Homework 3
CS659, Spring 2015
Due April 7, 2015 in class

Given the two datasets found here and here, do the following

1. Cluster each dataset using K-means,EM, and DBScan. Try different values of k in the interval [2,5] for K-Means.
2. Determine, using SSE which is the best value of K. Compare with what DBScan finds.
3. For your best clustering, in all cases (K-means, EM, and DBScan), and for both datasets, do the following
    a. Measure the clustering validity using correlation
    b. Produce a graphic that shows how the clustering information and the similarity matrix 'correlate' with each other (order the points according to their cluster labels and visualize their similarity).
    c. Compute the silhouette coefficient