Homework 1- R

GurtejSingh 24216077

Introduction

The iris dataset is a classic example used in both statistics and machine learning. It contains measurements of sepal and petal sizes for 150 iris flowers, divided evenly across three species: setosa, versicolor, and virginica.

Dataset Overview

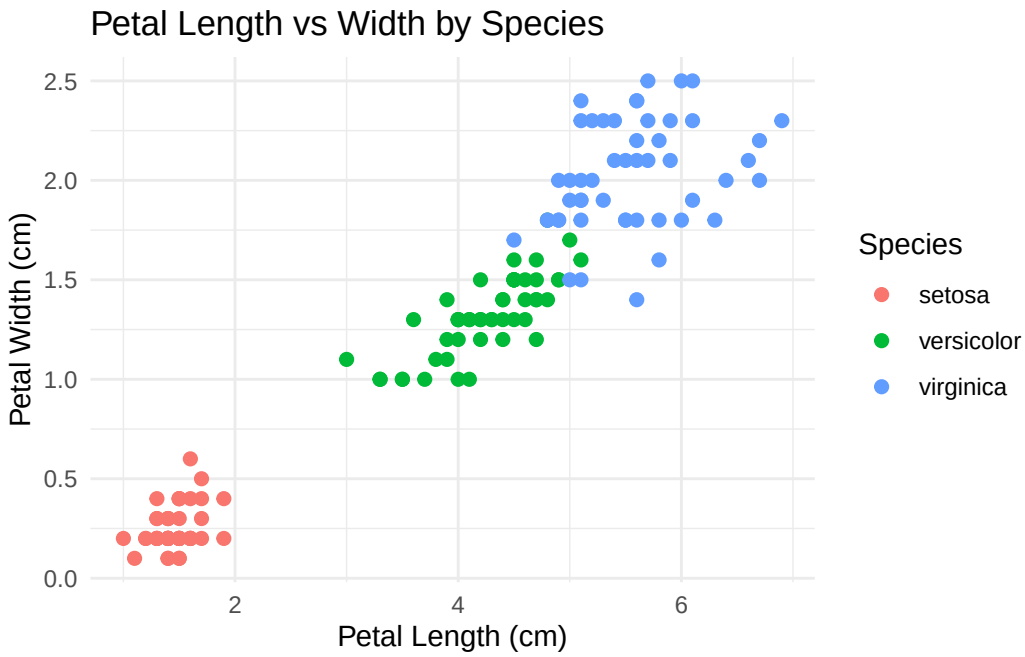
```
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

This dataset comes built into R and includes five variables:

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

There are 150 observations in total, with 50 flowers from each of the three species.

Plot: Petal Dimensions by Species



Conclusion

The scatter plot shows clear differences between the three iris species based on their petal size. Setosa flowers have the smallest petals, Virginica the largest, and Versicolor sits in the middle, with some overlap. This pattern suggests that petal length and width are useful traits for telling the species apart, both visually and in predictive models.