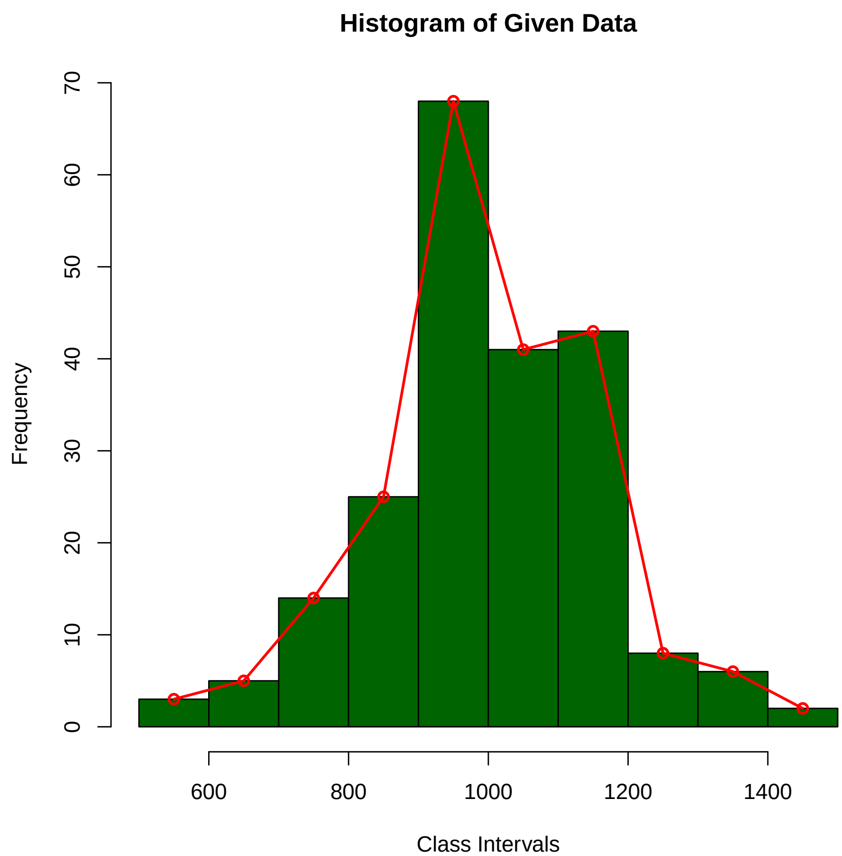
Registration Number: 24MDT0184
Name: Tufan Kundu
Slot: L23+L24
Course Code: PMDS503P
Course Title: Statistical Inference Lab
DA 1

Q1-Draw the Histogram and Frequency polygon for the following data:

C.I.	Frequency
500-600	3
600-700	5
700-800	14
800-900	25
900-1000	68
1000-1100	41
1100-1200	43
1200-1300	8
1300-1400	6
1400-1500	2

```r
# Defining the upper bound, lower bound and frequency
lower_bounds<- c(500,600,700,800,900,1000,1100,1200,1300,1400)
upper_bounds<- c(600,700,800,900,1000,1100,1200,1300,1400,1500)
frequencies <- c(3, 5, 14, 25, 68, 41, 43, 8, 6, 2)

# calculating the midpoints of the class intervals
midpoints <-  (lower_bounds+upper_bounds)/ 2

# Plotting the histogram
hist(rep(midpoints, frequencies),col = "darkgreen",
     main = "Histogram of Given Data", xlab = "Class Intervals", ylab = "Frequency",
     border = "black")

# Plotting frequency polygon on top of the histogram
lines(midpoints, frequencies, type = "o", col = "red", lwd = 2)
```

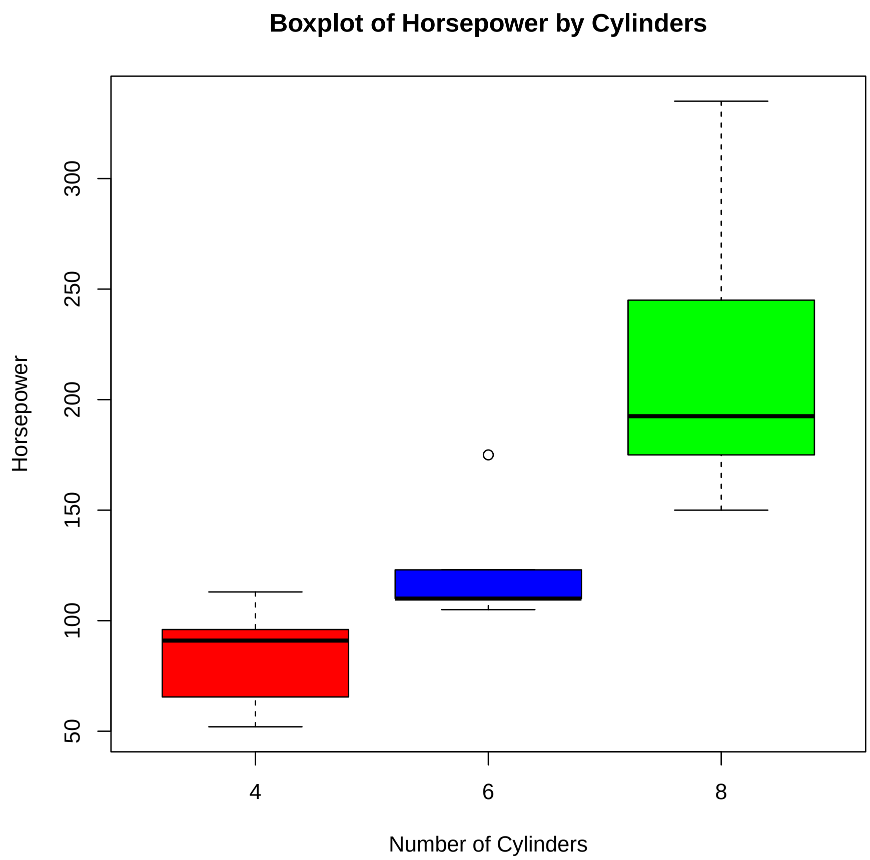


Q2.Import a Multivariable dataset from the datasets/MASS package and plot the following:

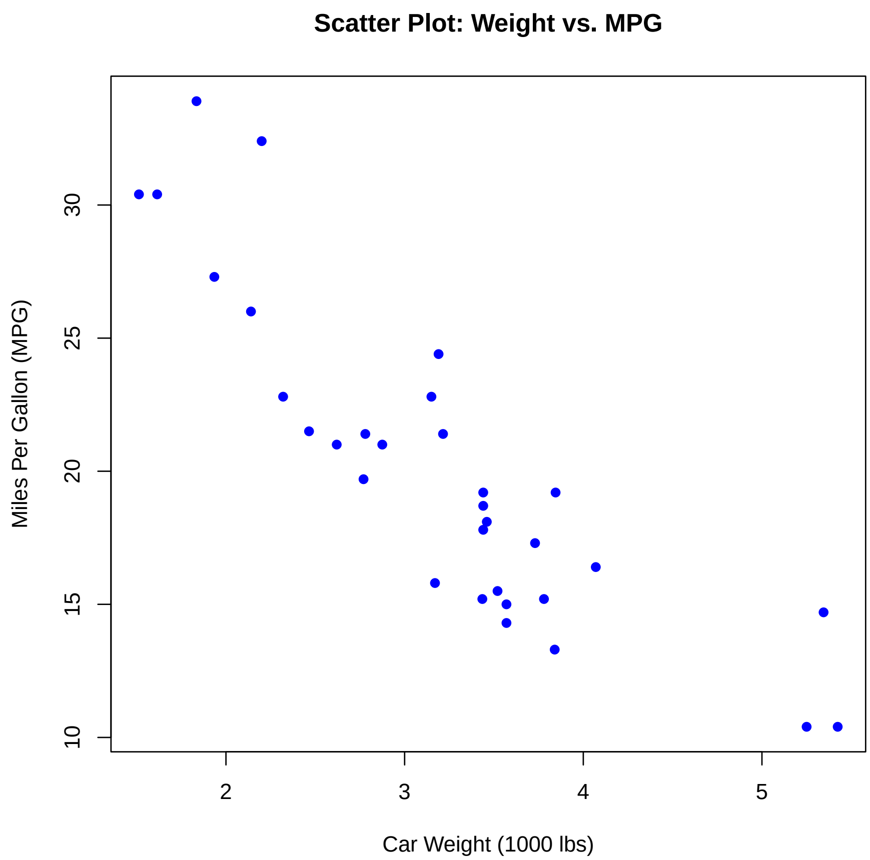
(i) Display the number of variables in the dataset
(ii) Draw a box plot for any two variables.
(iii) Scatterplot for any two variables
(iv) Multiple bar diagram(with different color)
(v) Write your observations.
Note: Add the labels for X-Axis, Y-Axis and Title of the diagram.

```
library(datasets)

# Loading the mtcars dataset
data(mtcars)

# Display number of variables and variable names
num_variables <- ncol(mtcars)  # Number of variables
```

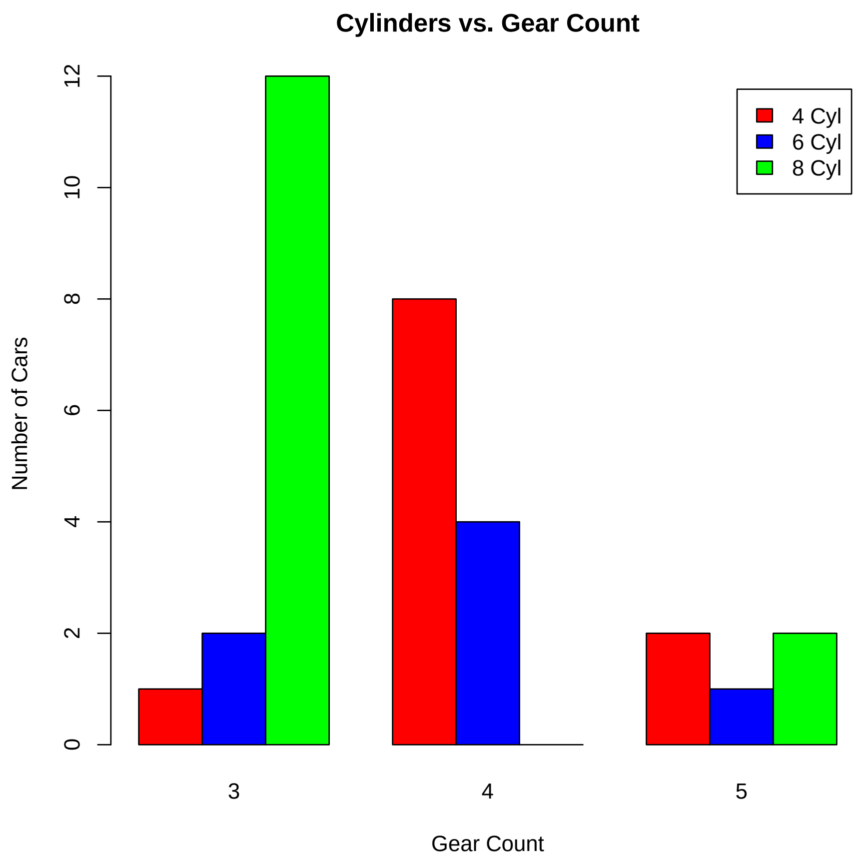
```
variable_names <- colnames(mtcars)  # Display variable names

# ---------------------------------------
# (i) Display the number of variables in the dataset
# ---------------------------------------
print(paste("Number of variables:", num_variables))

## [1] "Number of variables: 11"

print("Variable names:")

## [1] "Variable names:"

print(variable_names)

##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"

# display the first few rows of the dataset
print(head(mtcars))

##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1

# ---------------------------------------
# (ii) Draw a box plot for any two variables.
# Boxplot: Horsepower vs. Cylinders
# ---------------------------------------
boxplot(hp ~ cyl, data = mtcars,
        col = c("red", "blue", "green"),
        main = "Boxplot of Horsepower by Cylinders",
        xlab = "Number of Cylinders",
        ylab = "Horsepower")
```




```r
# ----------------------------------------
# (iii) Scatterplot for any two variables
# Scatter Plot: Weight vs. Miles Per Gallon
# ----------------------------------------
plot(mtcars$wt, mtcars$mpg,
     col = "blue", pch = 16,
     main = "Scatter Plot: Weight vs. MPG",
     xlab = "Car Weight (1000 lbs)",
     ylab = "Miles Per Gallon (MPG)")
```




```
# ----------------------------------------------
# (iv) Multiple bar diagram: Cylinders vs. Gears
# ----------------------------------------------
barplot(table(mtcars$cyl, mtcars$gear), beside = TRUE,
        col = c("red", "blue", "green"),
        main = "Cylinders vs. Gear Count",
        xlab = "Gear Count",
        ylab = "Number of Cars",
        legend = c("4 Cyl", "6 Cyl", "8 Cyl"))
```




```
#-----------------------------------------------
# (v) Write your observations
#-----------------------------------------------
print("Observations for the mtcars dataset:")

## [1] "Observations for the mtcars dataset:"

print("1. The dataset contains 11 variables representing different characteristics of 32 car

## [1] "1. The dataset contains 11 variables representing different characteristics of 32 ca

print("2. Scatter Plot (Weight vs. MPG): As car weight increases, MPG decreases.")

## [1] "2. Scatter Plot (Weight vs. MPG): As car weight increases, MPG decreases."

print("3. Box Plot (Horsepower vs. Cylinders): 8-cylinder cars have the highest horsepower.'

## [1] "3. Box Plot (Horsepower vs. Cylinders): 8-cylinder cars have the highest horsepower.
```

```r
print("4. Multiple Bar Diagram (Cylinders vs. Gear Count):")

## [1] "4. Multiple Bar Diagram (Cylinders vs. Gear Count):"

print("   - Most 8-cylinder cars have 3 gears.")

## [1] "   - Most 8-cylinder cars have 3 gears."

print("   - 4-cylinder cars are more common with 4 or 5 gears.")

## [1] "   - 4-cylinder cars are more common with 4 or 5 gears."

print("   - 6-cylinder cars are mainly associated with 4 gears.")

## [1] "   - 6-cylinder cars are mainly associated with 4 gears."

print("5. Heavier cars tend to be less fuel-efficient.")

## [1] "5. Heavier cars tend to be less fuel-efficient."

print("6. Higher cylinder count is associated with more power but lower fuel efficiency.")

## [1] "6. Higher cylinder count is associated with more power but lower fuel efficiency."
```

Q3. From the data set given below, obtain the descriptive statistics for each variable and write your observations on the performance of students.

```r
# loading the data from csv file
df <- read.csv("D:\\study material\\VIT_Data_Science\\Winter_Sem\\Statistical_Inference_Lab\

# displaying the first few rows of the data
head(df)

##   DA Quiz.1 Quiz.2 CAT.1 CAT.2 FAT
## 1 10     14     15    25    16  16
## 2 10     14     15    11    26  16
## 3 10     14     14    11    19  16
## 4 10     12      8    16    20  16
## 5 10     14     12    25    25  16
## 6 10     15     15    11     7  16

# User defined function to calculate the descriptive statistics

# Minimum
min_value <- function(x)
{
  return(sort(x)[1])
}
```

```
# Maximum
max_value <- function(x)
{
  return(sort(x, decreasing = TRUE)[1])
}

# Mean
mean_value <- function(x)
{
  return(sum(x)/length(x))
}

# variance
variance_value <- function(x)
{
  return(sum((x-mean_value(x))^2)/(length(x)-1))
}

# Standard Deviation
sd_value <- function(x)
{
  return(sqrt(variance_value(x)))
}

# Quartiles
quartiles<- function(x)
{
  x<- sort(x)   # sorting the data
  n<- length(x) # length of the data

  # Finding the position of the quartiles
  q1_pos <- (n+1)/4
  q2_pos <- (n+1)/2
  q3_pos <- 3*(n+1)/4

  # Function to interpolate for non-integer positions
  interpolate <- function(pos)
  {
    lower <- floor(pos)
    upper <- ceiling(pos)
    if(lower==upper) # i.e if the position is an integer
    {
      return (x[lower])
    }
    else
```

```
    {
      return (x[lower]+(pos-lower)*(x[upper]-x[lower]))
    }
  }

  # computing the quartile values
  q1<- interpolate(q1_pos)
  q2<- interpolate(q2_pos)
  q3<- interpolate(q3_pos)

  return (c(Q1=q1,Median=q2,Q3=q3))
}

# Third moment
third_moment<- function(x)
{
  return(sum((x - mean_value(x))^3) / length(x))
}

# Fourth moment
fourth_moment<- function(x)
{
  return(sum((x - mean_value(x))^4) / length(x))
}

# Beta1
beta1<- function(x)
{
  return (third_moment(x)^2/(variance_value(x)^3))
}

# Beta2
beta2<- function(x)
{
  return (fourth_moment(x)/(variance_value(x)^2))
}

#Gamma1
gamma1<- function(x)
{
  return (sqrt(beta1(x)))
}

#Gamma2
gamma2<- function(x)
```

```r
{
  return (beta2(x)-3)
}

# Applying function to each numerical column
results_df <- data.frame(Measure = c("Min", "Max", "Mean", "Variance", "SD",
                                     "Q1", "Q2 (Median)", "Q3", "Third Moment",
                                     "Fourth Moment", "Beta1", "Beta2", "Gamma1", "Gamma2"))

# Looping through each column and calculating statistics
for (col in colnames(df)) {
  data_col <- df[[col]]

  # Compute the statistics
  stats <- round(c(min_value(data_col),
             max_value(data_col),
             mean_value(data_col),
             variance_value(data_col),
             sd_value(data_col),
             quartiles(data_col)["Q1"],
             quartiles(data_col)["Median"],
             quartiles(data_col)["Q3"],
             third_moment(data_col),
             fourth_moment(data_col),
             beta1(data_col),
             beta2(data_col),
             gamma1(data_col),
             gamma2(data_col)),2)

  # Append results as a new column
  results_df[[col]] <- stats
}

# Print the final table
print(results_df)
```

```
##          Measure  DA Quiz.1 Quiz.2    CAT.1    CAT.2 FAT
## 1            Min  10  10.00   0.00    10.00     7.00  16
## 2            Max  10  20.00  18.00    50.00    48.00  16
## 3           Mean  10  14.12  11.32    28.32    25.48  16
## 4       Variance   0   5.13   8.04    97.19    99.37   0
## 5             SD   0   2.27   2.84     9.86     9.97   0
## 6             Q1  10  12.00  10.00    20.50    18.50  16
## 7    Q2 (Median)  10  14.00  10.00    26.00    25.00  16
## 8             Q3  10  16.00  14.00    36.00    28.00  16
```

```
## 9   Third Moment   0   6.96 -10.60   136.93   669.08   0
## 10 Fourth Moment   0  88.91 351.66 20808.38 27141.12   0
## 11         Beta1 NaN   0.36   0.22     0.02     0.46 NaN
## 12         Beta2 NaN   3.37   5.43     2.20     2.75 NaN
## 13        Gamma1 NaN   0.60   0.46     0.14     0.68 NaN
## 14        Gamma2 NaN   0.37   2.43    -0.80    -0.25 NaN
```