612303050 Deshmukh Mehmood Rehan’s Mid Semester Practical Examination

#Q.1) For the following frequency distribution [5M]  
# x 1 2 3 4 5  
# f 7 11 9 8 3  
# Write a R code to find  
  
#a) Mean  
x = c(1,2,3,4,5)  
f = c(7,11,9,8,3)  
  
temporary = rep(x,f)  
  
print(mean(temporary))

## [1] 2.710526

#b) Median  
print(median(temporary))

## [1] 3

#c) Mode  
y = table(temporary)  
mode = names(y)[which(y == max(y))]  
print(mode)

## [1] "2"

#d) Seventh decile  
d7 = quantile(temporary, 0.7)  
print(d7)

## 70%   
## 3

#e) 29th percentile  
p29 = quantile(temporary, 0.29)  
print(p29)

## 29%   
## 2

#Q.2) Create a dataframe of the following two vectors: [ 2M]  
#Price 10 15 30 42 50 60  
#Qty 4 20 15 10 16 8  
#Also write a R code to add vector/variable named value=price\*qty in the created dataframe.  
  
price = c(10,15,30,42,50,60)  
qty = c(4,20,15,10,16,8)  
df = data.frame(price, qty)  
df = transform(df, "value"=price\*qty)  
print(df)

## price qty value  
## 1 10 4 40  
## 2 15 20 300  
## 3 30 15 450  
## 4 42 10 420  
## 5 50 16 800  
## 6 60 8 480

#Q.3) Suppose the age is a vector containing ages of 10 persons as 22,27,31,41,30,25,19,20,23,35 [5M]  
  
age = c(22,27,31,41,30,25,19,20,23,35)  
  
#a) Remove the 5th and 7th element from the vector.  
print(age[c(-5,-7)])

## [1] 22 27 31 41 25 20 23 35

#b) Create a new vector containing the ages of persons greater than 30.  
age30 = age[age>30]  
  
print(age30)

## [1] 31 41 35

#c) Extract the 4th to 6th element from the vector.  
print(age[4:6])

## [1] 41 30 25

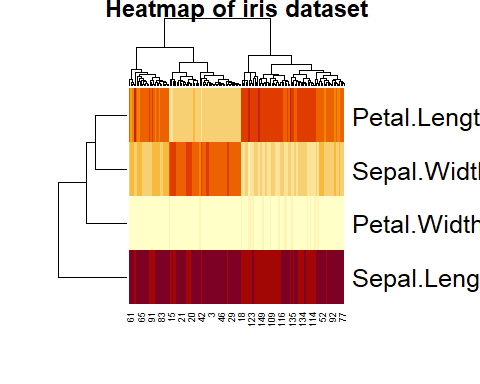
#d) Create a new vector containing the ages of persons between 20 and 25.  
  
age2 = age[age>20 & age<25]  
  
print(age2)

## [1] 22 23

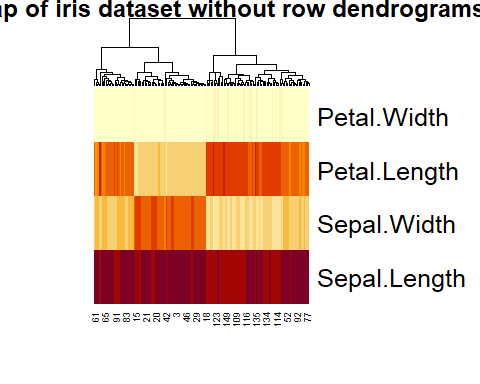
#e) Extract the last 3 elements from the vector.  
  
print(tail(age,3))

## [1] 20 23 35

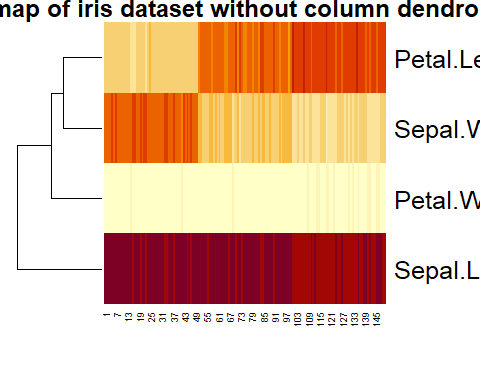
#Q.4) Create a heatmap using inbuilt iris datasets. [3M]  
data(iris)  
  
M = iris[,-5]  
  
heatmap(t(M), main="Heatmap of iris dataset", scale = "column")



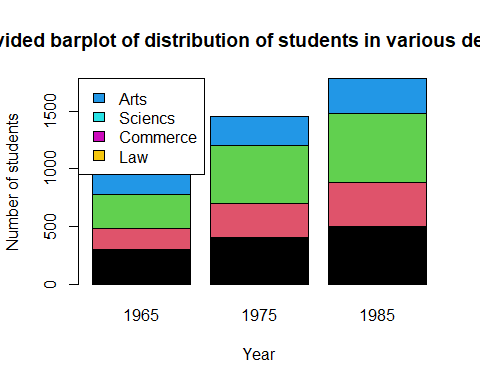
#Also write a R code for the following  
#a) Write a code to remove row dendrogram  
  
heatmap(t(M), main="Heatmap of iris dataset without row dendrograms", scale = "column", Rowv = NA)



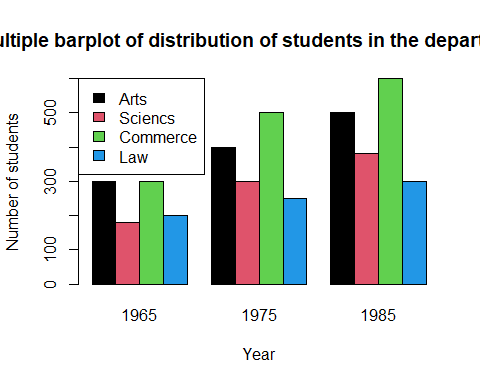
#b) Write a code to remove column dendrogram  
  
heatmap(t(M), main="Heatmap of iris dataset without column dendrograms", scale = "column", Colv = NA)



#Q.5) Create a subdivided barplot and a multiple barplot of the following data: [5M]  
# 1965 1975 1985  
#Arts 300 400 500  
#Sci 180 300 380  
#Com 300 500 600  
#Law 200 250 300  
  
years = c(1965,1975,1985)  
arts = c(300,400,500)  
science = c(180, 300, 380)  
commerce = c(300,500,600)  
law = c(200,250,300)  
student\_data = data.frame(arts,science,commerce,law)  
barplot(t(student\_data), main="Subdivided barplot of distribution of students in various departments", beside=FALSE, names.arg=years, xlab="Year", ylab="Number of students",col = 1:4)  
legend("topleft",c("Arts","Sciencs","Commerce","Law"),fill=4:8)



barplot(t(student\_data), main="Multiple barplot of distribution of students in the departments", beside=TRUE, names.arg=years, xlab="Year", ylab="Number of students",col = 1:4)  
legend("topleft",c("Arts","Sciencs","Commerce","Law"),fill=1:4)



# Q.6) Solve the following questions [5M]  
# a) Write a R program to create a Dataframes which contain details of 5 employees and display  
# summary of the data.  
# b) Write a R program to create a two-dimensional 5×3 array of sequence of even integers greater  
# than 50.  
# c) Write a R program to find the levels of factor of a given vector 1, 2, 3, 3, 4, NA, 3, 2, 4, 5,  
# NA, 5.  
# d) Write a R program to create a list containing strings, numbers, vectors and a logical values.  
# e) Write a R program to create a vector which contains 10 random integer values between -50  
# and +50.  
  
#a) Dataframe  
  
name = c("john", "doe", "jane", "smith", "james");  
age = c(33, 35, 36, 29, 40);  
salary = c(16000, 23000, 27000, 21000, 20000);  
df = data.frame(name, age, salary)  
print(df)

## name age salary  
## 1 john 33 16000  
## 2 doe 35 23000  
## 3 jane 36 27000  
## 4 smith 29 21000  
## 5 james 40 20000

summary(df)

## name age salary   
## Length:5 Min. :29.0 Min. :16000   
## Class :character 1st Qu.:33.0 1st Qu.:20000   
## Mode :character Median :35.0 Median :21000   
## Mean :34.6 Mean :21400   
## 3rd Qu.:36.0 3rd Qu.:23000   
## Max. :40.0 Max. :27000

#b) 2D array  
  
arr = array(seq(from = 52, length.out = 15, by = 2),dim = c(5,3))  
  
print(arr)

## [,1] [,2] [,3]  
## [1,] 52 62 72  
## [2,] 54 64 74  
## [3,] 56 66 76  
## [4,] 58 68 78  
## [5,] 60 70 80

#c) Factor levels  
  
vec = c( 1, 2, 3, 3, 4, NA, 3, 2, 4, 5, NA, 5)  
  
factor(vec)

## [1] 1 2 3 3 4 <NA> 3 2 4 5 <NA> 5   
## Levels: 1 2 3 4 5

#d) List  
  
vec = c(1,2,3)  
  
list("abc", 123, vec, TRUE)

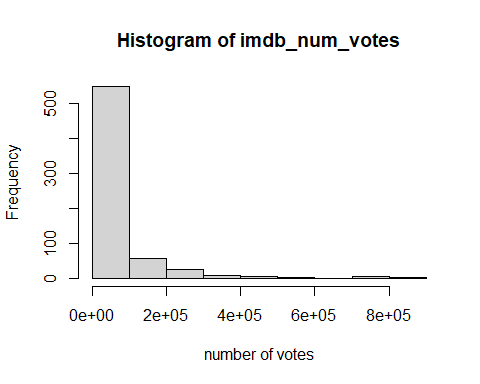
## [[1]]  
## [1] "abc"  
##   
## [[2]]  
## [1] 123  
##   
## [[3]]  
## [1] 1 2 3  
##   
## [[4]]  
## [1] TRUE

#e) Random vector  
  
vec1 = sample(-50 : 50, 10)  
   
print(vec1)

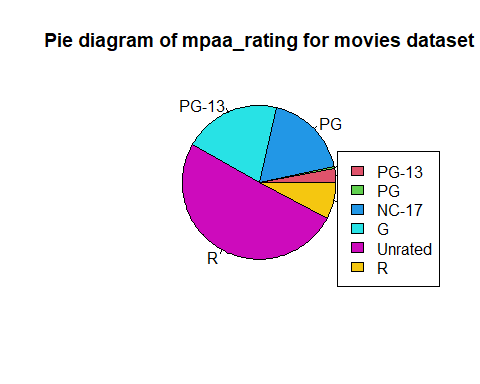
## [1] 43 22 -14 20 28 -25 -2 44 38 26

#Q7) Read the file moviesData.csv and solve the following questions. [5M]  
#) Use the moviesData. Create a histogram of the object named imdb num votes in this file.  
#b) Create a pie chart of the object mpaa rating.  
#c) Create a bar chart of critics score for the first 10 movies.  
#d) Create a scatter plot of imdb rating and imdb num votes to see their relation.  
#e) Create a boxplot for dvd rel day variable and also display labels.  
  
moviesData = read.csv('./moviesData.csv')

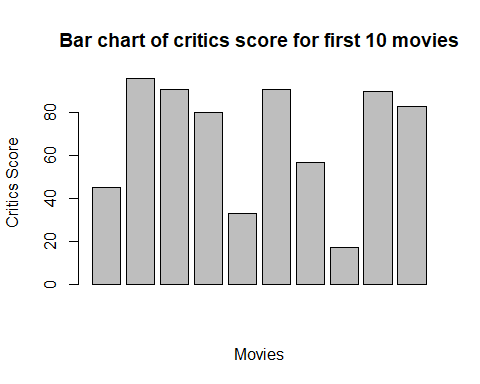
#a) Histogram  
imdb\_num\_votes = moviesData$imdb\_num\_votes  
hist(imdb\_num\_votes, main="Histogram of imdb\_num\_votes", xlab="number of votes")



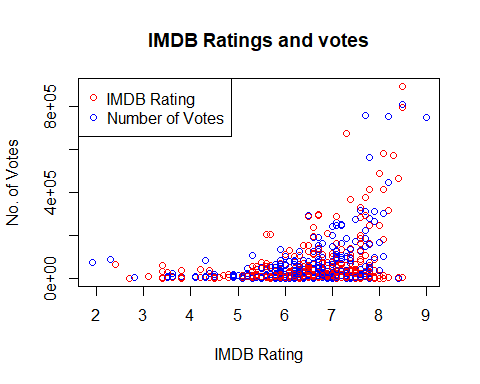
#b) pie diagram  
rating = table(moviesData$mpaa\_rating)  
pie(rating, main="Pie diagram of mpaa\_rating for movies dataset", col=10:16)  
legend("bottomright",legend=c("PG-13","PG","NC-17","G","Unrated","R"),fill=10:16)



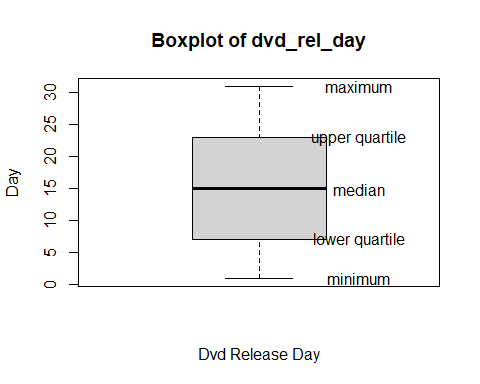
#c) Bar Chart  
  
critics\_score = moviesData$critics\_score  
barplot(critics\_score[1:10], main="Bar chart of critics score for first 10 movies", xlab="Movies", ylab="Critics Score")



#d) Scatter Plot  
imdb\_rating = moviesData$imdb\_rating  
plot(imdb\_rating, imdb\_num\_votes, col = c("red", "blue"), main = "IMDB Ratings and votes", xlab="IMDB Rating", ylab="No. of Votes")  
legend("topleft", legend = c("IMDB Rating", "Number of Votes"), col = c("red", "blue"), pch = 1)



#e) Boxplot  
dvd\_rel\_day = moviesData$dvd\_rel\_day  
boxplot(dvd\_rel\_day, main="Boxplot of dvd\_rel\_day", xlab="Dvd Release Day", ylab="Day")  
f = fivenum(dvd\_rel\_day)  
text(rep(1.3,5),f,labels=c("minimum","lower quartile","median","upper quartile","maximum"))



#Q8) For the following in built dataset CO2 in R.Write a R program taking the uptake variable and  
#calculate the following terms [5M]  
#a) Standard deviation.  
#b) Quartile deviation.  
#c) Range.  
#d) Mode.  
#e) Coefficient of range.  
  
data("CO2")  
  
uptake = CO2$uptake  
  
#a) Standard deviation  
  
print(sd(uptake))

## [1] 10.81441

#b) Quartile deviation  
  
print((quantile(uptake, 0.75) - quantile(uptake, 0.25))/2);

## 75%   
## 9.6125

#c) Range  
  
print(max(uptake)-min(uptake))

## [1] 37.8

#d) Mode  
  
y = table(uptake)  
mode = names(y)[which(y == max(y))]  
print(mode)

## [1] "17.9" "32.4"

#e) Coefficient of range  
  
print((max(uptake)-min(uptake))/(max(uptake)+min(uptake)))

## [1] 0.7105263