Visualizing Fisher's Iris Data

s1250172 Yuto Kimura

I. Explanation on Target Data

・The data set consists of 50 samples of three types (Iris setosa, Iris virginica and Iris versicolor). From each sample, the length and width of the scale, and the length and width of the petals are measured in centimeters.

II. Visualization Results

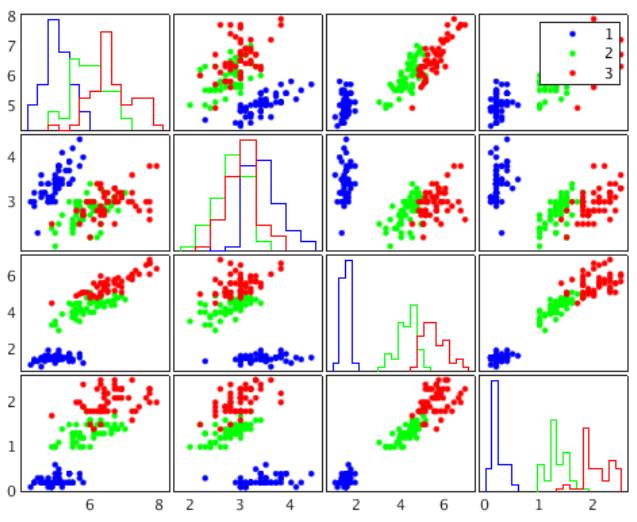

Fig. 1: Scatterplot matrix

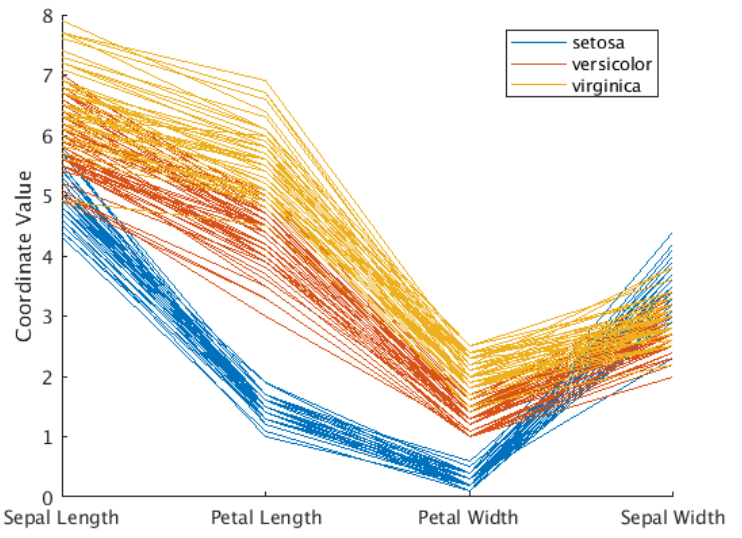

Fig. 2: Parallel coordinate plots

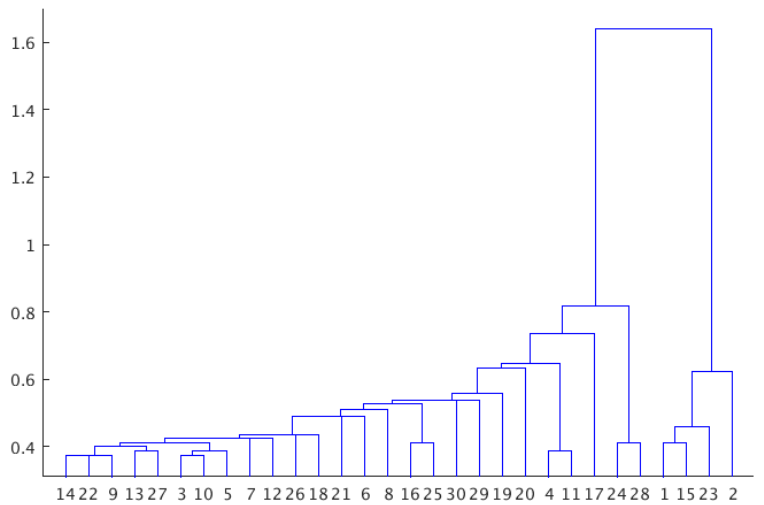

Fig. 3: Hierarchical clustering

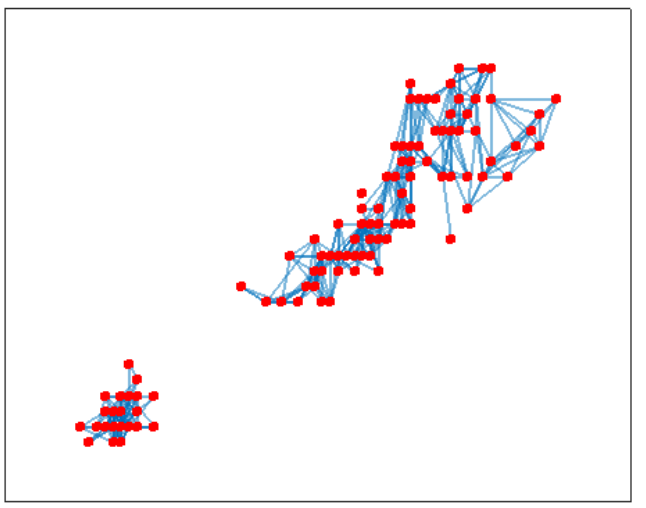

Fig. 4: Proximity graph

III. Methods and Discussion

- I used matlab respectively.
- Scatter plot matrices are a great way to roughly determine if there is a linear correlation between multiple variables. This is particularly useful for identifying specific variables that may have similar correlations with genomic and proteome data.
- Parallel coordinate plots is that A more compact representation for multivariate data samples is wanted. Parallel coordinate plots are ideal for comparing many variables and verifying the relationships between them.
- In data mining and statistics, hierarchical clustering is a method of cluster analysis that seeks to build a hierarchy of clusters.
- A proximity graph is a simply a graph in which two vertices are connected by an edge if and only if the vertices satisfy particular geometric requirements.