

Cluster Analysis: Dail Voting Data

Clustering, which is used for identifying the distribution of patterns and intrinsic correlations, can point out similarity and divide data into various classes. This is the case for Dail voting dataset in the Irish Parliament where items are recorded as binary variables. In order to detect the similarities and differences between politicians as well as their voting behaviors, this paper will make a comparison between k-means and hierarchical clustering.

➤ K-means Clustering

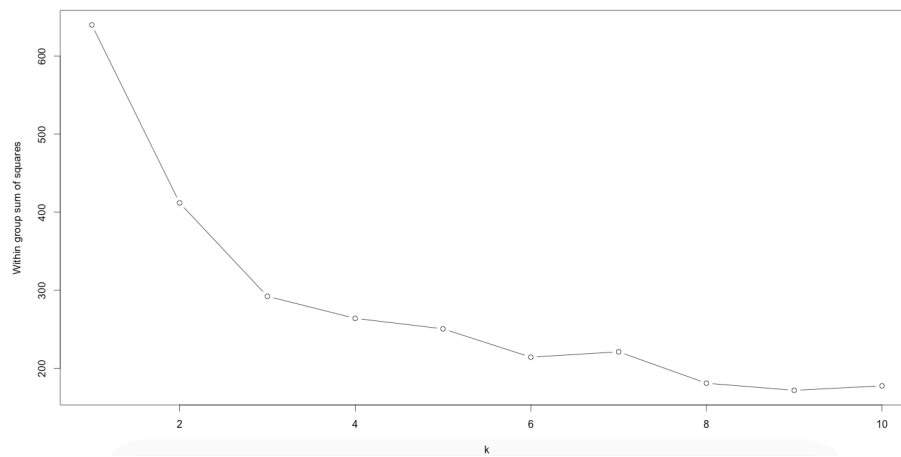


Figure 1. K-means Clustering

We generally run the k-means over a range of values of K. The first question we are interested in is 'how tightly packed the clusters are'. As shown in Figure 1, it records the within group sum of squares for each value of k and then plot k versus WGSS. It suggests k = 8 could be a good clustering solution where the graph flattens quickly.

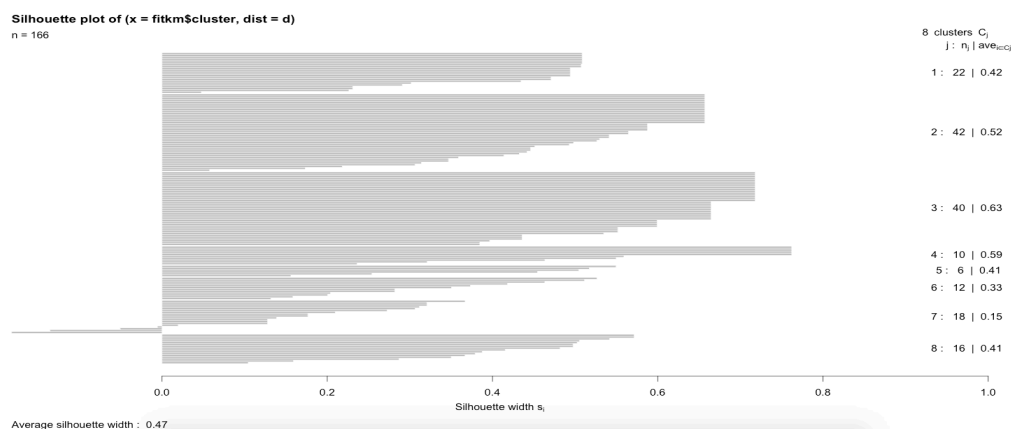


Figure 2. Assessing clustering results

Cluster Dendrogram

Height

0 1 2 3

hclust ("average")

In Figure 4, values at branches are AU p-values and BP values, and cluster labels (bottom). Clusters are highly supported by the data with large p-values. Politicians in each cluster tend to have the similar voting behaviors (e.g., politicians in the first rectangle from the right have similar action for voting 1 & 2).