Assignment #14 Clustering by Joshua Troup

**Q1: Apply hierarchical clustering with Euclidean distance and Ward's method.**

**Make sure to normalize the data first.**

**How many clusters appear?**

4000 clusters turned into 30 sub clusters turned into 2 broad clusters.

**Q2: What would happen if the data were not normalized?**

The variables with the largest scale would dominate the measure including balance, bonus miles, flight miles 12mo, and days since enroll.

**Q3: Compare the cluster centroid to characterize the different clusters, and try to give**

**each cluster a label.**

Custer 1’s variables are lower so we can assume this category is the “occasional and minimal flyer”. Cluster 2’s variables are higher so we can assume this category is the “frequent and loyal travelers”.

**Q4: To check the stability of the clusters, remove a random 5% of the data (by taking**

**random sample of 95% of the records), and repeat the analysis. Does the same picture**

**emerge?**

Yes the picture is very similar.

**Q5: Use k-means clustering with the number of clusters that you found above. Does the**

**same picture emerge?**

Yes we get the similar results and visualization. Cluster 1 shows 1302. Custer 2 shows 2697. This indicates Cluster 1 is the “occasional and minimal flyer” category. Cluster 2 indicates the “frequent and loyal travelers” category.