DAY 2

Question 1 :

CODE:

(i)b<-c(22, 28, 10)
c<-c(20, 40, 40)
cov(b,c)

(ii)a<-c(18, 2, 20)
b<-c(22, 28, 10)
c<-c(20, 40, 40)
pre<-cbind(a,b,c)
cov(pre)

(iii).b<-c(22, 28, 10)
c<-c(20, 40, 40)
cor(b,c)

(iv)a<-c(18, 2, 20)
b<-c(22, 28, 10)
c<-c(20, 40, 40)
pre<-cbind(a,b,c)
cor(pre)

OUTPUT:

```
> b<-c(22, 28, 10)
> c<-c(20, 40, 40)
> cov(b,c)
[1] -20
> a<-c(18, 2, 20)
> b<-c(22, 28, 10)
> c<-c(20, 40, 40)
> pre<-cbind(a,b,c)
> cov(pre)
          a   b         c
a  97.33333 -74 -46.66667
b -74.00000  84 -20.00000
c -46.66667 -20 133.33333
> b<-c(22, 28, 10)
> c<-c(20, 40, 40)
> cor(b,c)
[1] -0.1889822
> a<-c(18, 2, 20)
> b<-c(22, 28, 10)
> c<-c(20, 40, 40)
> pre<-cbind(a,b,c)
> cor(pre)
           a          b          c
a  1.0000000 -0.8183918 -0.4096440
b -0.8183918  1.0000000 -0.1889822
c -0.4096440 -0.1889822  1.0000000
```

Question 2 :

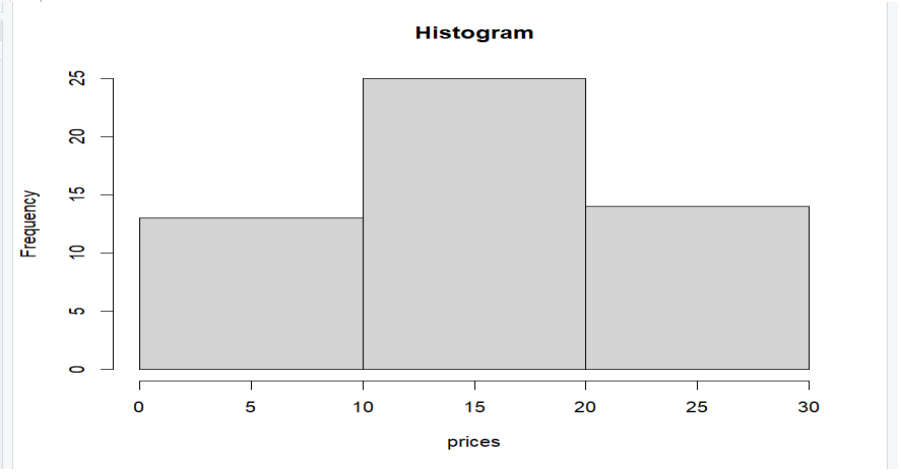
CODE :

data<-c(1, 1, 5, 5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18, 18, 18, 20, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 25, 25, 25, 25, 25, 28, 28, 30, 30, 30)
bin<-length(data)/3

```
bins<-cut(data, breaks = c(-Inf, quantile(data, probs = seq(0,1, 1/3)),Inf), include.lowest = TRUE)
tapply(data,bins,mean)
tapply(data, bins,function(x) c(min(x),max(x)))

hist(data,breaks = 3, main = "Histogram", xlab = "prices")
```

OUTPUT:

```
Console   Terminal   Background Jobs
R 4.2.2 · ~/
> data<-c(1, 1, 5, 5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18, 18, 20, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 25, 25, 25, 25, 25, 28, 28, 30, 30, 30)
> bin<-length(data)/3
> bins<-cut(data, breaks = c(-Inf, quantile(data, probs = seq(0,1, 1/3)),Inf), include.lowest = TRUE)
> tapply(data,bins,mean)
 [-Inf,1]    (1,15]   (15,20]   (20,30] (30, Inf]
  1.00000  10.71429  18.93333  25.35714        NA
> tapply(data, bins,function(x) c(min(x),max(x)))
$`[-Inf,1]`
[1] 1 1

$`(1,15]`
[1]  5 15

$`(15,20]`
[1] 18 20

$`(20,30]`
[1] 21 30

$`(30, Inf]`
NULL

>
> hist(data,breaks = 3, main = "Histogram", xlab = "prices")
>
```



QUESTION 3 :

CODE:

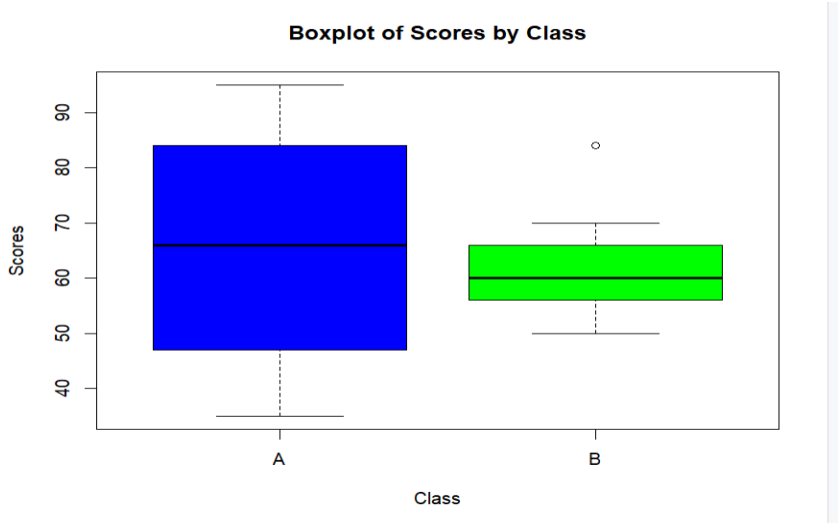
```
A <- c(76, 35, 47, 64, 95, 66, 89, 36, 84)
B <- c(51, 56, 84, 60, 59, 70, 63, 66, 50)

mean_A <- mean(A)
median_A <- median(A)
range_A <- max(A) - min(A)
mean_B <- mean(B)
median_B <- median(B)
range_B <- max(B) - min(B)
```

```
combined_data <- data.frame(Class = c(rep("A", length(A)), rep("B", length(B))), Score = c(A, B))
boxplot(Score ~ Class, data = combined_data, col = c("blue", "green"), xlab = "Class", ylab = "Scores", main = "Boxplot of Scores by Class")
```

(II)

```
combined_data <- data.frame(Class = c(rep("A", length(A)), rep("B", length(B))), Score = c(A, B))

boxplot(Score ~ Class, data = combined_data, col = c("blue", "green"), xlab = "Class", ylab = "Scores", main = "Boxplot of Scores by Class")
```

OUTPUT:



QUESTION 4 :

CODE:

```
data <- c(200, 300, 400, 600, 1000)

min_value <- 50000
max_value <- 100000
v <- 80
min_max_normalized <- (v - min_value) / (max_value - min_value)
min_max_normalized
mean_value <- mean(data)
standard_deviation <- sd(data)
z_score_normalized <- (v - mean_value) / standard_deviation
z_score_normalized
```

OUTPUT :

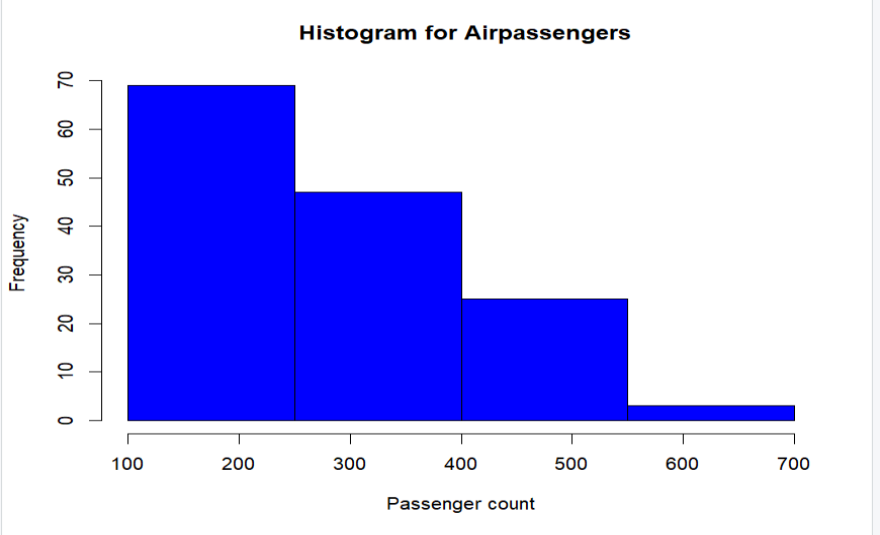
```
> data <- c(200, 300, 400, 600, 1000)
>
> min_value <- 50000
> max_value <- 100000
> v <- 80
> min_max_normalized <- (v - min_value) / (max_value - min_value)
> min_max_normalized
[1] -0.9984
> mean_value <- mean(data)
> standard_deviation <- sd(data)
> z_score_normalized <- (v - mean_value) / standard_deviation
> z_score_normalized
[1] -1.328157
```

QUESTION 5 :

CODE:

data("AirPassengers")
hist(AirPassengers, breaks = seq(100, 700, by = 150), col = "blue", main=" Histogram for Airpassengers", xlab = "Passenger count", ylab = "Frequency")

OUTPUT:

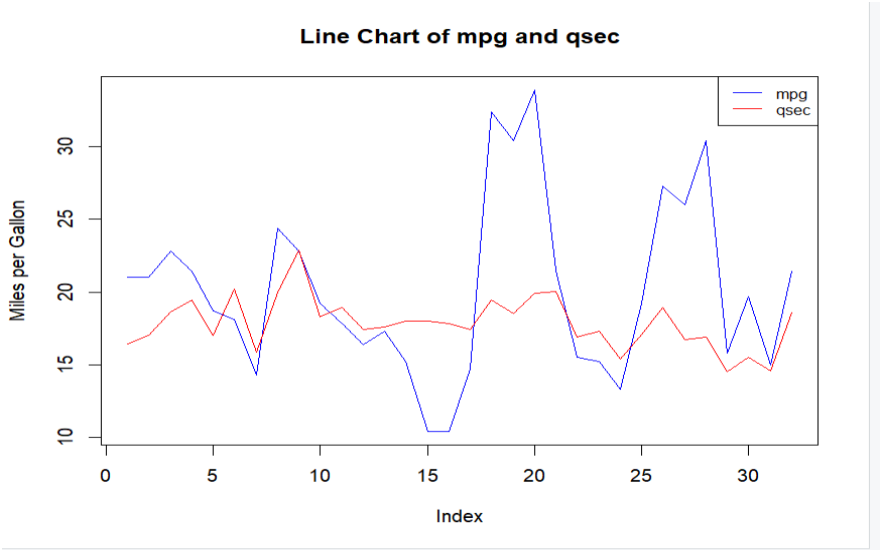


QUESTION 6 :

CODE :

data("mtcars")
plot(mtcars$mpg, type = "l", col = "blue", xlab = "Index", ylab = "Miles per Gallon", main = "Line Chart of mpg and qsec")
lines(mtcars$qsec, col = "red")
legend("topright", legend = c("mpg", "qsec"), col = c("blue", "red"), lty = 1, cex = 0.8)

OUTPUT :



QUESTION 7 :

CODE :

```
data("iris")
str(iris)
plot(iris$Sepal.Length, iris$Petal.Length, main = "Scatter plot of Sepal.Length vs. Petal.Length",xlab = "Sepal.Length", ylab = "Petal.Length", col = "blue", pch = 16)
model <- lm(Petal.Length ~ Sepal.Length, data = iris)
abline(model, col = "red")
new_data <- data.frame(Sepal.Length = 5.5)
predicted_Petal_Length <- predict(model, newdata = new_data)
predicted_Petal_Length
```

OUTPUT:

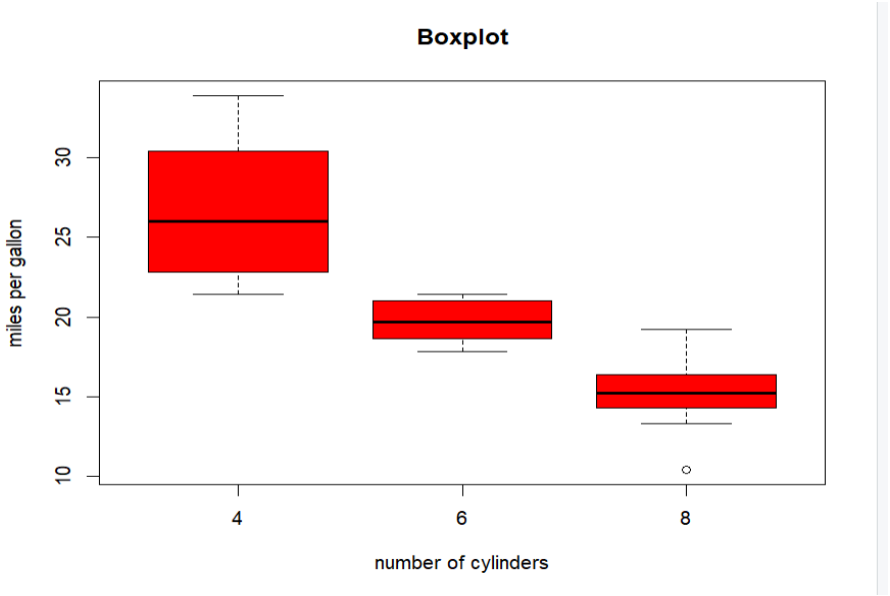
```
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
> plot(iris$Sepal.Length, iris$Petal.Length, main = "Scatter plot of Sepal.Length vs. Petal.Lengt
h",xlab = "Sepal.Length", ylab = "Petal.Length", col = "blue", pch = 16)
> model <- lm(Petal.Length ~ Sepal.Length, data = iris)
> abline(model, col = "red")
> new_data <- data.frame(Sepal.Length = 5.5)
> predicted_Petal_Length <- predict(model, newdata = new_data)
> predicted_Petal_Length
       1
3.119938
>
```

QUESTION 8 :

CODE :

```
data("mtcars")
boxplot(mpg ~ cyl, data = mtcars, main = "Boxplot", xlab = "number of cylinders", ylab =
"miles per gallon", col= "red")
```

OUTPUT :

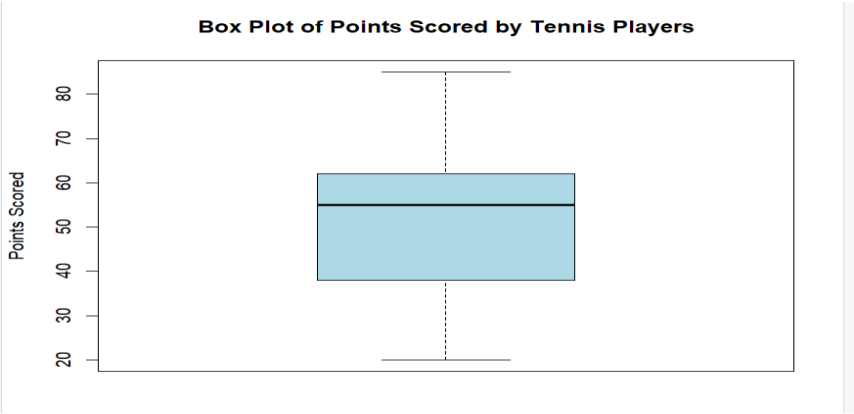


QUESTION 9 :

CODE :

```
score <- c(20, 25, 30, 32, 35, 38, 40, 45, 50, 52, 55, 56, 58, 59, 60, 62, 65, 70, 75, 80, 85)

boxplot(score, col = "lightblue", main = "Box Plot of Points Scored by Tennis Players", ylab =
"Points Scored")
```

OUTPUT :



QUESTION 10 :

CODE :

```
dia<-read.csv("C://Users//FLORENCIA ABEL//OneDrive//Documents//diabetes.csv")
View(dia)

plot(dia$Age, dia$BloodPressure, xlab = "Age", ylab = "Blood Pressure", main = "Blood Pressure vs. Age", col = "blue",pch = 16)

age_group_labels <- cut(dia$Age, breaks = c(0, 35, 55, Inf), labels = c("Young", "Middle-aged", "Elderly"))

age_group_avg_bp <- tapply(dia$BloodPressure, age_group_labels, mean)

barplot(age_group_avg_bp, main = "Average Blood Pressure by Age Group",xlab = "Age Group",ylab = "Average Blood Pressure", col = "steelblue", ylim = c(0, max(age_group_avg_bp) * 1.2))
```