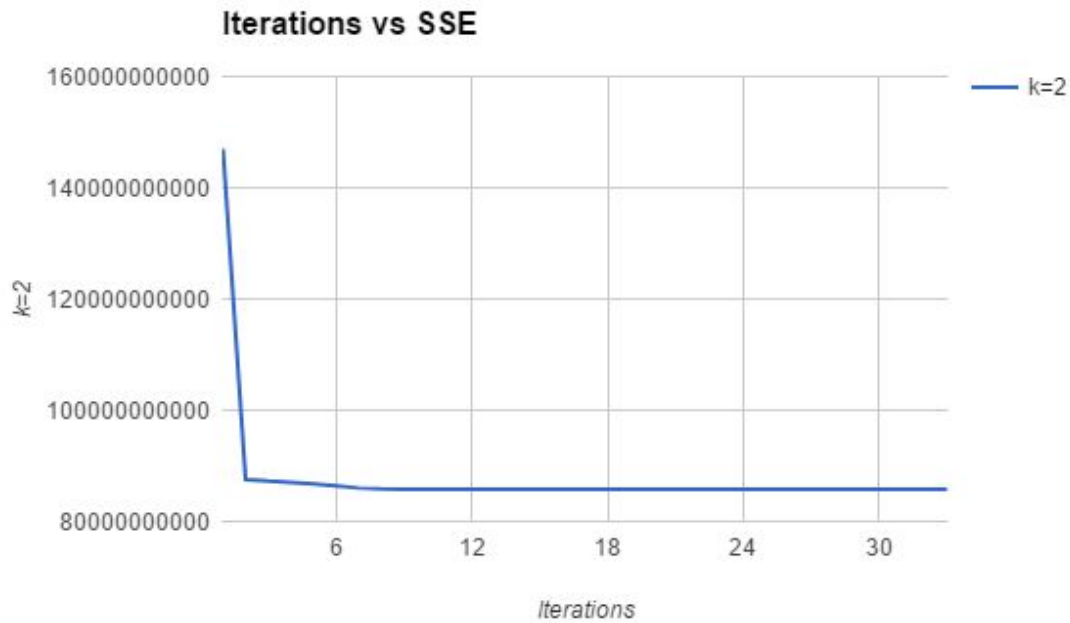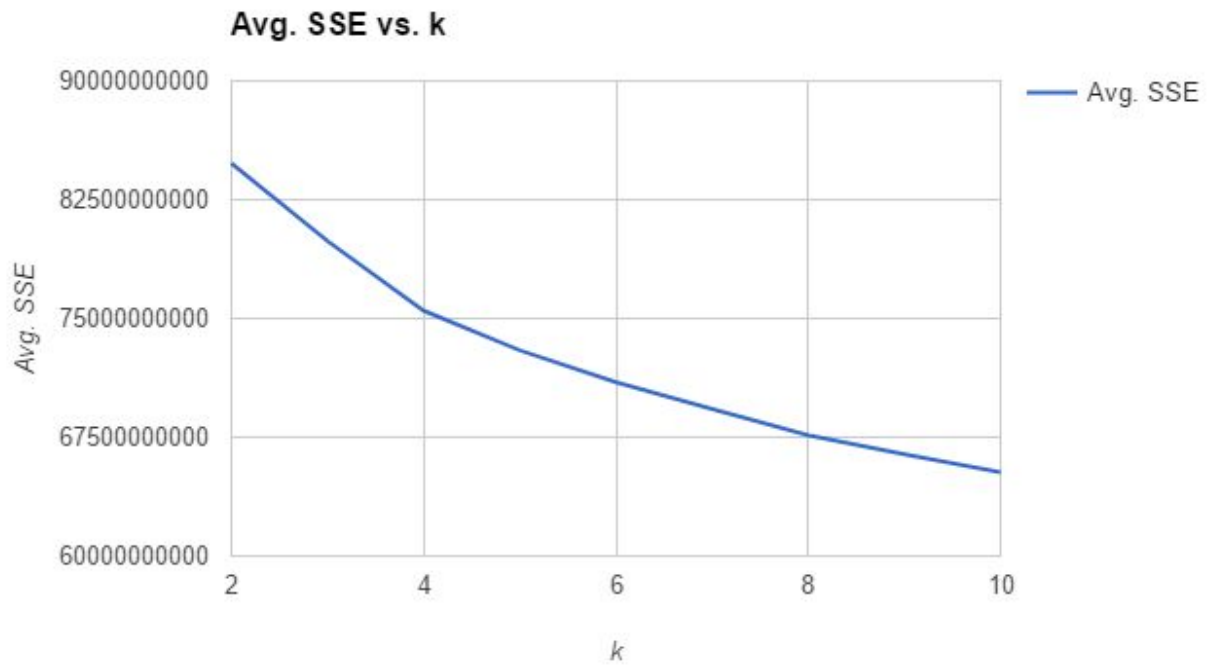Nathan Brahmstadt
Jordan Crane
Assignment 4

Part 1: K-Means

1. Iterations vs SSE

2. K vs. SSE

Avg. SSE vs. k

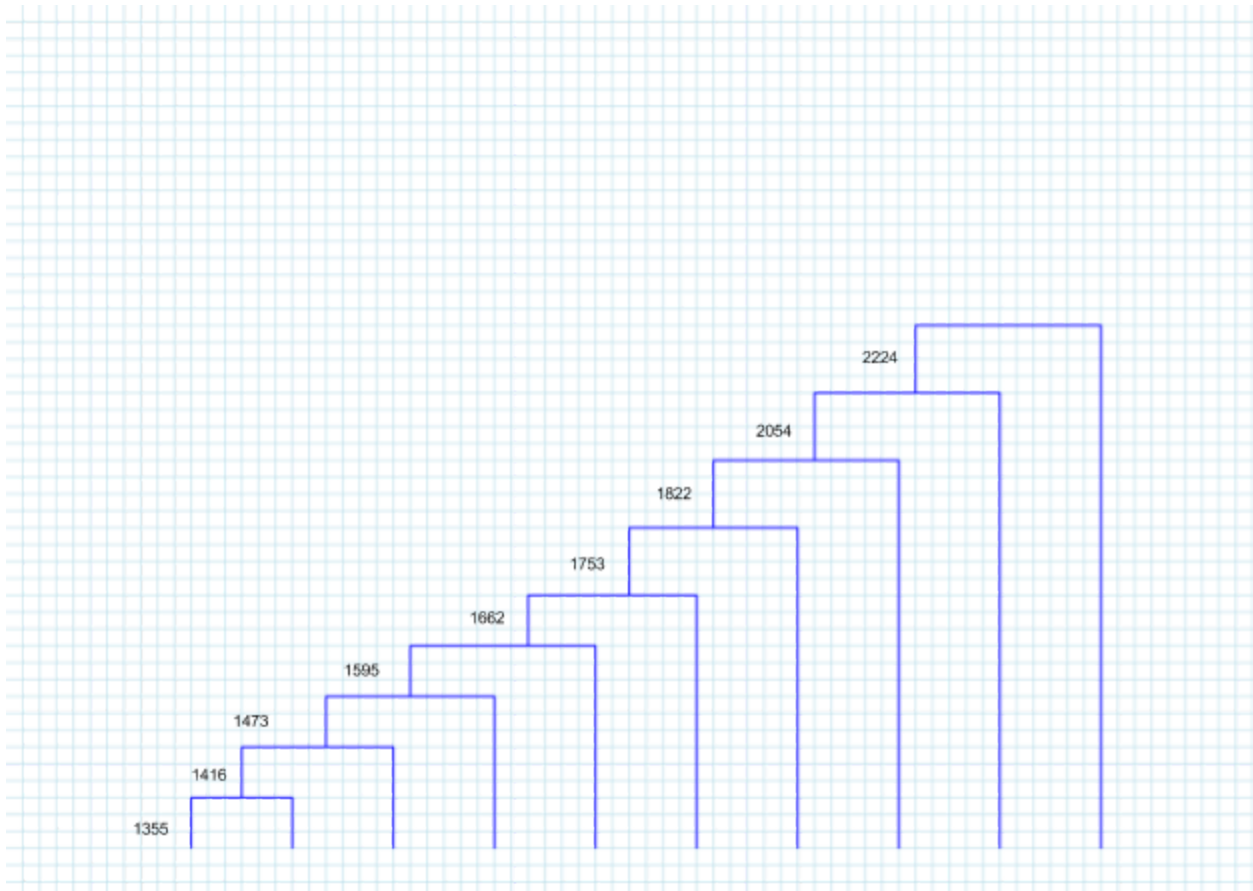We think the proper k is 4, based on the graph. The slope of the line changes at 4, and this "elbow" can represent overfitting occurring.
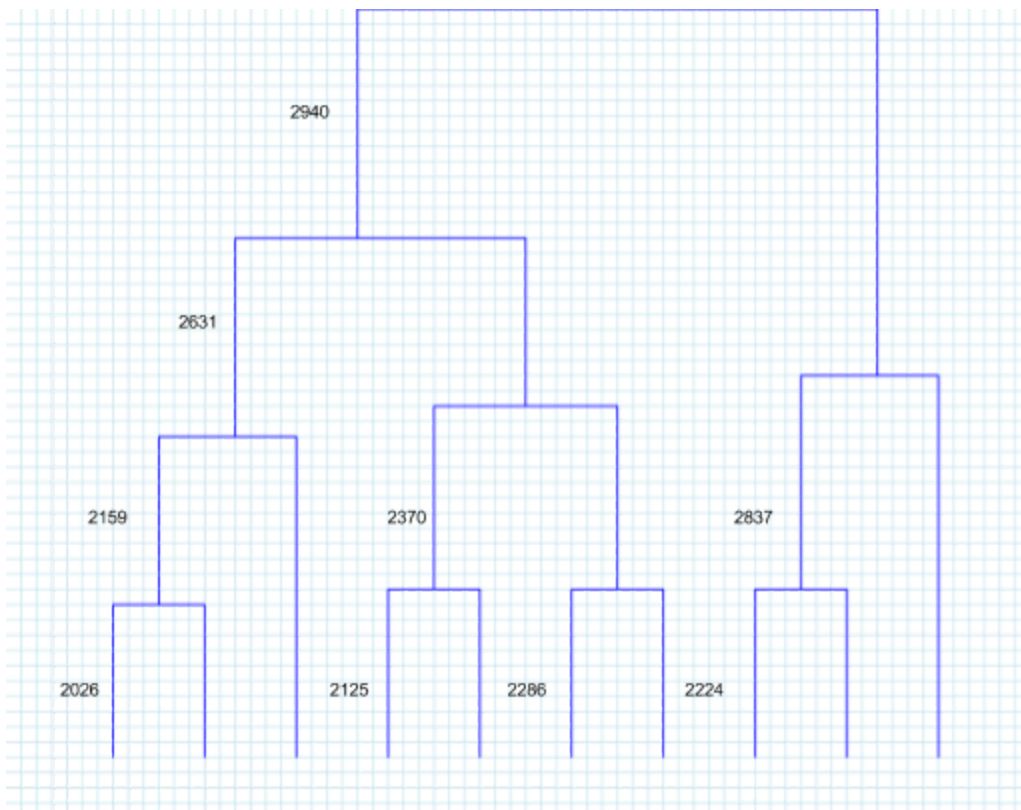
Part 2: HAC

1. Single Link

It seems impossible to chose a number of clusters based on the single-link clustering. The "chaining" effect that single-link has prevents us from getting meaningful data from the dendrogram.

2. Complete Link

Looking at the dendogram, there are three splits with slightly higher distances than the ones before. There is a 300+ unit jump in distance from the other splits. The distances are 2631, 2837, and 2940. These splits signify that there are 4 clusters found with the complete link algorithm.