Statistical Programming Project-2 for Final Exam

Student Name- Surname	Student Number
Esmanur deli	191805056
Kübra Uçar	191805067
Kevser Öztürk	191805054
Rabia Yıldırım	191805043
Görkem Avcı	191805013

We took the dataset from our folder. Here are our codes for this:

```
getwd()
setwd('C:/Users/kubra/OneDrive - Aydin Adnan Menderes University/Belgeler/RFinal/Rfinal')
data<-read.table("DatasetNA1.txt",header = TRUE)
View(data)
```

Question1:

We wrote these functions and We took outputs for every different var values :

Number of Observation :

```
numberofobservation <- function(column){
 score = 0
 for(x in column){
  if(is.na(x) == TRUE){
   score <- score + 0
  }
  else{
   score <- score + 1}}
 return(score)}
numberofobservation(data$Var3)
```

Answer for var3 : 98 Observation

Minumum :

```
minimum <- function(column){
 min = Inf
 for (i in seq_along(column)) {
  if(is.na(column[i]) == FALSE) {
   if (column[i] < min){
    min = column[i] } }}
  return(min)
}
minimum(data$Var2)
```

Answer for Var2 : 16.16

Maximum :

```
maximun <- function(column){
 na.omit(column)
 max = column[1]
 for (i in seq_along(column)) {
  if(is.na(column[i]) == FALSE) {
   if (column[i] > max){
    max = column[i] }} }
 return(max)}
maximun(data$Var2)
```

Answer for Var2 : 25.11

Range :

```
range <- function(column){
 cat(minimum(column) ,"-" ,maximun(column))
}
range(data$Var2)
```

Answer for Var2 : 16.16 - 25.11

Sum :

```
sumfunc <- function(column) {
 sum = 0
 for (i in seq_along(column)){
  if(is.na(column[i]) == FALSE){
   sum = sum + column[i]} }
 sum}
sumfunc(data$Var6)
```

Answer for Var6: 7084.41

Mean :

```
meanfunc <- function(column){
 my_average = sumfunc(column)/length(na.omit(column))
 my_average
}
meanfunc(data$Var1)
```

Answer for Var1 : 3.988384

Median :

```
medianfunc <- function(column, na.rm = FALSE) {
 a <- length(na.omit(column))
 b <- sort(na.omit(column))
 ifelse(a%%2==1,b[(a+1)/2],meanfunc(b[a/2+0:1]))
}
medianfunc(data$Var1)
```

Answer for Var1 : 3.96

Sum of Squares :

```
SumOfSquaresfunc <- function(column){
  difference <-  column - meanfunc(column)
  sum_squares <- sumfunc(difference^2)
  output <- sum_squares
  return(output)
}
SumOfSquaresfunc(data$Var1)
```

Answer for Var1 : 7.974941

Variance :

```
varfunc <- function(column){
  variance = SumOfSquaresfunc(column)/(length(na.omit(column))-1)
  return(variance)}
varfunc(data$Var1)
```

Answer for Var1 : 0.08137695

Standard deviation :

```
sdfunc <- function(x){
  return(sqrt(varfunc(x)))}
sdfunc(data$Var1)
```

Answer for Var1 : 0.2852665

Cross-products :

```
crossprodfunc <- function(x,y){
  a = t(na.omit(x)) %*% na.omit(y)
  return(a)}
crossprodfunc(data$Var1,data$Var2)
```

Answer for Var1-2 : 8285.273

Covariance :

```
covfunc <- function(x,y){
  xx <-  na.omit(x) - meanfunc(x)
  yy <-  na.omit(y) - meanfunc(y)
if(length(xx)==length(yy)){
    r = sumfunc(xx*yy)/(length(yy)-1)
  }
  else print("vectors are not the same length")
  return(r)}
covfunc(data$Var1,data$Var2)
```

Answer for Var1-2 : 0.1225436

Correlations :

```
corr <- function(x,y){
  r = covfunc(x,y)/(sdfunc(x)*sdfunc(y))
  return(r)
}

corr(data$Var1,data$Var2)
```

Answer for Var1-2 : 0.2167875

Question 2 :

We convert char value to numeric value in Gender (‘female ‘= 1 and ‘male ’= 2)

```
data2 = data
data2$Gender <- gsub('Female', 1, data2$Gender)
data2$Gender <- gsub('Male', 2, data2$Gender)
data2$Gender <- as.numeric(as.character(data2$Gender))
summary(data2$Gender)
```

We convert char value to numeric value in Group(‘Group 1‘= 1 and ‘Group2 ’= 2 and ‘Group3’=3 and ‘Group4 = 4’)

```
data2$Group <- gsub('Group1', 1, data2$Group)
data2$Group <- gsub('Group2', 2, data2$Group)
data2$Group <- gsub('Group3', 3, data2$Group)
data2$Group <- gsub('Group4', 4, data2$Group)
data2$Group <- as.numeric(as.character(data2$Group))
summary(data2$Group)
```

We select only female=1 and create subset1 data for only female. We select only male=2 and create subset1 data for only male.

```
subset1 =  data2[floor(data2$Gender) == 1,]
subset2 =  data2[floor(data2$Gender) == 2,]
```

We select only group 1-2-3-4 and create subsetgroup 1-2-3-4 data for only group 1-2-3-4

```
subsetgroup1 =  data2[floor(data2$Group) == 1,]
subsetgroup2 =  data2[floor(data2$Group) == 2,]
subsetgroup3 =  data2[floor(data2$Group) == 3,]
subsetgroup4 =  data2[floor(data2$Group) == 4,]
```

We select groups by female=1 and by male=2 create group$byfemale and group$bymale .

```
group1byfemale =  subsetgroup1[floor(subsetgroup1$Gender) == 1,]
group2byfemale =  subsetgroup2[floor(subsetgroup2$Gender) == 1,]
group3byfemale =  subsetgroup3[floor(subsetgroup3$Gender) == 1,]
group4byfemale =  subsetgroup4[floor(subsetgroup4$Gender) == 1,]
```

```
group1bymale =  subsetgroup1[floor(subsetgroup1$Gender) == 2,]
group2bymale =  subsetgroup2[floor(subsetgroup2$Gender) == 2,]
group3bymale =  subsetgroup3[floor(subsetgroup3$Gender) == 2,]
group4bymale =  subsetgroup4[floor(subsetgroup4$Gender) == 2,]
```

We calculated the functions, you can write any var value to see other results.

Only for gender :

```
minimum(subset1$Var1)
minimum(subset2$Var1)


maximun(subset1$Var1)
maximun(subset2$Var1)

range(subset1$Var1)
range(subset2$Var1)


sumfunc(na.omit(subset1$Var1))
sumfunc(na.omit(subset2$Var2))

meanfunc(subset1$Var1)
meanfunc(subset2$Var1)


medianfunc(na.omit(subset1$Var1))
medianfunc(na.omit(subset2$Var1))

SumOfSquaresfunc(na.omit(subset1$Var1))
SumOfSquaresfunc(na.omit(subset2$Var1))


varfunc(na.omit(subset1$Var1))
varfunc(na.omit(subset2$Var1))

sdfunc(na.omit(subset1$Var1))
sdfunc(na.omit(subset2$Var1))
```

Only for Groups :

minimum(subsetgroup1$Var1)
minimum(subsetgroup2$Var1)
minimum(subsetgroup3$Var1)
minimum(subsetgroup4$Var1)

maximun(subsetgroup1$Var1)
maximun(subsetgroup2$Var1)
maximun(subsetgroup3$Var1)
maximun(subsetgroup4$Var1)

range(subsetgroup1$Var1)
range(subsetgroup2$Var1)
range(subsetgroup3$Var1)
range(subsetgroup4$Var1)

sumfunc(na.omit(subsetgroup1$Var1))
sumfunc(na.omit(subsetgroup2$Var1))
sumfunc(na.omit(subsetgroup3$Var1))
sumfunc(na.omit(subsetgroup4$Var1))

meanfunc(na.omit(subsetgroup1$Var1))
meanfunc(na.omit(subsetgroup1$Var1))
meanfunc(na.omit(subsetgroup1$Var1))
meanfunc(na.omit(subsetgroup1$Var1))

medianfunc(na.omit(subsetgroup1$Var1))
medianfunc(na.omit(subsetgroup2$Var1))
medianfunc(na.omit(subsetgroup3$Var1))
medianfunc(na.omit(subsetgroup4$Var1))

SumOfSquaresfunc(na.omit(subsetgroup1$Var1))
SumOfSquaresfunc(na.omit(subsetgroup2$Var1))
SumOfSquaresfunc(na.omit(subsetgroup3$Var1))
SumOfSquaresfunc(na.omit(subsetgroup4$Var1))

varfunc(na.omit(subsetgroup1$Var1))
varfunc(na.omit(subsetgroup2$Var1))
varfunc(na.omit(subsetgroup3$Var1))
varfunc(na.omit(subsetgroup4$Var1))

sdfunc(na.omit(subsetgroup1$Var1))
sdfunc(na.omit(subsetgroup2$Var1))
sdfunc(na.omit(subsetgroup3$Var1))
sdfunc(na.omit(subsetgroup4$Var1))

Only factor of gender and group by gender factor combination: (You can try other variables.)

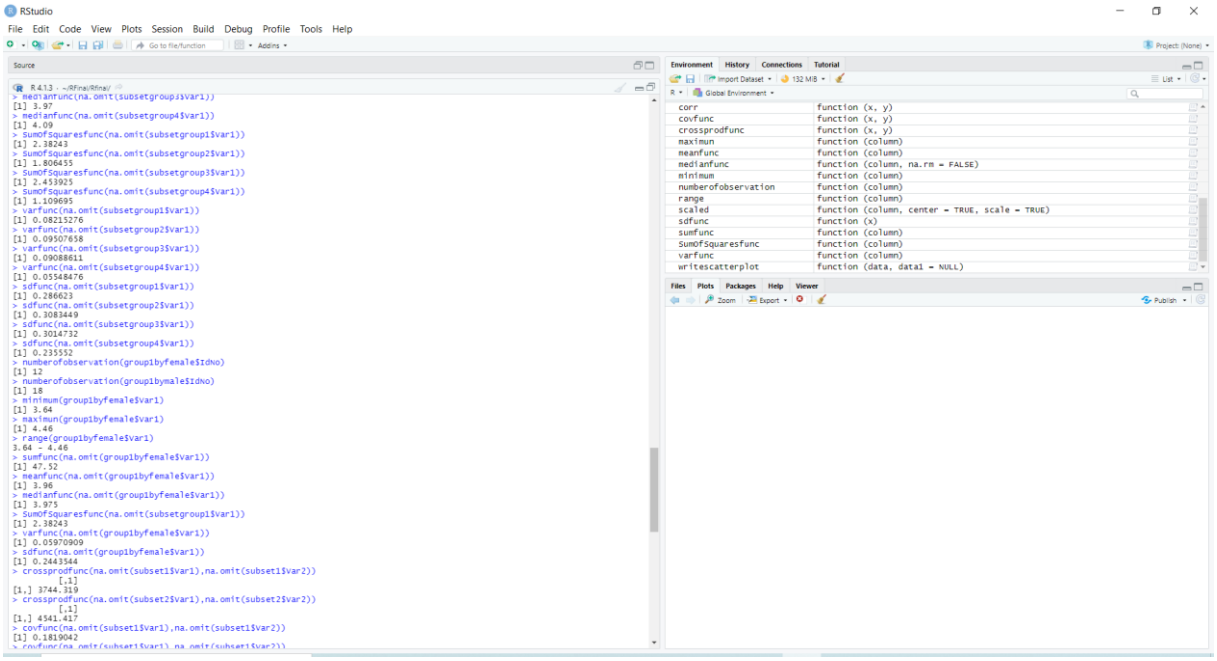
minimum(group1byfemale$Var1)
maximun(group1byfemale$Var1)
range(group1byfemale$Var1)
sumfunc(na.omit(group1byfemale$Var1))
meanfunc(na.omit(group1byfemale$Var1))
medianfunc(na.omit(group1byfemale$Var1))
SumOfSquaresfunc(na.omit(subsetgroup1$Var1))
varfunc(na.omit(group1byfemale$Var1))
sdfunc(na.omit(group1byfemale$Var1))

crossprodfunc(na.omit(subset1$Var1),na.omit(subset1$Var2))
crossprodfunc(na.omit(subset2$Var1),na.omit(subset2$Var2))

covfunc(na.omit(subset1$Var1),na.omit(subset1$Var2))
covfunc(na.omit(subset1$Var1),na.omit(subset1$Var2))

corr(na.omit(subset1$Var1),na.omit(subset1$Var2))
corr(na.omit(subset2$Var1),na.omit(subset2$Var2))

Some output by functions:

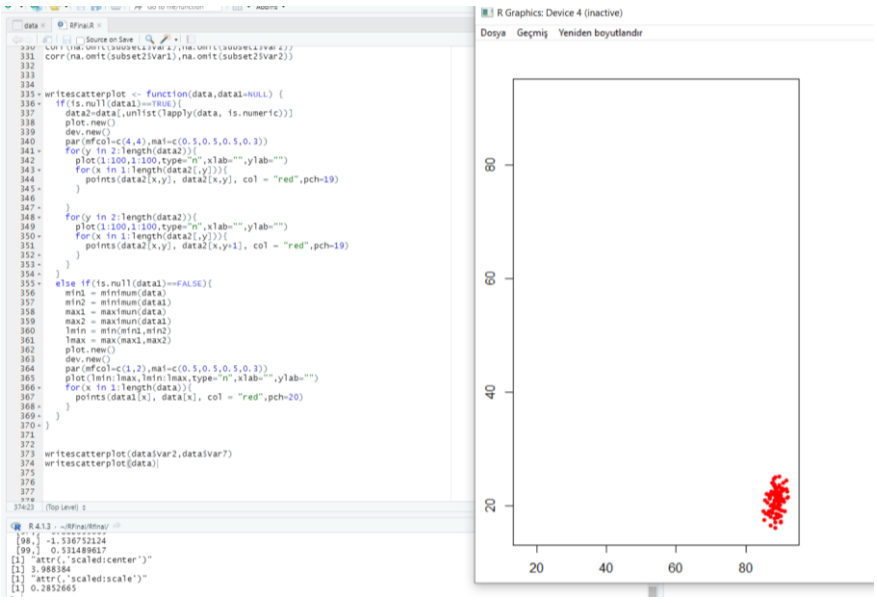


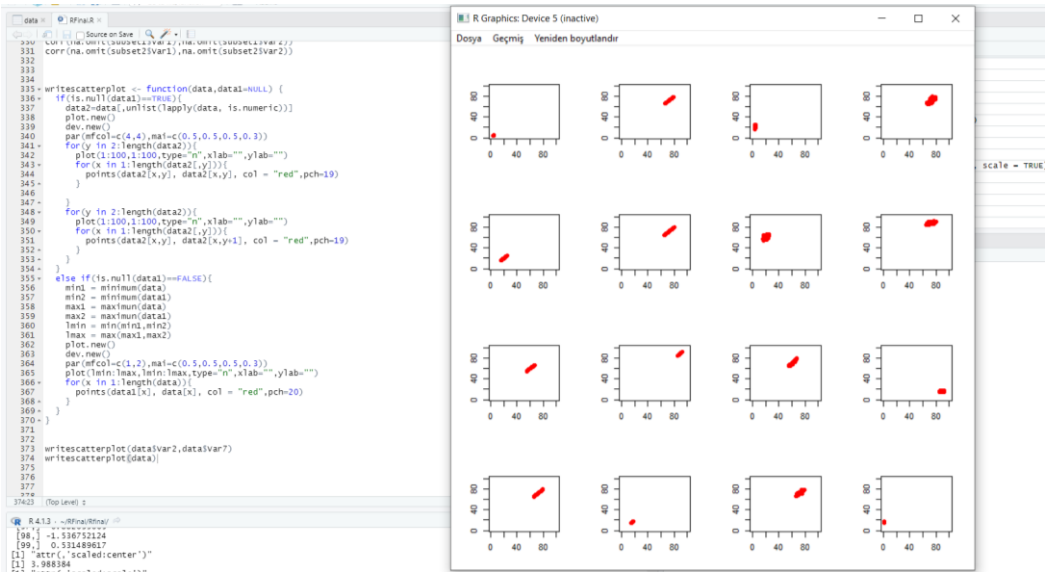
Question 3 :

We drew scatterplot and scatterplot matrix with our own functions:

```
writescatterplot <- function(data,data1=NULL) {
 if(is.null(data1)==TRUE){
   data2=data[,unlist(lapply(data, is.numeric))]
   plot.new()
   dev.new()
   par(mfcol=c(4,4),mai=c(0.5,0.5,0.5,0.3))
   for(y in 2:length(data2)){
    plot(1:100,1:100,type="n",xlab="",ylab="")
    for(x in 1:length(data2[,y])){
     points(data2[x,y], data2[x,y], col = "red",pch=19)
    }   }
   for(y in 2:length(data2)){
    plot(1:100,1:100,type="n",xlab="",ylab="")
    for(x in 1:length(data2[,y])){
     points(data2[x,y], data2[x,y+1], col = "red",pch=19)
    }}  }
 else if(is.null(data1)==FALSE){
   min1 = minimum(data)
```

```
    min2 = minimum(data1)
    max1 = maximun(data)
    max2 = maximun(data1)
    lmin = min(min1,min2)
    lmax = max(max1,max2)
    plot.new()
    dev.new()
    par(mfcol=c(1,2),mai=c(0.5,0.5,0.5,0.3))
    plot(lmin:lmax,lmin:lmax,type="n",xlab="",ylab="")
    for(x in 1:length(data)){
      points(data1[x], data[x], col = "red",pch=20)
    } }}
writescatterplot(data$Var2,data$Var7)
writescatterplot(data)
```

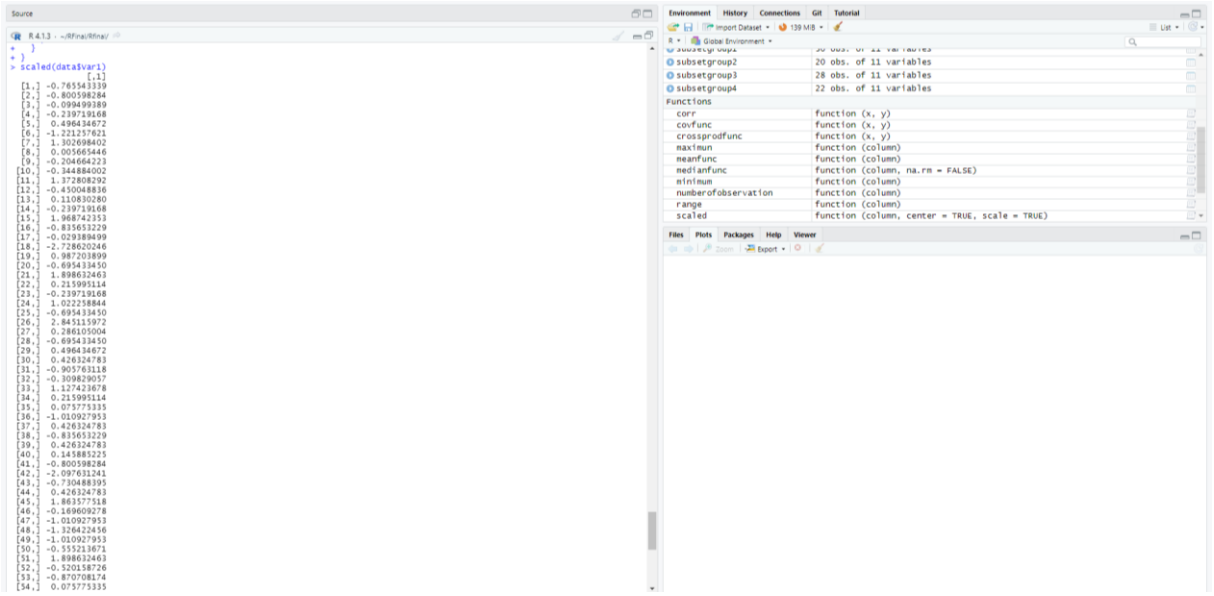




Question 4 :

We write our own function to scale variables in a data frame:

```
scaled <- function(column, center = TRUE, scale = TRUE){
  mylist <- c()
  for(x in column){
    if(is.na(x)==TRUE){
      mylist <- append(mylist,x)
    }
    else{
      sc = x - mean(na.omit(column))
      sca = sc / sdfunc(column)
      mylist <- append(mylist,sca)
    }
  }
  print(matrix(mylist,ncol = 1))
  if(center==TRUE){
    print("attr(,'scaled:center')")
    mean = meanfunc(na.omit(column))
    print(mean)
  }
  if(scale == TRUE){
    print("attr(,'scaled:scale')")
    sd = sdfunc(na.omit(column))
    print(sd)
  }
}
scaled(data$Var1)
```


Source
R 4.1.3 · ~/RFinal/Rfina/
+ }
+ }
> scaled(data$var1)
[,1]
[1,] -0.765543339
[2,] -0.800598284
[3,] -0.099499389
[4,] -0.239719168
[5,] 0.496434672
[6,] -1.221257021
[7,] 1.302698402
[8,] 0.005665446
[9,] -0.204664223
[10,] -0.344884002
[11,] 1.372808292
[12,] -0.450048836
[13,] 0.110830280
[14,] -0.239719168
[15,] 1.968742353
[16,] -0.835653229
[17,] -0.029389499
[18,] -2.728620246
[19,] 0.987203899
[20,] -0.695433450
[21,] 1.898632463
[22,] 0.215995114
[23,] -0.239719168
[24,] 1.022258844
[25,] -0.695433450
[26,] 2.845115972
[27,] 0.286105004
[28,] -0.695433450
[29,] 0.496434672
[30,] 0.426324783
[31,] -0.905763118
[32,] -0.309829057
[33,] 1.127423678
[34,] 0.215995114
[35,] 0.075775335
[36,] -1.010927953
[37,] 0.426324783
[38,] -0.835653229
[39,] 0.426324783
[40,] 0.145885225
[41,] -0.800598284
[42,] -2.097631241
[43,] -0.730488395
[44,] 0.426324783
[45,] 1.863577518
[46,] -0.169609278
[47,] -1.010927953
[48,] -1.326422456
[49,] -1.010927953
[50,] -0.555213871
[51,] 1.898632463
[52,] -0.520158726
[53,] -0.870708174
[54,] 0.075775335
Environment
History
Connections
Git
Tutorial
Import Dataset
139 MiB
List
R
Global Environment
subsetgroup2
20 obs. of 11 variables
subsetgroup3
28 obs. of 11 variables
subsetgroup4
22 obs. of 11 variables
Functions
corr
function (x, y)
covfunc
function (x, y)
crossprodfunc
function (x, y)
maximun
function (column)
meanfunc
function (column)
medianfunc
function (column, na.rm = FALSE)
minimum
function (column)
numberofobservation
function (column)
range
function (column)
scaled
function (column, center = TRUE, scale = TRUE)
Files
Plots
Packages
Help
Viewer
Zoom
Export