Advanced R Programming - Homework 1

Navya Zacharia - 24209002

Introduction

I have chosen `iris` dataset for this homework. The `iris` dataset is one of the most popular datasets in data science. It has measurements of 150 iris flowers such as *sepal length*, *sepal width*, *petal length* and *petal width*. There are three species: *setosa*, *versicolor*, and *virginica*.

In this document, I summarize the iris dataset using descriptive statistics and create visualizations to explore differences between the species.

Summary Statistics

I have used the `dplyr` and `knitr` packages for creating a summary table showing the average measurements for each iris species.

Table 1: Average Measurements and Count for Each Species

Species	Count	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	50	5.01	3.43	1.46	0.25
versicolor	50	5.94	2.77	4.26	1.33
virginica	50	6.59	2.97	5.55	2.03

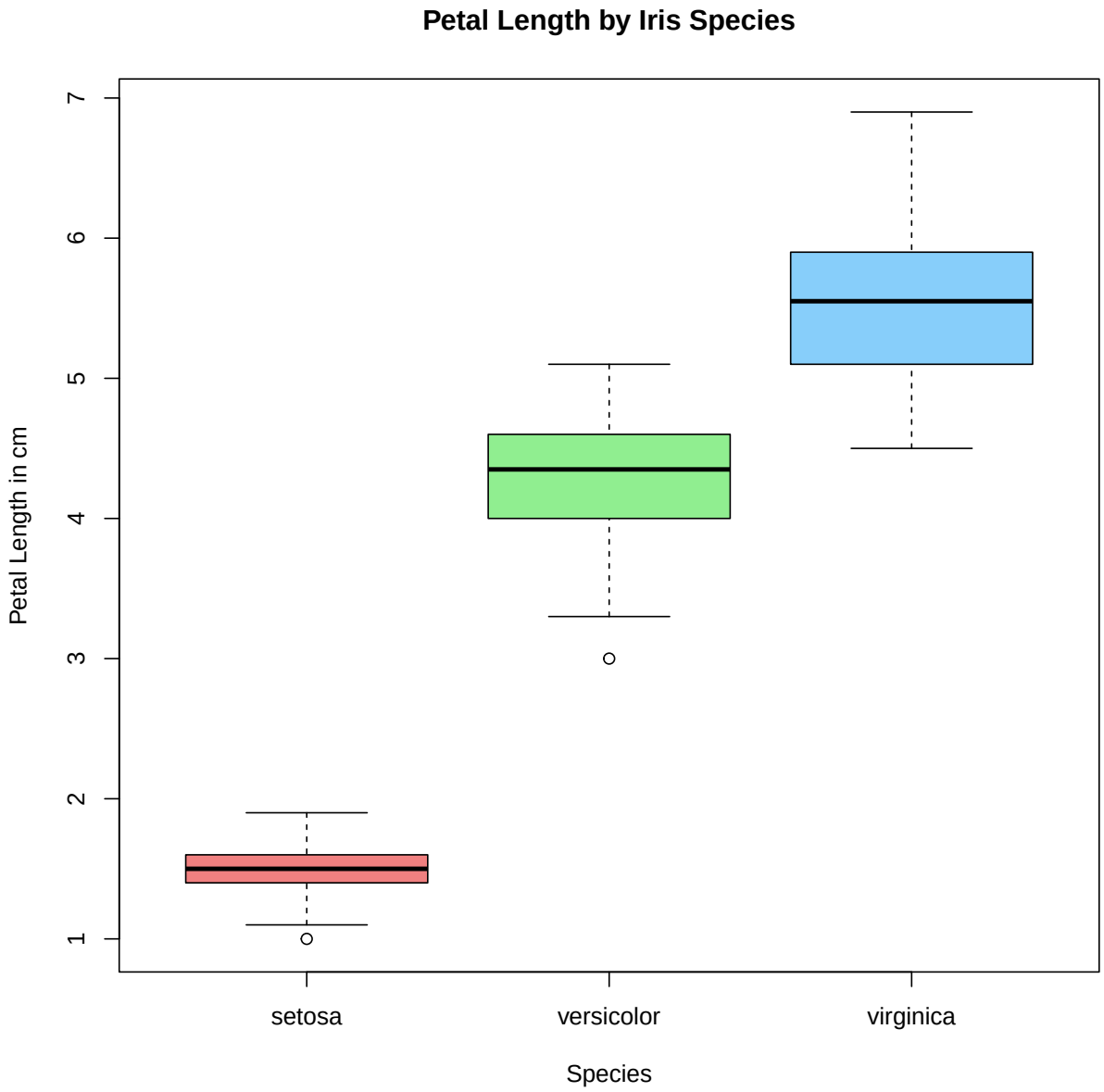
All three iris species *setosa*, *versicolor*, and *virginica* have equal number of observations.

- Setosa has the smallest petal dimensions, with an average petal length of 1.46 cm and petal width of 0.25 cm and both sepal length and width are also relatively small, though its sepal width 3.43 cm which is slightly higher than the other two species. This species is **smaller** in flower size compared to the others.

- Versicolor has medium-sized petals, with an average petal length of 4.26 cm and width of 1.33 cm. Its sepals are also moderate in size around 5.94 cm long and 2.77 cm wide. Overall, it falls between setosa and virginica in terms of flower size and shape. i.e **medium** in flower size.
- Virginica has the largest petal dimensions, with a petal length of 5.55 cm and width of 2.03 cm. It also has the longest sepals, averaging 6.59 cm in length and 2.97 cm in width. This species is clearly the **largest** among the three species.

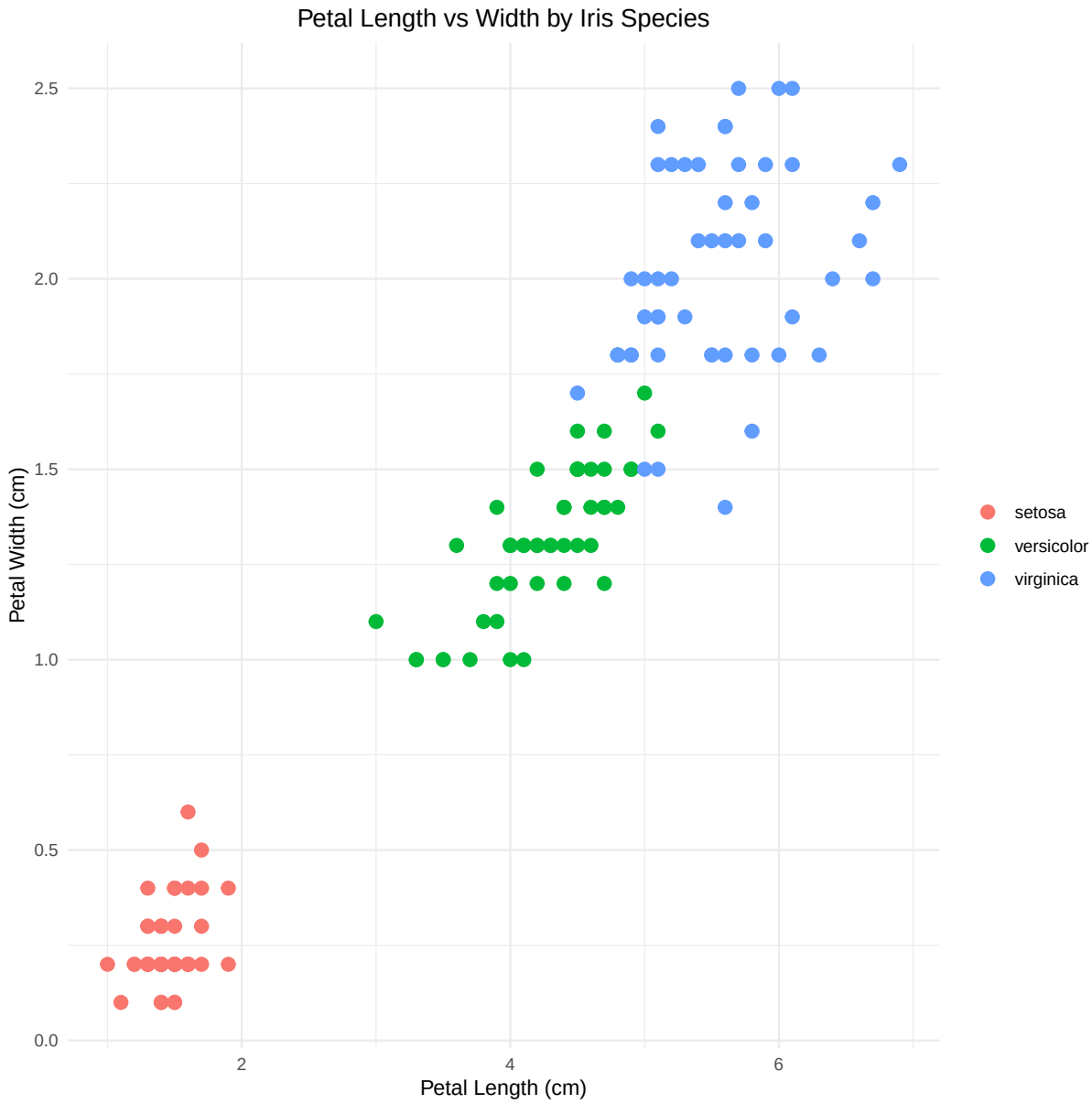
Plots

Box Plot



The boxplot shows clear differences in petal length across the three iris species. **Setosa** has the shortest petals, all tightly grouped around 1.5 cm. **Versicolor** falls in the middle, with more variation and a few values dipping below 3 cm. **Virginica** has the longest petals, reaching up to 7 cm and showing a wider spread. Overall, petal length is a strong feature for classifying the species, especially for identifying setosa.

Scatter Plot



The scatter plot shows how petal length and width vary across the three iris species. **Setosa** (red) is clearly separated in the lower-left corner, with very short and narrow petals. **Versicolor** (green) lies in the middle with moderate values and overlaps slightly with virginica. **Virginica** (blue) appears in the top-right, having the longest and widest petals. The three groups are visually distinct, especially setosa, making these two features excellent for species classification.