Nicole Kayambu
4341-02
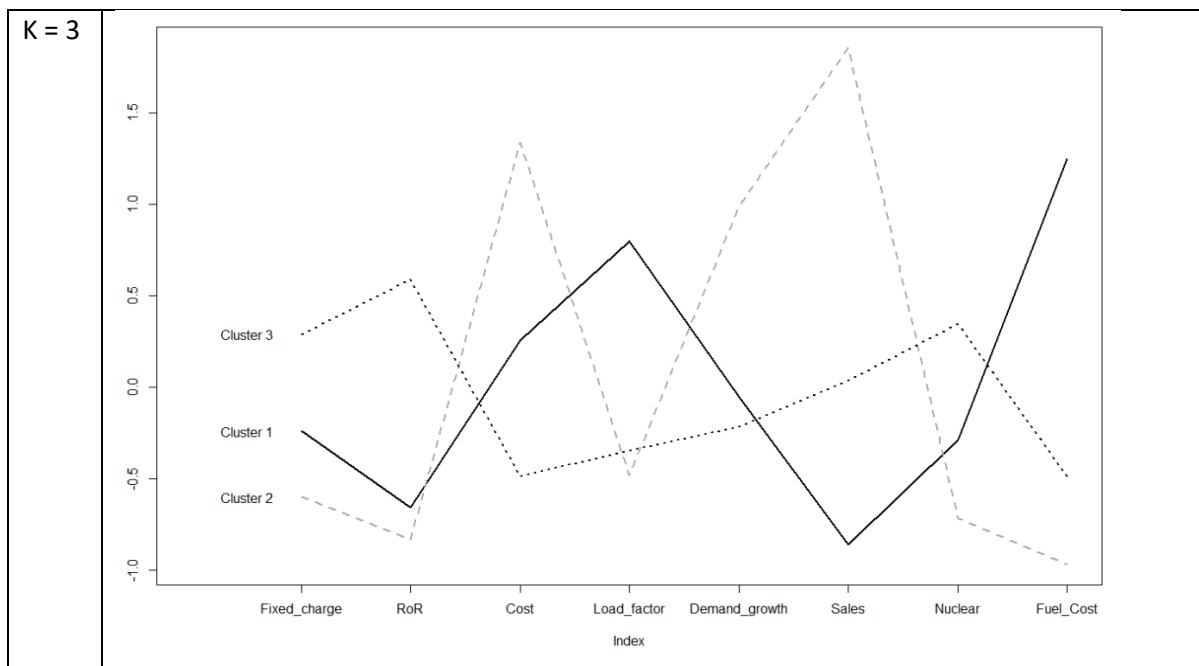
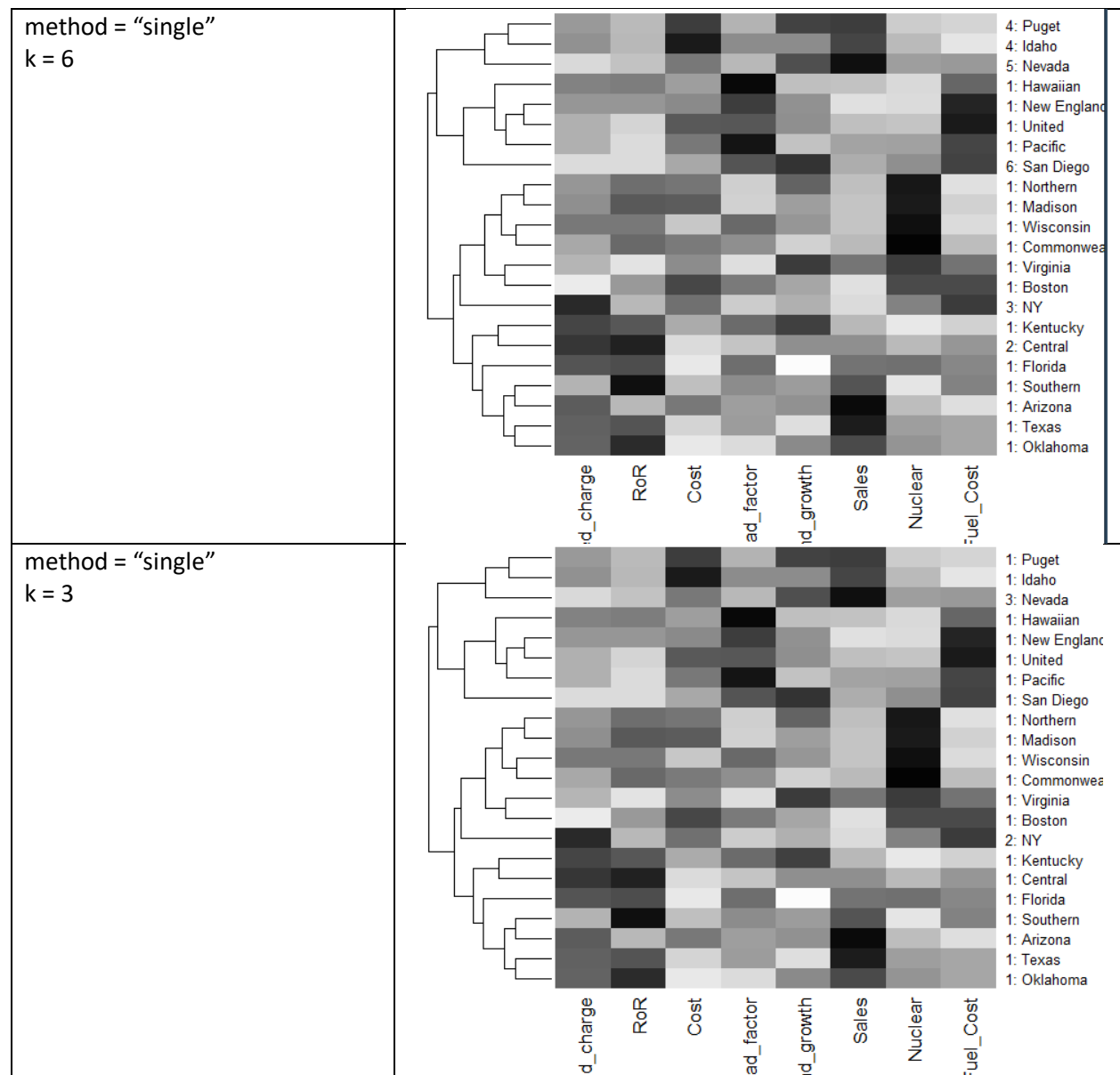Homework 5 Machine Learning - Clustering
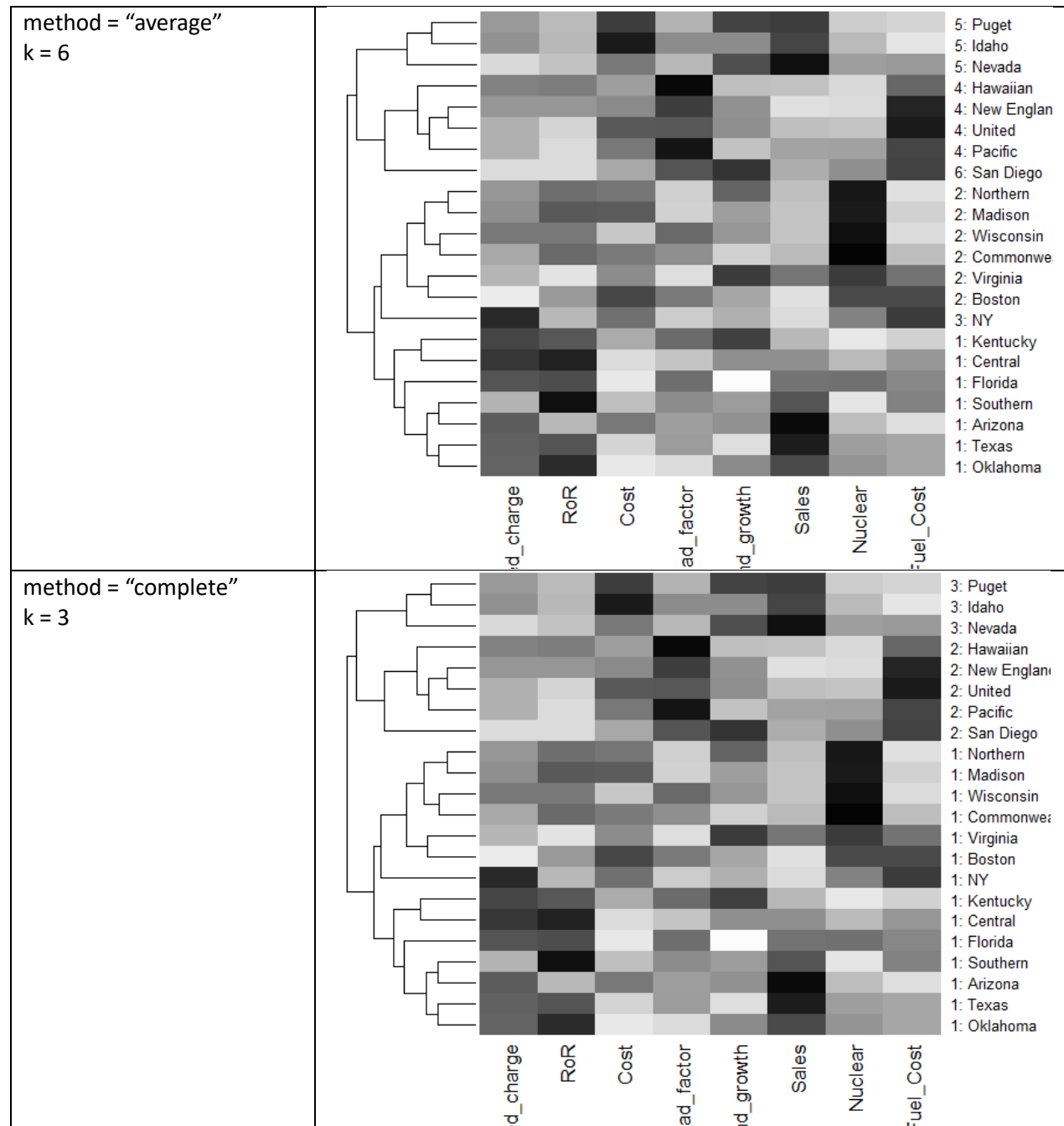
Part 1.

a. K-means

Nicole Kayambu
4341-02
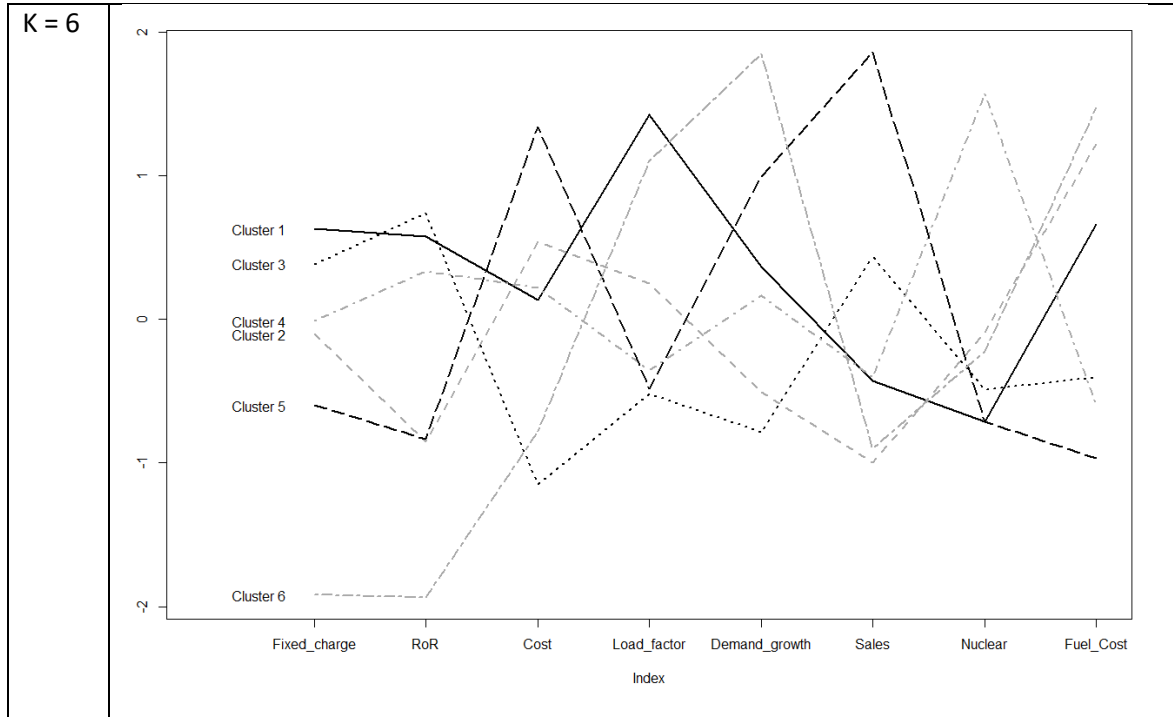
When k = 3 it shows the trends more generally than if k = 6, while still retaining key attributes, such as the different peaks in the data, such as the load factor, like in clusters 3 and 4 in the k = 6 plot compared to cluster 3 in the k = 3 chart. The same can be said of clusters 1 and 6 in k = 6 and cluster 1 in k = 3.

b. Hierarchical Clustering

method = "single" k = 6	
method = "single" k = 3	

Nicole Kayambu
4341-02

method = "average" k = 6	
method = "complete" k = 3	

With both methods, some cities were placed in different clusters when compared to the k means clusters.

Nicole Kayambu
4341-02

Part 2.

a.

Plot of Clusters
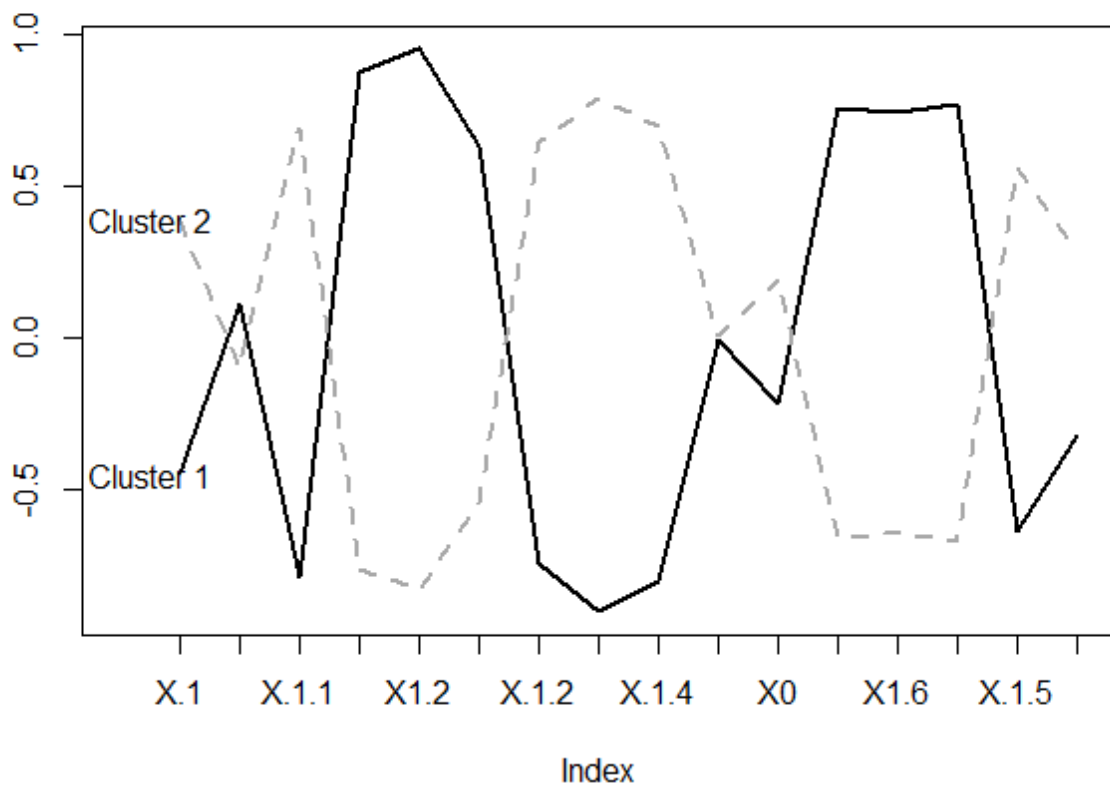
Cluster 1	
Democrats	74
Republicans	128
Total Size	202

Cluster 2	
Democrats	87
Republicans	145
Total Size	232

Overall, in congress in 1984 was majority Republican. In the clusters the proportion of Democrats and Republicans is almost the same, with cluster 1 being 37% Democrat and cluster 2 being 38% Democrat. Despite this, the clusters seem to vote almost entirely opposite of each other on each of the issues that

Nicole Kayambu
4341-02

were being voted on. Of all the issues voted on the only one that seems to be voted the same on for both clusters are x1.4, which is immigration, and x.1 which is water project cost sharing.