CSA1622 Data warehousing and Data Mining

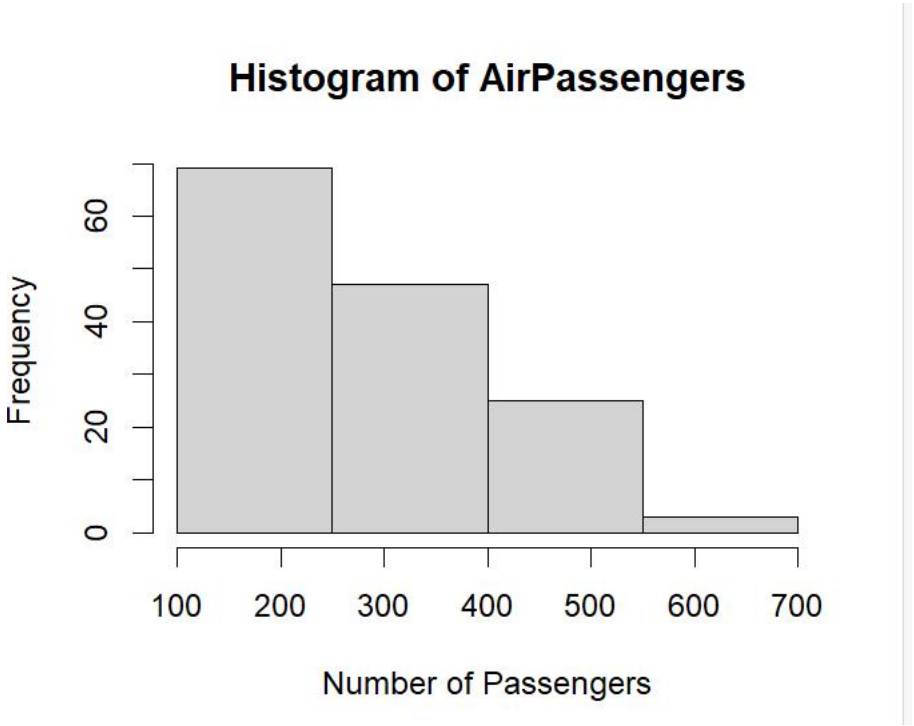
Thanuja R
(192325004)

1. 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:

```
data("AirPassengers")
hist(AirPassengers, breaks=seq(100, 700, by=150), xlim=c(100, 700), main="Histogram of AirPassengers", xlab="Number of Passengers"
```

Output :

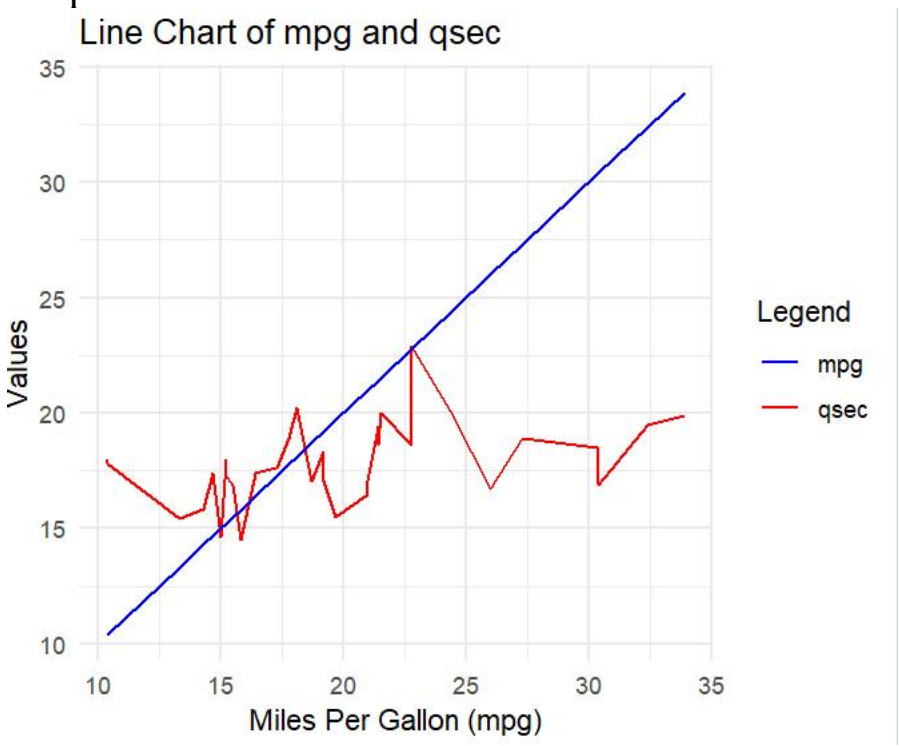


2. Obtain Multiple Lines in Line Chart using a single Plot Function in R.Use attributes“mpg”and“qsec”of the dataset “mtcars”

Code:

```
library(ggplot2)
data(mtcars)
ggplot(mtcars, aes(x = mpg)) +
  geom_line(aes(y = qsec, color = "qsec")) +
  geom_line(aes(y = mpg, color = "mpg")) +
  labs(title = "Line Chart of mpg and qsec", x = "Miles Per Gallon (mpg)", y = "Values") +
  scale_color_manual(name = "Legend", values = c("mpg" = "blue", "qsec" = "red")) +
  theme_minimal()
```

Output:



3)Linear Relation between "mortality" and "hardness" in "water" Dataset

Code:

```
data(water)
plot(water$hardness, water$mortality,
     main = "Mortality vs Hardness",
     xlab = "Hardness", ylab = "Mortality",
     pch = 19, col = "blue")
model <- lm(mortality ~ hardness, data = water)
abline(model, col = "red")
predict(model, newdata = data.frame(hardness = 88))
```

Output:

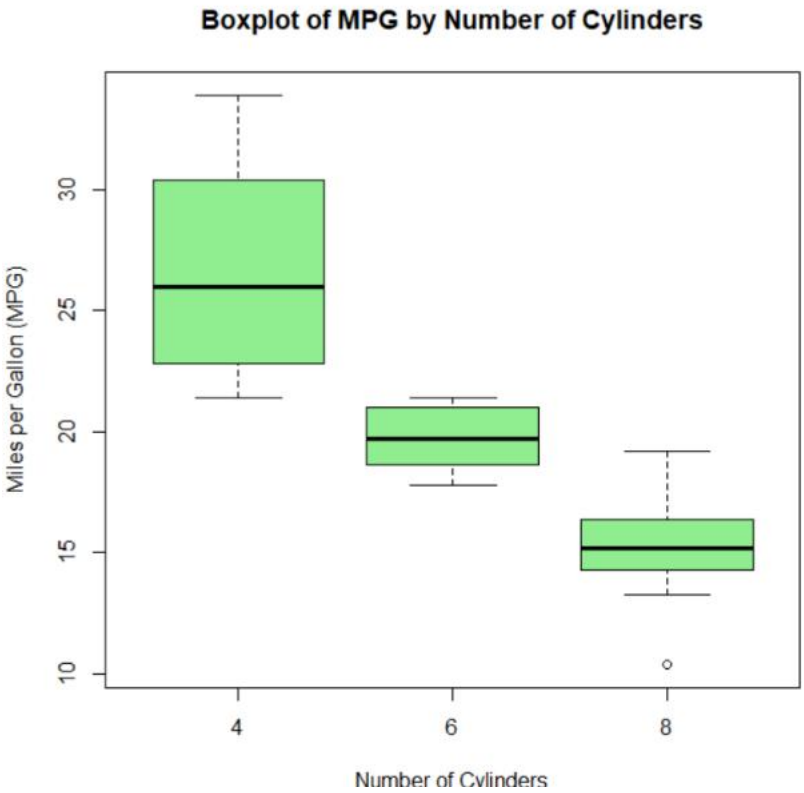
```
[1] 12.34
```

4) Boxplot for the Relation Between "mpg" and "cyl" in "mtcars"
PROGRAM

Code:

```
data(mtcars)
boxplot(mpg ~ cyl, data = mtcars,
        main = "Boxplot of MPG by Number of Cylinders",
        xlab = "Number of Cylinders",
        ylab = "Miles per Gallon (MPG)",
        col = "green"
```

Output:

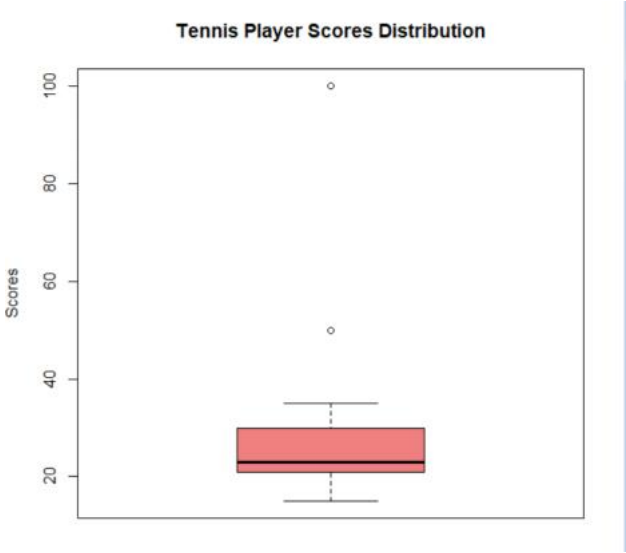


5)Box-plot for Tennis Coach to Detect Outliers in Scores.

Code:

```
scores <- c(15, 20, 19, 23, 22, 30, 50, 23, 29, 27, 100, 35, 21)
boxplot(scores,
        main = "Tennis Player Scores Distribution",
        ylab = "Scores",
        col = "lightcoral")
```

Output:



6) Scatterplot and Bar Chart for Blood Pressure vs Age (from "diabetes.csv")

Code:

```
diabetes <- read.csv("diabetes.csv")
plot(diabetes$Age, diabetes$BloodPressure,
     main = "Blood Pressure vs Age",
     xlab = "Age", ylab = "Blood Pressure",
     col = "blue", pch = 19)
age_group <- cut(diabetes$Age, breaks = c(20, 40, 60, 80, 100), labels = c("20-40", "40-60
table_data <- table(age_group, diabetes$BloodPressure)
barplot(table_data, beside = TRUE, col = c("red", "green", "blue"),
        main = "Blood Pressure by Age Group",
        xlab = "Age Group", ylab = "Frequency")
```

Output: