```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

Descriptive Statistics

The `mtcars` dataset is made up of 11 numeric variables which form the columns and 32 observations which form the rows.

```
mtcars
```

```
##                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4          21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag      21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360         14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230           22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280           19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C          17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE         16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL         17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC        15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4  8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial  14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128           32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic        30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla     33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona      21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger   15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin        15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28         13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird   19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9          27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2      26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa       30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L     15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino       19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora      15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E         21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
# Calculating the mean for hp and mpg
mean_hp = mean(mtcars$hp)
mean_mpg = mean(mtcars$mpg)

mean_hp
```

```
## [1] 146.6875
```

```
mean_mpg
```

```
## [1] 20.09062
```

```
# Calculating the standard deviation for hp and mpg
std_hp = sd(mtcars$hp)
std_mpg = sd(mtcars$mpg)

std_hp
```

```
## [1] 68.56287
```

```
std_mpg
```

```
## [1] 6.026948
```

```
# Getting the five-number summary for hp and mpg
fivenum(mtcars$hp)
```

```
## [1]  52  96 123 180 335
```

```
fivenum(mtcars$mpg)
```

```
## [1] 10.40 15.35 19.20 22.80 33.90
```

Interpretation

The standard deviations of both variables are lower than their means, thus suggesting a decreased level of variability and dispersion of the data points. The variables can then be described to be generally stable as the data points are not really that far spread out from the mean.

The medians of the variables are also lower than their standard deviations. This suggests a positively skewed distribution when plotted on a graph thus showing that there are a few large values/ data points in the variables.

The minimum values and first quartiles if the variables seem to behaving in the opposite manner as they are higher than the standard deviation.This prooves the previous observations in regards to the medians and standard deviation. Lower standard deviations when compared to the minimum values and first quartiles suggest that a significant part of the data is concentrated in the lower value range, thus few large values, and low variability.

Pearsons Correlation Analysis

This type of correlation measures the relationship between two quantitative continuous variables that are seen to have a linear association.

```
correlation_coeffs = cor(mtcars)
correlation_coeffs["hp","mpg" ]
```

```
## [1] -0.7761684
```

Interpretation

A correlation coefficient of -0.7762 suggests a strong negative correlation with negative association between the two variables; `hp` and `mpg` . As one decreases, the other increases.

Simple Regression Model

```
regression_model = lm(mpg ~ hp, data = mtcars)
regression_model
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = mtcars)
##
## Coefficients:
## (Intercept)           hp
##    30.09886     -0.06823
```

```
summary(regression_model)
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7121 -2.1122 -0.8854  1.5819  8.2360
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
## hp          -0.06823    0.01012  -6.742 1.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.863 on 30 degrees of freedom
## Multiple R-squared:  0.6024, Adjusted R-squared:  0.5892
## F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

Regression Equation

Taking the equation y = mx + c, which is a representation of a simple linear regression model equation, lets replace y with mpg and have the x be hp, thus creating the equation:

```
          Miles Per Gallon = 30.09886 - 0.06823 (Gross horsepower)
```

Interpretation

Fitting the regression model for the data produces an output that suggests a negative association because the the slope/gradient(-0.06823) is negative, thus showing a negative correlation, therefore, a negative association. This then suggests as the x variable increases, the y variable decreases, meaning that the miles covered per gallon of fuel decreases the horsepower used by the observed automobiles increases.

Plotting a Scatter Plot together with a Regression Line

To achieve this plot, functions from the `ggplot2` library/package were used where `geom_point()` plotted the scatter plot and `geom_smooth()` plotted the regression line better known as the trend line/ slope.

```
regression_plot = ggplot(mtcars, aes(x = hp, y = mpg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "darkgreen") +
  labs(title = "Scatter Plot with Regression Line",
       x = "Horsepower",
       y = "Miles Per Gallon")

regression_plot
```

```
## `geom_smooth()` using formula = 'y ~ x'
```