Experiment 2

1. Summary Statistics for a Dataset

- Dataset: Built-in mtcars dataset (Car Specifications)
- Compute summary statistics (mean, median, standard deviation, etc.).
- Understand the distribution of miles per gallon (mpg) and horsepower (hp).

[1]:
```
data("mtcars")
summary(mtcars)
```

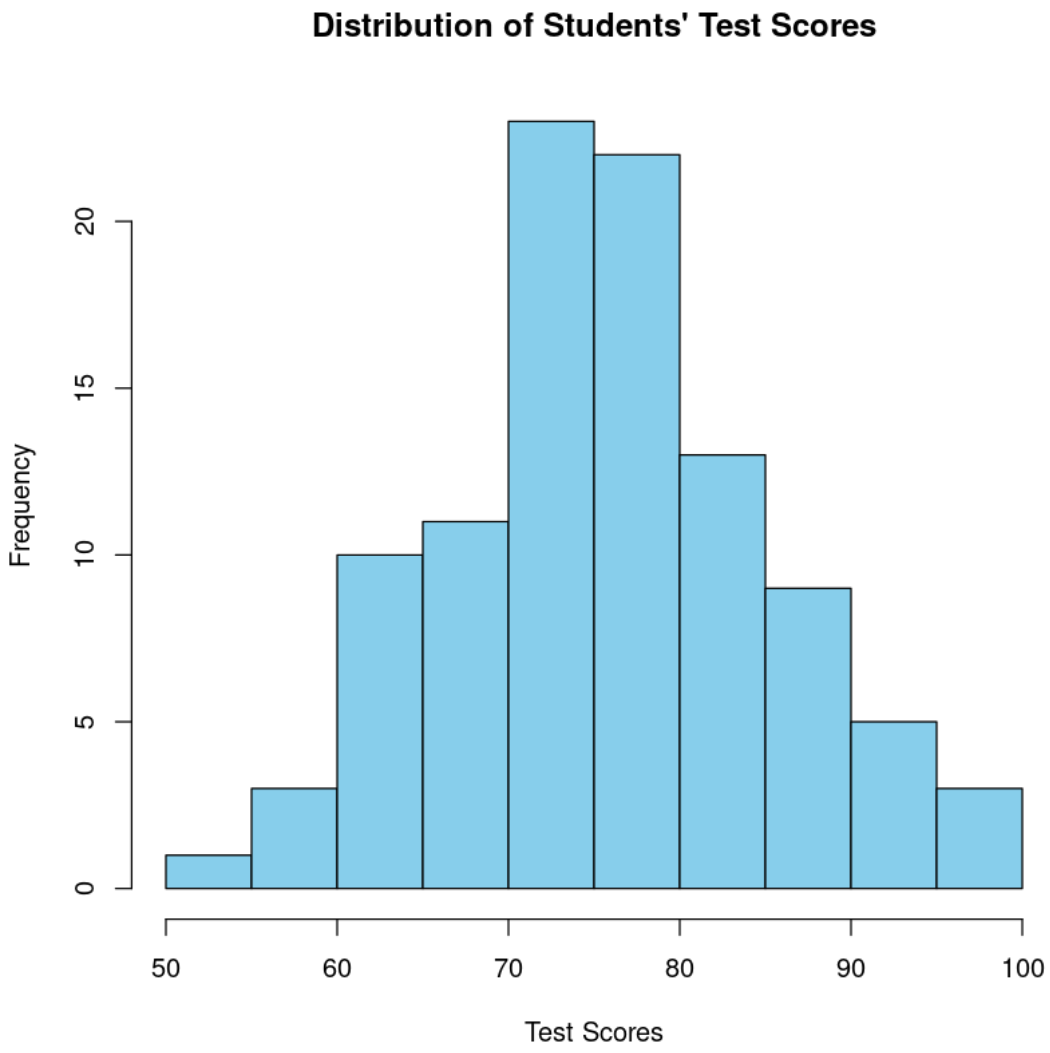
```
      mpg             cyl             disp             hp       
 Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
 Median :19.20   Median :6.000   Median :196.3   Median :123.0  
 Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
 Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
      drat             wt             qsec             vs        
 Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
 Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
 Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
 Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
       am              gear            carb      
 Min.   :0.0000   Min.   :3.000   Min.   :1.000  
 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
 Median :0.0000   Median :4.000   Median :2.000  
 Mean   :0.4062   Mean   :3.688   Mean   :2.812  
 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
 Max.   :1.0000   Max.   :5.000   Max.   :8.000  
```

2. Create a Histogram

- Generate a random dataset of students test scores
- Visualize data distribution using histograms.
- Understand skewness and spread of data.

[2]:
```
# Generate random test scores (normal distribution, mean=75, sd=10)
set.seed(123)  # For reproducibility
scores <- rnorm(100, mean = 75, sd = 10)
# Create a histogram
hist(scores,
     main = "Distribution of Students' Test Scores",
     xlab = "Test Scores",
```

```
     col = "skyblue",
     border = "black",
    )
# Check skewness and spread
# library(moments)
# skewness(scores)  # Measure skewness
```



[3]:
```
summary(scores)    # Summary statistics
dev.off()
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  51.91   70.06   75.62   75.90   81.92   96.87
```

null device: 1

3. Scatterplot to Explore Relationships Dataset: Built-in iris dataset (Flower Measurements) The iris dataset contains sepal and petal lengths and widths of three flower species.

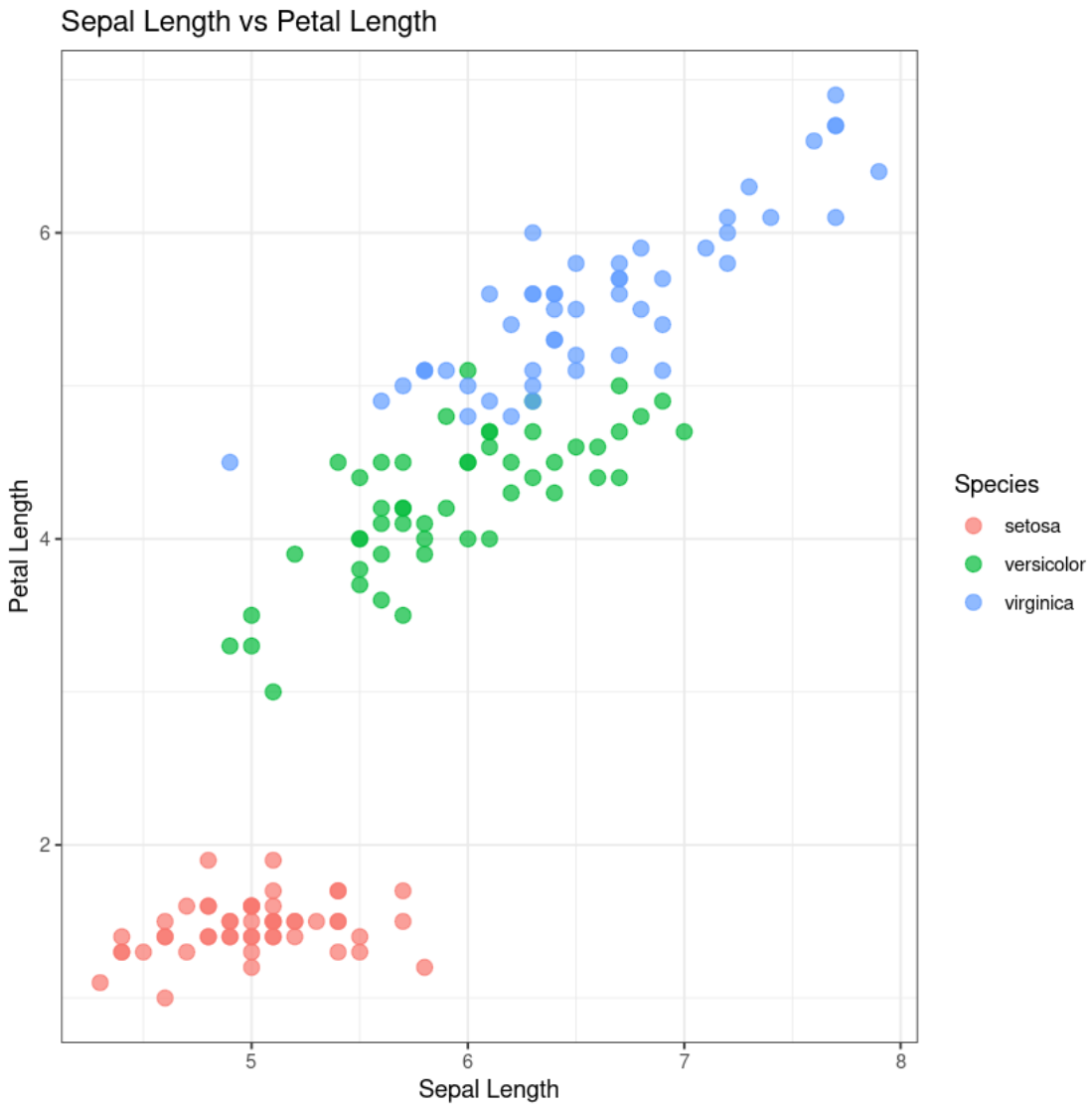
- Create a scatterplot to explore relationships between variables.
- Use colors to distinguish species.

[4]:
```r
library(ggplot2)

# Load the iris dataset
data(iris)

# Scatterplot: Sepal Length vs. Petal Length
ggplot(iris, aes(x = Sepal.Length, y = Petal.Length, color = Species)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(title = "Sepal Length vs Petal Length",
       x = "Sepal Length",
       y = "Petal Length") +
  theme_bw()

# Save the plot
ggsave("iris_scatterplot.png", width = 8, height = 6)
```



4. Boxplot for Detecting Outliers

- Dataset: Simulated monthly sales data for a store. Generate random monthly sales data to analyze outliers.
- Create a boxplot to detect outliers.
- Understand quartiles and interquartile range (IQR).

[5]:
```
# Generate random monthly sales data (e.g., 12 months)
set.seed(42)
monthly_sales <- rnorm(12, mean = 50000, sd = 10000)

# Create a boxplot
boxplot(monthly_sales, col = "lightgreen", main = "Monthly Sales Data",
        ylab = "Sales", outline = TRUE)
```

```r
# Add a horizontal line for the mean
abline(h = mean(monthly_sales), col = "red", lty = 2)
legend("topright", legend = "Mean", col = "red", lty = 2)
```