# Iris Data Set - Summary of Findings

To summarise, the data set consists of four measurements, the length and width of the petals and the length and width of the sepals. In total there are 150 Iris flowers from three species: [*Iris setosa*](https://en.wikipedia.org/wiki/Iris_setosa), [*Iris virginica*](https://en.wikipedia.org/wiki/Iris_virginica) and [*Iris versicolor*](https://en.wikipedia.org/wiki/Iris_versicolor)*.*

## Statistical Analysis

These are some parametric and non parametric statistics of our dataset. Parametric: mean, std, min, max, count. Non Parametric: 25%, 50%, 75%

sepal\_length sepal\_width petal\_length petal\_width

count 150 150 150 150

mean 5.843333 3.054000 3.758667 1.198667

std 0.828066 0.433594 1.764420 0.763161

min 4.300000 2.000000 1.000000 0.100000

25% 5.100000 2.800000 1.600000 0.300000

50% 5.800000 3.000000 4.350000 1.300000

75% 6.400000 3.300000 5.100000 1.800000

max 7.900000 4.400000 6.900000 2.500000

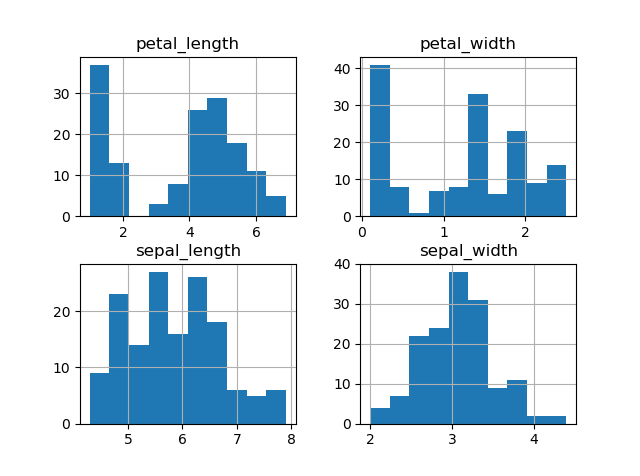
<https://stackoverflow.com/questions/49970309/how-do-i-calculate-the-mean-of-each-species-of-the-iris-data-set-in-python>

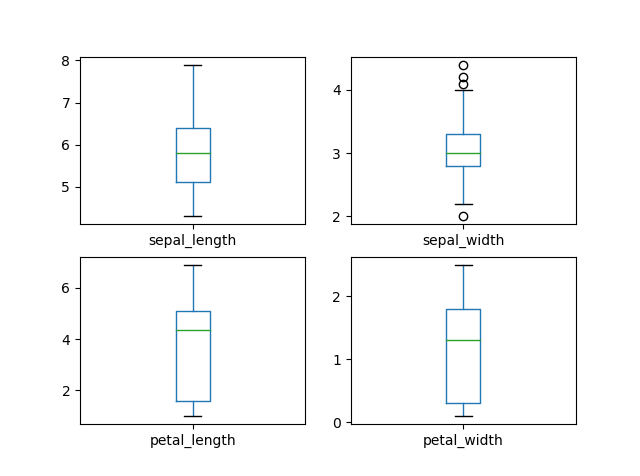
Adapting the code above, I was able to get the Mean, Median, Standard Deviation, Max Value, Min Value and Variance of each column, by species of flower (See Stats.py).

## Univariate Plot

Univariate Plot: shows the plots of each individual variable. I created a box plot and histogram for all the four input variables which shows the distribution of input variables.

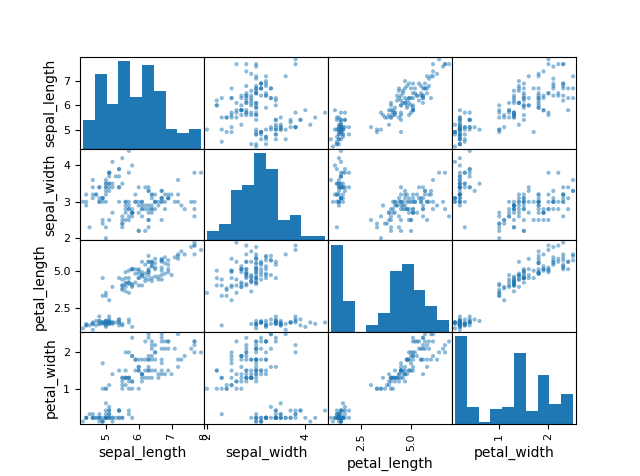
In the histogram representation of the univariate plots for each measurement, the **petal length** and **petal width** seem to **diverge** from the **normal distribution.**



The below boxplot is in the univariate form for each measurement. This highlights that it is the sepal length and sepal width that have normal distribution.

## Multivariate Plot

Scatterplots of all pairs of attributes can be helpful to spot structured relationships between input variables



The diagonal grouping of some pairs of attributes suggests a high correlation and a predictable relationship among the different species of flowers.

## Scatterplot

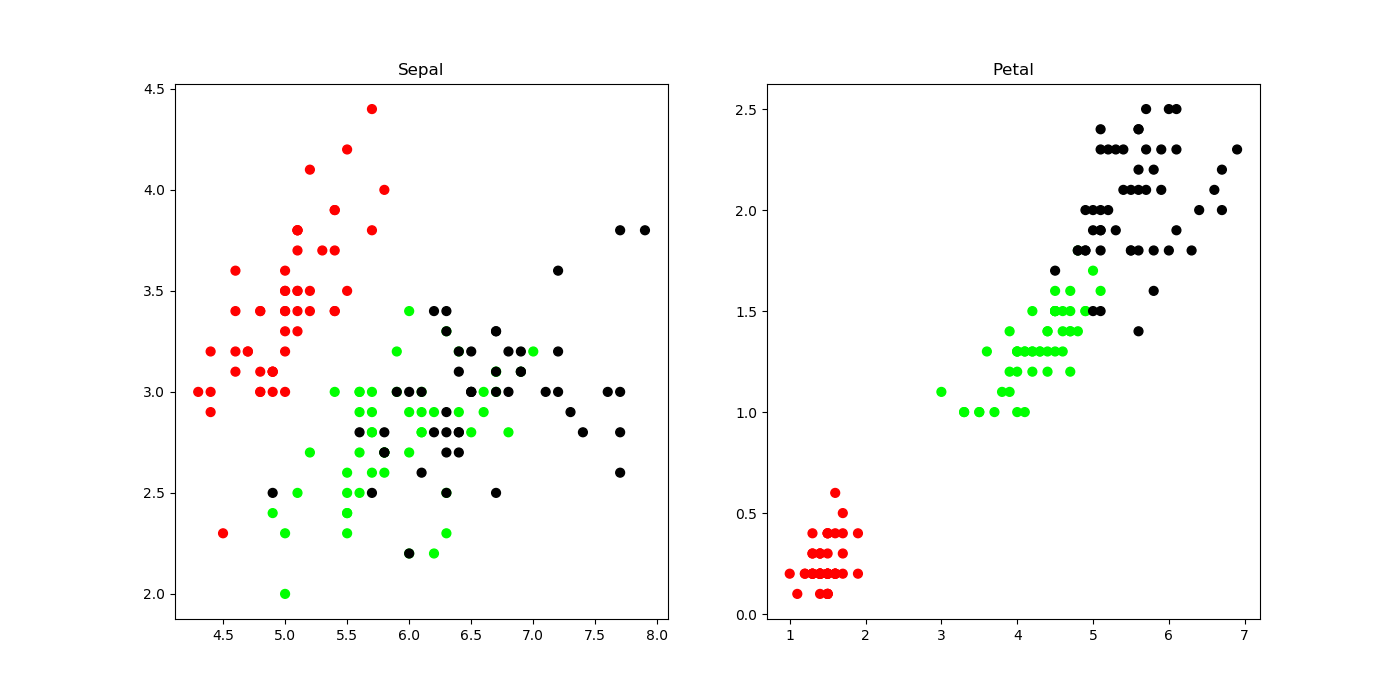
The below scatterplot plots the Sepal Width & Sepal Length, and the Petal Width & Petal Length. Using different colours it is clear that the three species have very different petal sizes

*Species:*

Iris Setosa

Iris Versicolour

Iris Virginica



*General Observations*

The Iris Virginica has the longest sepals compared to the other classes

The Iris Setosa has the widest sepals compared to the other classes

The Iris-Virginica has the longest petals compared to the other classes

The Iris-Virginica has the widest petals when compared to the other classes

*Sepal Length & Sepal Width*

Here, we can distinctly distinguish Iris Setosa, but Iris Versicolor and Iris Virginica can't be distinguished based on their sepal width and sepal length.

*Petal Length & Petal Width*

The petal length and petal width are highly correlated over all species. The Iris Setosa has the shortest petal width and length. The Iris Virginica has the widest petal width and length. The Iris Versicolour sits in middle when it comes to petal width and length.