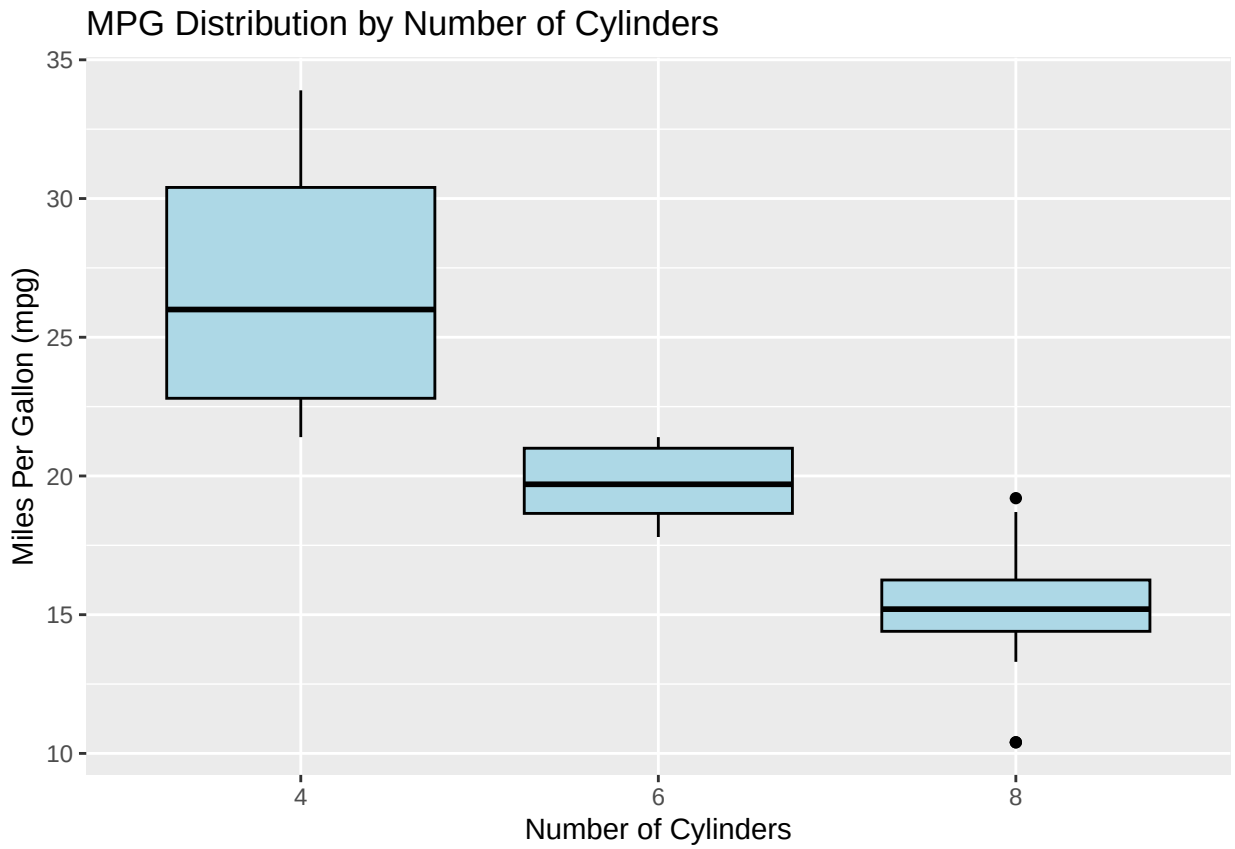
Assignment

R V Abhishek

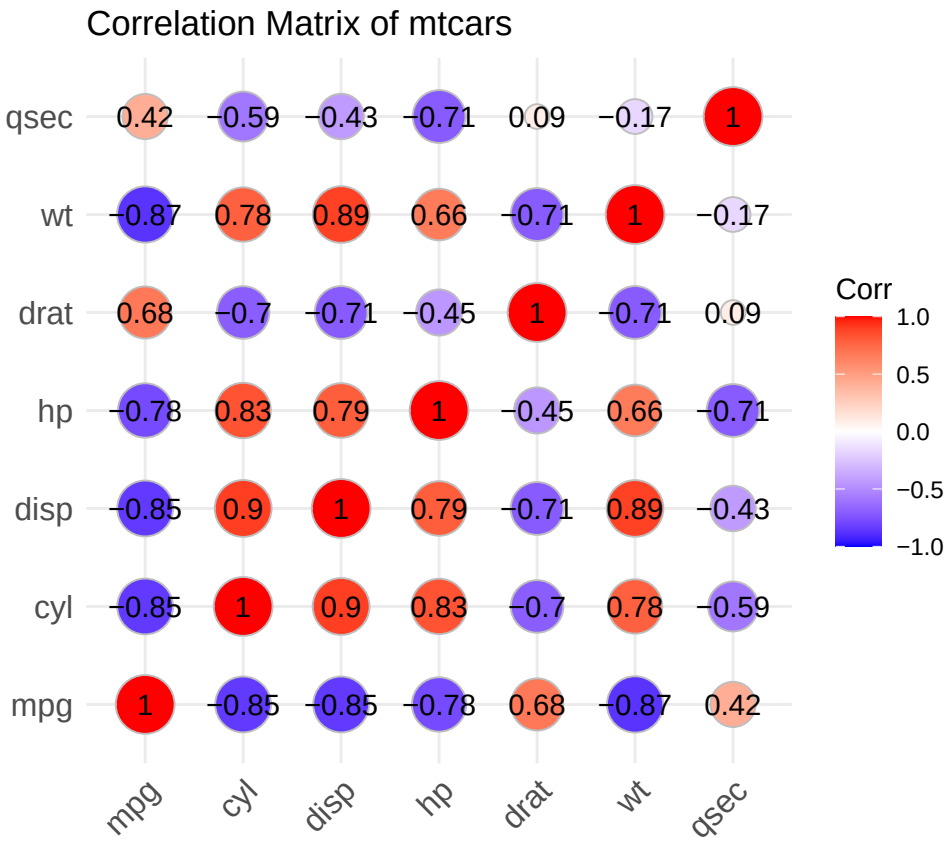
2025-09-09

Statistical Analysis of the 'mtcars' Dataset

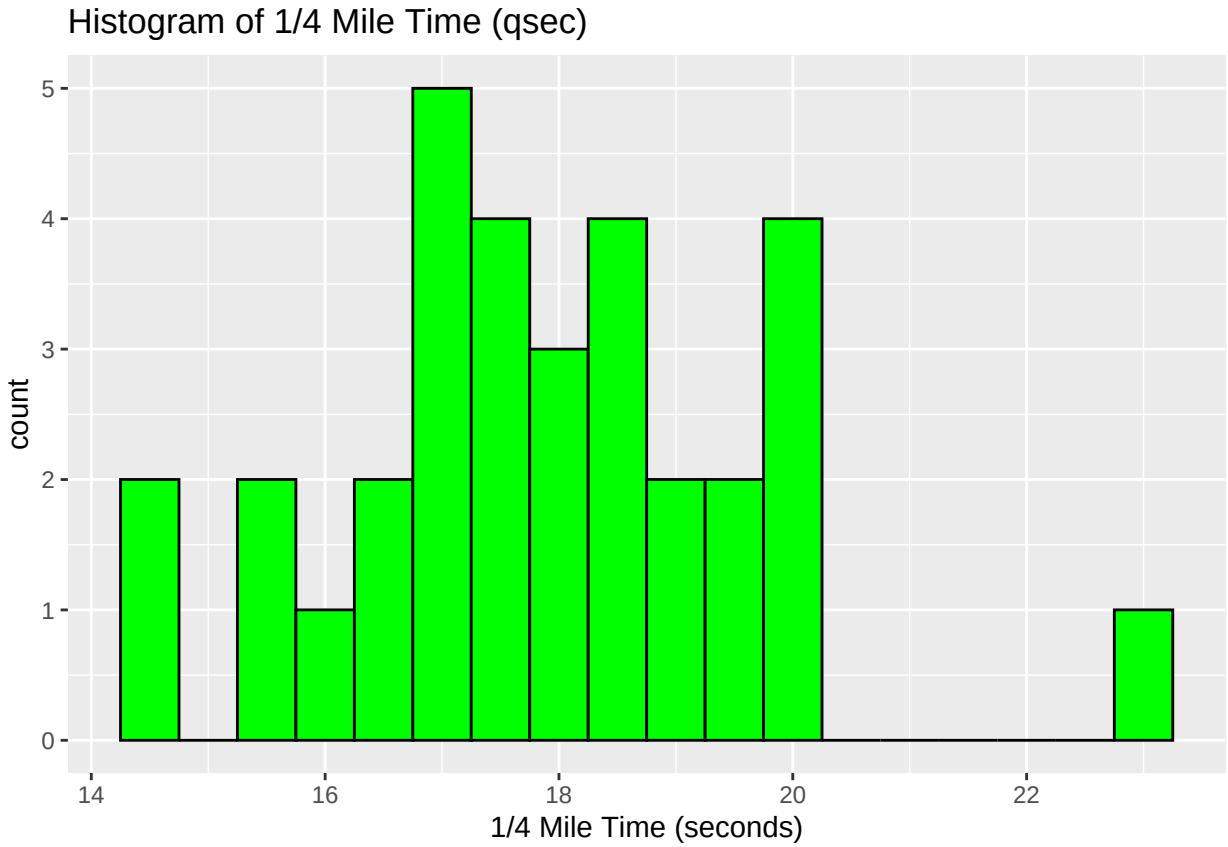
```
# Load necessary libraries
library(ggplot2)
library(ggcorrplot)

# Load the built-in mtcars dataset
data(mtcars)

# Calculate the mean and standard deviation for the mpg variable
mean_mpg <- mean(mtcars$mpg)
sd_mpg <- sd(mtcars$mpg)

# Create a frequency table for the cyl variable
cyl_counts <- table(mtcars$cyl)

# Perform a t-test to compare the mean mpg for different cylinder counts
mpg_4cyl <- mtcars$mpg[mtcars$cyl == 4]
mpg_6cyl <- mtcars$mpg[mtcars$cyl == 6]
t_test_mpg <- t.test(mpg_4cyl, mpg_6cyl)

# Generate a boxplot to visualize the distribution of mpg by cylinder count
ggplot(mtcars, aes(x = factor(cyl), y = mpg)) +
  geom_boxplot(fill = "lightblue", color = "black") +
  labs(title = "MPG Distribution by Number of Cylinders",
       x = "Number of Cylinders",
       y = "Miles Per Gallon (mpg)")
```




```r
# Compute the correlation matrix for selected numeric variables
numeric_mtcars <- mtcars[, c("mpg", "cyl", "disp", "hp", "drat", "wt", "qsec")]
correlation_matrix <- cor(numeric_mtcars)

# Perform a t-test to compare the mean weight of manual and automatic cars
wt_auto <- mtcars$wt[mtcars$am == 0]
wt_manual <- mtcars$wt[mtcars$am == 1]
t_test_wt <- t.test(wt_auto, wt_manual)

# Visualize the correlation matrix using a specialized plotting function
ggcorrplot(correlation_matrix,
           method = "circle",
           lab = TRUE,
           title = "Correlation Matrix of mtcars")
```

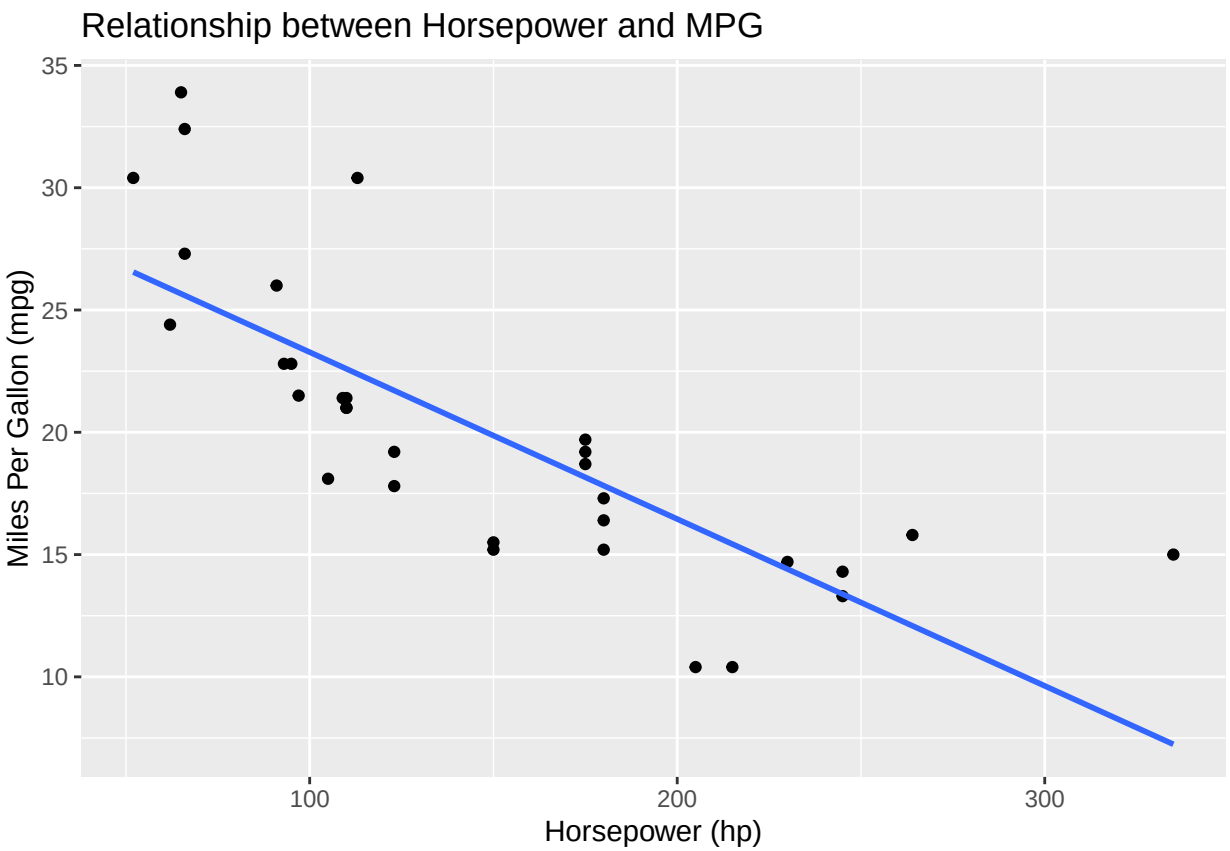



```r
# Calculate the mean and standard deviation of horsepower for 8-cylinder cars
hp_8cyl <- mtcars$hp[mtcars$cyl == 8]
mean_hp_8cyl <- mean(hp_8cyl)
sd_hp_8cyl <- sd(hp_8cyl)

# Generate a histogram to visualize the distribution of qsec
ggplot(mtcars, aes(x = qsec)) +
  geom_histogram(binwidth = 0.5, fill = "green", color = "black") +
  labs(title = "Histogram of 1/4 Mile Time (qsec)",
       x = "1/4 Mile Time (seconds)")
```




```
# Create a scatter plot to visualize the relationship between horsepower and mpg
ggplot(mtcars, aes(x = hp, y = mpg)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Relationship between Horsepower and MPG",
       x = "Horsepower (hp)",
       y = "Miles Per Gallon (mpg)")
```

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




```
# Perform Z-score analysis to identify outliers in selected variables
z_scores_df <- scale(mtcars)
outliers <- which(abs(z_scores_df) > 3, arr.ind = TRUE)

# Display results
print("Mean MPG:")
```

```
[1] "Mean MPG:"
```

```
print(mean_mpg)
```

```
[1] 20.09062
```

```
print("Standard Deviation of MPG:")
```

```
[1] "Standard Deviation of MPG:"
```

```
print(sd_mpg)
```

```
[1] 6.026948
```

```r
print("Cylinder Counts:")
```

```
[1] "Cylinder Counts:"
```

```r
print(cyl_counts)
```

```
 4  6  8
11  7 14
```

```r
print("T-test Results for MPG (4 vs 6 cylinders):")
```

```
[1] "T-test Results for MPG (4 vs 6 cylinders):"
```

```r
print(t_test_mpg)
```

```

    Welch Two Sample t-test

data:  mpg_4cyl and mpg_6cyl
t = 4.7191, df = 12.956, p-value = 0.0004048
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  3.751376 10.090182
sample estimates:
mean of x mean of y
 26.66364  19.74286
```

```r
print("Mean Horsepower for 8-cylinder cars:")
```

```
[1] "Mean Horsepower for 8-cylinder cars:"
```

```r
print(mean_hp_8cyl)
```

```
[1] 209.2143
```

```r
print("Standard Deviation of Horsepower for 8-cylinder cars:")
```

```
[1] "Standard Deviation of Horsepower for 8-cylinder cars:"
```

```r
print(sd_hp_8cyl)
```

```
[1] 50.97689
```

```r
print("T-test Results for Weight (Manual vs Automatic):")
```

```
[1] "T-test Results for Weight (Manual vs Automatic):"
```

```r
print(t_test_wt)
```

```
    Welch Two Sample t-test

data:  wt_auto and wt_manual
t = 5.4939, df = 29.234, p-value = 6.272e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.8525632 1.8632262
sample estimates:
mean of x mean of y
 3.768895  2.411000
```

```r
print("Outliers based on Z-scores:")
```

```
[1] "Outliers based on Z-scores:"
```

```r
print(outliers)
```

```
              row col
Maserati Bora  31  11
```