Iris Data Set

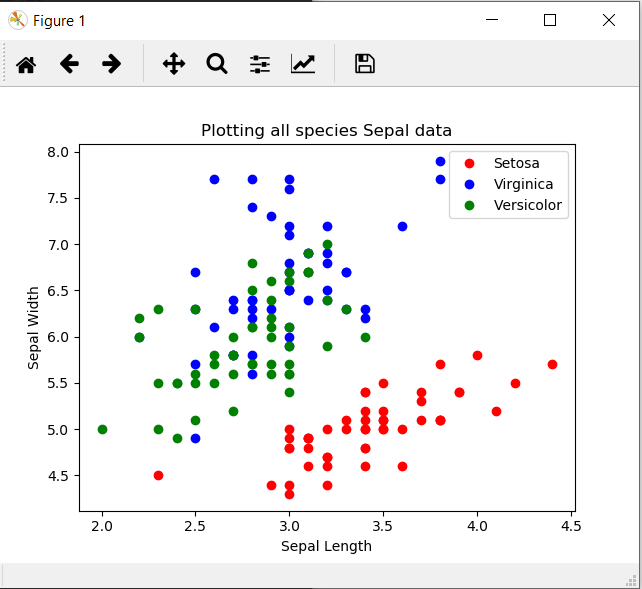
# Background Information

The Iris flower data set, or Fisher's Iris data set is a multivariate data set multivariate introduced by the British statistician and biologist Ronald in his 1936 paper The use of multiple measurements in taxonomic problems as an example of linear discriminant analysis.

The data set consists of 50 samples from each of three species of Iris (Setosa, Versicolor and virginica. Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

The dataset contains a set of 150 records under five attributes - petal length, petal width, sepal length, sepal width and species.

# Summary of Investigations



From the two figures above, you can see the differences between the three iris species. Setosa is red, Virginica is blue and Versicolor is green. The Figure on the left is for the species Petal data and the figure on the right is for the species Sepal data.

After plotting all the species in the two figures above it is clear that the setosa flower is clearly different from the other two. The setosa in red is shown away from the other two species.

In the figure on the right the similarity between the virginica and versicolor is clearly shown as the two are intertwined.