# Load Breast cancer data file CSV

cancerData <- read.csv('C:/Users/prabh/Desktop/Stevens/fall\_2020/kdd/assignment2/breast-cancer-wisconsin.data.csv',na.string="?")

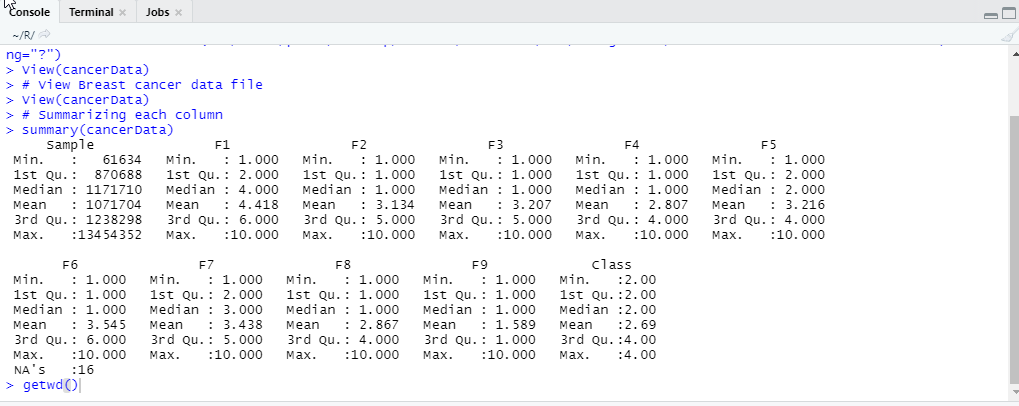
# View Breast cancer data file

View(cancerData)

I. Summarizing each column (e.g. min, max, mean)

summary(cancerData)

Output:

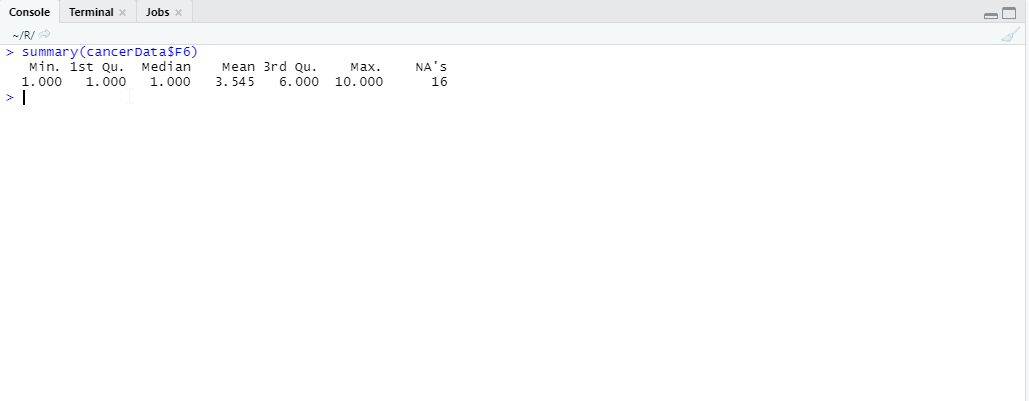


II.Identifying missing values

cancerData[(cancerData == "?")] <- NA

summary(cancerData$F6)

summary(cancerData)



III.Replacing the missing values with the the “mean” of the column

for(i in 1:ncol(cancerData)){

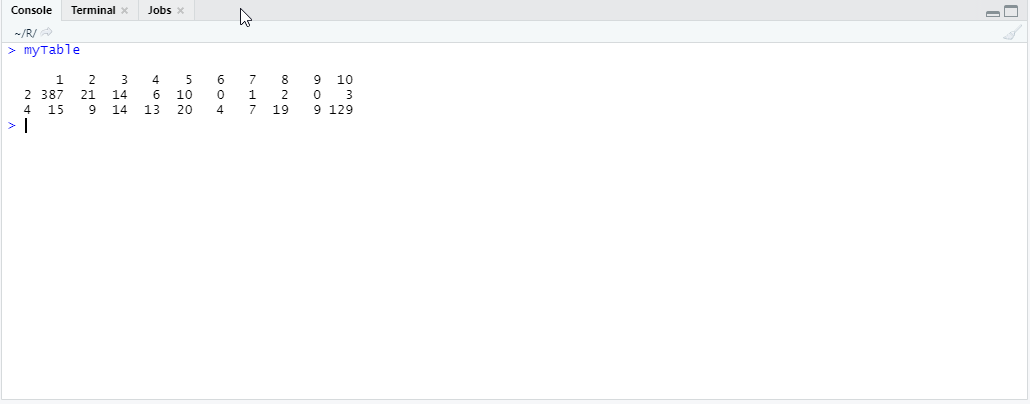
cancerData[is.na(cancerData[,i]), i] <- mean(cancerData[,i], na.rm = TRUE)

IV. Displaying the frequency table of “Class” vs. F6

attach(cancerData)

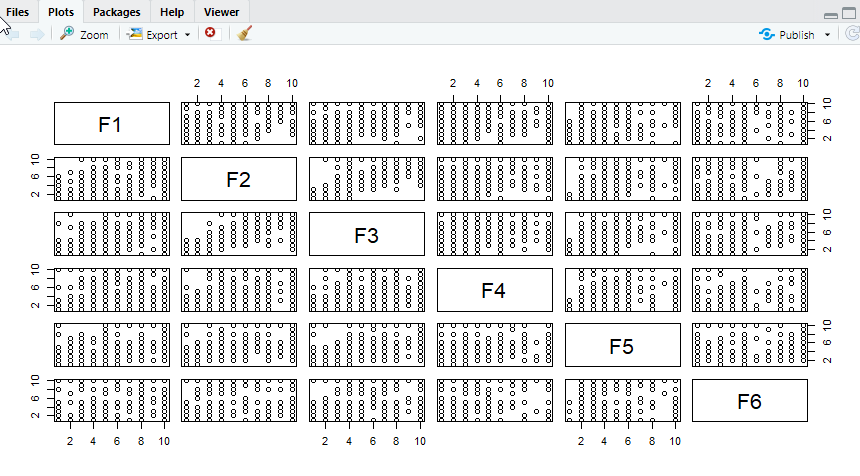
myTable<- table(F6,Class)

myTable



V. Displaying the scatter plot of F1 to F6, one pair at a time

plot(cancerData[,2:7])



VI. Show histogram box plot for columns F7 to F9

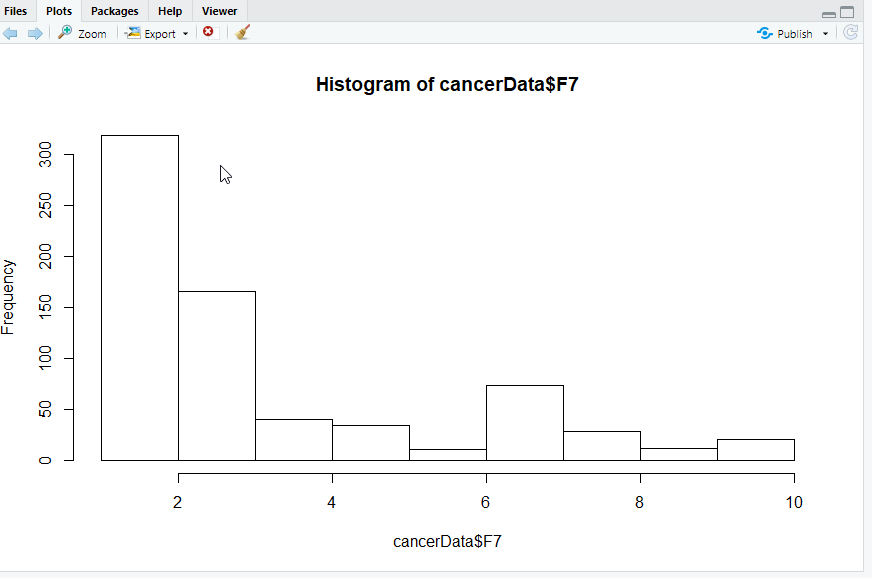
par(mfrow=c(2,2))

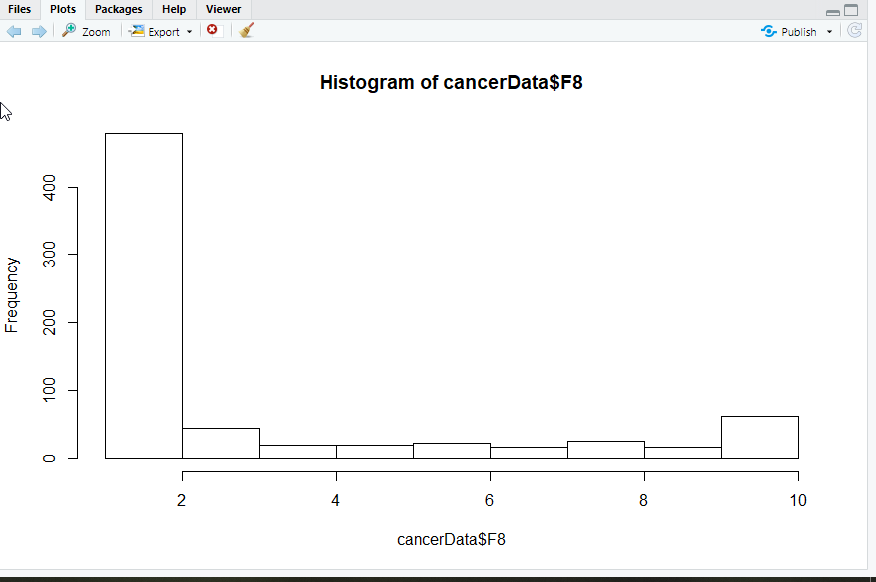
hist(cancerData$F7)

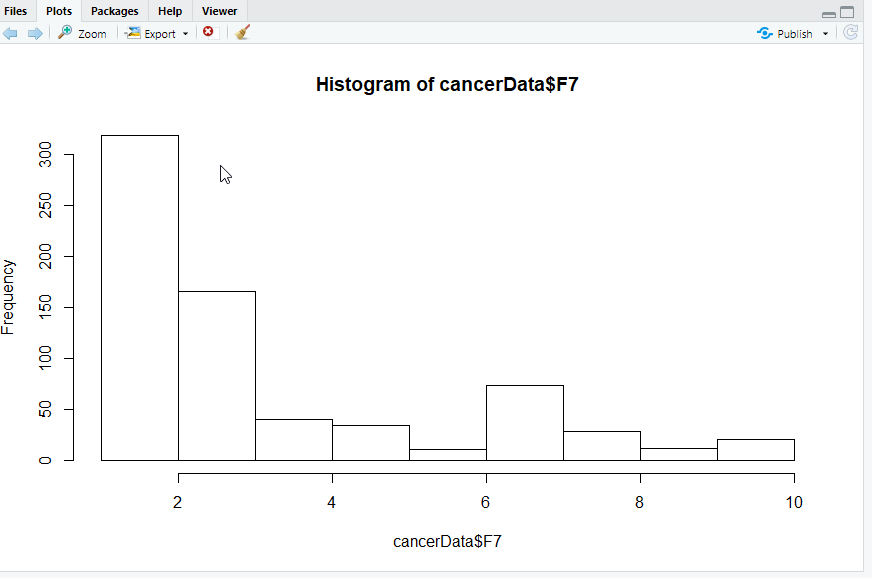
hist(cancerData$F8)

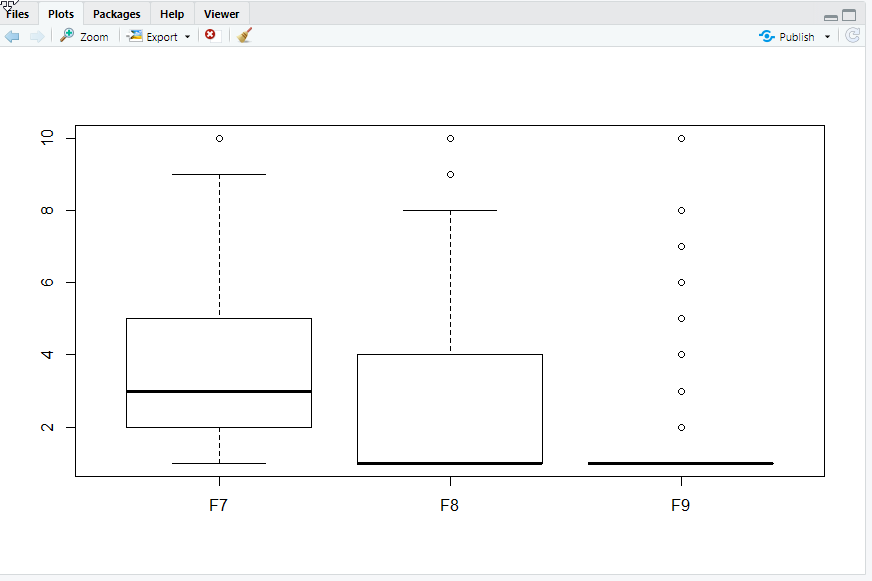
hist(cancerData$F9)

boxplot(cancerData[8:10])









2- Delete all the objects from your R- environment. Reload the “breast-cancer-wisconsin.data.csv” from canvas into R. Remove any row with a missing value in any of the columns.

rm(list = ls())

cancerData<-read.csv("C:/Users/prabh/Desktop/Stevens/fall\_2020/kdd/assignment2/breast-cancer-wisconsin.data.csv",header = TRUE,na.strings=c("?"))

cancerData

#Remove any row with a missing value in any of the columns

cs <- na.omit(cancerData)

cs

summary(cs)

