accessibility to radial highways, more crime

plot(Boston$tax, Boston$crim)

# Higher tax rate, more crime

plot(Boston$ptratio, Boston$crim)

# Higher pupil:teacher ratio, more crime



# (d)

par(mfrow=c(1,3))

hist(Boston$crim[Boston$crim>1], breaks=25)

# most cities have low crime rates, but there is a long tail: 18 suburbs appear

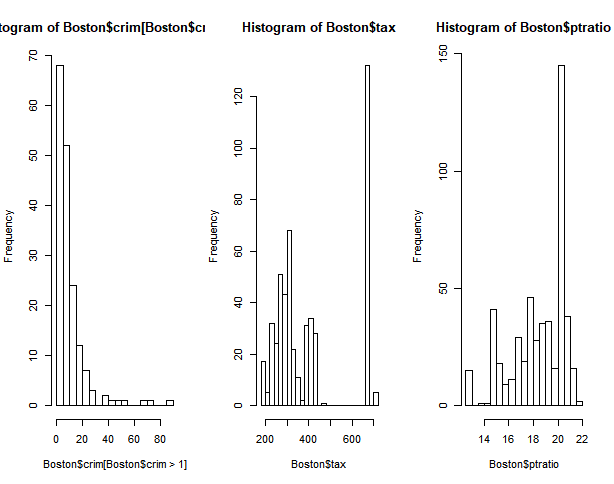
# to have a crime rate > 20, reaching to above 80

hist(Boston$tax, breaks=25)

# there is a large divide between suburbs with low tax rates and a peak at 660-680

hist(Boston$ptratio, breaks=25)

# a skew towards high ratios, but no particularly high ratios



(e)

dim(subset(Boston, chas == 1))



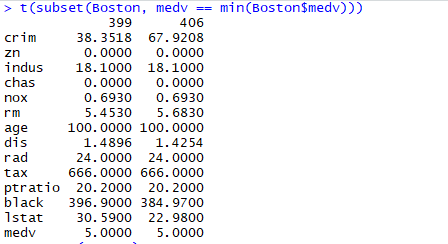
(f)

median(Boston$ptratio)

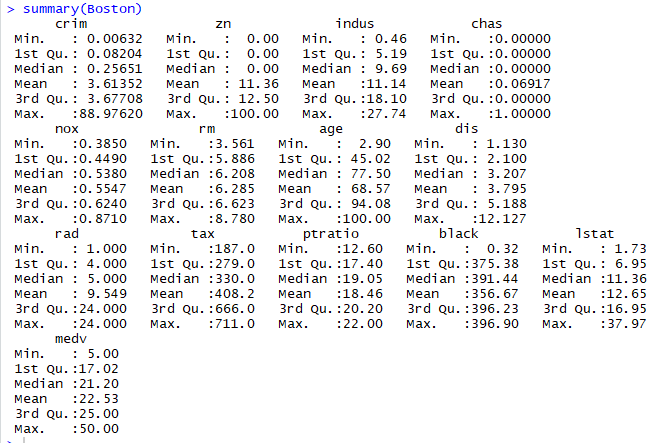


(g)

t(subset(Boston, medv == min(Boston$medv)))



summary(Boston)



(h)

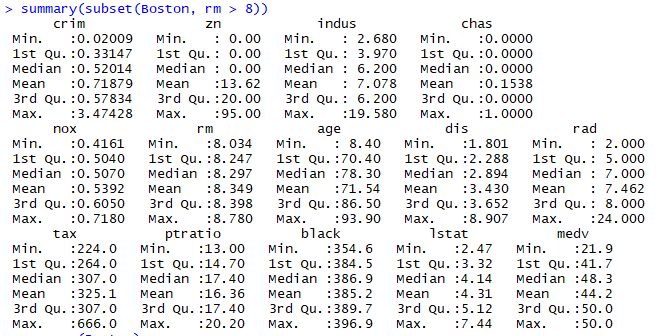
dim(subset(Boston, rm > 7))



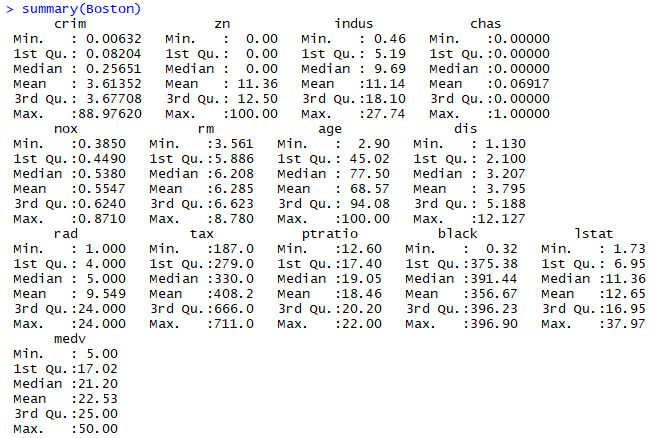
dim(subset(Boston, rm > 8))



summary(subset(Boston, rm > 8))



summary(Boston)



Github Link:

https://github.com/Yuvesh95/R-programming