Homework 1

24226196

Dataset: *iris*

The iris dataset is a built-in dataset in R that contains information about 150 iris flowers, split evenly across three species: *setosa*, *versicolor*, and *virginica*. It records flower measurements in centimeters, including:

- Sepal.Length
- Sepal.Width
- Petal.Length
- Petal.Width
- Species

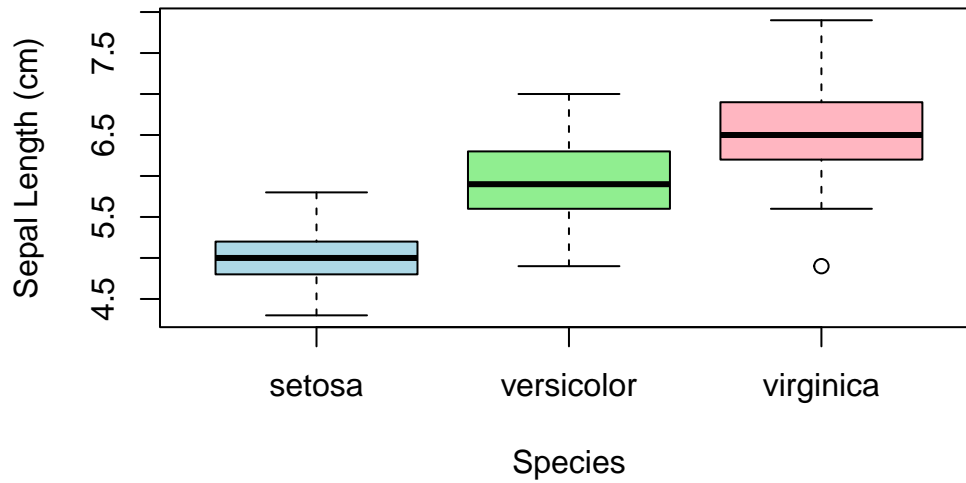We will visualize the distribution of **Sepal Length** across different species using a boxplot.

Descriptive Summary

Here is a summary of the dataset (code hidden in output):

Boxplot of Sepal Length by Species

The boxplot below shows how Sepal Length varies across different iris species:

Sepal Length by Species

The boxplot displays the distribution of Sepal Length for each of the three species: setosa, versicolor, and virginica.

Each box represents the middle 50% of the data.

The line inside the box is the median of the values.

The whiskers show the full range of typical values.

Points outside the whiskers are outliers.

The boxplot compares sepal length across iris species, showing that *setosa* has the shortest, *virginica* the longest, and *versicolor* is intermediate with more variation.