**DATA SCIENCE AND MACHINE LEARNING**

**ASSIGNMENT**

Submitted by,

Pranav P K

S3 RMCA-B

Rollno:22

**Data visualization – Histogram, Quartile plot, Distribution chart, Scatter plot, Bubble chart, Density chart**

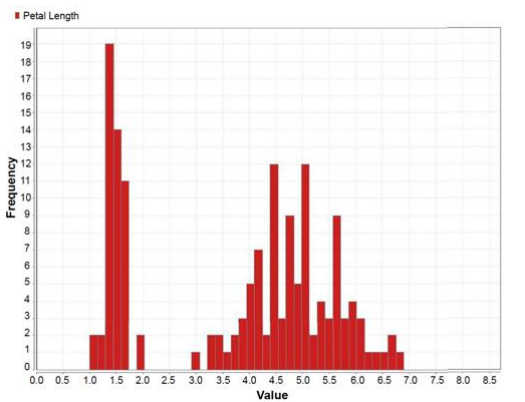
Visualizing data is one of the most important techniques of data discovery and exploration

• Comprehension of dense information: A simple visual chart can easily include thousands of data points. By using visuals, the user can see the big picture, as well as longer term trends that are extremely difficult to interpret purely by expressing data in numbers

• Relationships: Visualizing data in Cartesian coordinates enables exploration of the relationships between the attributes.

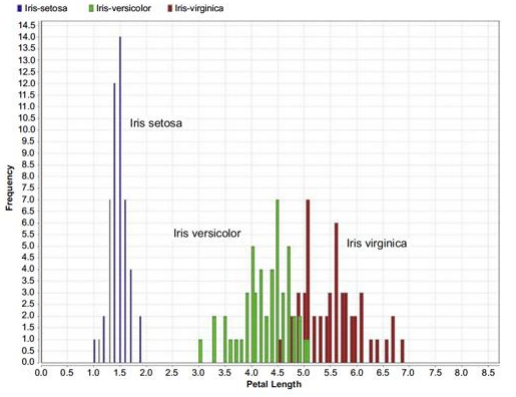
Histogram

It shows the distribution of the data by plotting the frequency of occurrence in a range. In a histogram, the attribute under inquiry is on the horizontal axis and the frequency of occurrence is on the vertical axis.

****

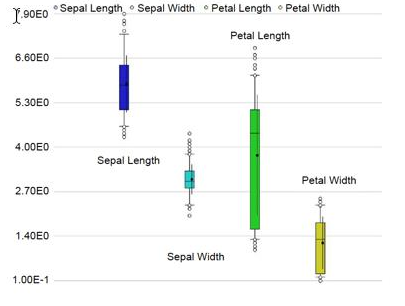
Class stratified Histogram.

A histogram can be stratified to include different classes in order to gain more insight

****

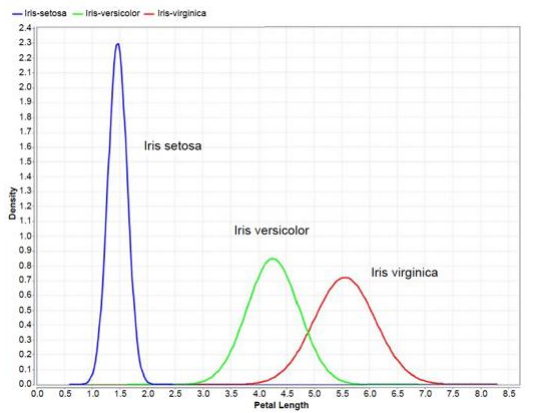
Quantile plot

A box whisker plot is a simple visual way of showing the distribution of a continuous variable with information such as quartiles, median, and outliers, overlaid by mean and standard deviation.

****

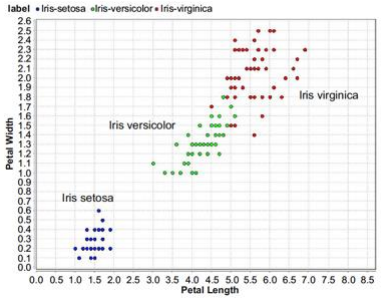
Distribution plot

For continuous numeric attributes like petal length, instead of visualizing the actual in the sample, its normal Distribution function can be visualized.

****

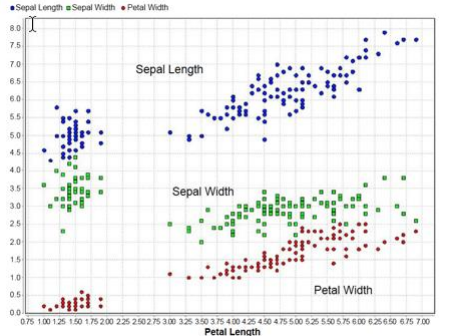
Scatter plot

The multivariate visual exploration considers more than one attribute in the same visual. the dataset aligned with the coordinates. The attributes are of continuous data type.

****

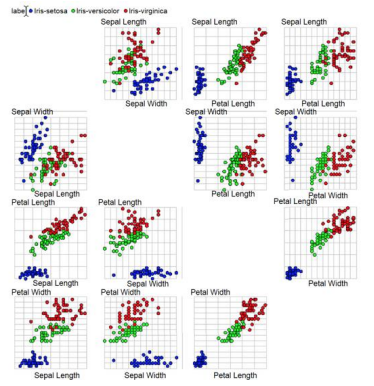
Scatter multiple

A scatter multiple is an enhanced form of a simple scatterplot where more than two dimensions can be included in the chart and studied simulta neously. The primary attribute is used for the x-axis coordinate. The secondary axis is shared with more attributes or dimensions.

****

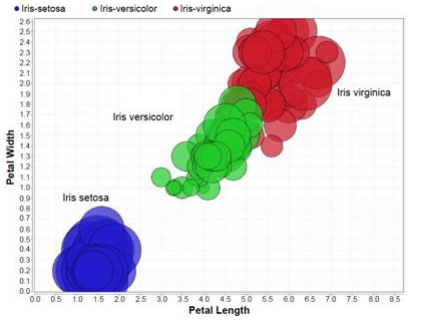
Multiple Scatter matrix

If the dataset has more than two attributes, it is important to look at combi nations of all the attributes through a scatterplot. A scatter matrix solves this need by comparing all combinations of attributes with individual scatterplots and arranging these plots in a matrix.

****

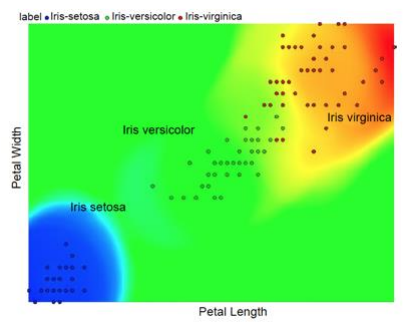
Bubble plot

A bubble chart is a variation of a simple scatterplot with the addition of one more attribute, which is used to determine the size of the data point.

****

Density chart

Density charts are similar to the scatterplots, with one more dimension included as a background color. The data point can also be colored to visual ize one dimension, and hence, a total of four dimensions can be visualized in a density chart.

****