Figure 1. Sepal Length Distribution

Observation

- On Figure 1, we can see that the Distribution of the Sepal Length is approximately symmetric and the data is distributed equally as shown on the Box Plot and the Density Line of the Histogram Plot. Moreover, the skewness of this column is (0.31) near zero, which means that the data is likely to be normally distributed.

- On Q-Q Plot, despite the fact that the data may be slightly non-normal because of the deviation or departure of normality of the tails, most of the data tend to follow the 45-degree or the red line. Thus, it can be considered that the data is likely to be normally distributed.

- In this column, the kurtosis value is 2.43 which means that the data may have flatter peak and shorter tails as shown on the Histogram Plot. This type of kurtosis is called Platykurtic or Negative Kurtosis. Moreover, this means that there are less chances of the presence of outliers as shown on the Box Plot.
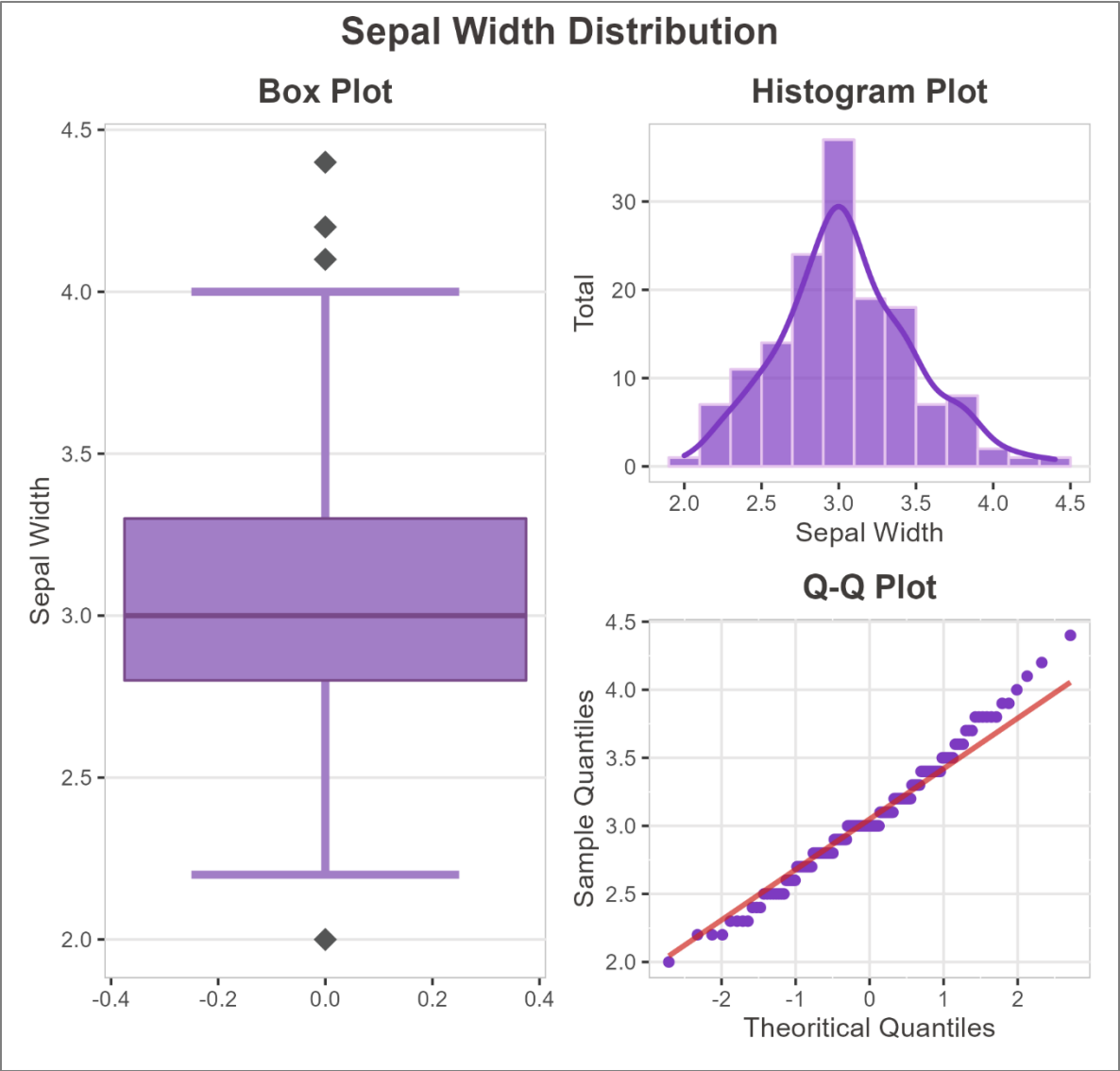
Figure 2. Sepal Width Distribution

Observation

- On Figure 2, we can see that the Distribution of the Sepal Width is approximately symmetric as shown on the Histogram Plot. Moreover, the skewness of this column is (0.32) near zero, which means that the data is likely to be normally distributed. This is supported by the Shapiro Wilk Normality Test which shows a p-value of 0.10 which means that we assume that the data is normally distributed.

- On Box Plot, we can see that the median has a value of 3 and is near the Lower Quantile Q1, which means that the distribution is Right or Positively Skewed as shown on the Histogram Plot. This could also mean that outliers might appear more above the median. As shown on the Box Plot, there are three outliers above the median, mainly 4.1, 4.2 and 4.4 and below the median there is one outlier that is a value of 2.

- On Q-Q Plot, there is a little to no deviation on the tails and some of those deviations are because of the outliers. Moreover, it is shown that most of the data follow the 45-degree line. Thus, the data is likely to be normally distributed.

- In this column, the kurtosis value is 3.18 which means that the data may have long tails because of outliers as shown on the Histogram Plot. This type of kurtosis is called Leptokurtic or Positive Kurtosis. Moreover, this means that there are more chances of the presence of outliers as shown on the Box Plot.
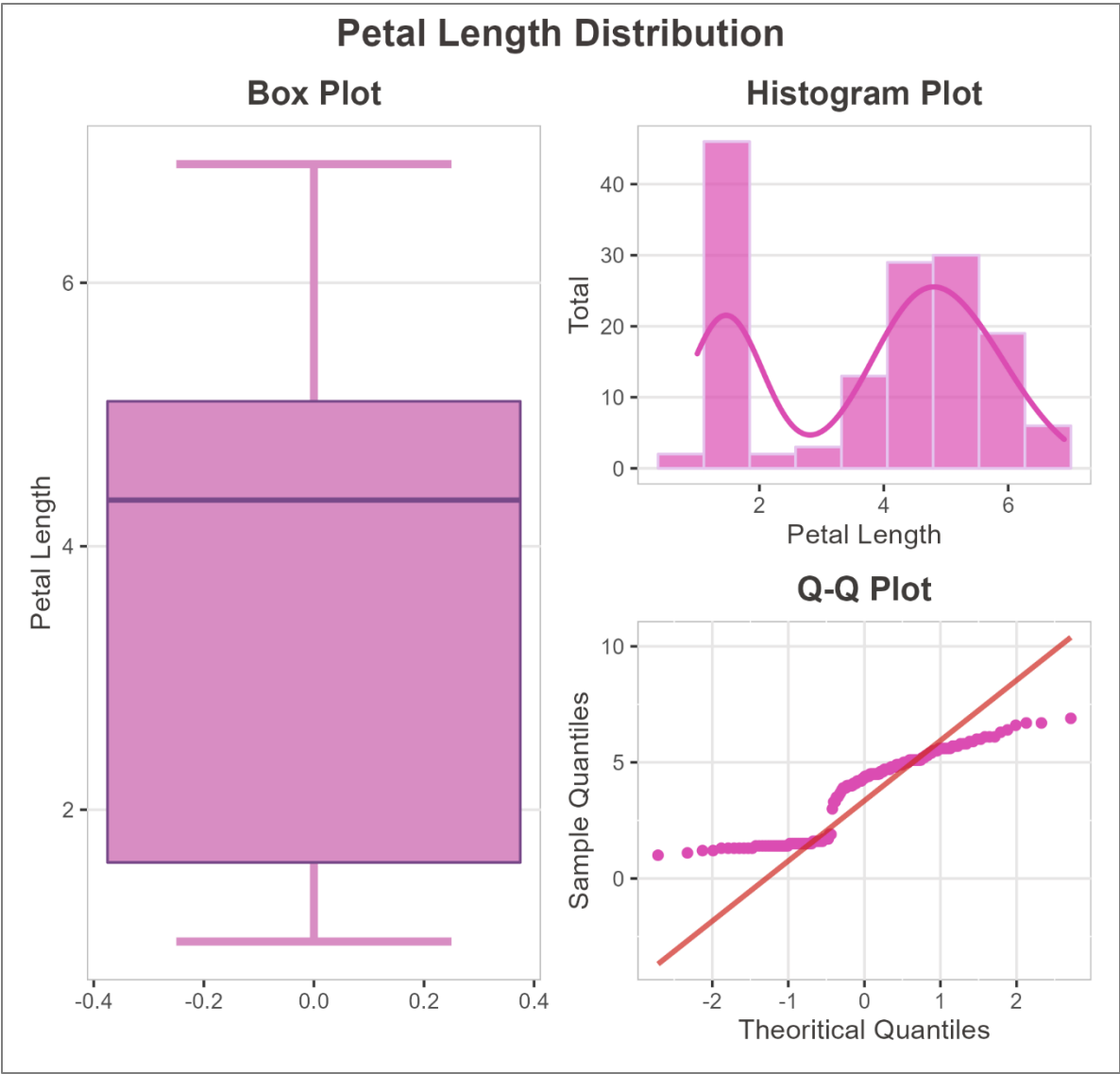
Figure 3. Petal Length Distribution

Observation

- On Figure 3, the Distribution of the Petal Length is a Bimodal Distribution or having two peaks because this column has 2 different groups or there are 2 data points where the data stop increasing and start decreasing as shown on the Histogram Plot. Moreover, the skewness of this column (-0.27). Seeing the contents of this dataset, we can assume that the two peaks indicate two differing groups from the type of Iris Species that has indistinguishable pattern, which also means that we might be easily identify the Species based on this column or data.

- On Box Plot, we can see that there is no outlier and the median has a value above 4 and is near the Upper Quantile Q3. Moreover, the kurtosis of this column is 1.60 that is a Platykurtic or Negative Kurtosis.

- On Q-Q Plot, we can that there is a gap between 45-degree line with the upper and bottom part of Q-Q Plot, this might be because of Bimodal Distribution of the data.
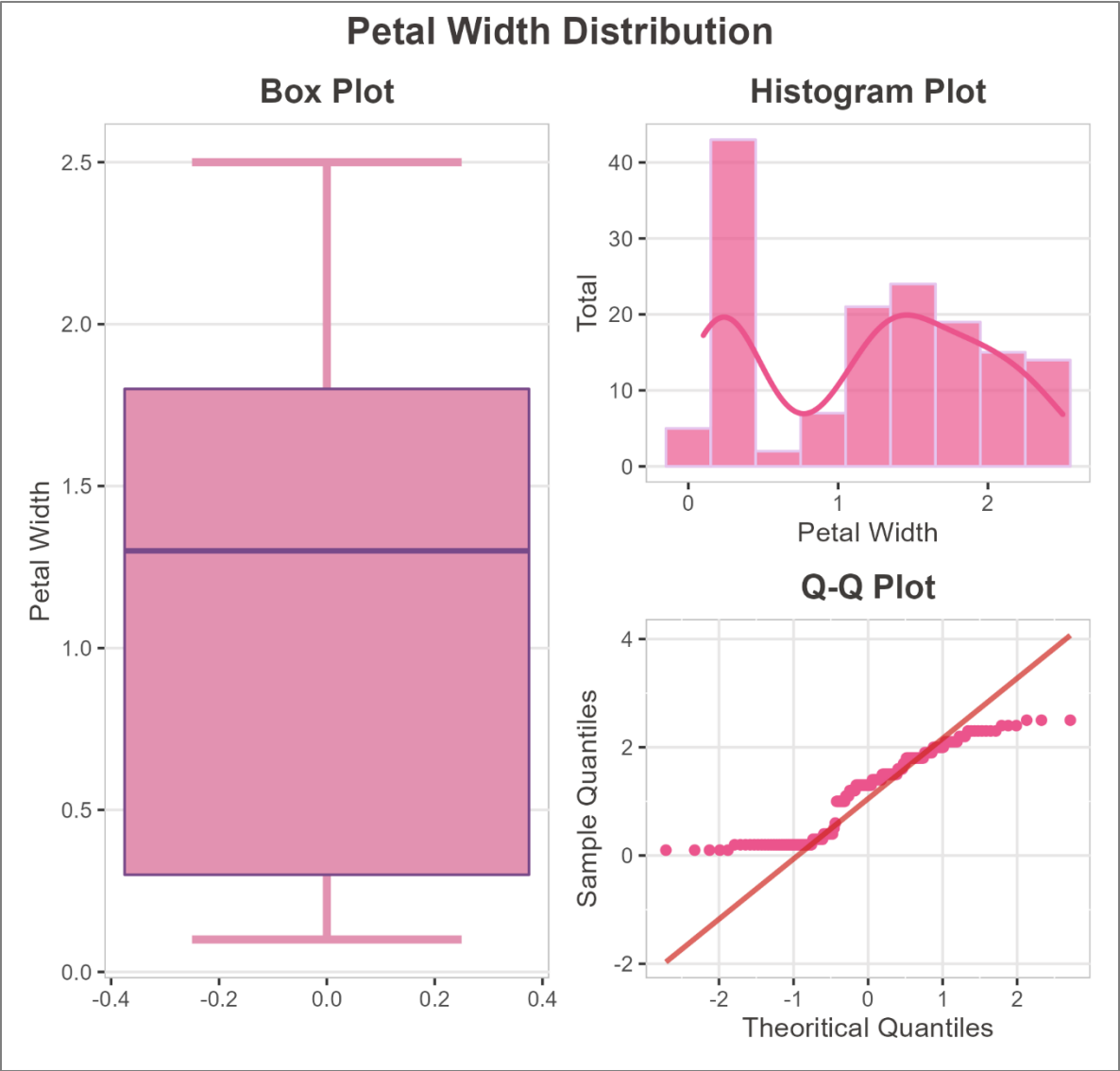
Figure 4. Petal Width Distribution

Observation

- On Figure 4, the same as Figure 3, the Distribution of the Petal Width is Bimodal as shown on the Histogram Plot. Moreover, the skewness of this column (-0.10). Seeing the contents of this dataset, we can assume that the two peaks indicate two differing groups from the type of Iris Species that has indistinguishable pattern, which also means that we might be easily identify the Species based on this column or data.

- On Box Plot, we can see that there is no outlier and the median has a value above 1 and is near the Upper Quantile Q3. Moreover, the kurtosis of this column is 1.66 that is a Platykurtic or Negative Kurtosis.

- On Q-Q Plot, we can that there is a gap between 45-degree line with the upper and bottom part of Q-Q Plot, this might be because of Bimodal Distribution of the data.
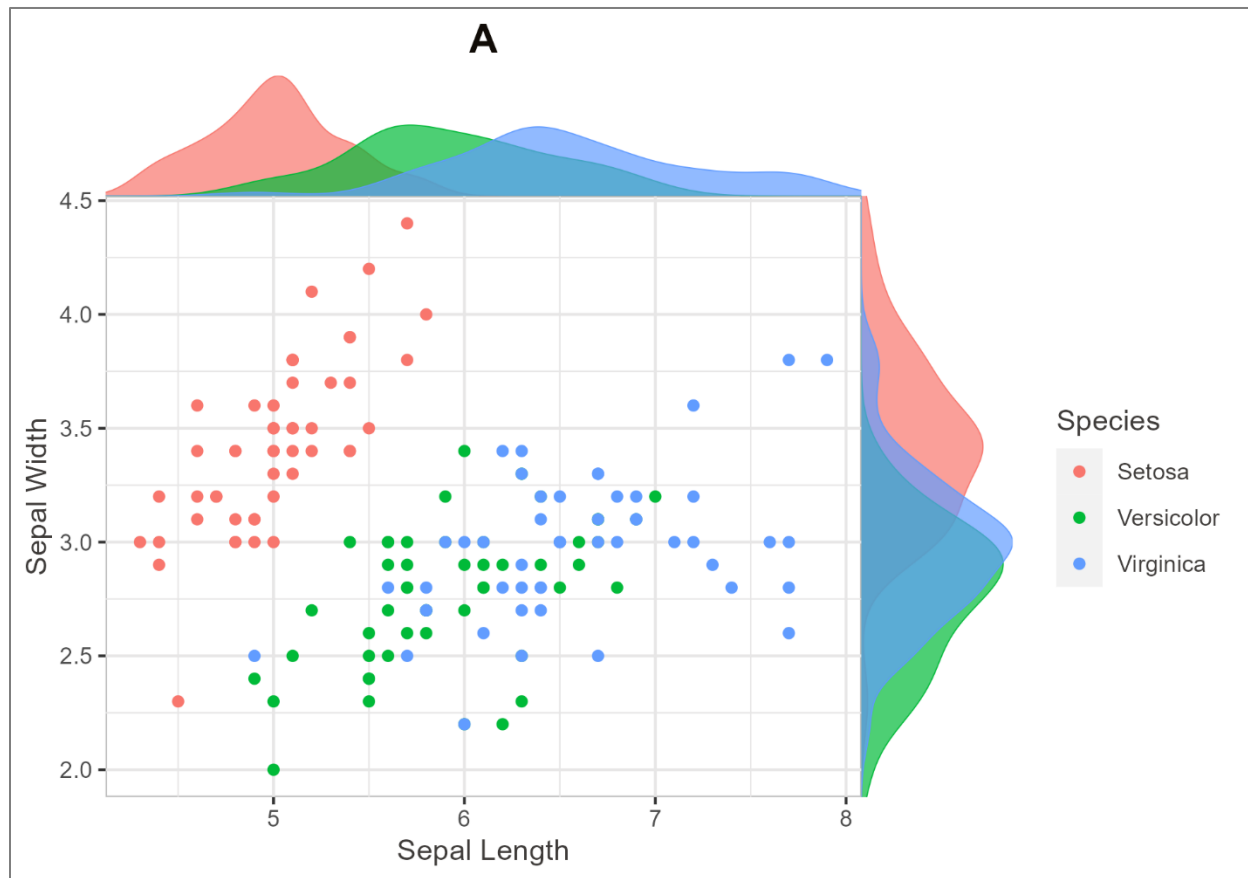
Figure 5. Scatter Plot of Sepal Width and Length

Observation

- In Figure 5, we can discern that Setosa, one of the Species of Iris has smaller length of sepal but has wider or higher width. Thus, Setosa are well separated from others which is easily distinguishable. However, Versicolor and Virginica has overlapping length and width of Sepals which makes it hard to identify each other. However, Versicolor has short sepal length but smaller sepal width than Virginica.

- The top and right of the Scatter Plot, they are Density Plot which shows the Data Distribution of each Species. The top is the Density Plot of Sepal Length and the right is for the Sepal Width. Moreover, the same as explain above, Setosa can be separated with some overlaps. As for the case for Species Versicolor and Virginica, you can see that there are really some overlaps between the two especially the Sepal Length which makes it hard for the machine to differentiate the two.
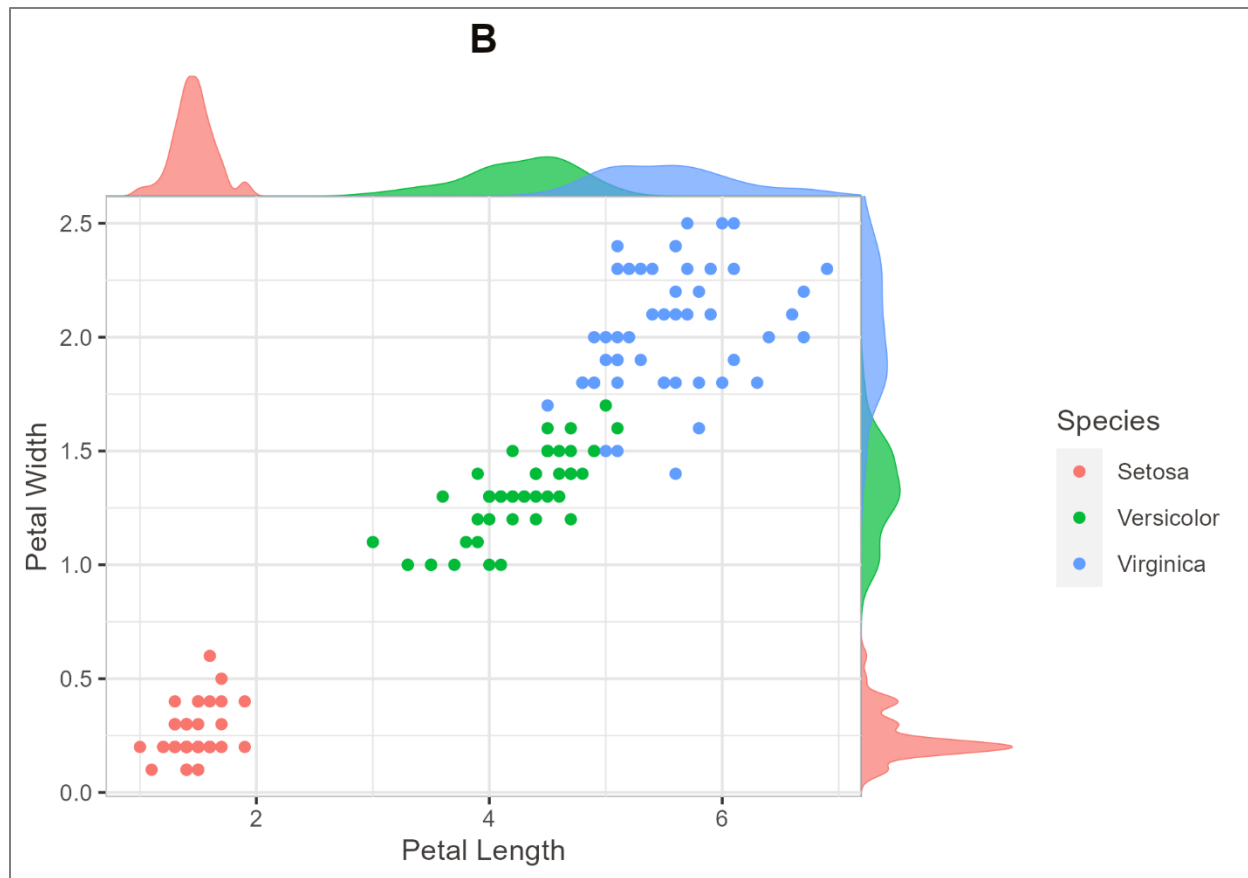
Figure 6. Scatter Plot of Petal Width and Length

Observation

- In Figure 6, like A, you can easily discern Setosa than other Species because of its smaller width and length of the petals. While Versicolor has an average petal length and width in which the Virginica has the highest length and width of the petals.

- Moreover, as shown by the plot on Figure 3 and 4, the Petal Length and Width Distribution, it is indeed that there are two groups that makes the plot a Bimodal Distribution as shown on the Density Plot of this Figure.
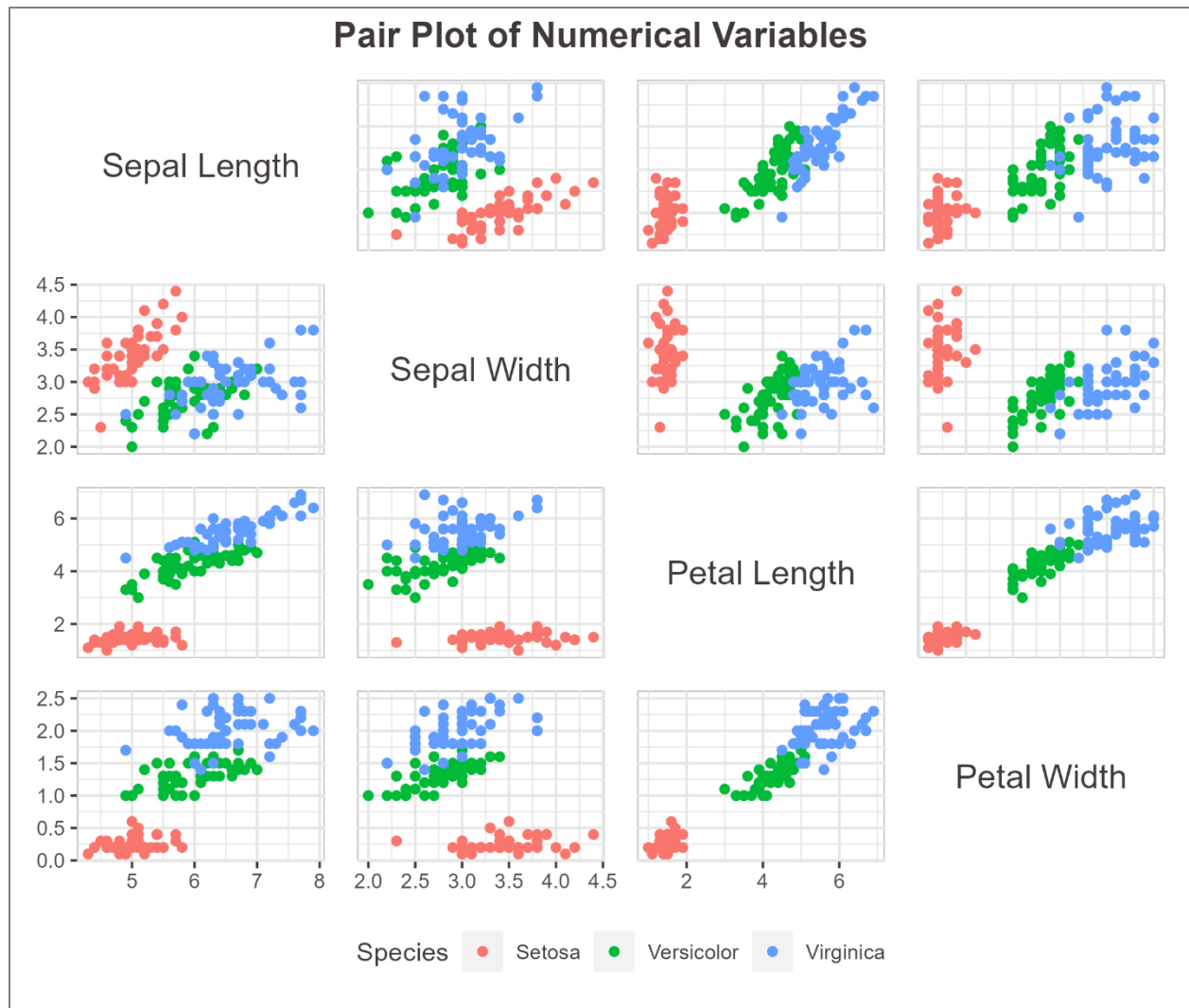
Figure 7. Pair Plot of Numerical Variables

Observation

- We expand the Scatter Plot on Figure 5 and 6 by making a Pair Plot to provide more insights about the data. Moreover, like the Scatter Plot of Sepal Length and Width on Figure 5, Iris Setosa can be separated than on the other Species which makes it easier to identify by the machine. Moreover, on this plot, we can see that all almost all plot separates the Setosa Species particularly when the plot is with Petal Length or Width.

- In Petal Length and Width, when one of the two are involved on the group or both, all the species are quite separable to each other with little overlaps. Which means that it may be the useful feature to distinguish the type of Species of Iris. On the other hand, Setosa species can easily be identified both petal and sepal.
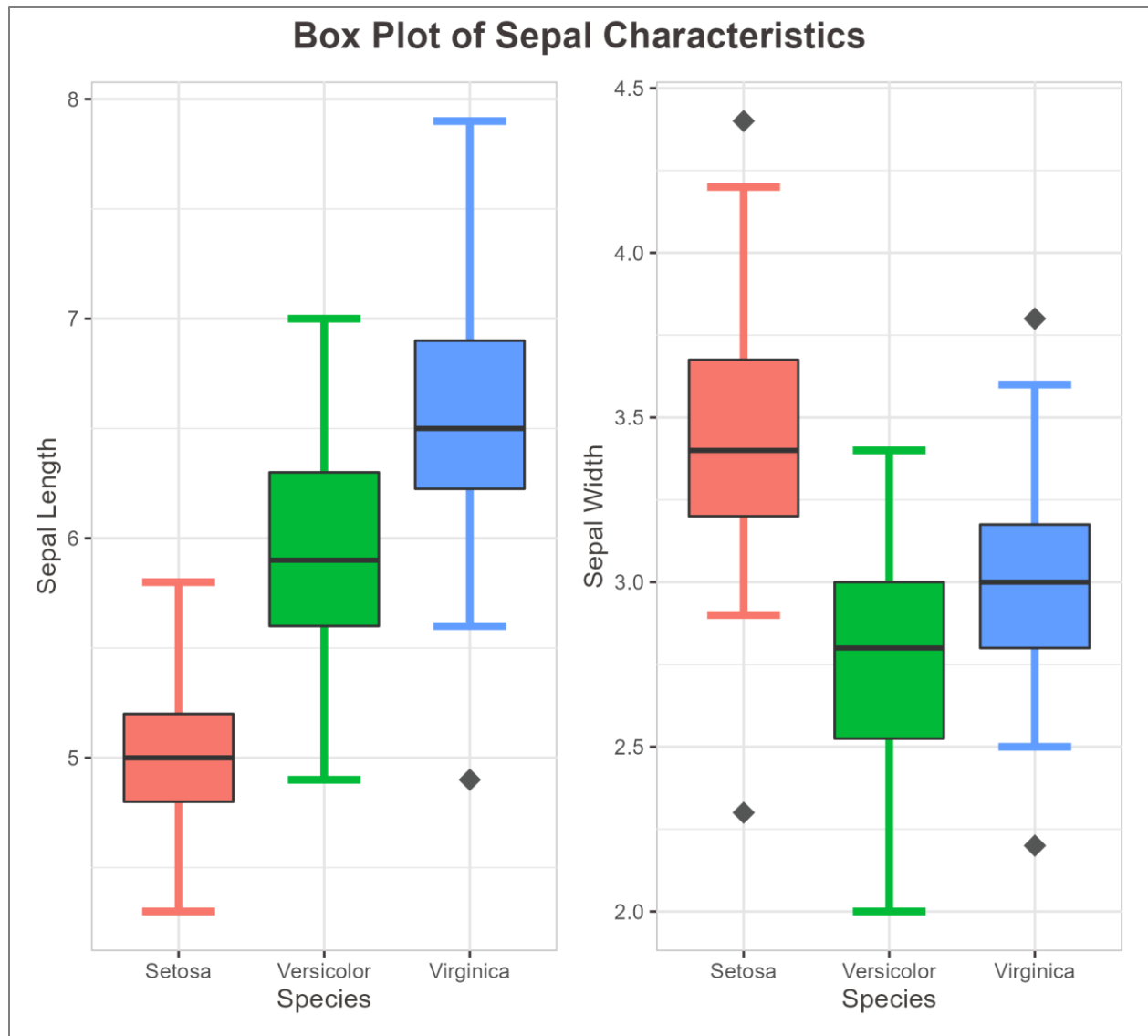
Figure 8. Box Plot of Sepal Characteristics and Species

Observation

- On this Figure, we can see that on Sepal Length, we can see that their Ranges or the whole body of the Box Plot seem to overlap a little. Also, Upper Quantile of Versicolor seem to overlap the Lower Quantile but just a little. Moreover, Virginica has an outlier with a value of 4.9.

- On the other hand, the Box Plot of Sepal Width, the ranges of all the species seem to overlap than Sepal Length Box Plot. Especially between the Versicolor and Virginica in which the median of Versicolor overlaps with the Lower Quantile of the Virginica. Moreover, Setosa have an outlier with a value 2.3 and 4.4, the Virginica also have an outlier with a value 2.2 and two 3.8.
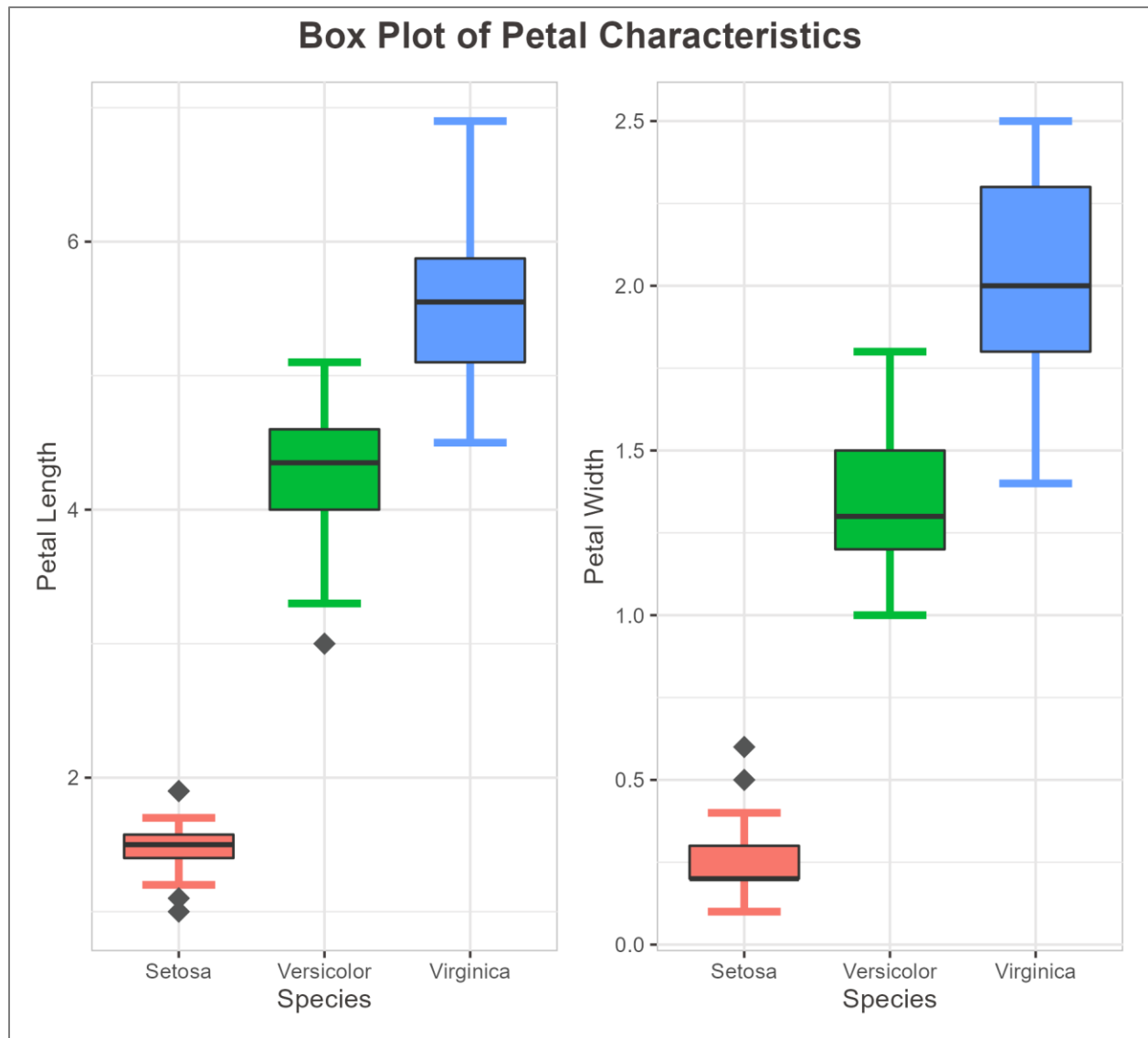
Figure 9. Box Plot of Petal Characteristics and Species

Observation

- On this Figure on Petal Length, there seems to be a quite big difference between the Species and there is a little overlap on them. Moreover, there are outliers on Setosa with a value, 1, 1.1 and two 1.9 and Versicolor also have an outlier with a value of 3.

- On the Petal Width, the Box Plot seem to be similar on the Petal Length. However, values of Petal Width seem to be higher than on the Petal Length. Moreover, only the Setosa have an outlier with a value 0.5 and 0.6.
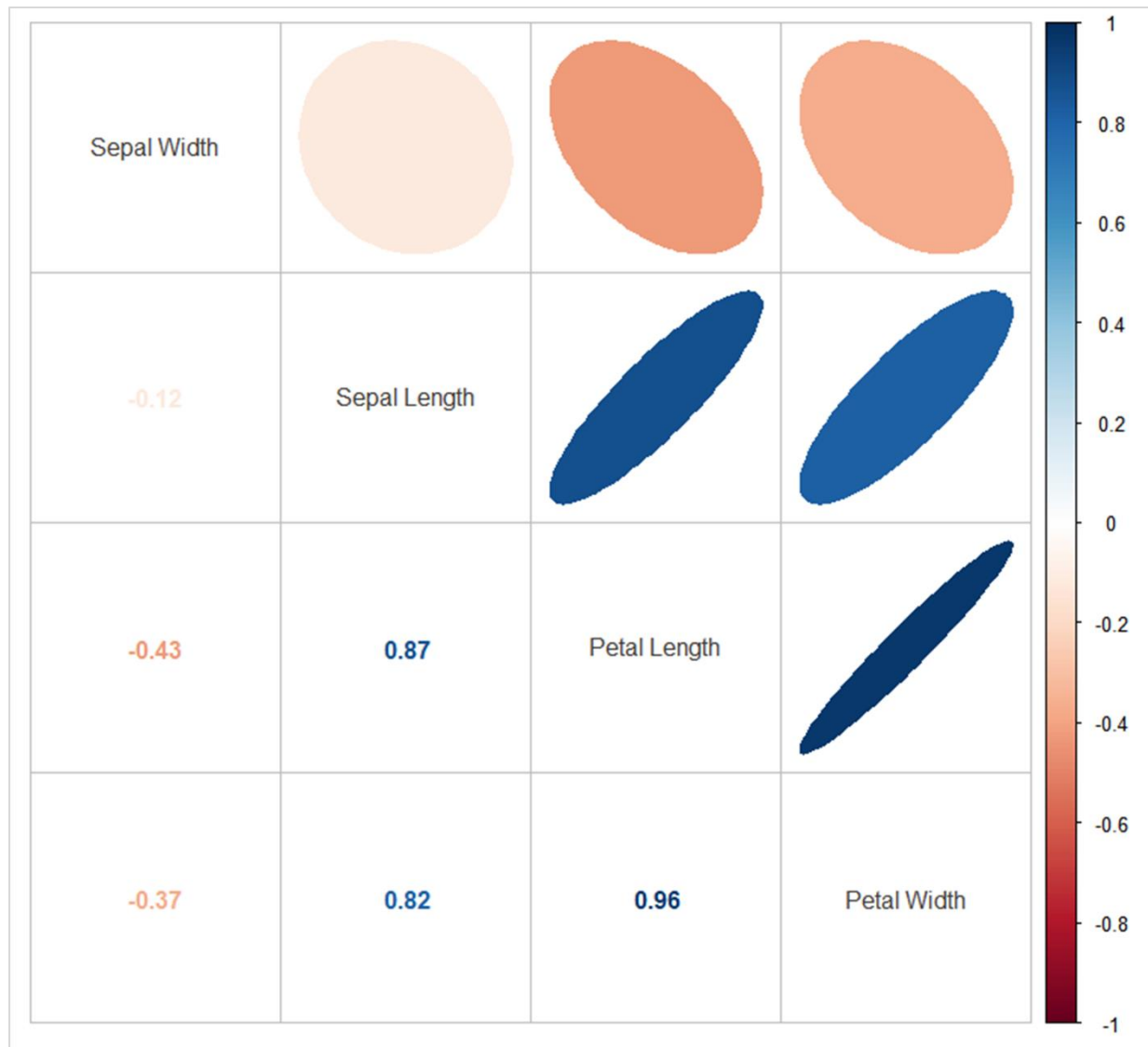
Figure 10. Correlation Plot of Numerical Variables

Observation

- On this Figure, we can see that the Petal Length and Petal Width have high correlation as suspected on earlier plots. Moreover, plot on Sepal Length and Petal Length, Sepal Length and Petal Width have also good correlation. On the other hand, Plots that involves with Sepal Width have Negative Correlation.