NAME : DEEBAK KUMAR K

REG NO : 192324064

DW&DM

LAB PROGRAMS - 2

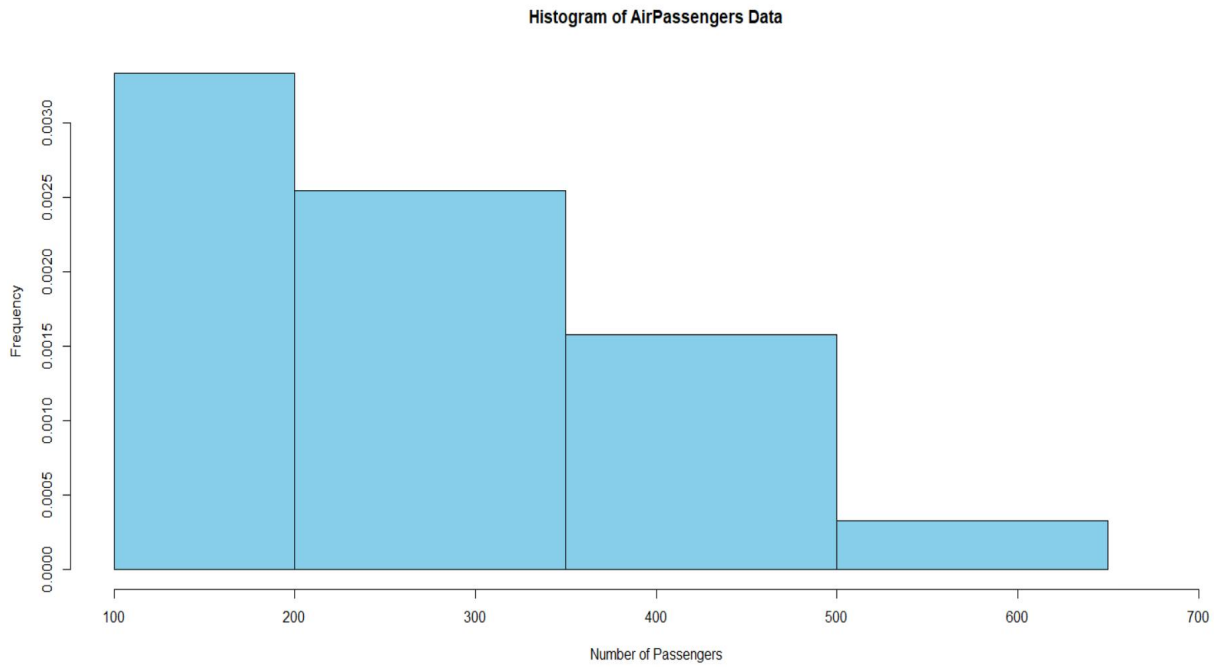
LAB PROGRAMS - 2

12 .Make a histogram for the “AirPassengers “dataset, start at 100 on the x-axis, and from values 200 to 700, make the bins 150 wide

CODE :

```
# Load the dataset
data("AirPassengers")

# Convert time series to a numeric vector
passenger_data <- as.numeric(AirPassengers)

# Define break points for the histogram
breaks <- c(100, seq(200, 700, by = 150))

# Create the histogram
hist(passenger_data, breaks = breaks, col = "skyblue", border = "black",
    main = "Histogram of AirPassengers Data",
    xlab = "Number of Passengers", ylab = "Frequency",
    xlim = c(100, 700))
```

OUTPUT :

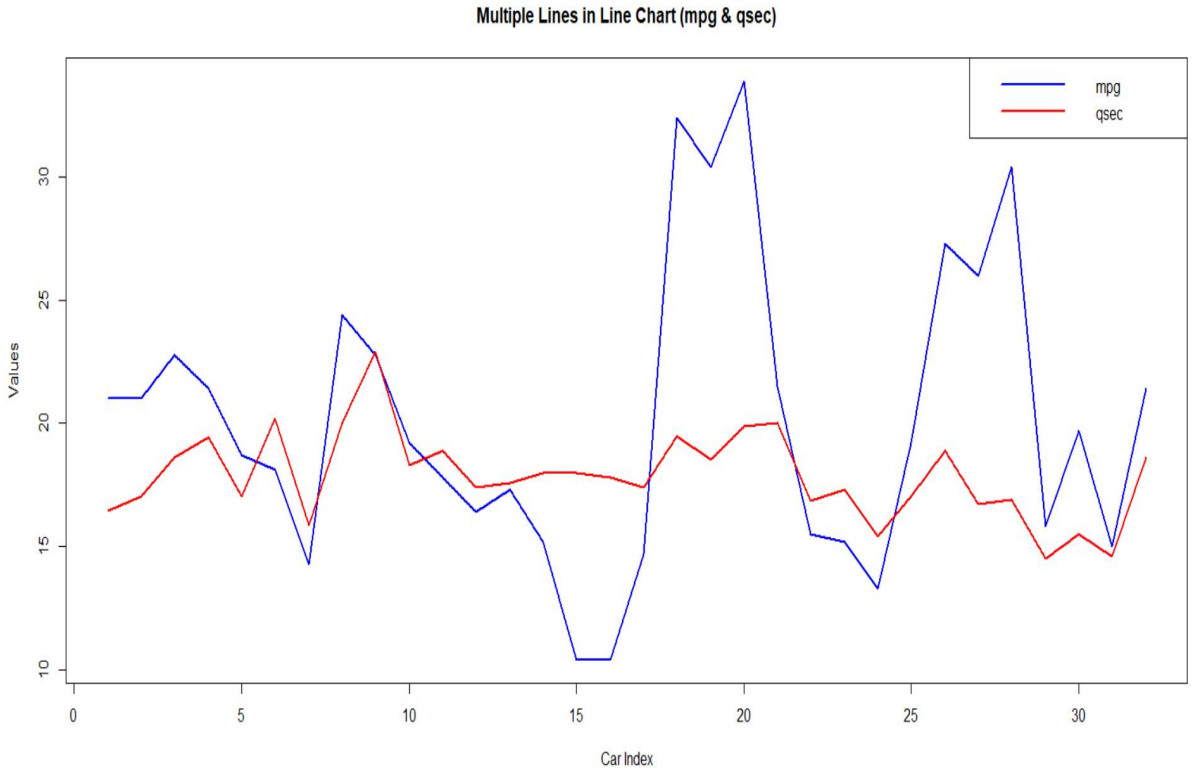


13. Obtain Multiple Lines in Line Chart using a single Plot Function in R.Use attributes“mpg”and“qsec”of the dataset “mtcars”

CODE :

```
# Load the mtcars dataset
data(mtcars)

# Define x-axis (Car Index)
x_values <- 1:nrow(mtcars)  # Using row index as x-axis

# Plot 'mpg' as the first line
plot(x_values, mtcars$mpg, type = "l", col = "blue", lwd = 2,
    xlab = "Car Index", ylab = "Values",
    main = "Multiple Lines in Line Chart (mpg & qsec)")

# Add 'qsec' as the second line
lines(x_values, mtcars$qsec, type = "l", col = "red", lwd = 2)

# Add a legend
legend("topright", legend = c("mpg", "qsec"), col = c("blue", "red"), lwd = 2)
```

OUTPUT :



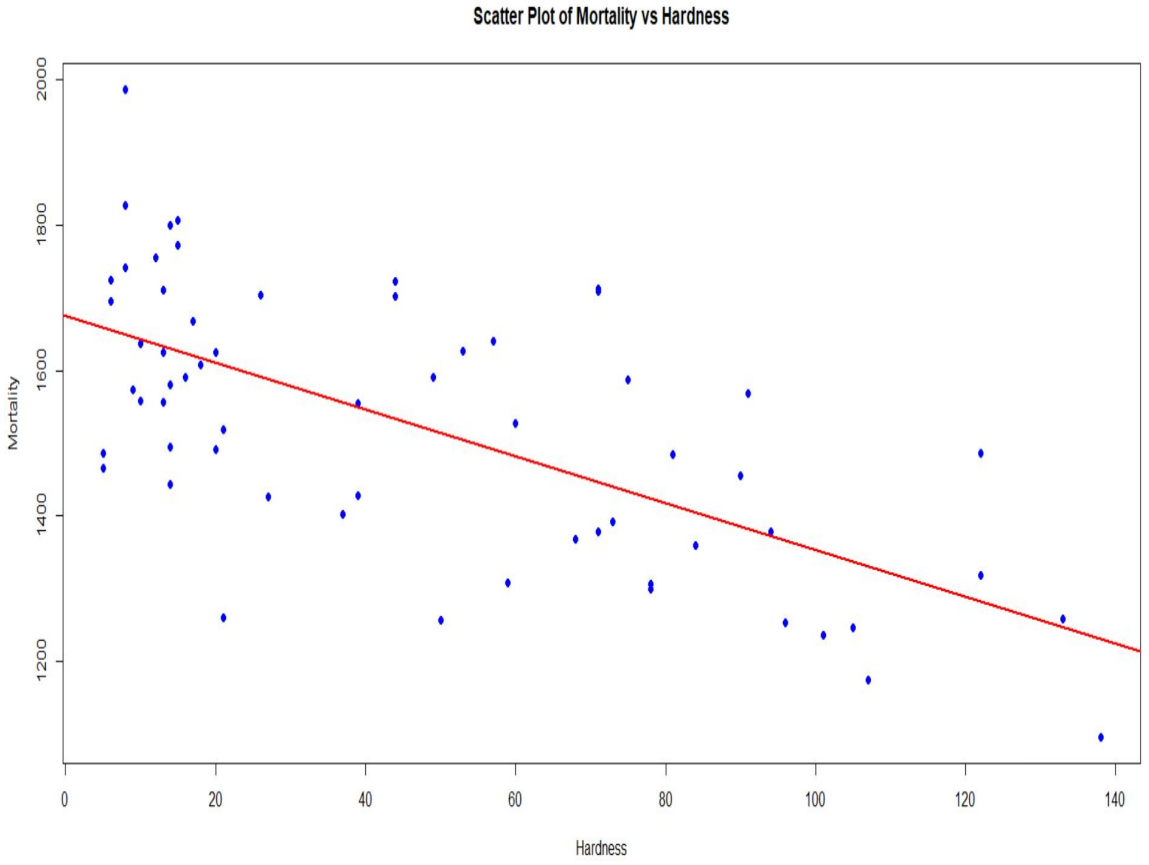
14. Download the Dataset "water" From R dataset Link.Find out whether there is a linear relation between attributes"mortality" and"hardness" by plot function.Fit the Data into the Linear Regression model.Predict the mortality for the hardness=88.

CODE :

```
# Install and load the required package to access the "water" dataset
install.packages("HSAUR")  # Install only if not already installed
library(HSAUR)

# Load the dataset
data("water")

# View the structure of the dataset
str(water)

# Scatter plot to check the linear relationship
plot(water$hardness, water$mortality, col = "blue", pch = 16,
    xlab = "Hardness", ylab = "Mortality",
    main = "Scatter Plot of Mortality vs Hardness")

# Fit a linear regression model
model <- lm(mortality ~ hardness, data = water)

# Add regression line to the plot
abline(model, col = "red", lwd = 2)

# Predict mortality for hardness = 88
new_data <- data.frame(hardness = 88)
predicted_mortality <- predict(model, new_data)
print(paste("Predicted Mortality for Hardness 88:", round(predicted_mortality, 2)))
```

OUTPUT :

```
> # View the structure of the dataset
> str(water)
'data.frame':   61 obs. of  4 variables:
 $ location : Factor w/ 2 levels "North","South": 2 1 2 1 1 1 1 2 1 2 ...
 $ town     : chr  "Bath" "Birkenhead" "Birmingham" "Blackburn" ...
 $ mortality: int  1247 1668 1466 1800 1609 1558 1807 1299 1637 1359 ...
 $ hardness : int  105 17 5 14 18 10 15 78 10 84 ...
>
> # Scatter plot to check the linear relationship
> plot(water$hardness, water$mortality, col = "blue", pch = 16,
+      xlab = "Hardness", ylab = "Mortality",
+      main = "Scatter Plot of Mortality vs Hardness")
>
> # Fit a linear regression model
> model <- lm(mortality ~ hardness, data = water)
>
> # Add regression line to the plot
> abline(model, col = "red", lwd = 2)
>
> # Predict mortality for hardness = 88
> new_data <- data.frame(hardness = 88)
> predicted_mortality <- predict(model, new_data)
> print(paste("Predicted Mortality for Hardness 88:", round(predicted_mortality$
[1] "Predicted Mortality for Hardness 88: 1392.46"
>
```

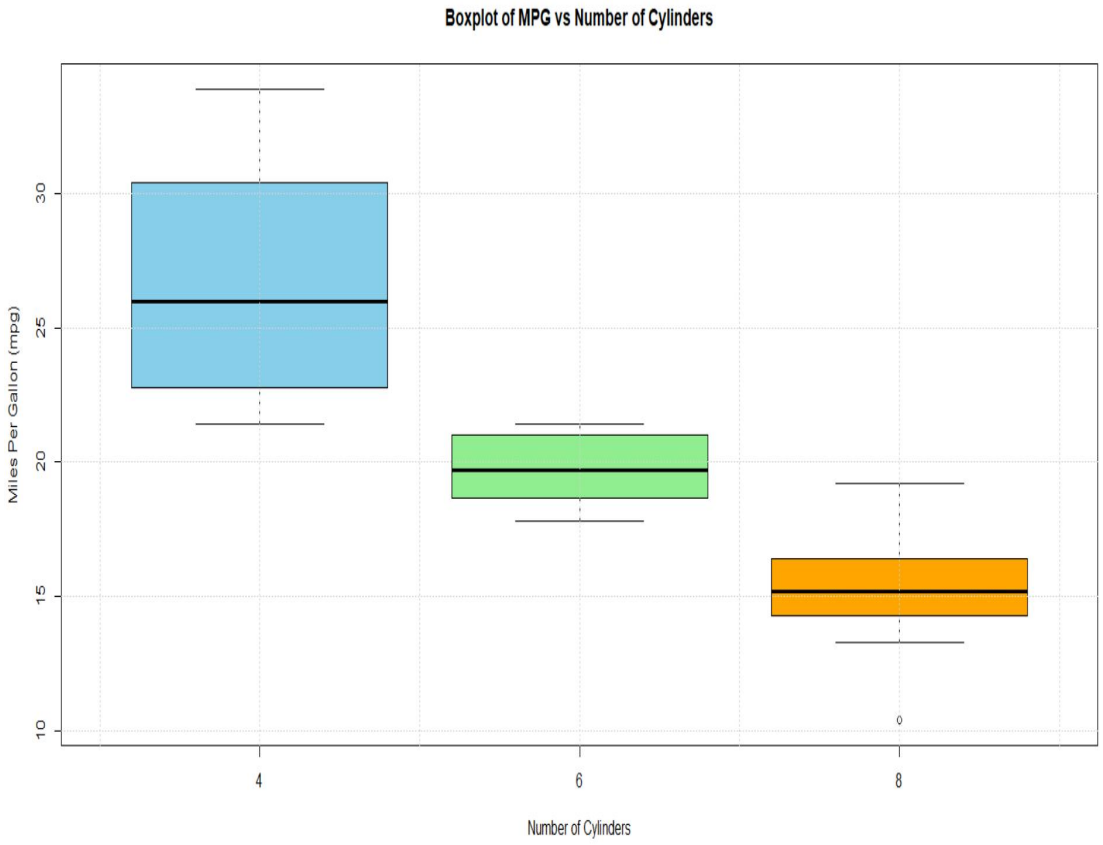


15. Create a Boxplot graph for the relation between "mpg"(miles per galloon) and "cyl"(number of Cylinders) for the dataset "mtcars" available in R Environment.

CODE :

```
# Load the dataset
data(mtcars)

# Convert 'cyl' to a factor (categorical variable)
mtcars$cyl <- as.factor(mtcars$cyl)

# Create a boxplot
boxplot(mpg ~ cyl, data = mtcars,
        col = c("skyblue", "lightgreen", "orange"),
        xlab = "Number of Cylinders",
        ylab = "Miles Per Gallon (mpg)",
        main = "Boxplot of MPG vs Number of Cylinders")

# Add grid lines for better readability
grid()
```

OUTPUT :

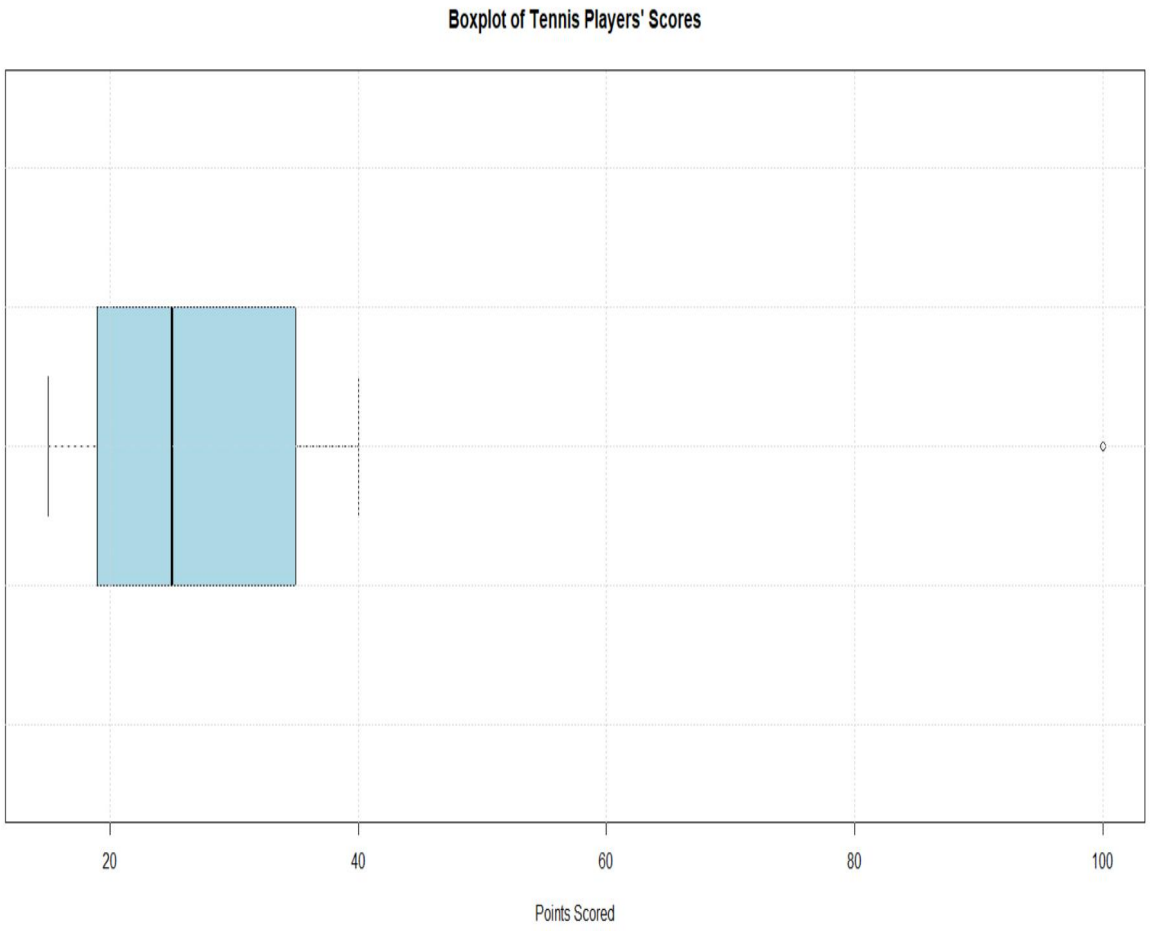


16. Assume the Tennis coach wants to determine if any of his team players are scoring outliers. To visualize the distribution of points scored by his players, then how can he decide to develop the box plot? Give suitable example using Boxplot visualization technique.

CODE :

```
# Simulated dataset: Points scored by Tennis players
player_scores <- c(15, 18, 21, 19, 25, 30, 35, 40, 100)  # 100 is an outlier

# Create a Boxplot
boxplot(player_scores, col = "lightblue", horizontal = TRUE,
        main = "Boxplot of Tennis Players' Scores",
        xlab = "Points Scored")

# Add grid lines for clarity
grid()
```

OUTPUT :



17. Implement using R language in which age group of people are affected byblood pressure based on the diabetes dataset show it using scatterplot and bar chart (that is BloodPressure vs Age using dataset “diabetes.csv”)

CODE :

```
# Load necessary packages
install.packages("ggplot2")  # Install if not already installed
library(ggplot2)

# Read the diabetes dataset
diabetes_data <- read.csv("diabetes.csv")

# View dataset structure
str(diabetes_data)

# Scatter plot: Blood Pressure vs. Age
ggplot(diabetes_data, aes(x = Age, y = BloodPressure)) +
  geom_point(color = "blue") +
  ggtitle("Scatter Plot of Blood Pressure vs Age") +
  xlab("Age") +
  ylab("Blood Pressure") +
  theme_minimal()

# Create age groups (e.g., 20-30, 30-40, etc.)
diabetes_data$AgeGroup <- cut(diabetes_data$Age,
                              breaks = seq(20, 80, by = 10),
                              labels = c("20-30", "30-40", "40-50", "50-60", "60-70", "70-80"))

# Bar chart: Average Blood Pressure by Age Group
ggplot(diabetes_data, aes(x = AgeGroup, y = BloodPressure, fill = AgeGroup)) +
  geom_bar(stat = "summary", fun = "mean") +
  ggtitle("Average Blood Pressure by Age Group") +
  xlab("Age Group") +
  ylab("Average Blood Pressure") +
  theme_minimal()
```

OUTPUT :