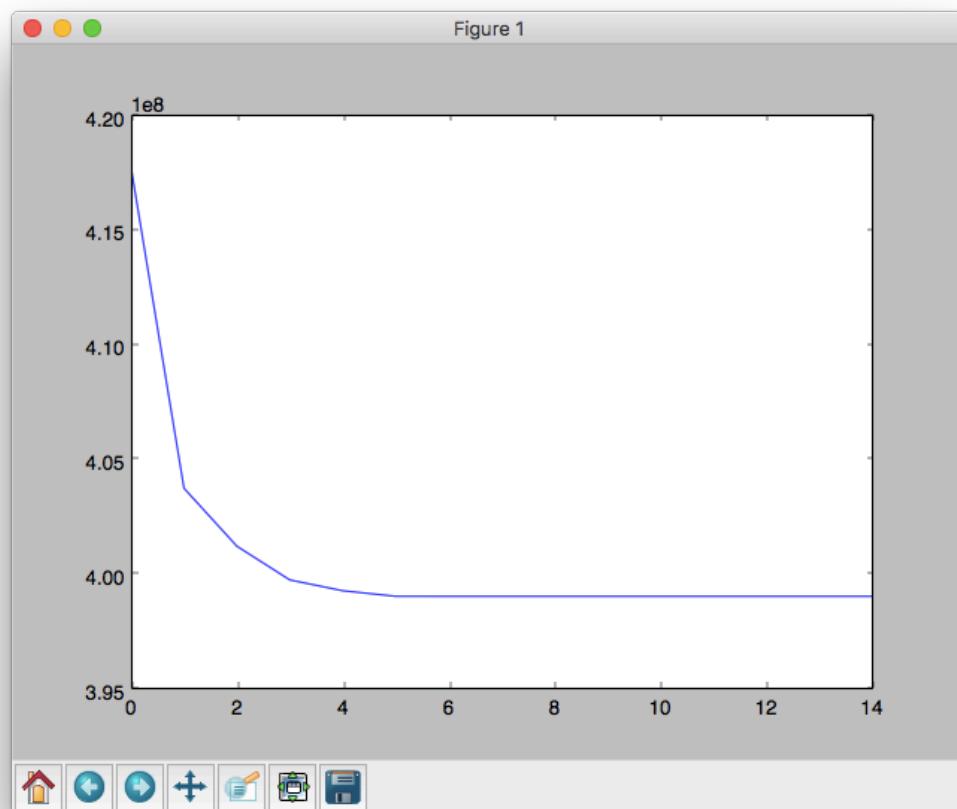


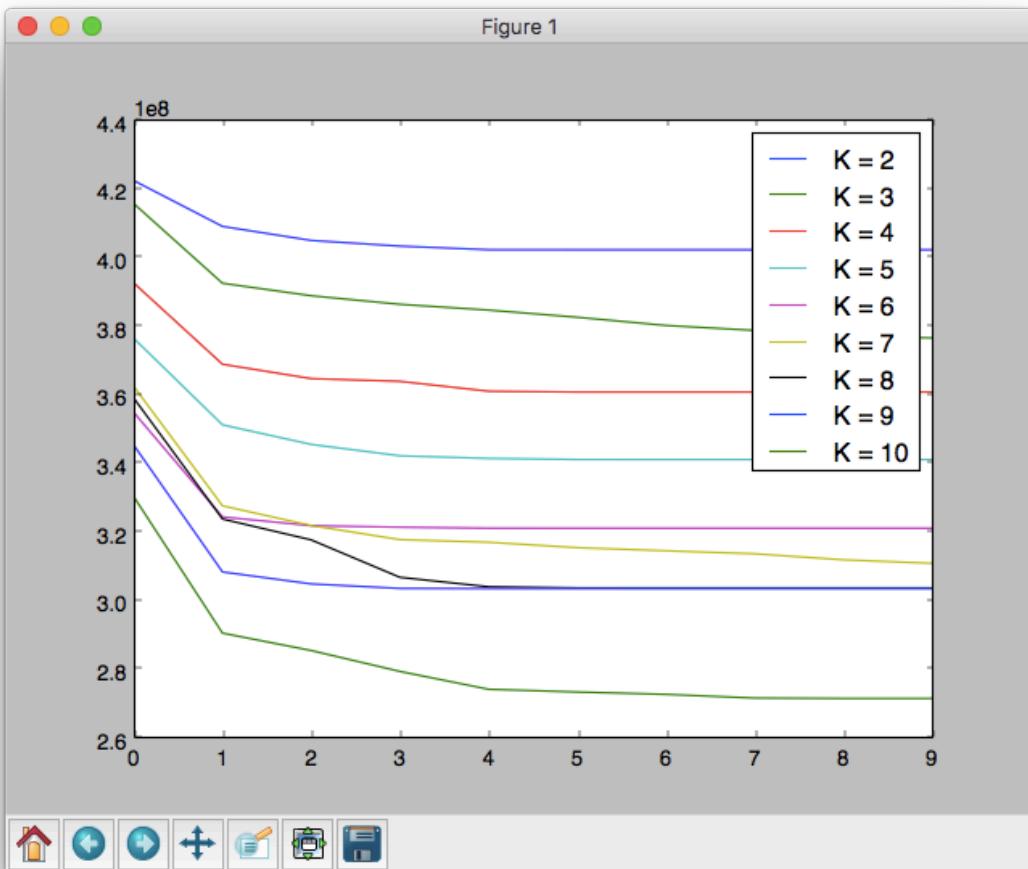
CS 434: Assignment 4

1. Non-hierarchical clustering – K-means algorithm

- a. We implemented the K-means algorithm for k=2 as you can see from our code in part_2_1() in main.py. We can see from the below graph that we've plotted, that the algorithm indeed converges and it converges fairly quickly as it bottoms out at around 5-6 iterations.

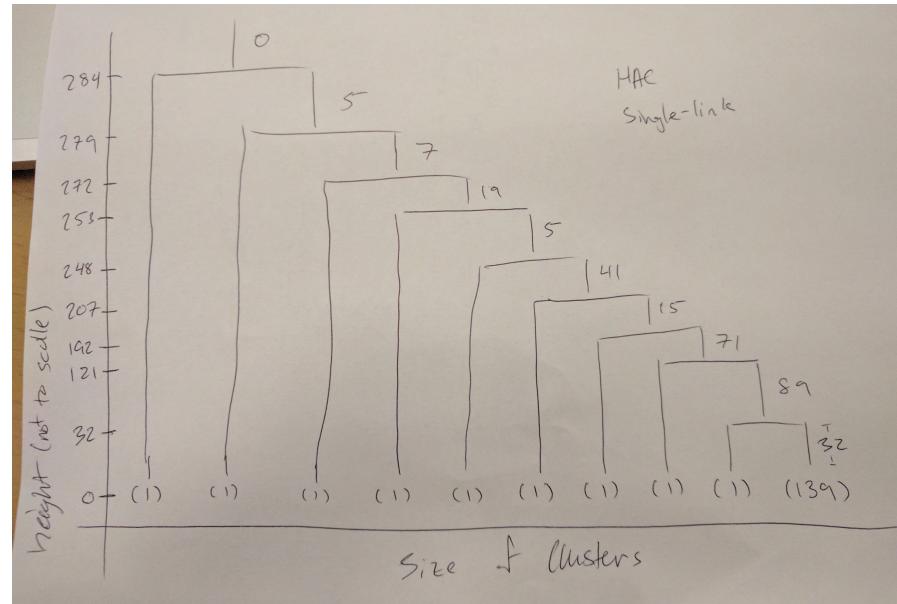


- b. You can see the results of multiple different $K=n$. To choose the best $k=n$ we would want the k that converges the fastest which would be $k = 6$. This would mean that our data fits most neatly into our k distinct groups.



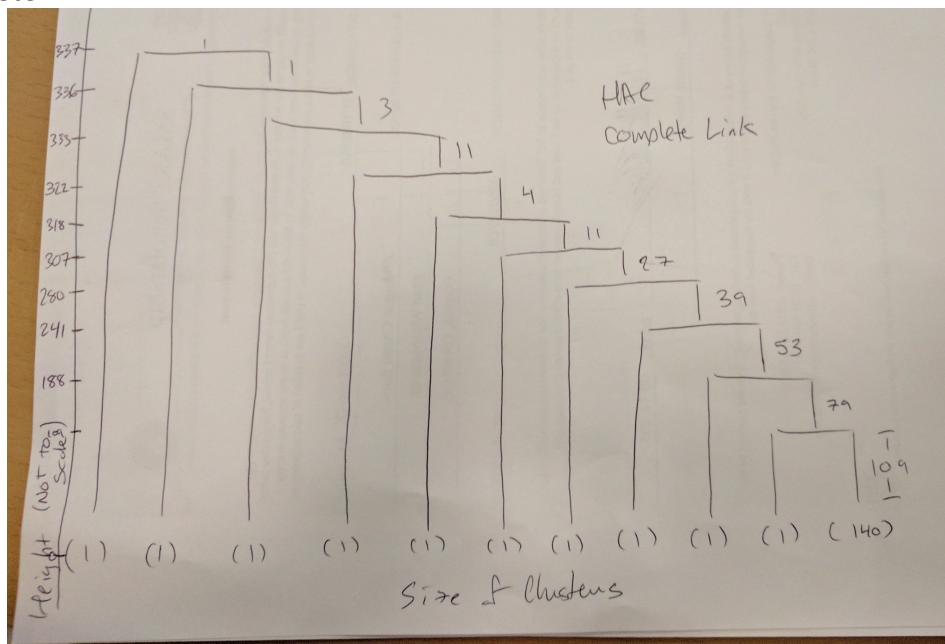
2. Hierarchical agglomerative clustering (HAC)

a. Single Link



i.

b. Complete Link



i.

- c. Looking at both single link and complete link dendrograms, there is only one real cluster at the 10-cluster mark. There other clusters shown are single data points that are yet to be merged. This may mean that our clusters could have converged before the 10-cluster mark or a lot the data is heavily centered around one point.