Pands-Project

Programming and Scripting

Anna Kozakiewicz

2022

Table of Contents

[1. Readingdata.py 1](#_Toc102144038)

[2. Summaryofdata.py 1](#_Toc102144039)

[3. Plots 2](#_Toc102144040)

[3.1. Hist.py 2](#_Toc102144041)

[3.2.0. Boxplot.py 3](#_Toc102144042)

[3.2.1. Boxplot.py 4](#_Toc102144043)

[3.2.2. Boxplot.py 5](#_Toc102144044)

[3.2.3. Boxplot.py 6](#_Toc102144045)

[3.2.4. Boxplot.py 7](#_Toc102144046)

[3.3.1. Scatter.py 8](#_Toc102144047)

[3.3.2. Scatter.py 9](#_Toc102144048)

[3.4.1. Violinplot.py 10](#_Toc102144049)

[3.4.2. Violinplot.py 11](#_Toc102144050)

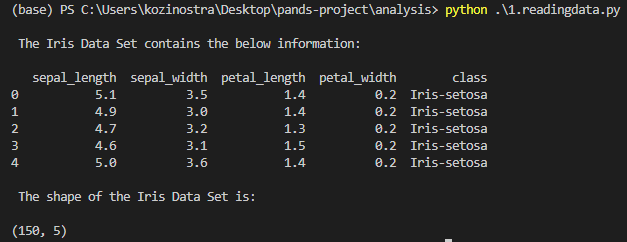
[3.4.3. Violinplot.py 12](#_Toc102144051)

[3.4.4. Violinplot.py 13](#_Toc102144052)

[4. Statistics.py 13](#_Toc102144053)

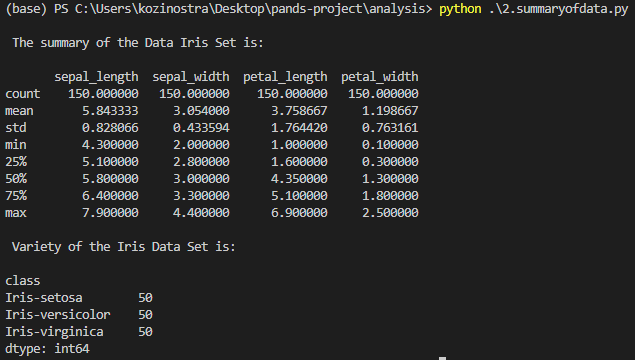
# Readingdata.py

The below shows what columns the Iris Data Set includes and what is the shape of the data (the shape is 150 rows and 5 columns)



# Summaryofdata.py

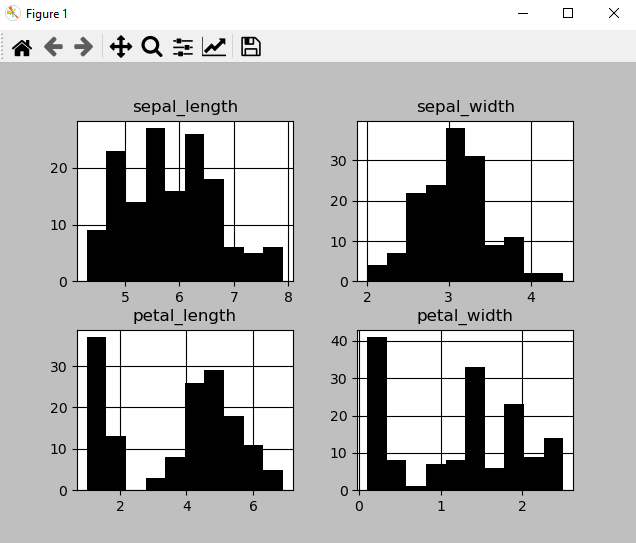
The below summary of data shows the basic statistics summary of each of 4 variables and how the data is split by each variety/class (there is 50 rows per each Iris class)



# Plots

## 3.1. Hist.py

The below histograms show the measurement of each variable in the Iris Data Set



The histogram for the sepal length shows a variation of shorter and longer sepals.

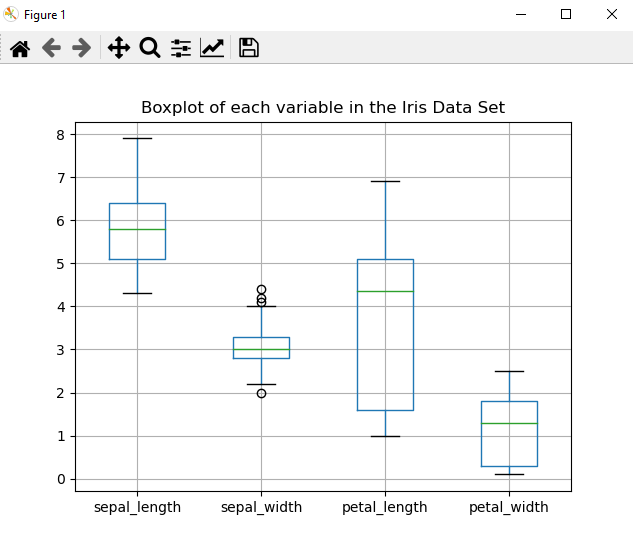
The histogram for the sepal width is very condensed in around 3 cm.

The histogram for the petal length is very divided into much shorter petals and longer ones.

The histogram for the petal width is similar to petal length histogram showing the split between the more narrow and more wide petals.

## 3.2.0. Boxplot.py

The below plot is a boxplot of each variable in the Iris Data Set.



Sepal length is between 5 and 6,5 cm with majority around 6cm, where sepal width is concentrated around 3 cm. Petal length vary between 1,5 and 5 cm with majority around 4.25 cm where petal width is between 0.25 and 2 cm with majority around 1.25 cm.

## 3.2.1. Boxplot.py

The below plot is a boxplot of petal width for 3 Iris classes



The Iris Setosa has much smaller petals than the other two classes, when the Iris Virginica has the biggest petals.

## 3.2.2. Boxplot.py

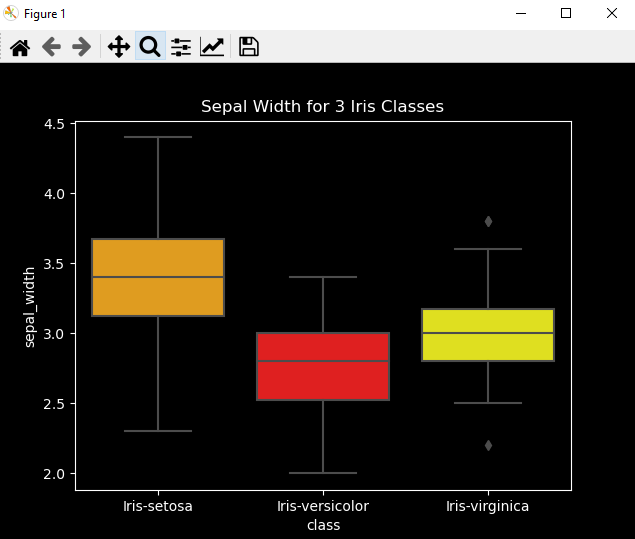
The below plot is a boxplot of petal length for 3 Iris classes



The same with the length of the petal – The iris Setosa has the smallest petals when Iris Virginica has the biggest ones.

## 3.2.3. Boxplot.py

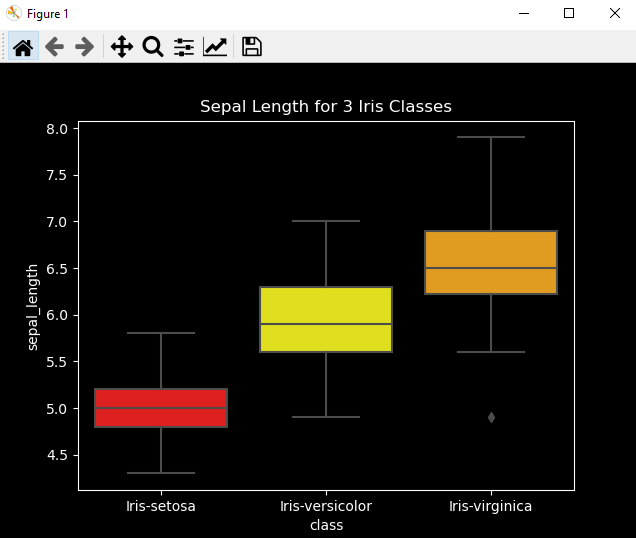
The below plot is a boxplot of sepal width for 3 Iris classes



In terms of sepal width, The Iris Setosa has the widest sepal when compared to other two classes.

## 3.2.4. Boxplot.py

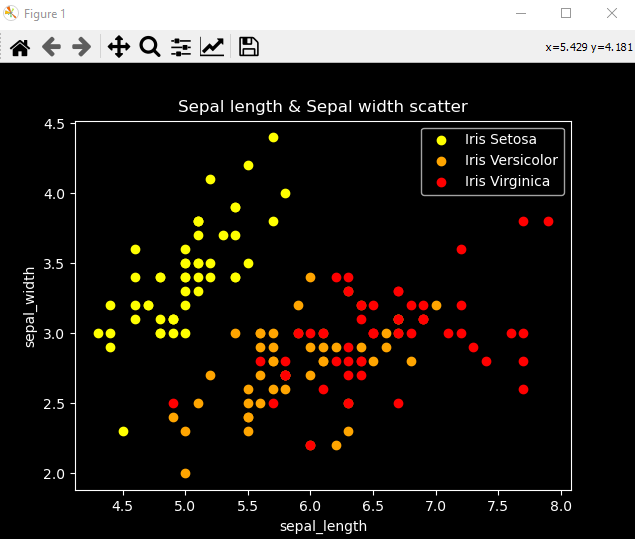
The below plot is a boxplot of sepal length for 3 Iris classes



The Iris Setosa has the shortest sepal and the Iris Virginica has the longest sepal from the three classes.

## 3.3.1. Scatter.py

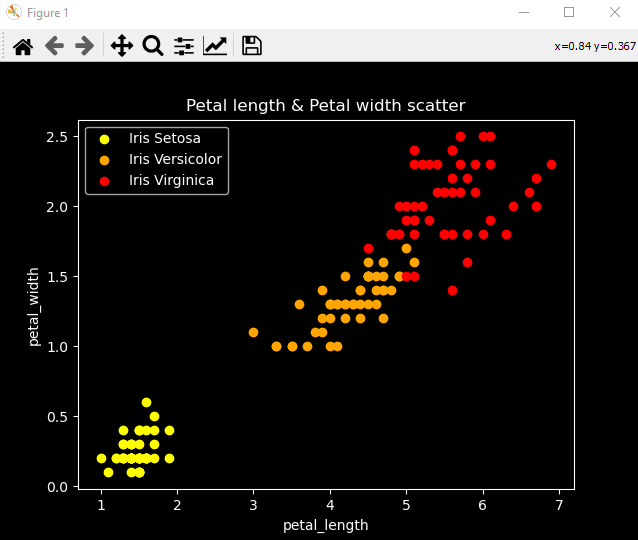
The below plot is a scatter plot of sepal length and sepal width as a one pair of variables



The above shows that sepal length and width vary in all 3 Iris classes.

## 3.3.2. Scatter.py

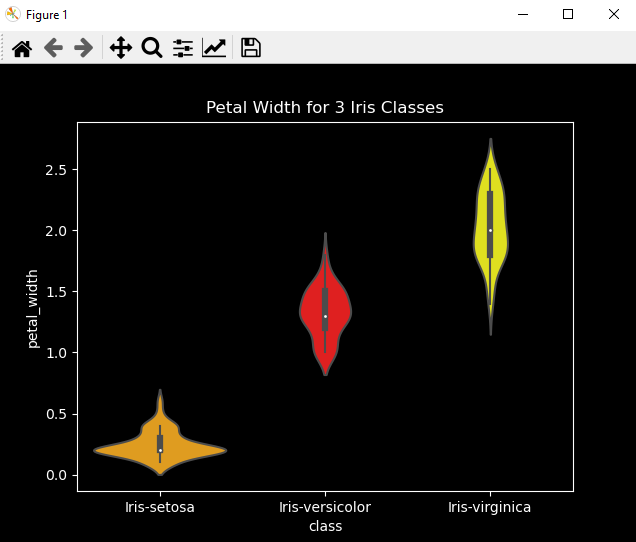
The below plot is a scatter plot of petal length and petal width as a second pair of variables



The above shows that Iris Setosa stands out of the other two classes in terms of petal length and width.

## 3.4.1. Violinplot.py

The below plot is a violin plot of petal width for 3 Iris classes



Similar like in a boxplot this violin plot shows that the Iris Setosa has much smaller petals than the other two classes, when the Iris Virginica has the biggest petals.

## 3.4.2. Violinplot.py

This code outputs a violin plot of petal length for 3 Iris classes



The same with the length of the petal and similar like in the boxplot – this violin plot shows that the iris Setosa has the smallest petals when Iris Virginica has the biggest ones

## 3.4.3. Violinplot.py

This code outputs a violin plot of sepal width for 3 Iris classes



Similar like in the boxplot this violin plot shows that the Iris Setosa has the widest sepal when compared to other two classes.

## 3.4.4. Violinplot.py

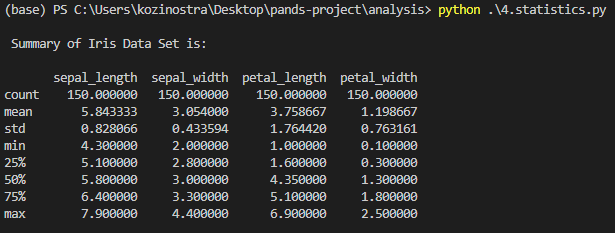
This code outputs a violin plot of sepal length for 3 Iris classes



Similar like in the boxplot this violin plot shows that the Iris Setosa has the shortest sepal and the Iris Virginica has the longest sepal from the three classes.

# Statistics.py

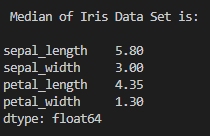
The below summary shows the basic statistics per each variable



Sepal length has the highest mean, min and max, where petal length has the highest std.

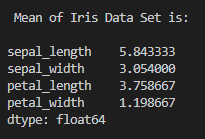
**Median** – is the middle value in the data set.

The below shows the median of each variable in the Iris Data Set

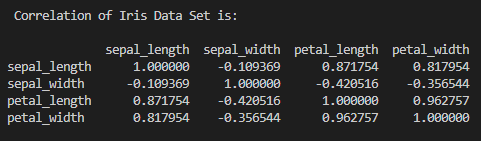


**Mean** – is the average value of the data set.

The below shows the mean of each variable in the Iris Data Set

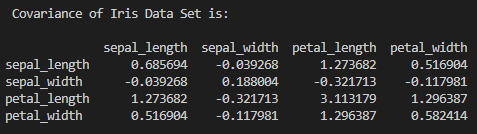


**Correlation** – measures the relationship between two variables



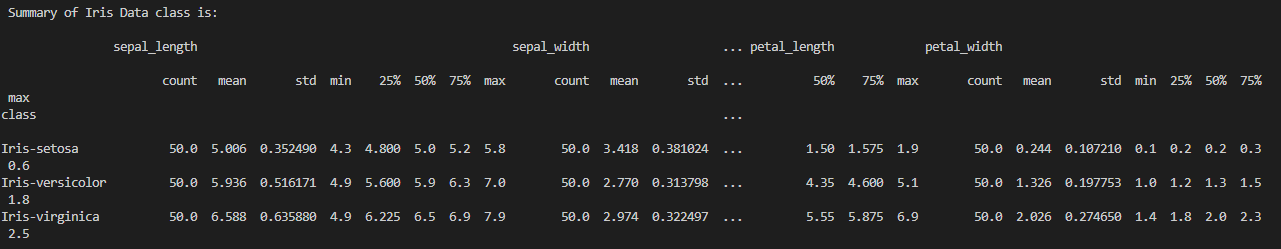
There is a big correlation of 0.87 between petal and sepal length and 0.82 between sepal length and petal width. There are negative correlations between sepal length and width, petal length and sepal width; and sepal and petal width.

**Covariance** – indicates the direction of the linear relationship between variables

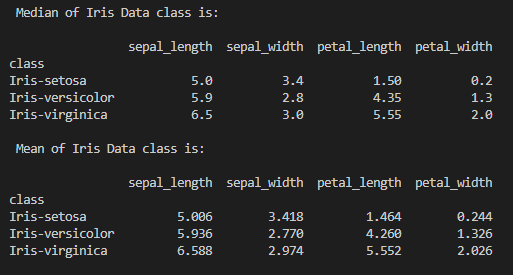


Similar like in correlation there is a big covariance of 1.27 between petal and sepal length.

The below summary shows the summary statistics for each class of the Iris Data Set

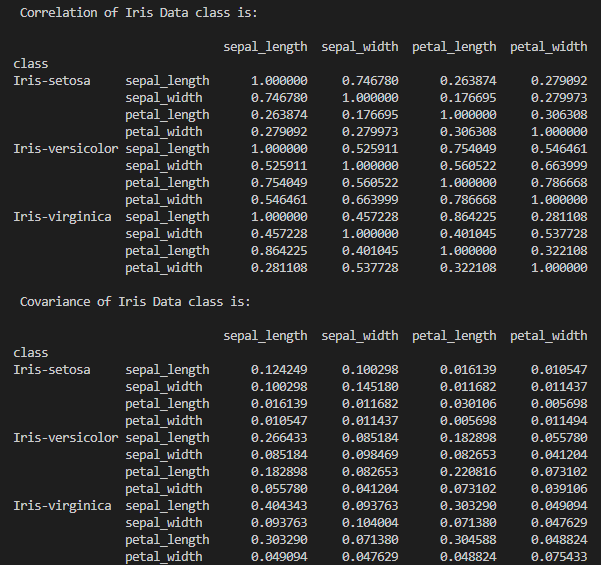


**Median & Mean** of each class



The highest median has Iris Virginica in sepal length and the smallest has Iris Setosa in petal width. The same occurs in mean.

**Correlation & Covariance** of each class



There is a strong correlation of 0.86 between sepal length and petal lengths in Iris Virginica and a high correlation of 0.75 between sepal length and width in Iris Setosa. There is also a high correlation of 0.75 in Iris Versicolor between sepal and petal length. Similar results are in covariance by each class.