Homework - 1

Bhupendra Singh Chaudhary - 23200179

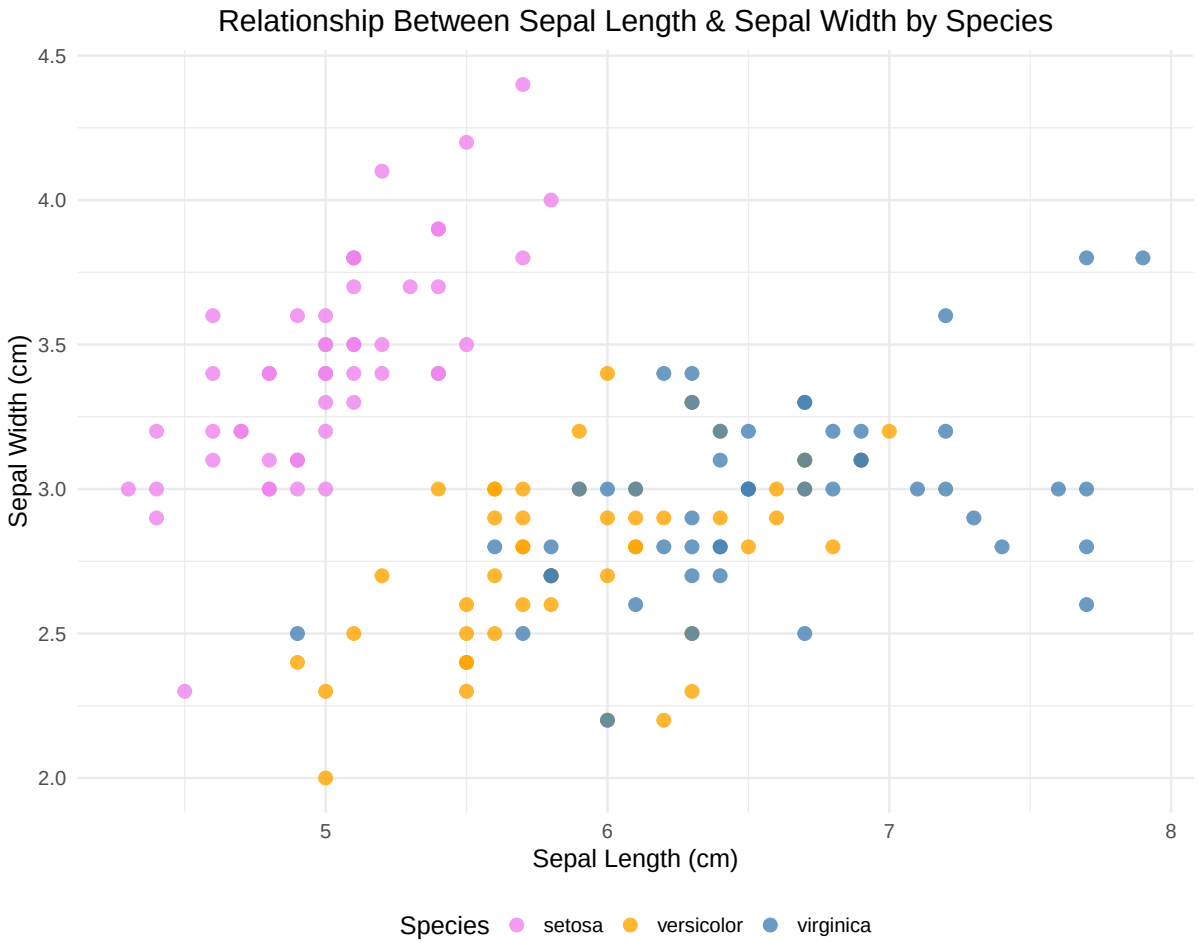
June 26, 2024

Introduction

The document discusses the iris dataset, which you can found in the datasets package in R. It includes data on several characteristics of various species of iris flowers.

I made a scatter plot using the `ggplot2` package in R to explore how Sepal Length and Sepal Width relate in iris flowers. The dataset I worked with includes 150 observations, divided equally among three species—setosa, versicolor, and virginica, with each species represented by 50 observations.

Plot



The scatter plot visualizes the relationship between the Sepal Length and Sepal Width of iris flowers, broken down by species: **setosa**, **versicolor**, and **virginica**.

From the graph, we can observe several trends and characteristics:

1. **Setosa** (pink dots) generally shows a wider Sepal Width relative to its Sepal Length compared to the other two species. The setosa points cluster tightly in the range of `4.5 to 5.5 cm` for Sepal Width and `4.5 to 6 cm` for Sepal Length.
2. **Versicolor** (orange dots) shows an intermediate range of Sepal Width and Length. The Sepal Width for versicolor tends to be around `2.5 to 3.5 cm`, and the Sepal Length spans from roughly `4.5 to 7 cm`. There's a bit more spread in the versicolor data compared to setosa, indicating more variability in these dimensions for the versicolor species.

3. **Virginica** (blue dots) tends to have longer sepals with relatively moderate widths. This species generally ranges from about `2.5 to 3.5 cm` in Sepal Width and extends further in Sepal Length from `5 to 8 cm`. The spread of the virginica data suggests considerable diversity in Sepal Length.

Conclusion

The overview provides understanding of the variations in sepal sizes across different iris species. The scatter plot clearly shows that each species exhibits unique sepal length and width, which can be helpful in identifying them. The way the species group distinctly on the plot emphasizes how sepal measurements can serve as a reliable method for classifying different types of iris flowers. This visual representation helps underscore the practical use of these dimensions in distinguishing between species.