Progressive Education Society's

Modern College of Engineering, Pune

MCA Department

A.Y.2023-24

(410908) Data Science Laboratory

****

Class : SY-MCA Shift / Div : S3/B Roll Number : 52147

Name : Nisha Harish Parekh Assignment No : 5 Date of Implementation : 8/11/2023

****

Q1) Find the mean, median, Mode, Range, Interquartile Range IQR and normal distribution of the physical-fitness scores. Third graders at Roth Elementary School were given a physical-fitness test. Their scores were:
a. 12 22 6 9 2 9 5 9 3 5 16 1 22 18
b. 6 12 21 23 9 10 24 21 17 11 18 19 17 5
c. 14 16 19 19 18 3 4 21 16 20 15 14 17 4
d. 5 22 12 15 18 20 8 10 13 20 6 9 2 17
e. 15 9 4 15 14 19 3 24

Program :

```
scores_a <- c(12, 22, 6, 9, 2, 9, 5, 9, 3, 5, 16, 1, 22, 18)
scores_b <- c(6, 12, 21, 23, 9, 10, 24, 21, 17, 11, 18, 19, 17, 5)
scores_c <- c(14, 16, 19, 19, 18, 3, 4, 21, 16, 20, 15, 14, 17, 4)
scores_d <- c(5, 22, 12, 15, 18, 20, 8, 10, 13, 20, 6, 9, 2, 17)
scores_e <- c(15, 9, 4, 15, 14, 19, 3, 24)

all_scores <- c(scores_a, scores_b, scores_c, scores_d, scores_e)

mean_score <- mean(all_scores)
cat("Mean:", mean_score, "\n")

median_score <- median(all_scores)
cat("Median:", median_score, "\n")

mode_score <- as.numeric(names(sort(table(all_scores), decreasing = TRUE)[1]))
cat("Mode:", mode_score, "\n")

range_score <- range(all_scores)
cat("Range:", range_score[2] - range_score[1], "\n")

q1 <- quantile(all_scores, 0.25)
q3 <- quantile(all_scores, 0.75)
iqr_score <- q3 - q1
cat("IQR:", iqr_score, "\n")

hist(all_scores, main="Histogram of Physical-Fitness Scores", xlab="Scores")

shapiro_test <- shapiro.test(all_scores)
cat("Normal distribution :", shapiro_test$p.value, "\n")
```

Progressive Education Society's

Modern College of Engineering, Pune

MCA Department

A.Y.2023-24

(410908) Data Science Laboratory

****

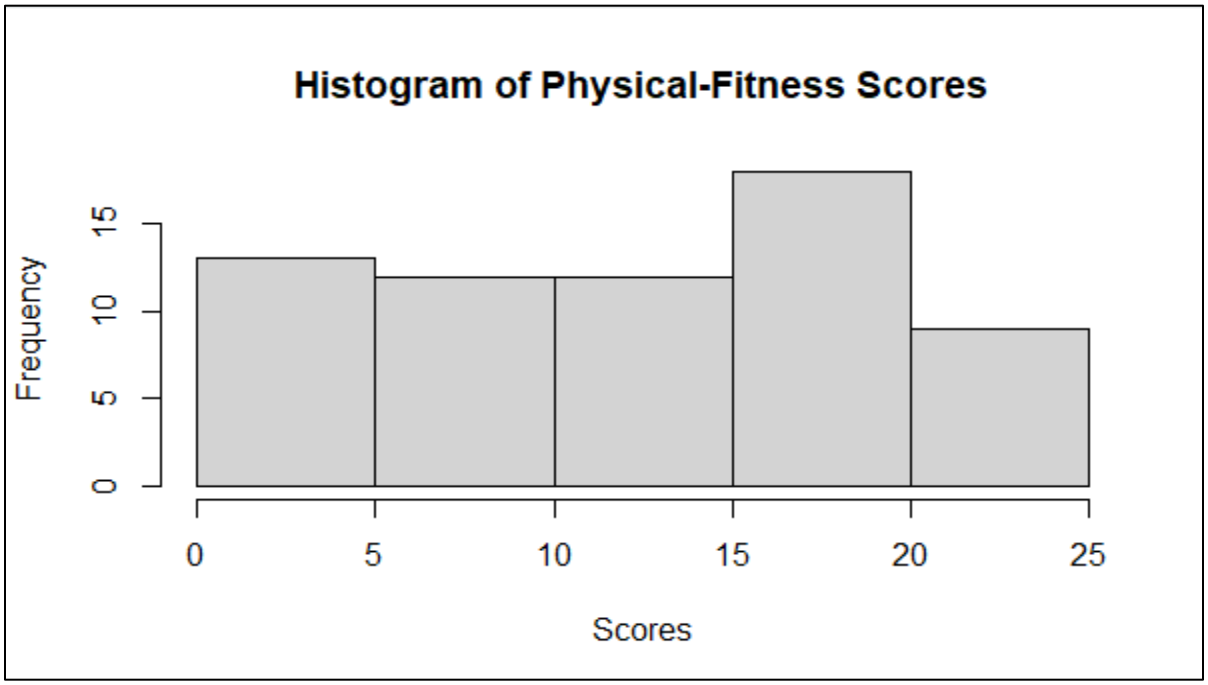
Class : SY-MCA Shift / Div : S3/B Roll Number : 52147

Name : Nisha Harish Parekh Assignment No : 5 Date of Implementation : 8/11/2023

****

Output :

```
> # Create a list of the physical-fitness scores for each group
> scores_a <- c(12, 22, 6, 9, 2, 9, 5, 9, 3, 5, 16, 1, 22, 18)
> scores_b <- c(6, 12, 21, 23, 9, 10, 24, 21, 17, 11, 18, 19, 17, 5)
> scores_c <- c(14, 16, 19, 19, 18, 3, 4, 21, 16, 20, 15, 14, 17, 4)
> scores_d <- c(5, 22, 12, 15, 18, 20, 8, 10, 13, 20, 6, 9, 2, 17)
> scores_e <- c(15, 9, 4, 15, 14, 19, 3, 24)
> all_scores <- c(scores_a, scores_b, scores_c, scores_d, scores_e)
> mean_score <- mean(all_scores)
> cat("Mean:", mean_score, "\n")
Mean: 13
>
> median_score <- median(all_scores)
> cat("Median:", median_score, "\n")
Median: 14
>
> mode_score <- as.numeric(names(sort(table(all_scores), decreasing = TRUE)[1]))
> cat("Mode:", mode_score, "\n")
Mode: 9
>
> range_score <- range(all_scores)
> cat("Range:", range_score[2] - range_score[1], "\n")
Range: 23
>
> q1 <- quantile(all_scores, 0.25)
> q3 <- quantile(all_scores, 0.75)
> iqr_score <- q3 - q1
> cat("IQR:", iqr_score, "\n")
IQR: 10.75
>
> hist(all_scores, main="Histogram of Physical-Fitness Scores", xlab="Scores")
>
> shapiro_test <- shapiro.test(all_scores)
> cat("Normal distribution:", shapiro_test$p.value, "\n")
Normal distribution: 0.008213226
>
```



Progressive Education Society's

Modern College of Engineering, Pune

MCA Department

A.Y.2023-24

(410908) Data Science Laboratory

****

Class : SY-MCA Shift / Div : S3/B Roll Number : 52147

Name : Nisha Harish Parekh Assignment No : 5 Date of Implementation : 8/11/2023

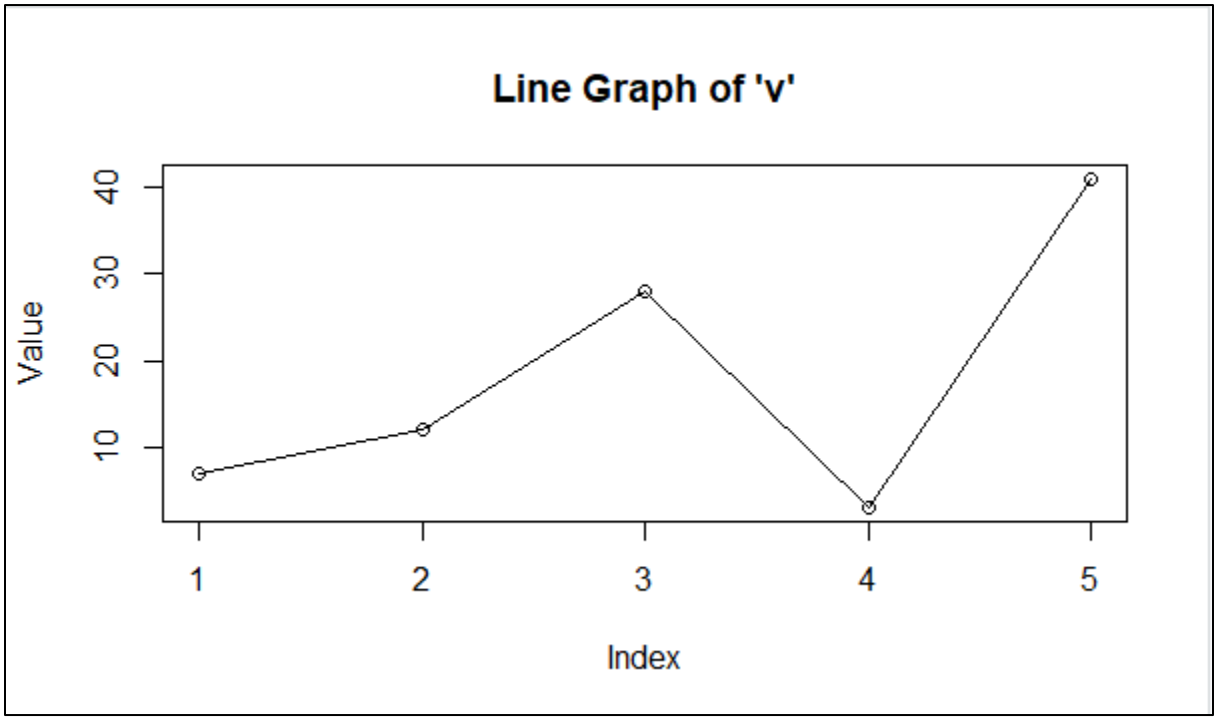
****

Q2) Plot the line graph using v<- c(7,12,28,3,41) and save the plot.

Program :

```
v <- c(7, 12, 28, 3, 41)
plot(v, type = "o", main = "Line Graph of 'v'", xlab = "Index", ylab = "Value")
png("line_plot.png", width = 800, height = 400)
plot(v, type = "o", main = "Line Graph of 'v'", xlab = "Index", ylab = "Value")
dev.off()
```

Output :



Progressive Education Society's

Modern College of Engineering, Pune

MCA Department

A.Y.2023-24

(410908) Data Science Laboratory

****

Class : SY-MCA Shift / Div : S3/B Roll Number : 52147

Name : Nisha Harish Parekh Assignment No : 5 Date of Implementation : 8/11/2023

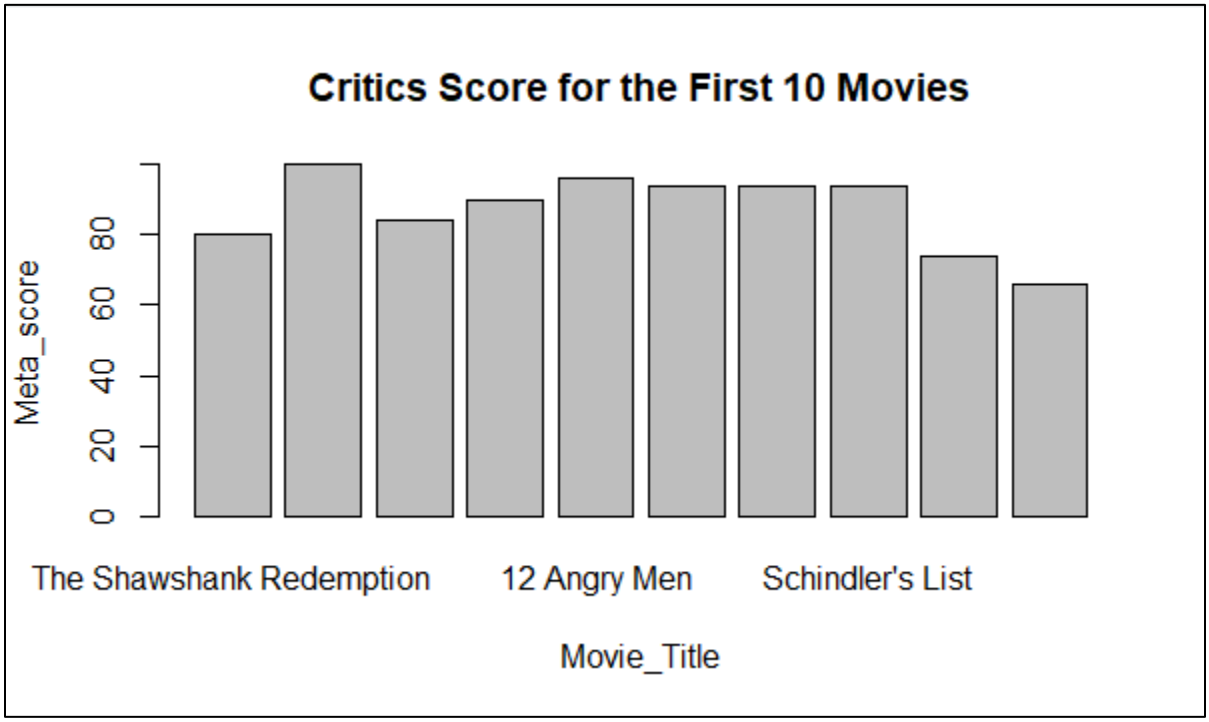
****

Q3) Read the file moviesData.csv create a bar chart of critics_score for the first 10 movies.Save the plot.

Program :

```
library(readr)
movies_data <- read.csv("E:\\MCA\\SY (1)\\Data Science\\Practical\\movies_data.csv")
top_10_critics_scores <- movies_data[1:10, "Meta_score"]
barplot(top_10_critics_scores, names.arg = movies_data[1:10, "Series_Title"],
        main = "Critics Score for the First 10 Movies", xlab = "Movie_Title", ylab = "Meta_score")
png("critics_score_bar_chart.png", width = 800, height = 400)
dev.off()
```

Output :



Progressive Education Society's

Modern College of Engineering, Pune

MCA Department

A.Y.2023-24

(410908) Data Science Laboratory

****

Class : SY-MCA Shift / Div : S3/B Roll Number : 52147

Name : Nisha Harish Parekh Assignment No : 5 Date of Implementation : 8/11/2023

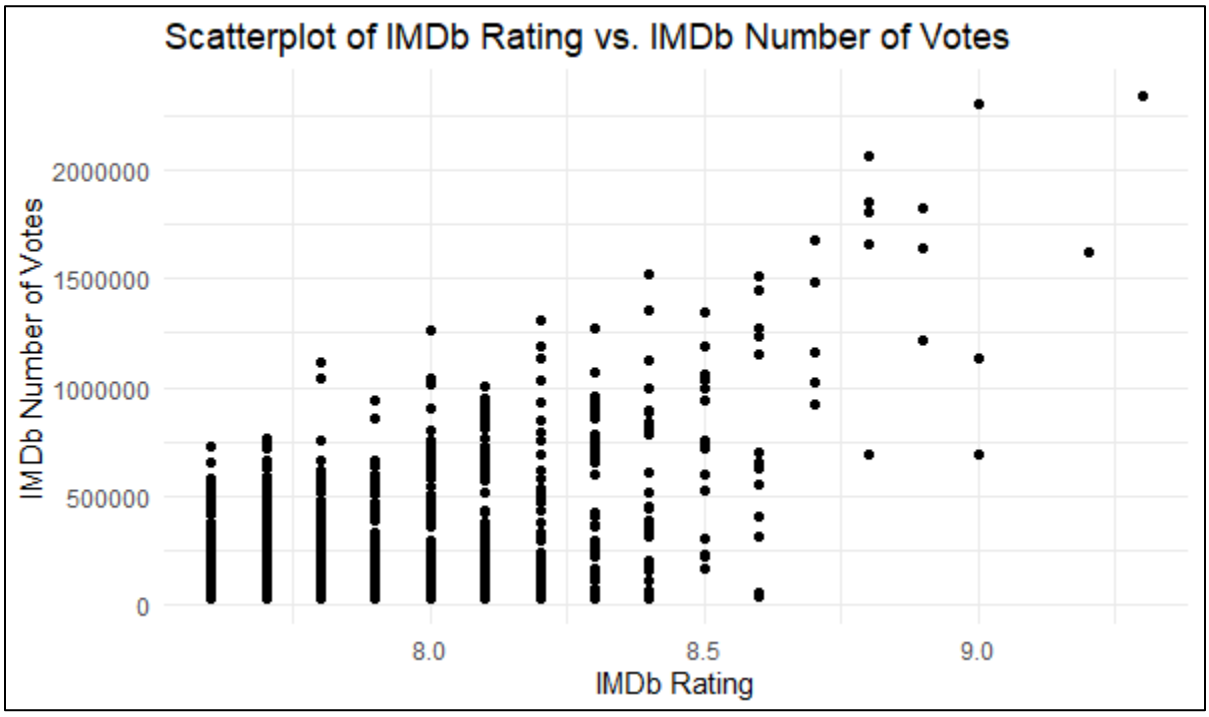
****

Q4) Create a scatterplot of imdb_rating and imdb_num_votes to see their relation and save the plot.

Program :

```
library(readr)
library(ggplot2)
movies_data <- read.csv("E:\\MCA\\SY (1)\\Data Science\\Practical\\movies_data.csv")
ggplot(movies_data, aes(x = IMDB_Rating, y = No_of_Votes)) +
  geom_point() +
  labs(title = "Scatterplot of IMDb Rating vs. IMDb Number of Votes",
       x = "IMDb Rating", y = "IMDb Number of Votes") +
  theme_minimal()
ggsave("imdb_rating_vs_votes_scatterplot.png", width = 800, height = 400)
```

Output :



Progressive Education Society's

Modern College of Engineering, Pune

MCA Department

A.Y.2023-24

(410908) Data Science Laboratory

****

Class : SY-MCA Shift / Div : S3/B Roll Number : 52147

Name : Nisha Harish Parekh Assignment No : 5 Date of Implementation : 8/11/2023

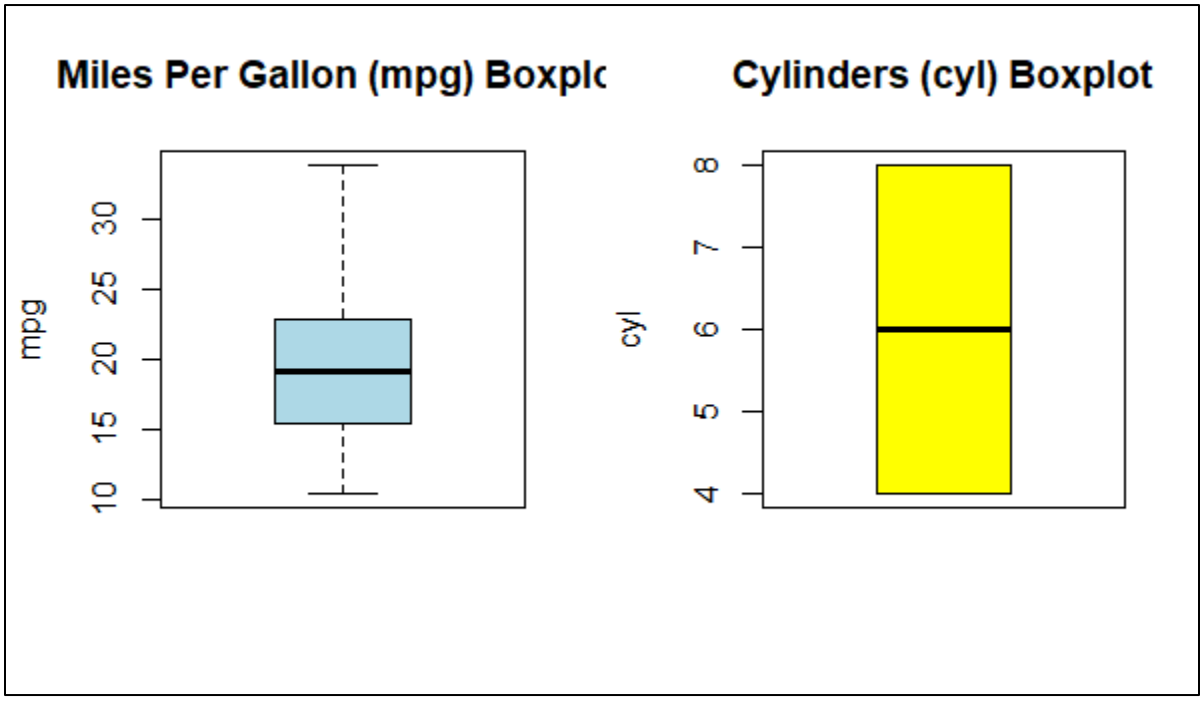
****

Q5) Use the data set “mtcars”and create boxplot for “mpg” and “cyl” columns.

Program :

```
data(mtcars)
par(mfrow=c(1,2))
boxplot(mtcars$mpg, main = "Miles Per Gallon (mpg) Boxplot",ylab = "mpg", col = "lightblue")
boxplot(mtcars$cyl, main = "Cylinders (cyl) Boxplot",ylab = "cyl", col = "yellow")
par(mfrow=c(1,1))
```

Output :



Progressive Education Society's

Modern College of Engineering, Pune

MCA Department

A.Y.2023-24

(410908) Data Science Laboratory

****

Class : SY-MCA Shift / Div : S3/B Roll Number : 52147

Name : Nisha Harish Parekh Assignment No : 5 Date of Implementation : 8/11/2023

****

Q6) Read the file movies Data.csv, create a histogram of the object named imdb_num_votes in this file. Save the plot.

Program :

```
movies_data <- read.csv("E:\\MCA\\SY (1)\\Data Science\\Practical\\movies_data.csv")
movies_data$IMDB_Rating <- as.numeric(movies_data$IMDB_Rating)
histogram <- hist(movies_data$IMDB_Rating,
                  main = "Histogram of IMDb Number of Votes",
                  xlab = "IMDb Number of Votes",
                  ylab = "Frequency")
png("imdb_num_votes_histogram.png", width = 800, height = 400)
plot(histogram, col = "lightblue", border = "black")
dev.off()
```

Output :