Iris Dataset Analysis

Sahil Chalkhure-23206052

May 31, 2024

Introduction

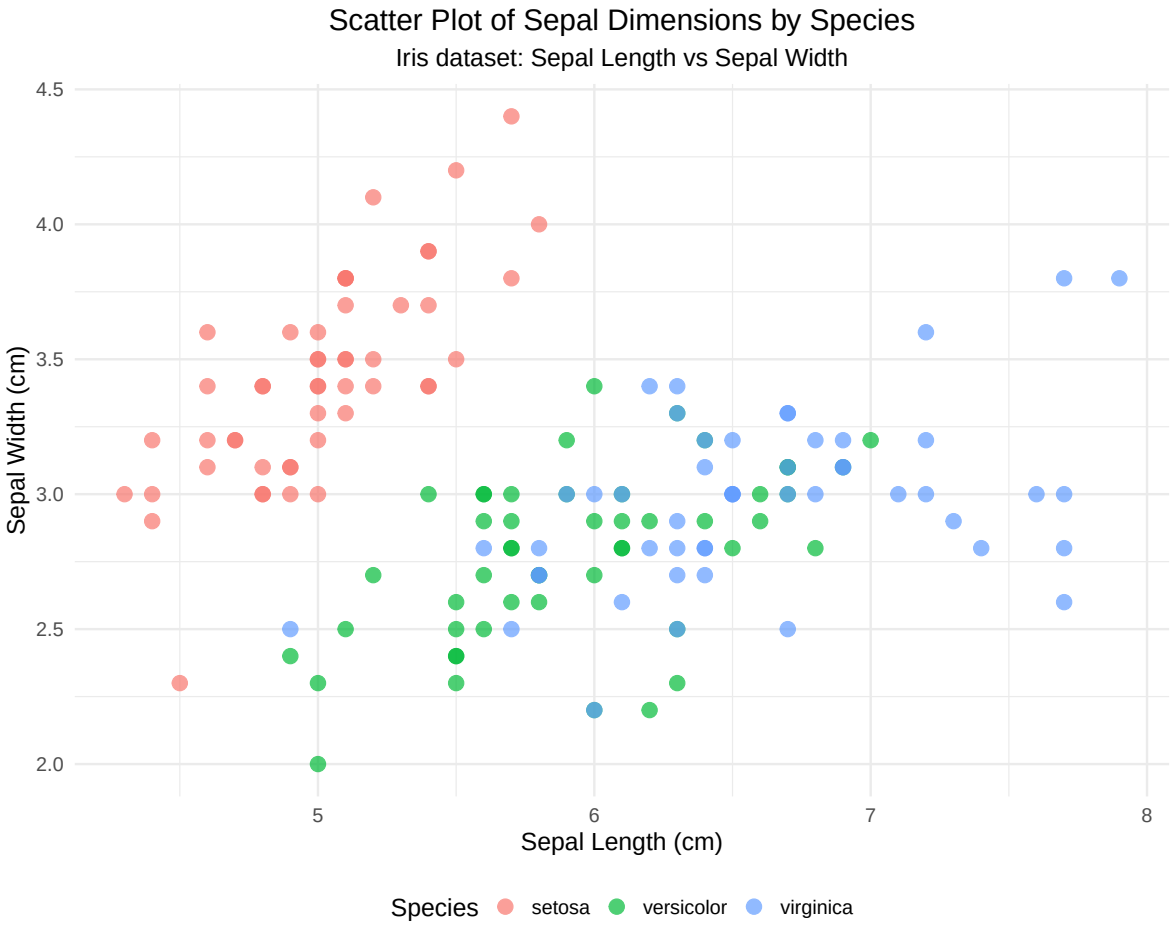
The dataset used in this document is the **iris** dataset, which is available in the `datasets` package in R. This dataset contains measurements of various attributes for different species of iris flowers.

Approach

I utilized the `ggplot2` package in R to create a scatter plot illustrating the relationship between the **Sepal Length** and **Sepal Width** of iris flowers. The `iris` dataset contains 150 observations of iris flowers, with 50 observations each for three different species: *setosa*, *versicolor*, and *virginica*.

Plot

Below is a plot showing the relationship between the **Sepal Length** and **Sepal Width** of the iris flowers.



Description

The scatter plot visualizes the relationship between sepal length and sepal width for three species of the **Iris dataset:** ***setosa***, ***versicolor***, *and* ***virginica***. Each point represents an individual flower, colored according to its species.

Key Observations:

1. **Setosa (red points)**:
   - Sepal length ranges from approximately 4.5 to 5.8 cm.
   - Sepal width ranges from about 2.3 to 4.5 cm.

- **Setosa** has the smallest sepal length compared to the other species and tends to have a wider sepal width.

2. **Versicolor (green points)**:
   - Sepal length ranges from approximately 4.9 to 7.0 cm.
   - Sepal width ranges from about 2.0 to 3.4 cm.
   - **Versicolor** has moderate sepal length and width, positioned between ***setosa*** and ***virginica***.
3. **Virginica (blue points)**:
   - Sepal length ranges from approximately 4.9 to 7.9 cm.
   - Sepal width ranges from about 2.2 to 3.8 cm.
   - **Virginica** generally has the longest sepals among the three species and varies significantly in sepal width.

Conclusion

This analysis provides a basic understanding of the differences in sepal dimensions among different species of iris flowers. The scatter plot highlights that each species has unique characteristics in terms of sepal length and width, which can help in identifying the species based on these attributes. The distinct clustering of species in the plot underscores the potential for using sepal dimensions as a criterion for species classification.