Name: Varun Sudhir

Reg No: 21BDS0040

Exploratory Data Analysis Lab Experiment 8.1

1) Load the in-build dataset mtcars from R environment.

Code:

```
# Varun Sudhir 21BDS0040

# Load the mtcars dataset
data("mtcars")

# View the first few rows
head(mtcars)
```

Output:

```
> # View the first few rows
> 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
>
```

2) Install the required packages and load in R. (psych - For statistical function, moments - Skewness and kurtosis, ggplot..)

Code:

```
# Varun Sudhir 21BDS0040

# Install the necessary packages if not already installed
install.packages(c("psych", "moments", "ggplot2"))

# Load the required packages
library(psych)    # For statistical functions
library(moments)  # For skewness and kurtosis
library(ggplot2)  # For plotting
```

3) Apply all the possibilities of 1D Statistical Data analysis
   a. A measure of Central tendency (All possibilities of mean, all possibilities of median - Quantile, deciles (ntile), percentiles)

1) Mean

```
# Varun Sudhir 21BDS0040

# Arithmetic Mean
mean_mpg <- mean(mtcars$mpg)
mean_mpg

# Trimmed mean
trimmed_mean_mpg <- mean(mtcars$mpg,trim=0.2)
trimmed_mean_mpg
```

```
>
> # Varun Sudhir 21BDS0040
>
> # Arithmetic Mean
> mean_mpg <- mean(mtcars$mpg)
> mean_mpg
[1] 20.09062
> # Trimmed mean
> trimmed_mean_mpg <- mean(mtcars$mpg,trim=0.2)
> trimmed_mean_mpg
[1] 19.22
>
```

2) Median

```
# Varun Sudhir 21BDS0040

# Median
median_mpg <- median(mtcars$mpg)
median_mpg
```

```
> # Varun Sudhir 21BDS0040
>
> # Median
> median_mpg <- median(mtcars$mpg)
> median_mpg
[1] 19.2
```

3) Quantile

```
# Quantiles (0% to 100%)
quantiles_mpg <- quantile(mtcars$mpg)
quantiles_mpg
```

```
> # Quantiles (0% to 100%)
> quantiles_mpg <- quantile(mtcars$mpg)
> quantiles_mpg
    0%    25%    50%    75%   100%
10.400 15.425 19.200 22.800 33.900
```

4) Deciles

```
# Deciles (using ntile function from dplyr package)
mtcars$mpg_deciles <- ntile(mtcars$mpg, 10)
table(mtcars$mpg_deciles)
```

```
> # Deciles (using ntile function from dplyr package)
> mtcars$mpg_deciles <- ntile(mtcars$mpg, 10)
> table(mtcars$mpg_deciles)

 1  2  3  4  5  6  7  8  9 10
 4  4  3  3  3  3  3  3  3  3
```

5) Percentiles

```
# Percentiles (0 to 100% by increments of 1%)
percentiles_mpg <- quantile(mtcars$mpg, probs = seq(0, 1, by =
0.01))
percentiles_mpg
```

```
> # Percentiles (0 to 100% by increments of 1%)
> percentiles_mpg <- quantile(mtcars$mpg, probs = seq(0, 1, by = 0.01))
> percentiles_mpg
    0%     1%     2%     3%     4%     5%     6%     7%     8%     9%    10%    11%    12%
10.400 10.400 10.400 10.400 11.096 11.995 12.894 13.470 13.780 14.090 14.340 14.464 14.588
   13%    14%    15%    16%    17%    18%    19%    20%    21%    22%    23%    24%    25%
14.709 14.802 14.895 14.988 15.054 15.116 15.178 15.200 15.200 15.200 15.239 15.332 15.425
   26%    27%    28%    29%    30%    31%    32%    33%    34%    35%    36%    37%    38%
15.518 15.611 15.704 15.797 15.980 16.166 16.352 16.607 16.886 17.165 17.380 17.535 17.690
   39%    40%    41%    42%    43%    44%    45%    46%    47%    48%    49%    50%    51%
17.827 17.920 18.013 18.112 18.298 18.484 18.670 18.830 18.985 19.140 19.200 19.200 19.200
   52%    53%    54%    55%    56%    57%    58%    59%    60%    61%    62%    63%    64%
19.260 19.415 19.570 19.765 20.168 20.571 20.974 21.000 21.000 21.000 21.088 21.212 21.336
   65%    66%    67%    68%    69%    70%    71%    72%    73%    74%    75%    76%    77%
21.400 21.400 21.400 21.408 21.439 21.470 21.513 21.916 22.319 22.722 22.800 22.800 22.800
   78%    79%    80%    81%    82%    83%    84%    85%    86%    87%    88%    89%    90%
23.088 23.584 24.080 24.576 25.072 25.568 26.052 26.455 26.858 27.261 28.168 29.129 30.090
   91%    92%    93%    94%    95%    96%    97%    98%    99%   100%
30.400 30.400 30.400 30.680 31.300 31.920 32.505 32.970 33.435 33.900
>
```

b) Measure of Dispersions (Range, IQR, Interdecile range, Deviation (mean and Standard deviation), Skewness, and Kurtosis)

1) Range

```
# Varun Sudhir 21BDS0040

# Range
range_mpg <- range(mtcars$mpg)
range_mpg
```

```
>
> # Varun Sudhir 21BDS0040
>
> # Range
> range_mpg <- range(mtcars$mpg)
> range_mpg
[1] 10.4 33.9
>
```

2) IQR

```
# Varun Sudhir 21BDS0040

# Interquartile Range (IQR)
iqr_mpg <- IQR(mtcars$mpg)
iqr_mpg
```

```
> # Varun Sudhir 21BDS0040\
>
> # Interquartile Range (IQR
> iqr_mpg <- IQR(mtcars$mpg)
> iqr_mpg
[1] 7.375
```

3) Interdecile Range

```
# Varun Sudhir 21BDS0040

# Interdecile Range
interdecile_range_mpg <- quantile(mtcars$mpg, 0.9) -
quantile(mtcars$mpg, 0.1)
interdecile_range_mpg
```

```
> # Varun Sudhir 21BDS0040
>
> # Interdecile Range
> interdecile_range_mpg <- quantile(mtcars$mpg, 0.9) - quantile(mtcars$mpg, 0.1)
> interdecile_range_mpg
  90%
15.75
>
```

4) Standard Deviation and Variance

```
# Varun Sudhir 21BDS0040

# Standard Deviation
sd_mpg <- sd(mtcars$mpg)
sd_mpg

# Variance
var_mpg <- var(mtcars$mpg)
var_mpg

# Mean Deviation
mean_deviation_mpg <- mean(abs(mtcars$mpg - mean_mpg))
mean_deviation_mpg
```

```
> # Varun Sudhir 21BDS0040
> # Standard Deviation
> sd_mpg <- sd(mtcars$mpg)
> sd_mpg
[1] 6.026948
> # Variance
> var_mpg <- var(mtcars$mpg)
> var_mpg
[1] 36.3241
> # Mean Deviation
> mean_deviation_mpg <- mean(abs(mtcars$mpg - mean_mpg))
> mean_deviation_mpg
[1] 4.714453
>
```

5) Skewness and Kurtosis

```
# Varun Sudhir 21BDS0040

# Skewness for mpg
skewness_mpg <- skewness(mtcars$mpg)
```

```
skewness_mpg

# Kurtosis for mpg
kurtosis_mpg <- kurtosis(mtcars$mpg)
kurtosis_mpg
```

```
> # Varun Sudhir 21BDS0040
>
> # Skewness for mpg
> skewness_mpg <- skewness(mtcars$mpg)
> skewness_mpg
[1] 0.6404399
> # Kurtosis for mpg
> kurtosis_mpg <- kurtosis(mtcars$mpg)
> kurtosis_mpg
[1] 2.799467
>
```

c) Frequency Distribution with necessary plots (Frequency Distribution, histogram, Relative frequency distribution, and cumulative frequency distribution)

1) Frequency Distribution and Relative frequency

```
# Varun Sudhir 21BDS0040

# Frequency distribution of mpg
mpg_cut <- cut(mtcars$mpg, breaks = 5)
table(mpg_cut)

# Relative Frequency Distribution
rel_freq_mpg <- prop.table(table(mpg_cut))
rel_freq_mpg
```

```
> # Varun Sudhir 21BDS0040
>
> # Frequency distribution of mpg
> mpg_cut <- cut(mtcars$mpg, breaks = 5)
> table(mpg_cut)
mpg_cut
(10.4,15.1] (15.1,19.8] (19.8,24.5] (24.5,29.2] (29.2,33.9]
          6          12           8           2           4
> # Relative Frequency Distribution
> rel_freq_mpg <- prop.table(table(mpg_cut))
> rel_freq_mpg
mpg_cut
(10.4,15.1] (15.1,19.8] (19.8,24.5] (24.5,29.2] (29.2,33.9]
     0.1875      0.3750      0.2500      0.0625      0.1250
>
```

2) Cumulative Frequency Distribution

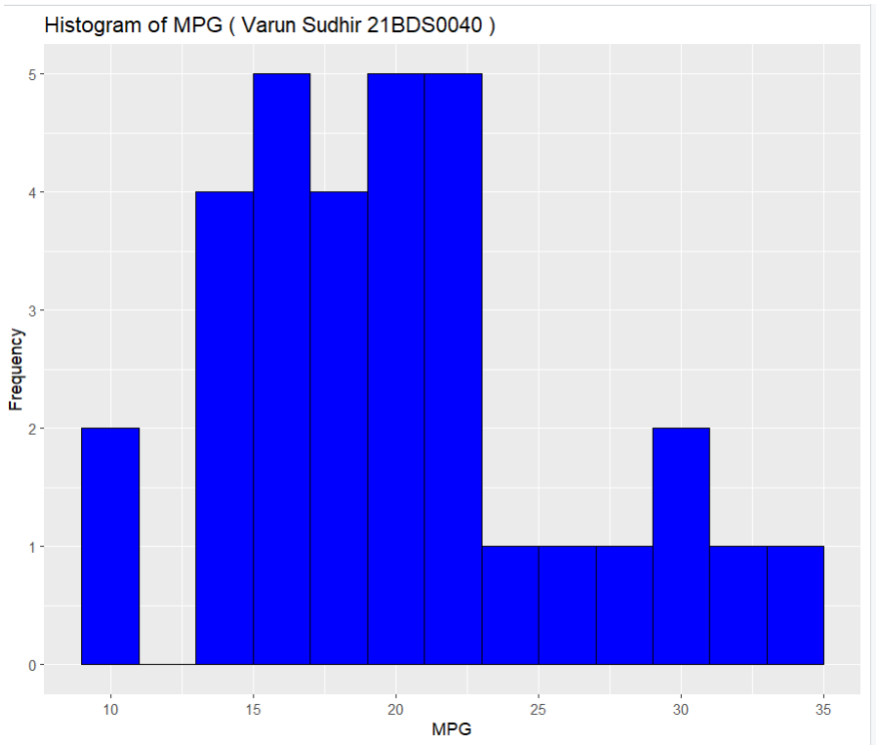
```
# Varun Sudhir 21BDS0040

# Cumulative Frequency Distributioncum_freq_mpg <-
cumsum(table(mpg_cut))
cum_freq_mpg
```

```
> # Varun Sudhir 21BDS0040
>
> # Cumulative Frequency Distribution
> cum_freq_mpg <- cumsum(table(mpg_cut))
> cum_freq_mpg
(10.4,15.1] (15.1,19.8] (19.8,24.5] (24.5,29.2] (29.2,33.9]
          6          18          26          28          32
>
```

3) Histogram

```
# Varun Sudhir 21BDS0040
# Histogram of mpg
ggplot(mtcars, aes(x = mpg)) +
  geom_histogram(binwidth = 2, fill = "blue", color = "black")
+
  labs(title = "Histogram of MPG ( Varun Sudhir 21BDS0040 )",
x = "MPG", y = "Frequency")
```



d) From the categorical variable (Pie plot and Stacked bar plot)

```
# Varun Sudhir 21BDS0040


# Pie chart for 'cyl' (number of cylinders)
cyl_count <- table(mtcars$cyl)
pie(cyl_count, labels = names(cyl_count), main = "Distribution of
Cylinders ( 21BDS0040) ")

# Stacked Bar plot for 'cyl' and 'gear' (number of gears)
ggplot(mtcars, aes(x = factor(cyl), fill = factor(gear))) +
  geom_bar(position = "stack") +
  labs(title = "Stacked Bar Plot of Cylinders and Gears (
21BDS0040 )", x = "Cylinders", y = "Count", fill = "Gears")
```