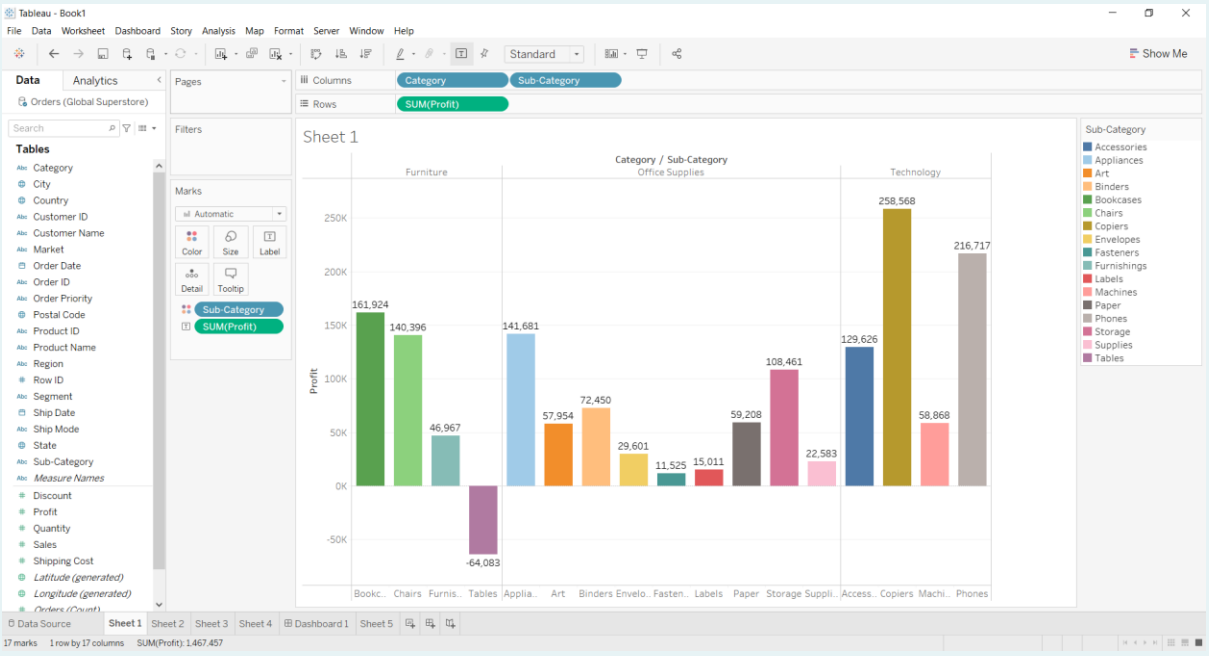
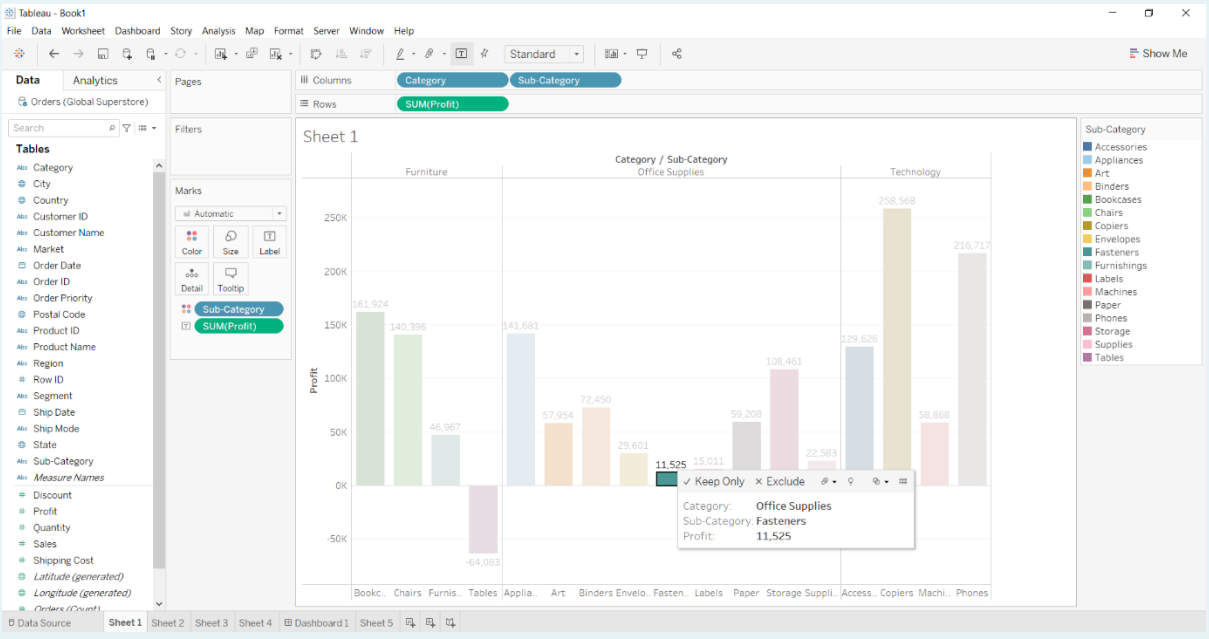
Aryaman Mishra

19BCE1027

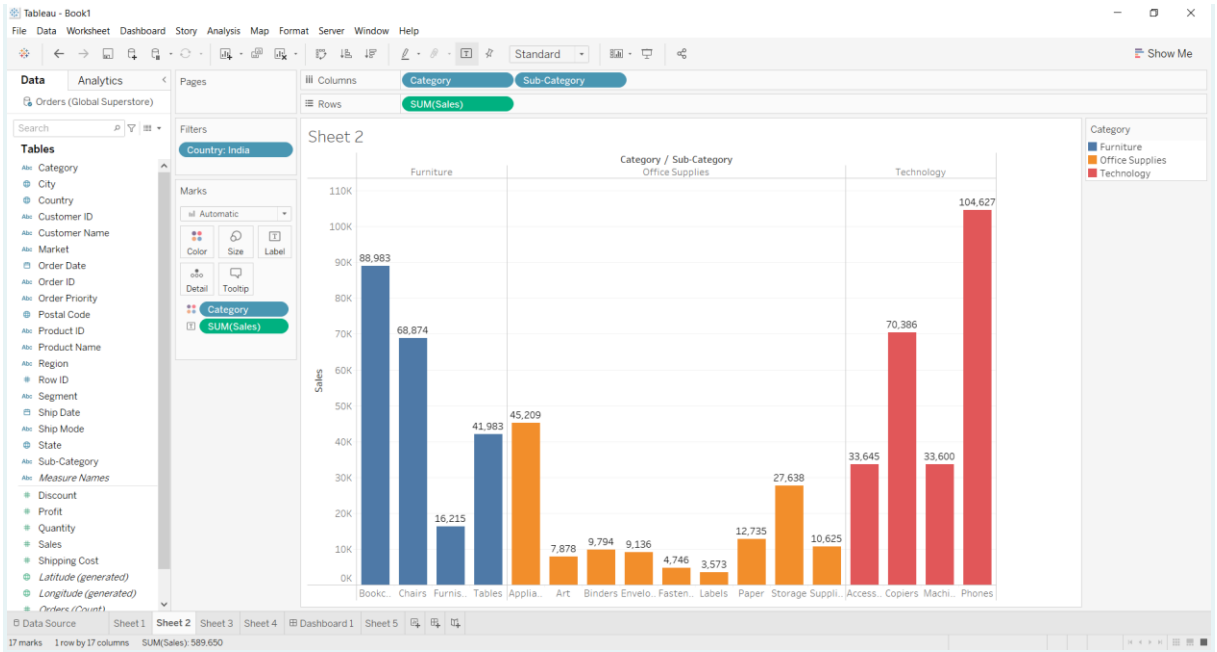
1. Using the Global *Superstore* dataset to do the visual analytics with respect to the following points using tableau. (25 marks)

a. Plot the sum of profits for each category and sub category. Which sub category is having $11,525 sum of profit across all sub categories?



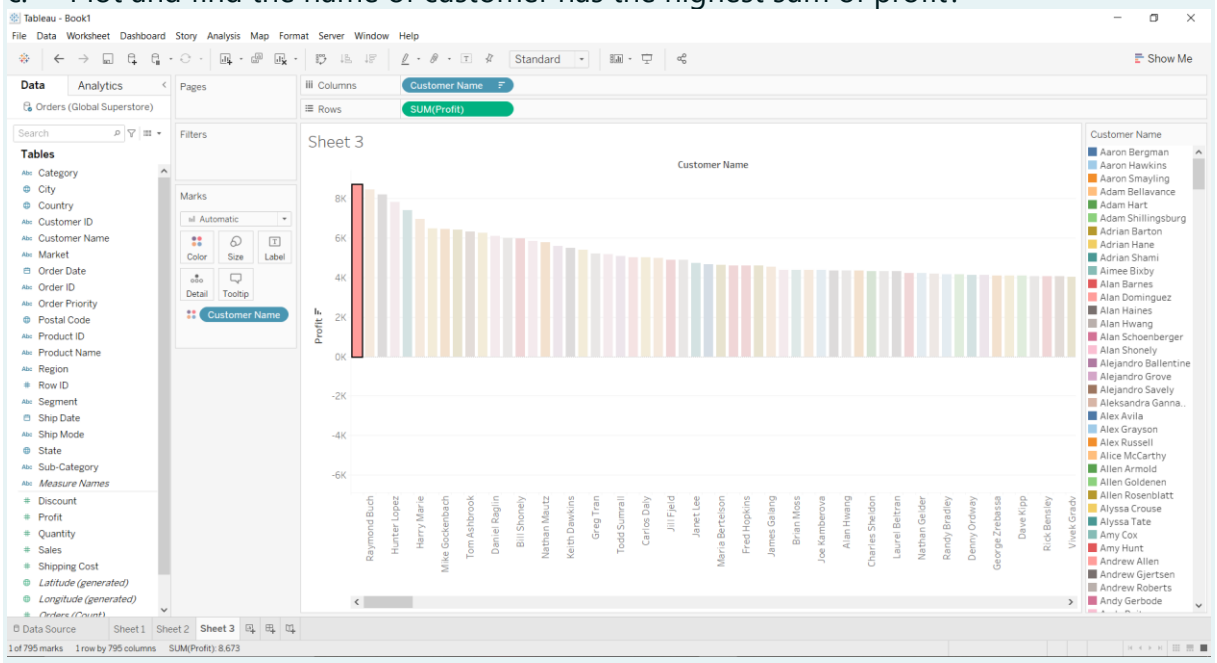


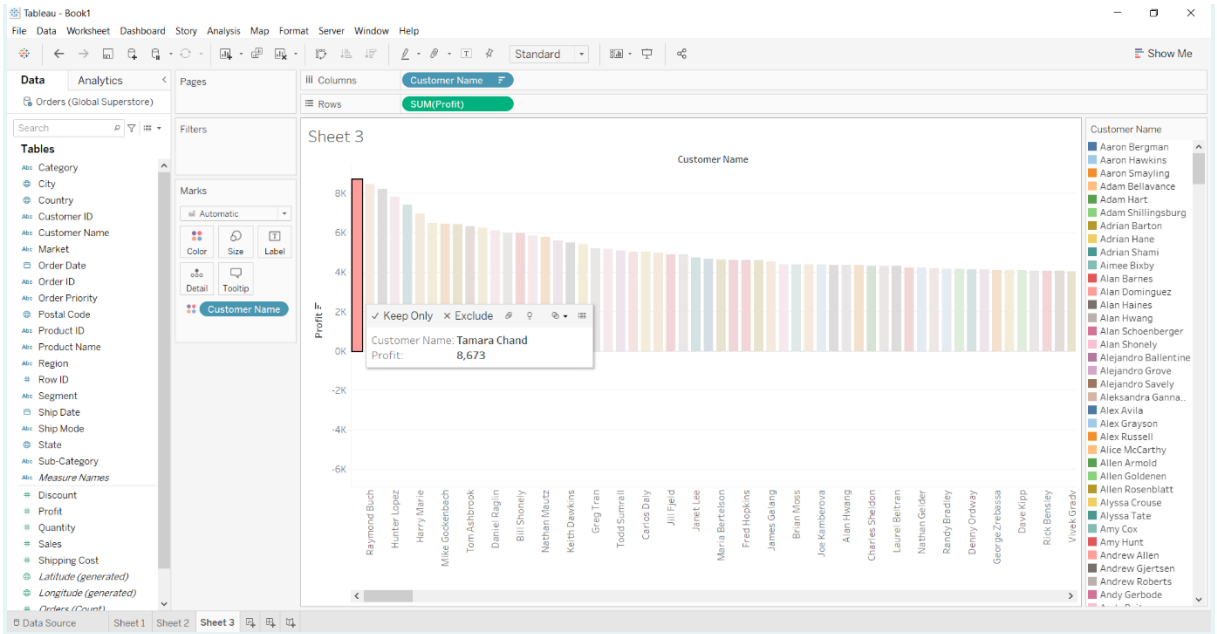
b. Plot the sum of sales for each category and sub category of India. Which sub category is having highest sum of sales in India?



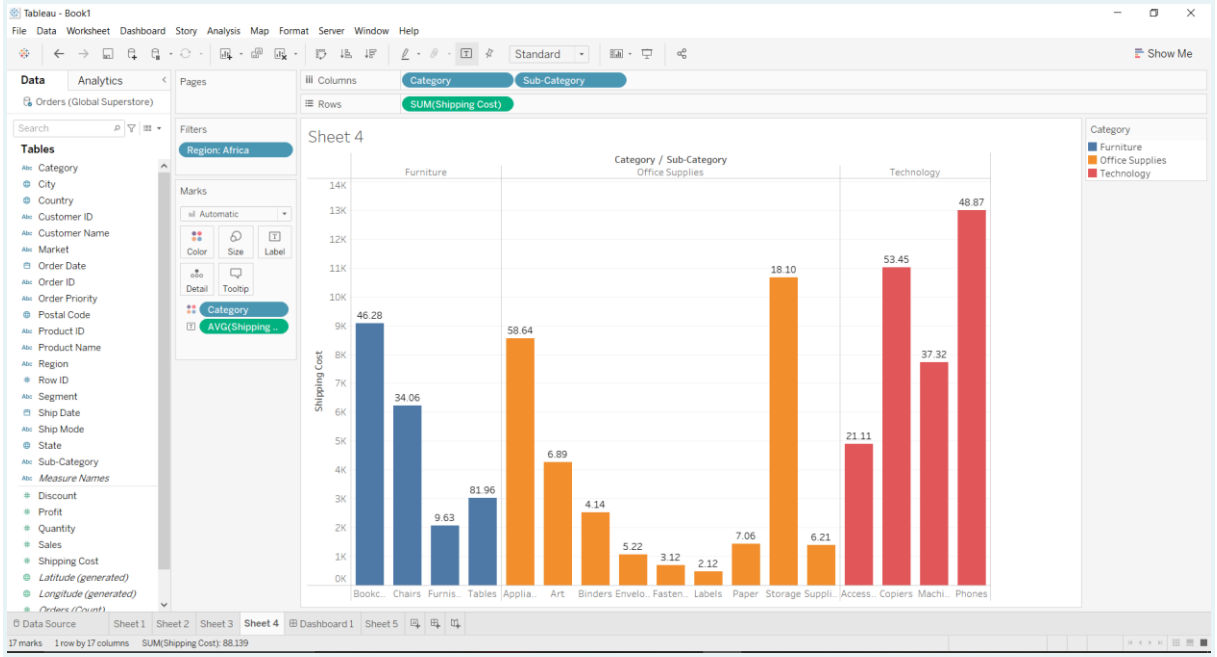
Phones in Technology Category has highest sum of sales.

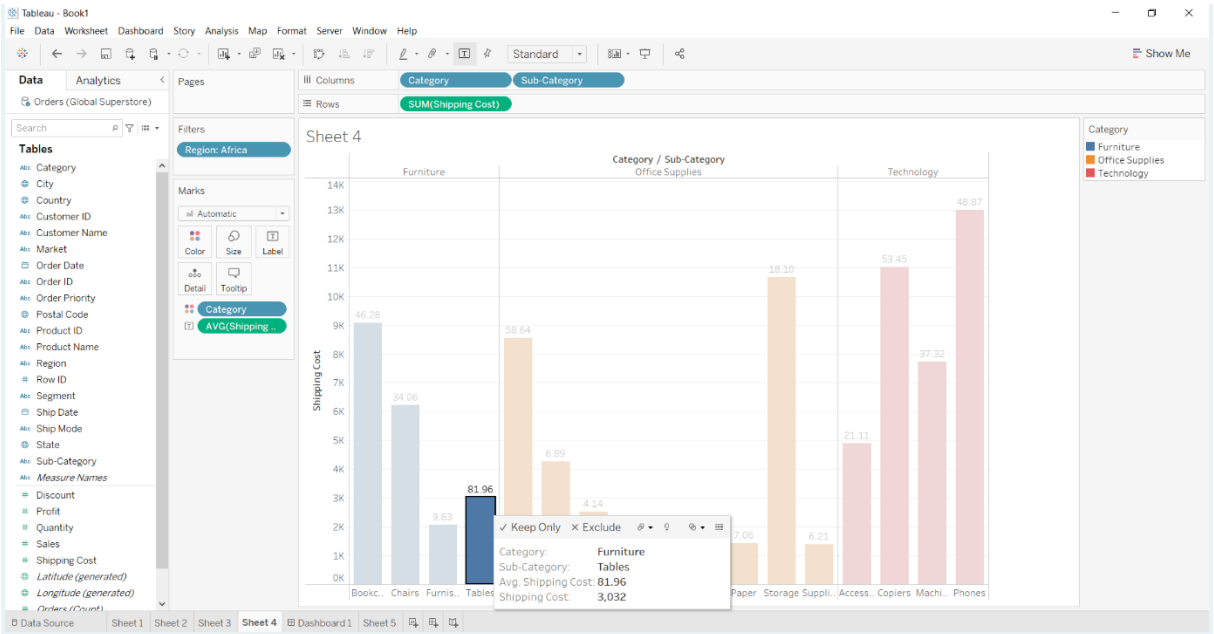
c. Plot and find the name of customer has the highest sum of profit?





d. Plot the sum of shipping cost for each category and sub category of Africa. Which sub category is having average shipping cost of $82.0?





Note: Apply labels, various colors and blue border in the appropriate plots

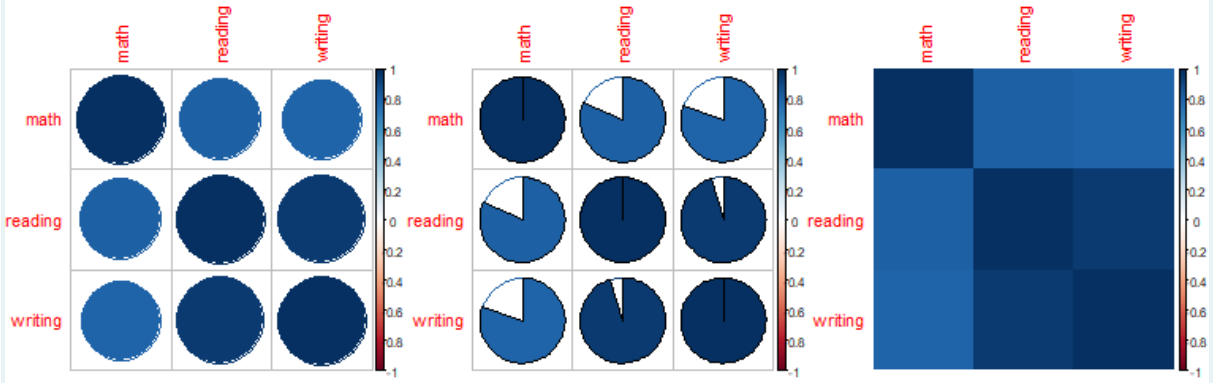
2. Write the R program for the following: (25 marks)

a. Using the student performance dataset apply the correct correlation techniques to identify the relationship among the variables for the dataset

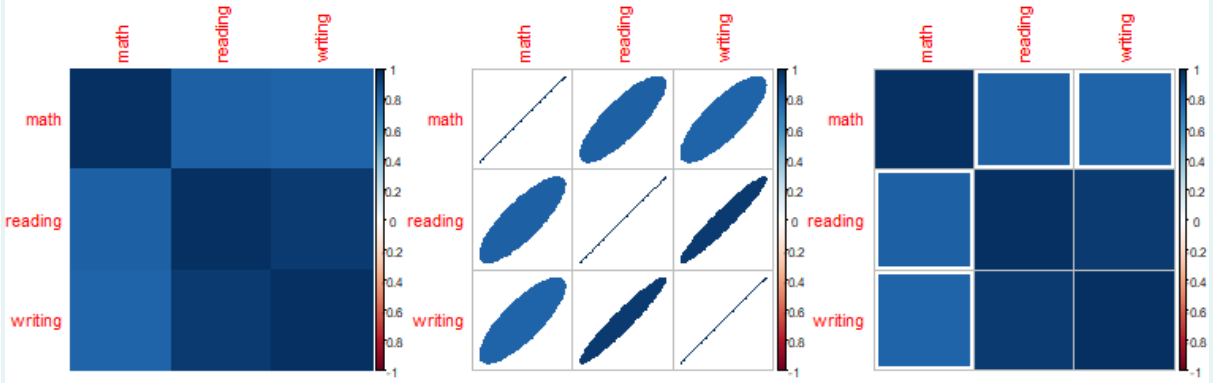
```
library("readxl")

library(corrplot)

df = read_excel("C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/StudentsPerformance_Dataset.xlsx")

print(ncol(df))

print(nrow(df))

cor(df$reading,df$writing,method="pearson")

cor(df$reading,df$writing,method="spearman")

cor(df$reading,df$writing,method="kendall")

cor.test(df$reading,df$writing,method="pearson")
```

```
> df = read_excel("C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/StudentsPerformance_Dataset.xlsx")
> print(ncol(df))
[1] 8
> print(nrow(df))
[1] 1000
> cor(df$reading,df$writing,method="pearson")
[1] 0.9545981
> cor(df$reading,df$writing,method="spearman")
[1] 0.9489525
> cor(df$reading,df$writing,method="kendall")
[1] 0.8200575
> cor.test(df$reading,df$writing,method="pearson")

        Pearson's product-moment correlation

data:  df$reading and df$writing
t = 101.23, df = 998, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9487506 0.9597921
sample estimates:
      cor
0.9545981
```

```
sapply(df, class)

M=cor(df)

print(M)

corrplot(M,method='number',type='upper')

corrplot(M,method='number',type='lower')

corrplot(M,method='number')

corrplot(M,method='circle')

corrplot(M,method='pie')

corrplot(M,method='shade')

corrplot(M,method='color')

corrplot(M,method='ellipse')

corrplot(M,method='square')
```

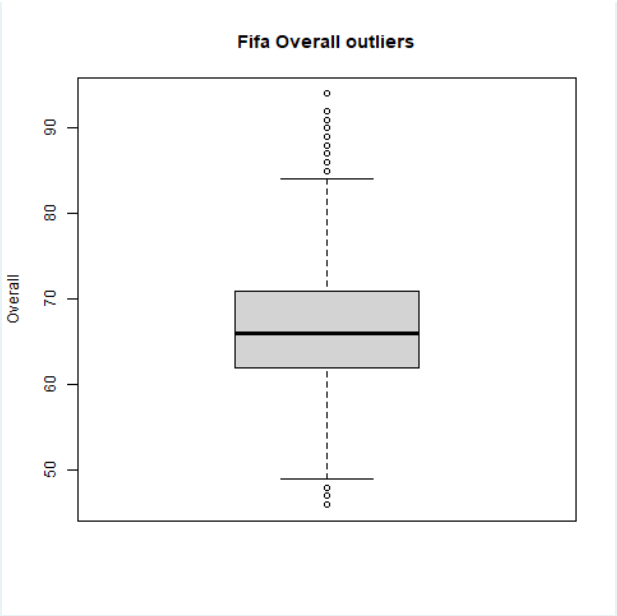

math
reading
writing
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


math
reading
writing
1.00
0.82
0.80
0.95
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

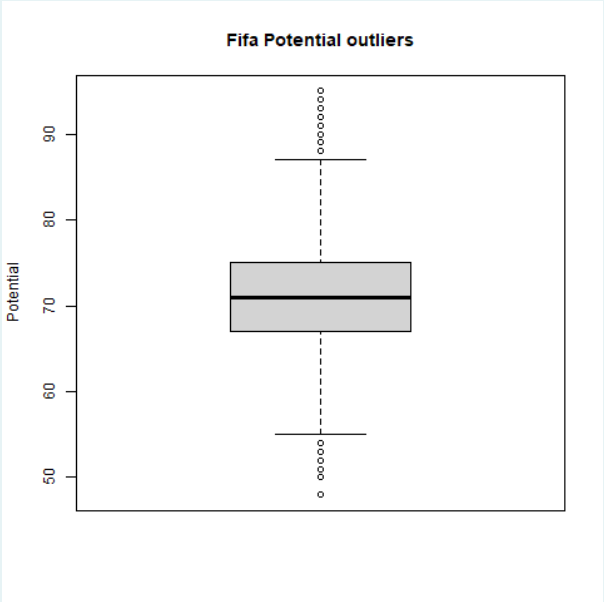


b. Using the players' performance, find the outliers based on the appropriate features

```
library("readxl")



fifa = read_excel("C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization
Lab/FAT/FIFA Players Data.xlsx")

head(fifa)

png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization
Lab/FAT/boxplot1.png")

boxplot(fifa$Overall, data = fifa,ylab = "Overall", main = "Fifa Overall outliers")

dev.off()
```

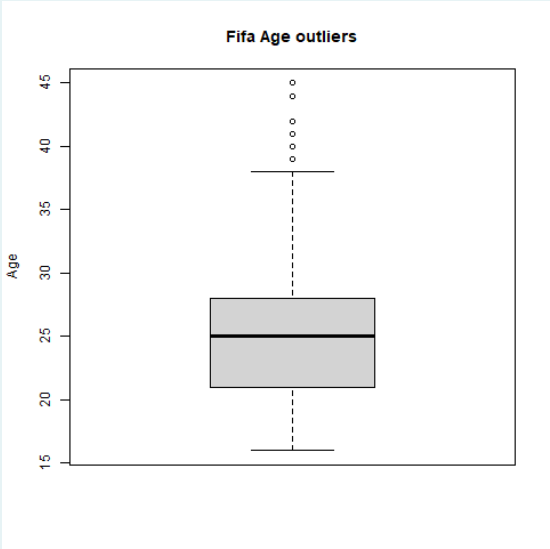



```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot2.png")

boxplot(fifa$Potential, data = fifa,ylab = "Potential", main = "Fifa Potential outliers")

dev.off()
```




```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot3.png")

boxplot(fifa$Age, data = fifa,ylab = "Age", main = "Fifa Age outliers")
```

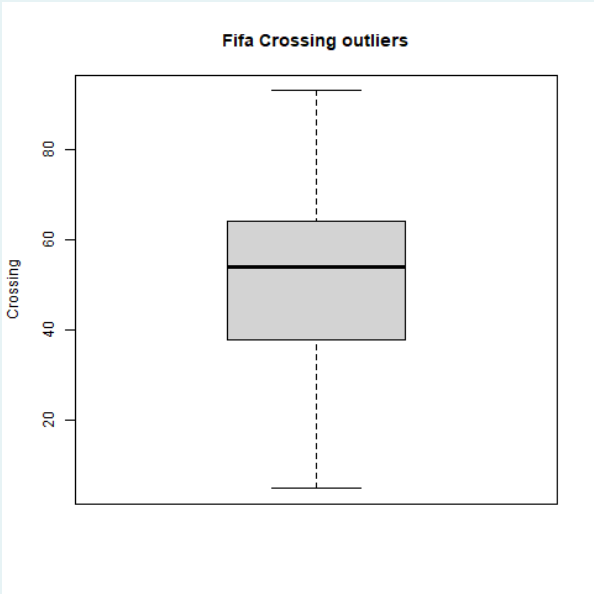
dev.off()



png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot4.png")

boxplot(fifa$Crossing, data = fifa,ylab = "Crossing", main = "Fifa Crossing outliers")

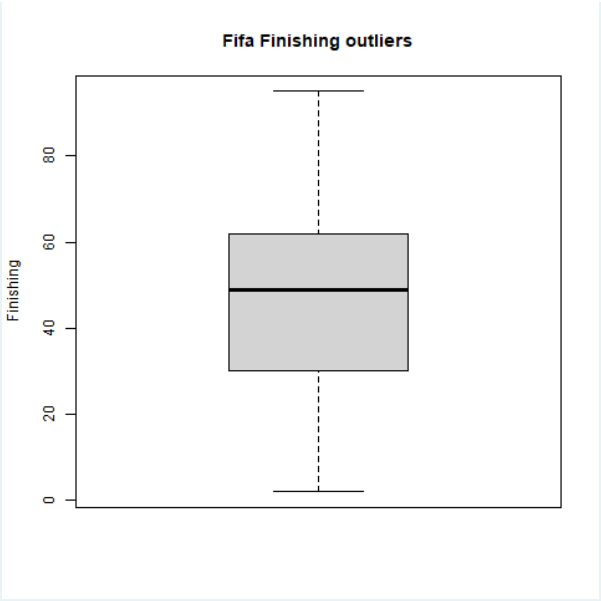
dev.off()



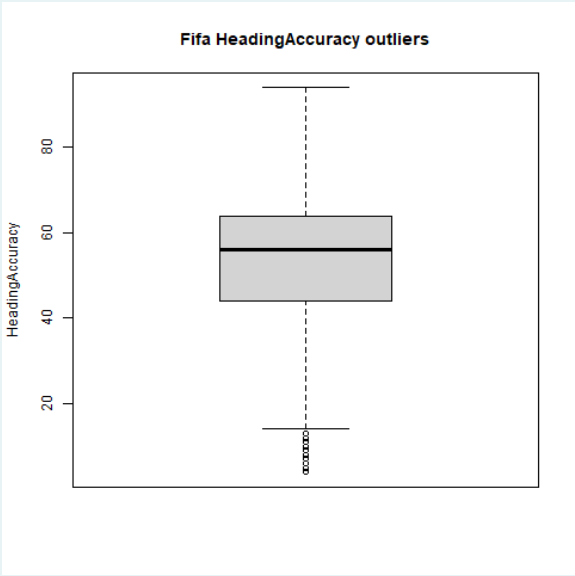
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot5.png")

boxplot(fifa$Finishing, data = fifa,ylab = "Finishing", main = "Fifa Finishing outliers")

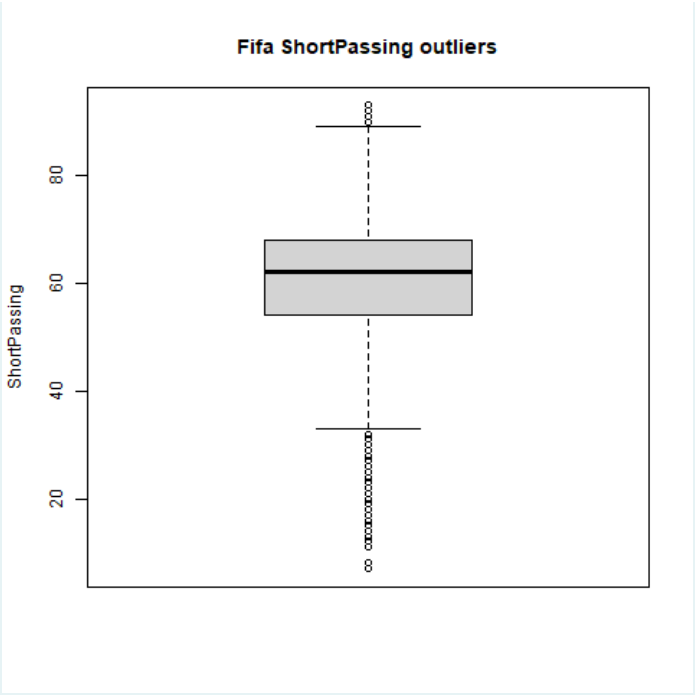
dev.off()



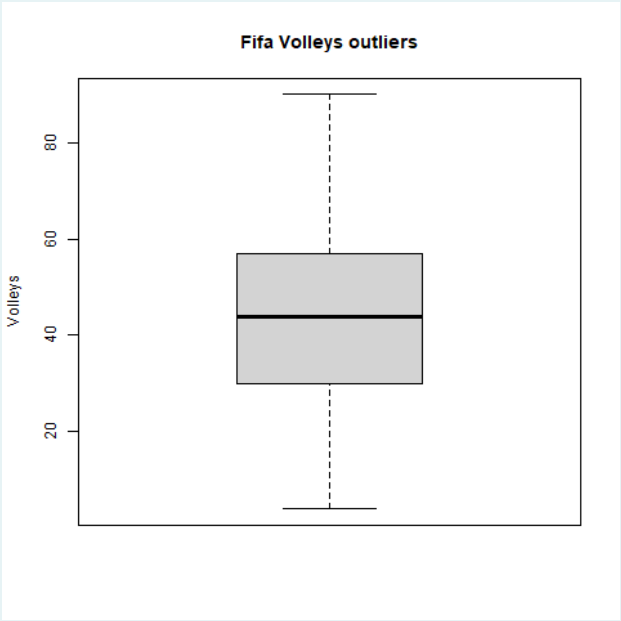

```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot6.png")

boxplot(fifa$HeadingAccuracy, data = fifa,ylab = "HeadingAccuracy", main = "Fifa HeadingAccuracy outliers")

dev.off()
```




```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot7.png")

boxplot(fifa$ShortPassing, data = fifa,ylab = "ShortPassing", main = "Fifa ShortPassing outliers")

dev.off()
```

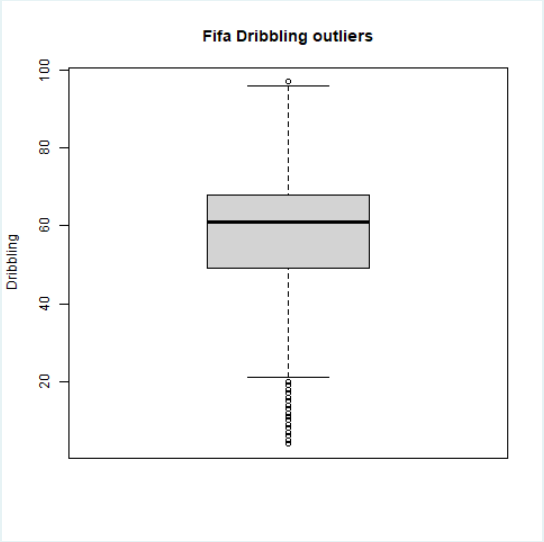



```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot8.png")

boxplot(fifa$Volleys, data = fifa,ylab = "Volleys", main = "Fifa Volleys outliers")

dev.off()
```




```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot9.png")

boxplot(fifa$Dribbling, data = fifa,ylab = "Dribbling", main = "Fifa Dribbling outliers")
```

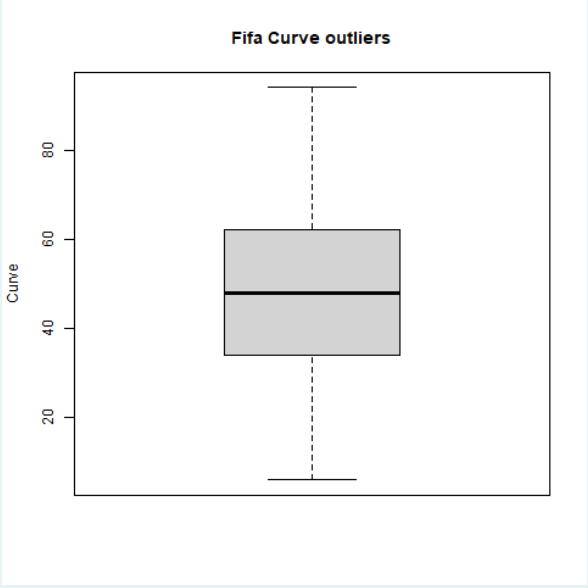
dev.off()



png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot10.png")

boxplot(fifa$Curve, data = fifa,ylab = "Curve", main = "Fifa Curve outliers")

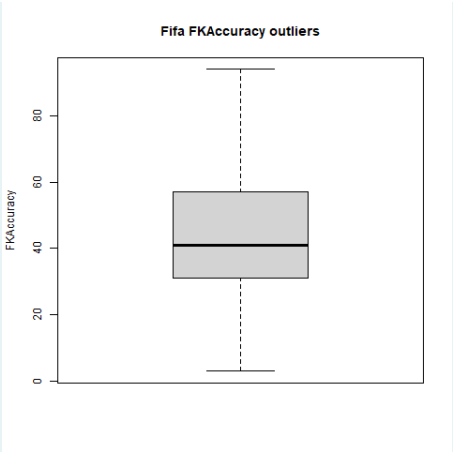
dev.off()



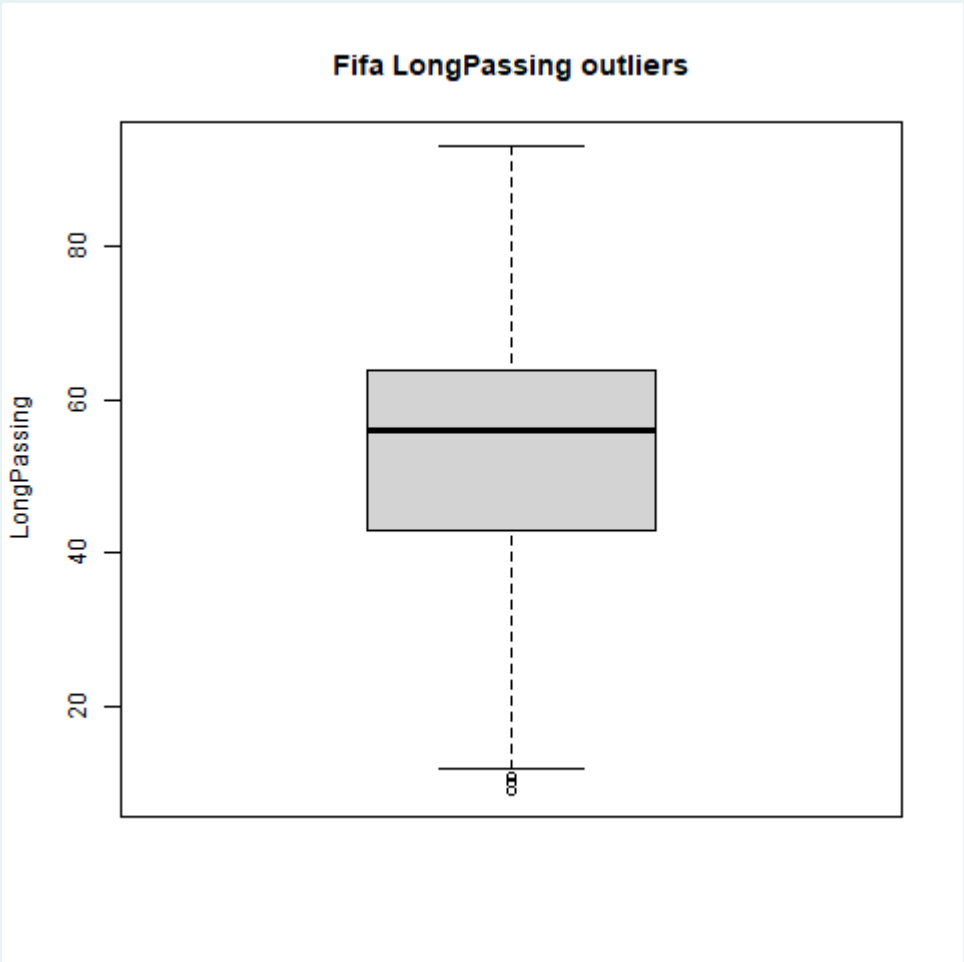
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot11.png")

boxplot(fifa$FKAccuracy, data = fifa,ylab = "FKAccuracy", main = "Fifa FKAccuracy outliers")

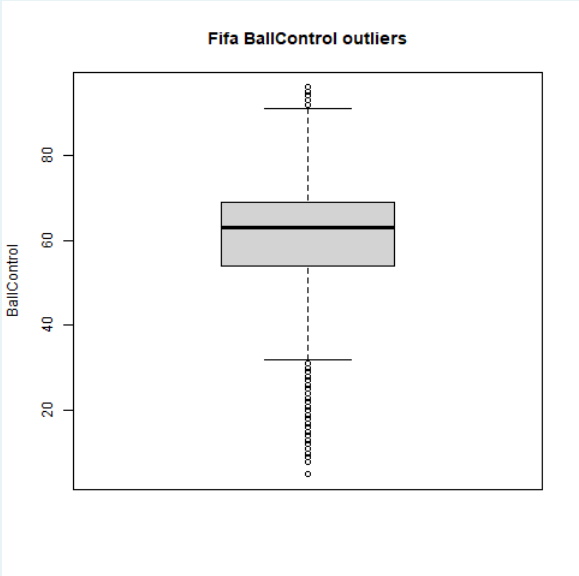
dev.off()



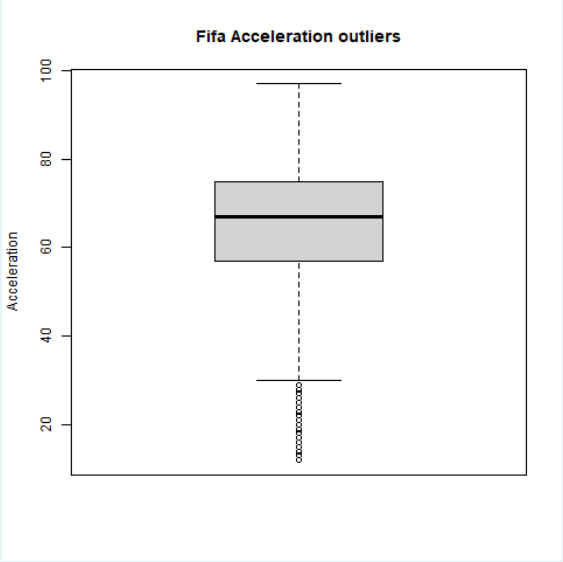

```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot12.png")

boxplot(fifa$LongPassing, data = fifa,ylab = "LongPassing", main = "Fifa LongPassing outliers")

dev.off()
```



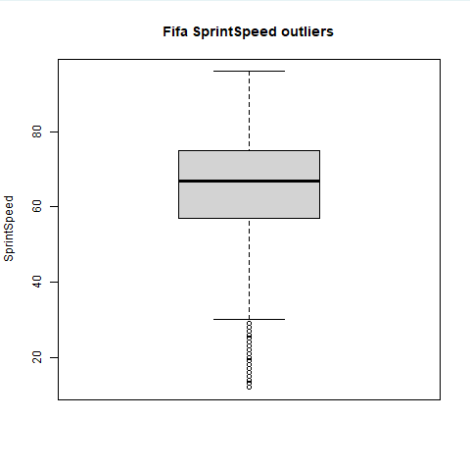
```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot13.png")

boxplot(fifa$BallControl, data = fifa,ylab = "BallControl", main = "Fifa BallControl outliers")

dev.off()
```



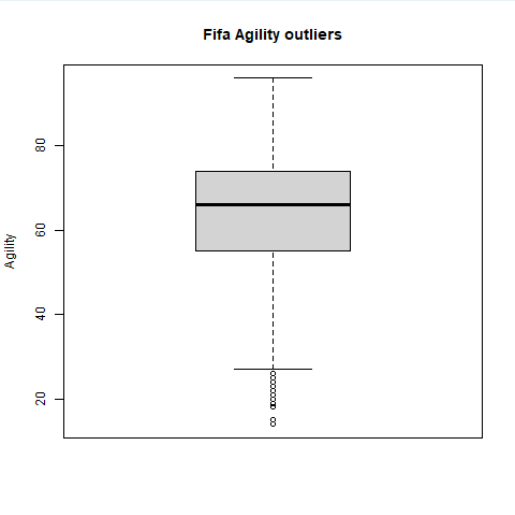

```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot14.png")

boxplot(fifa$Acceleration, data = fifa,ylab = "Acceleration", main = "Fifa Acceleration outliers")

dev.off()
```



```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot15.png")

boxplot(fifa$SprintSpeed, data = fifa,ylab = "SprintSpeed", main = "Fifa SprintSpeed outliers")

dev.off()
```

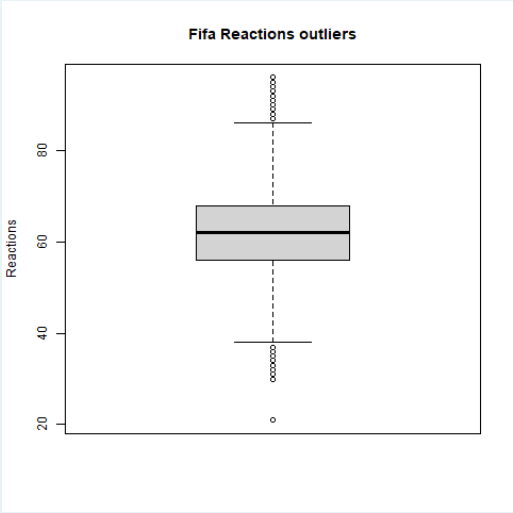



```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot16.png")

boxplot(fifa$Agility, data = fifa,ylab = "Agility", main = "Fifa Agility outliers")

dev.off()
```




```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot17.png")

boxplot(fifa$Reactions, data = fifa,ylab = "Reactions", main = "Fifa Reactions outliers")
```

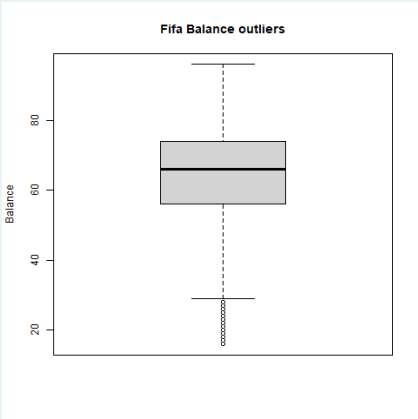
dev.off()



png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot18.png")

boxplot(fifa$Balance, data = fifa,ylab = "Balance", main = "Fifa Balance outliers")

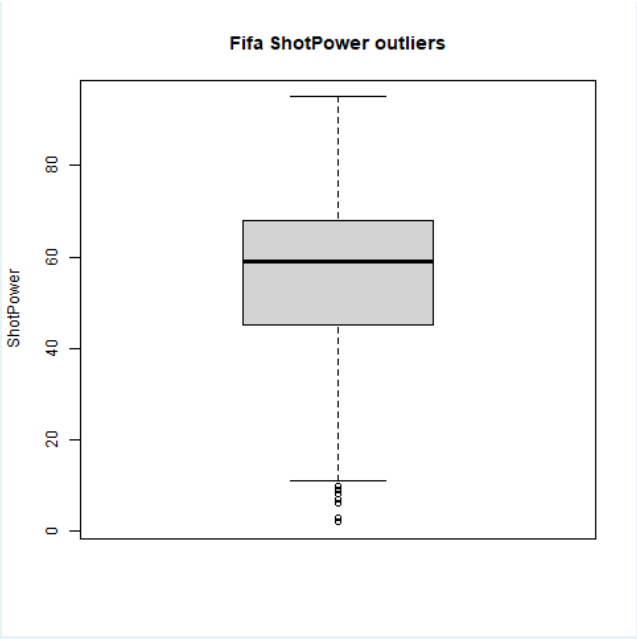
dev.off()



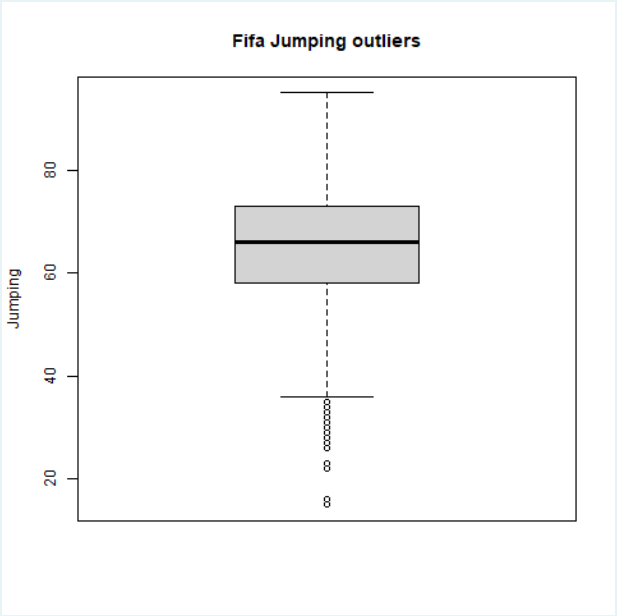
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot19.png")

boxplot(fifa$ShotPower, data = fifa,ylab = "ShotPower", main = "Fifa ShotPower outliers")

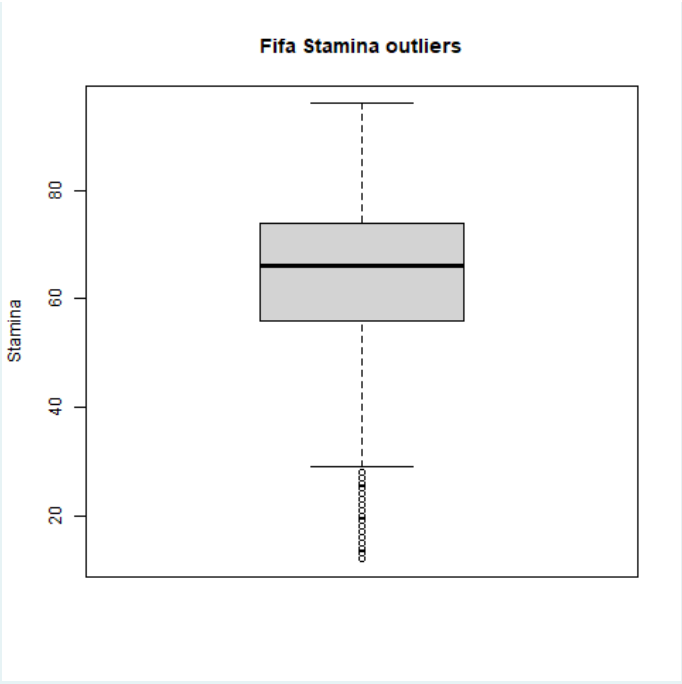
dev.off()



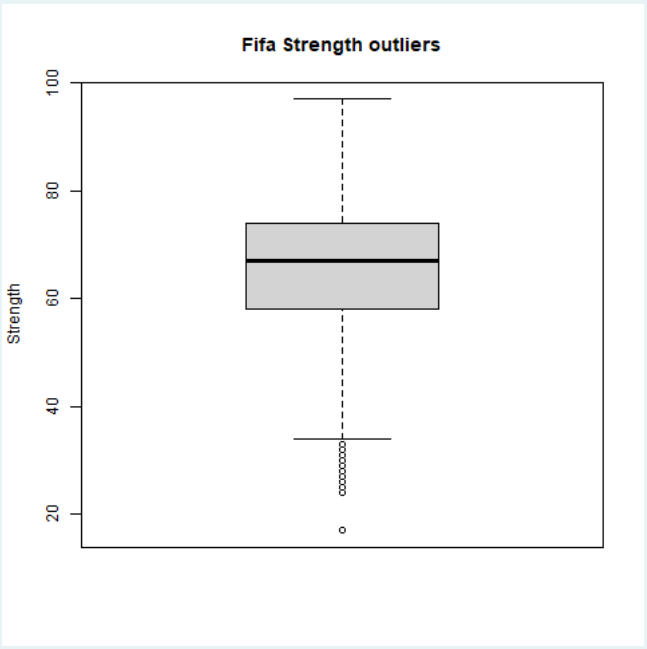

```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot20.png")

boxplot(fifa$Jumping, data = fifa,ylab = "Jumping", main = "Fifa Jumping outliers")

dev.off()
```




```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot21.png")

boxplot(fifa$Stamina, data = fifa,ylab = "Stamina", main = "Fifa Stamina outliers")

dev.off()
```



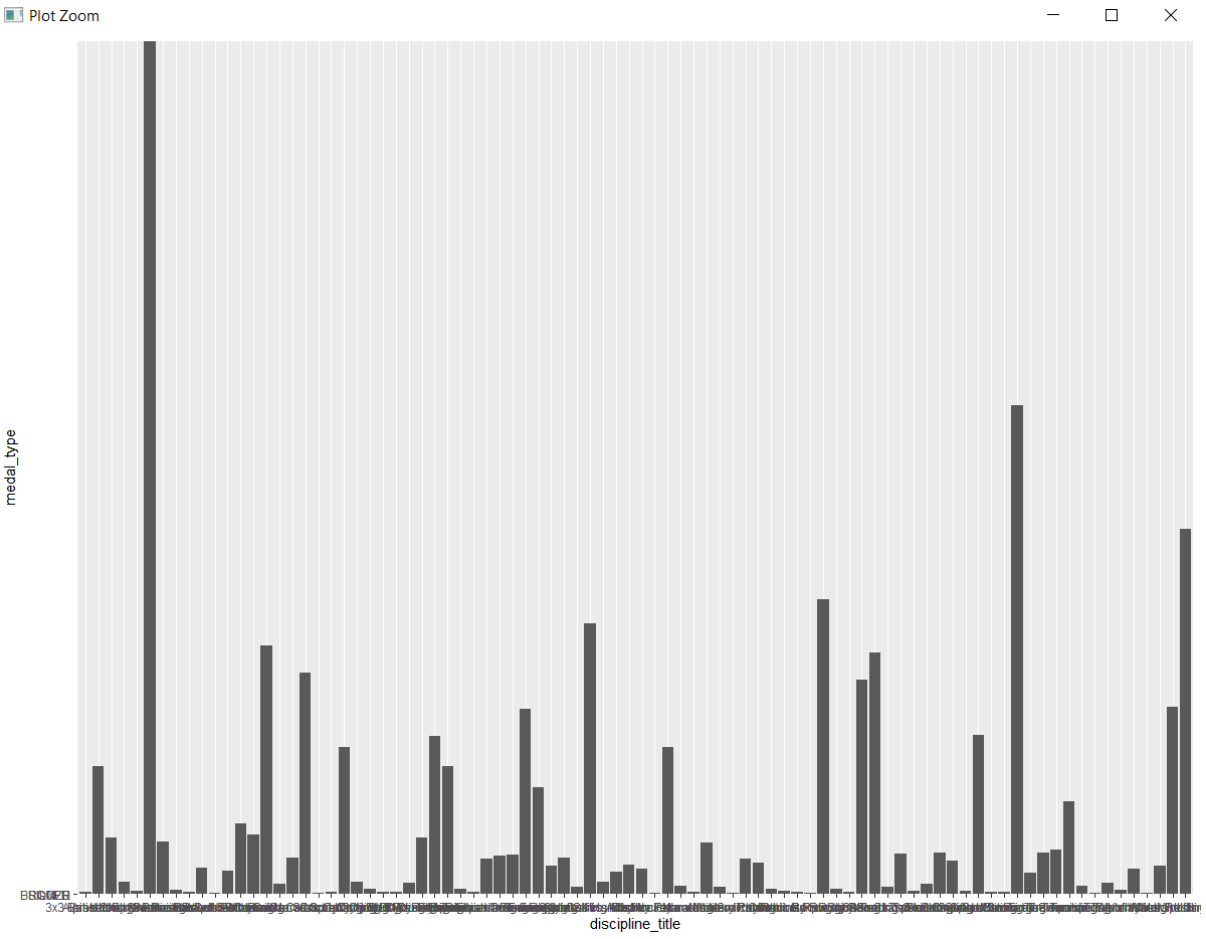

```
png(file = "C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization Lab/FAT/boxplot22.png")

boxplot(fifa$Strength, data = fifa,ylab = "Strength", main = "Fifa Strength outliers")

dev.off()
```



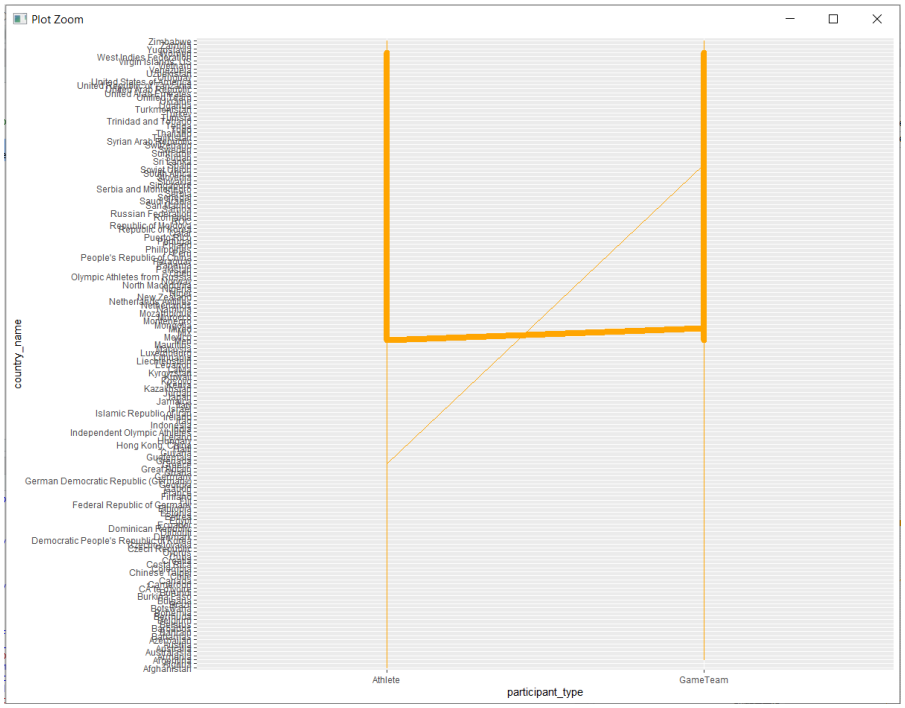
c. Using Olympics dataset apply GGplot2 toidentify the relationship among the variables.

```
library("readxl")
```

```
library(ggplot2)

library(ggvis)

df = read_excel("C:/Users/aryam/Desktop/Fall Sem 2021/Data Visualization
Lab/FAT/olympic_medals.xlsx")

ggplot(df, aes(x = discipline_title,y = medal_type))+ geom_bar(stat='identity')
```




```
ggplot(df, aes(x = participant_type,y = country_name, group = 1))+ geom_line(color = "orange")+

 geom_line( aes(y=event_gender), size=3, color="orange")
```




```
ggplot(df) + geom_bar(aes(x=discipline_title, y=medal_type),stat="identity")+
  geom_line(aes(x=slug_game, y=event_title,color="Bad Rate"),stat="identity",group = 1,
            size=1.3)
```