INF 552 Assingment 3 Q3
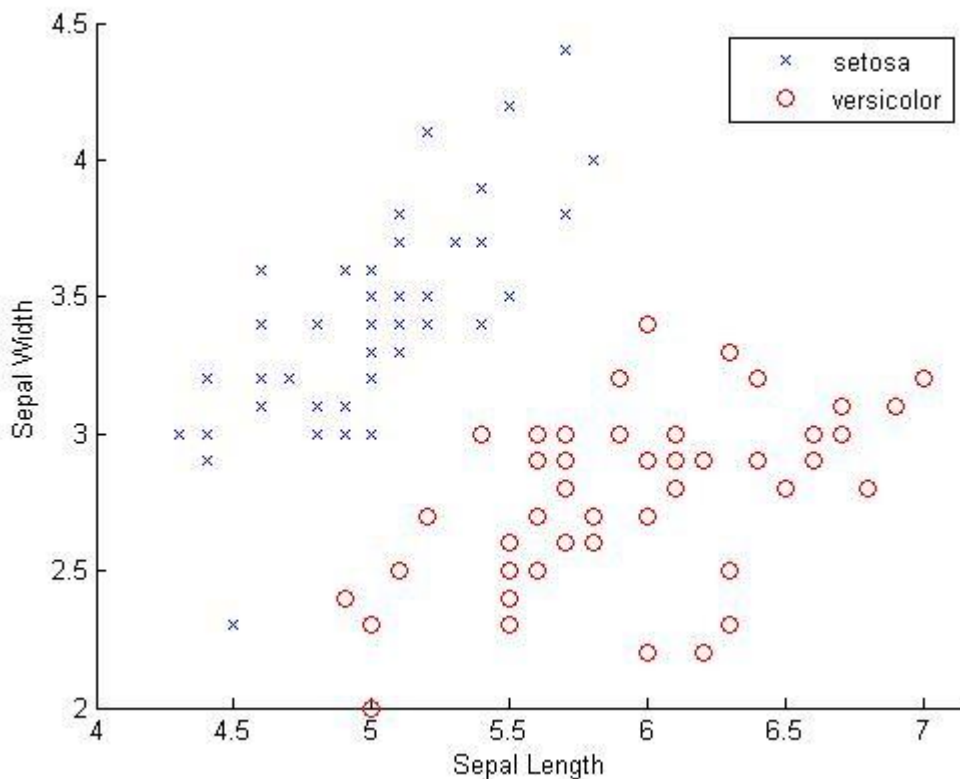
-Aamir Goriawala

1) Scatter Plot

>>load fisheriris

gscatter(x,y,species(1:100),'br','xo')

 xlabel('Sepal Length');

 ylabel('Sepal Width');

A scatter plot is a simple plot of one variable against another that reveals the relationship among variables. The scatter plot above shows the correlation between sepal length and sepal width in two species of Iris (setosa and versicolor).

From this plot, one can say that (i) setosa sepals tend to be shorter and wider than versicolor sepals and (ii) sepal lengths and widths appear to be directly proportional to each other

2) Gplotmatrix

>>load fisheriris

gplotmatrix(meas,meas,species)
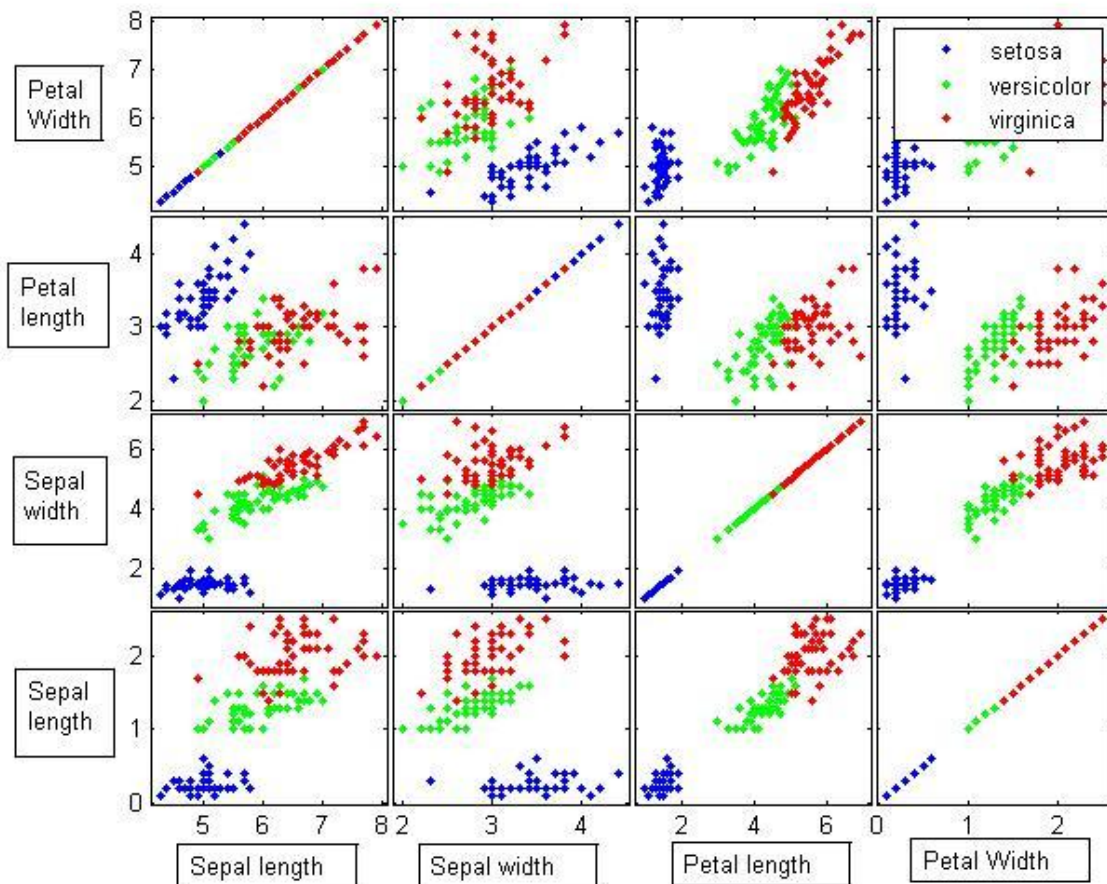
Gplotmatrix creates a matrix of scatter plots. Each individual set of axes in the resulting figure contains a scatter plot of a column of x against a column of y. All plots are grouped by the grouping variable group. The points in the graphs on the diagonal lie on a straight line since they are plotted against themselves so they are to be ignored.

The Gplotmatrix gives us multiple insights into the data. We can that versicolor and virginica flowers are more similar to each other in terms of sepal and petal widths and lengths as compared to setosa flowers. Moreover, setosa variables tend to correlate less with each other ie they do not form a sort of cluster. The Gplotmatrix provides more details than the simple scatterplot.

3) Box Plots

>>load fisheriris

boxplot([meas(1:50,3:3),meas(51:100,3:3),meas(101:150,3:3)],'labels',

{'setosa','versicolor','virginica'});
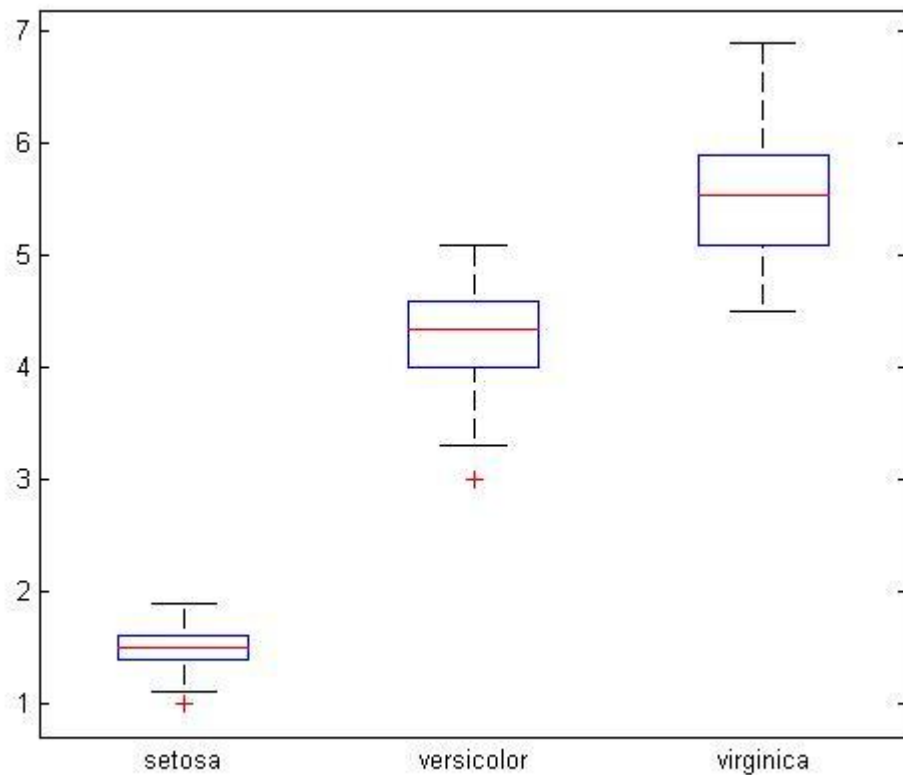
The boxplot shown above compares petal lengths in samples from the three species of iris. A box plot gives us the statistical distribution of data by displaying the max,min,median and inter quartiles.

- The tops and bottoms of each box are the first and third quartile respectively. The distance between them forms the inter- quartile range.
- The line in the middle of each box is median.
- The whiskers are lines extending to the furthest observations which are within 1.5 times of the inter quartile ranges from the top and bottom respectively. Any points outside the whisker are outliers