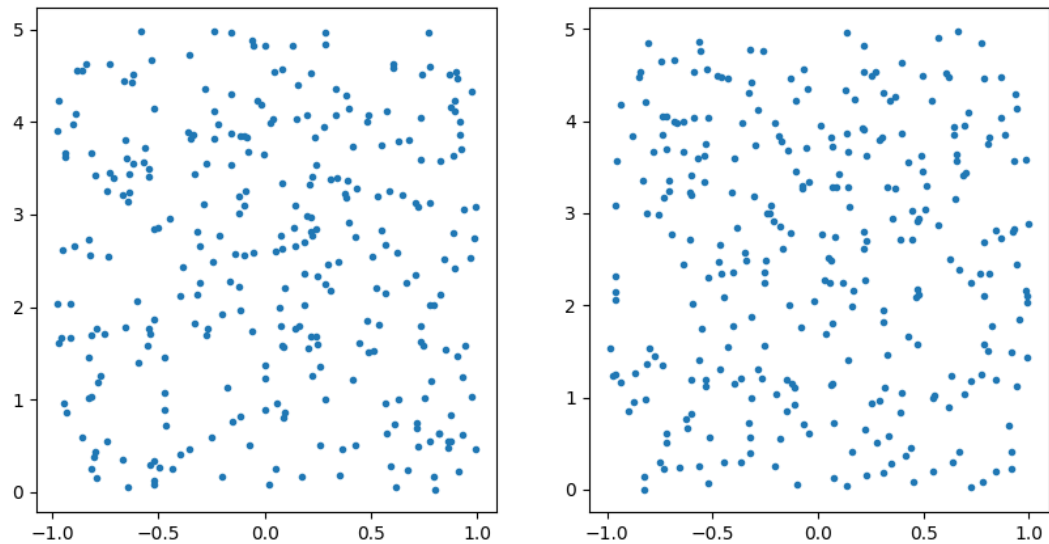


Homework #2: Similar items, Clustering, Community Detection

Problem 3

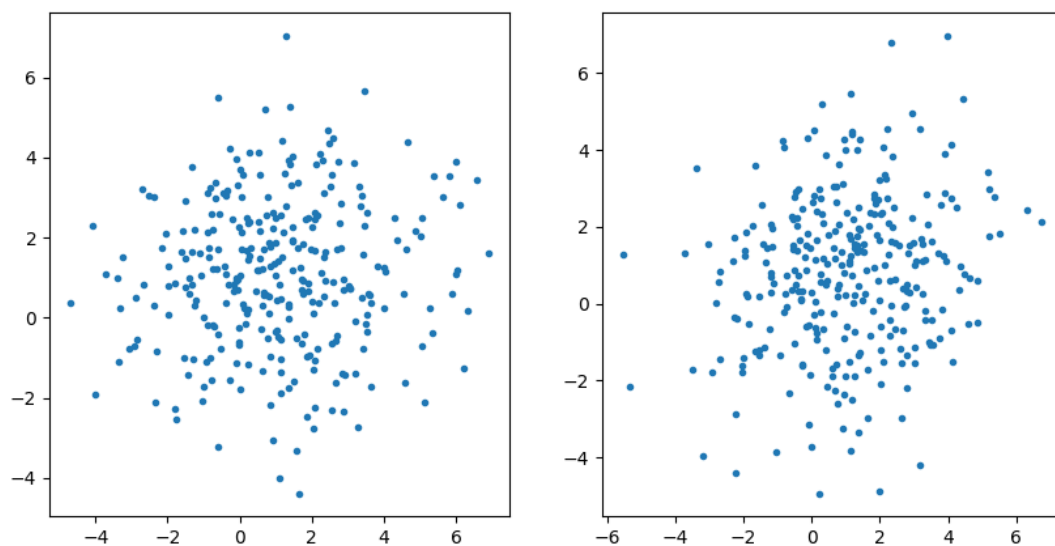
(a)

Uniform distribution, $x \in [-1, 1]$, $y \in [0, 5]$



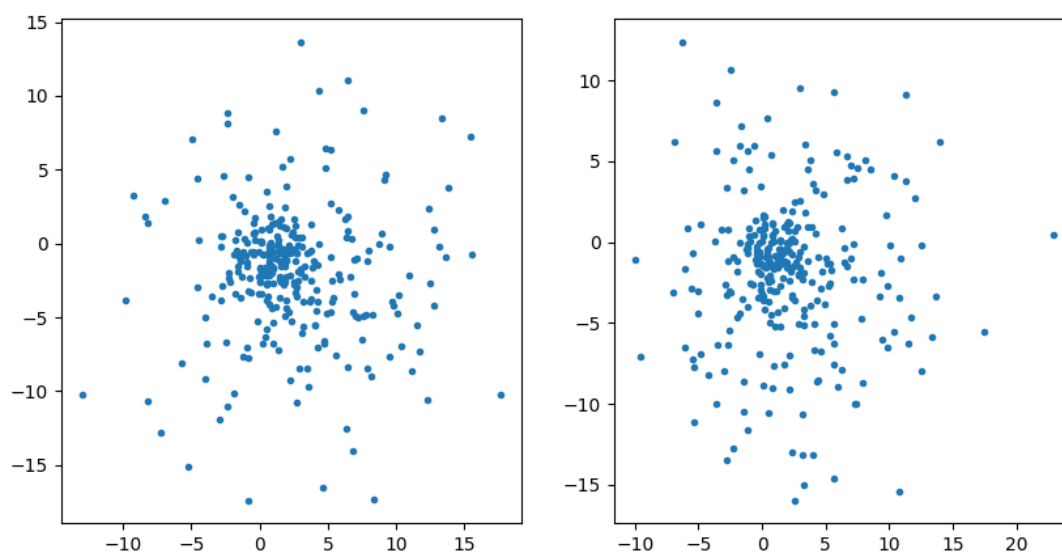
(b)

Gaussian with center at [1,1] and std=2



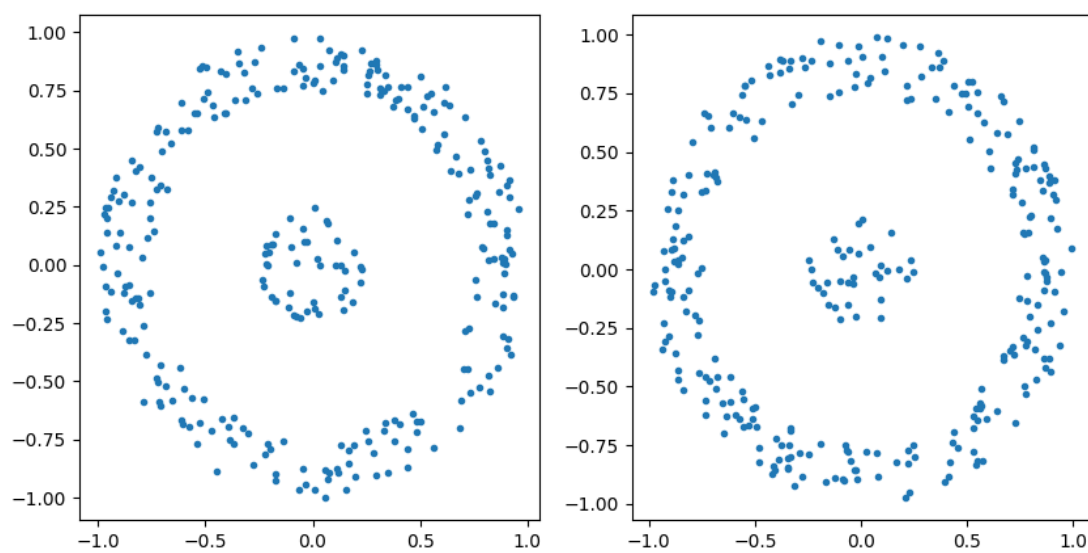
(c)

Three Gaussians



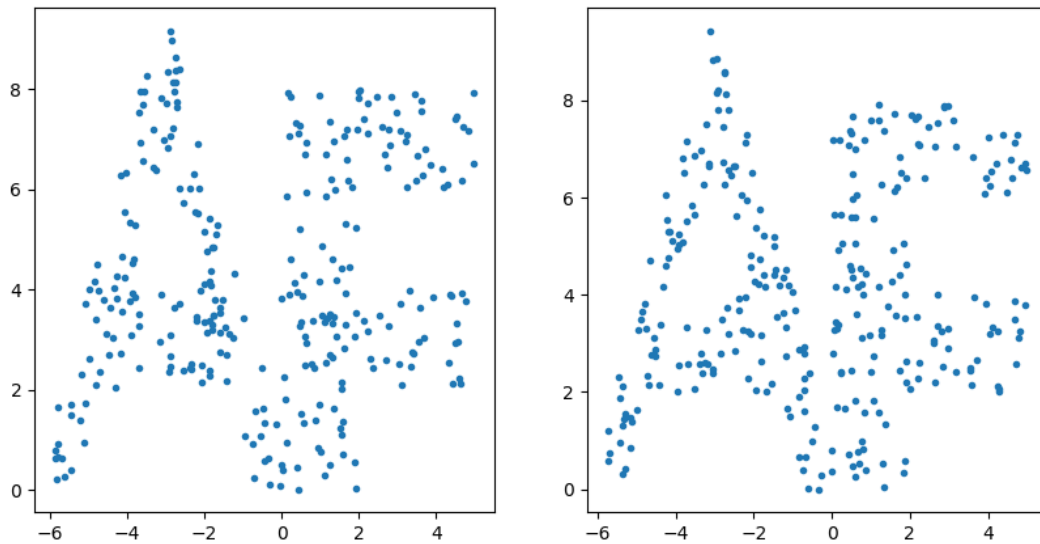
(d)

CURE Circles



(e)

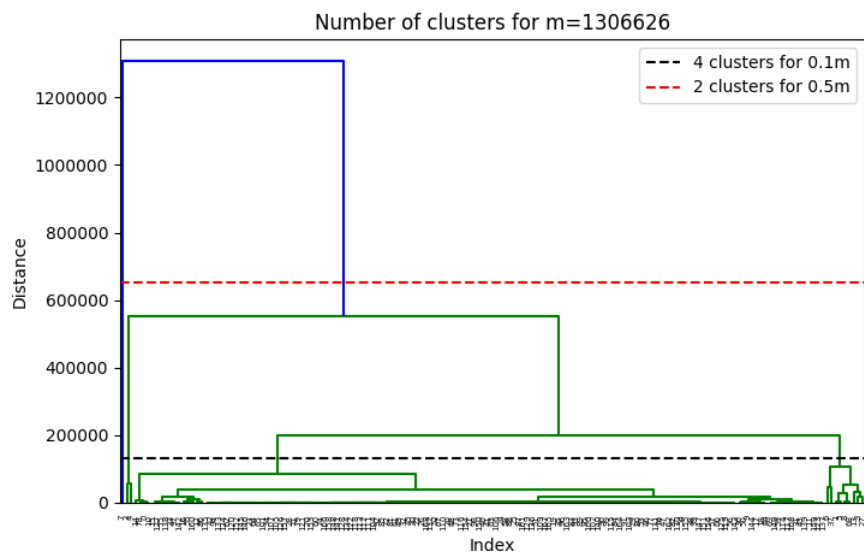
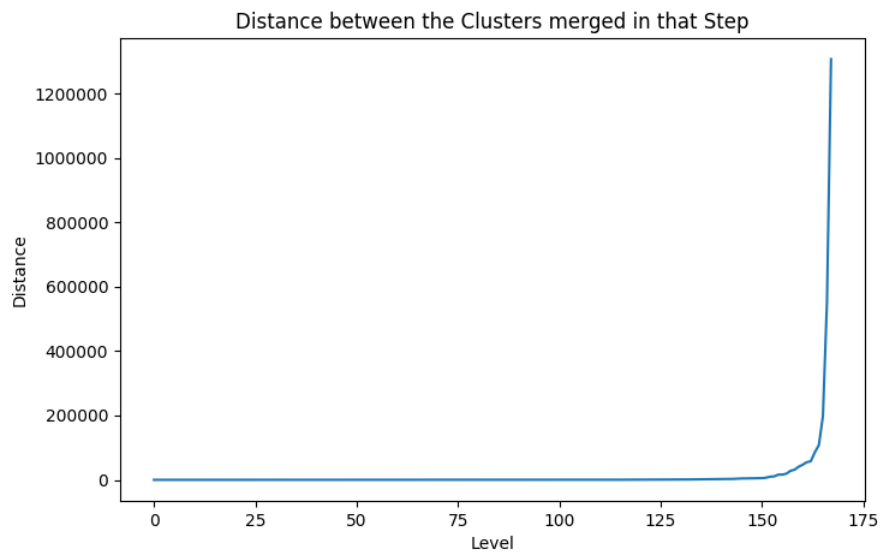
The first letter of both your first names - AF



Problem 4

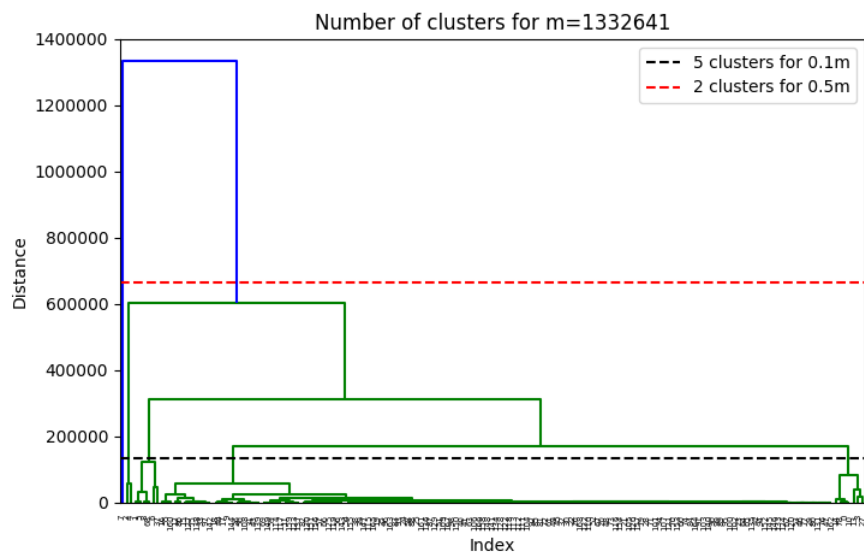
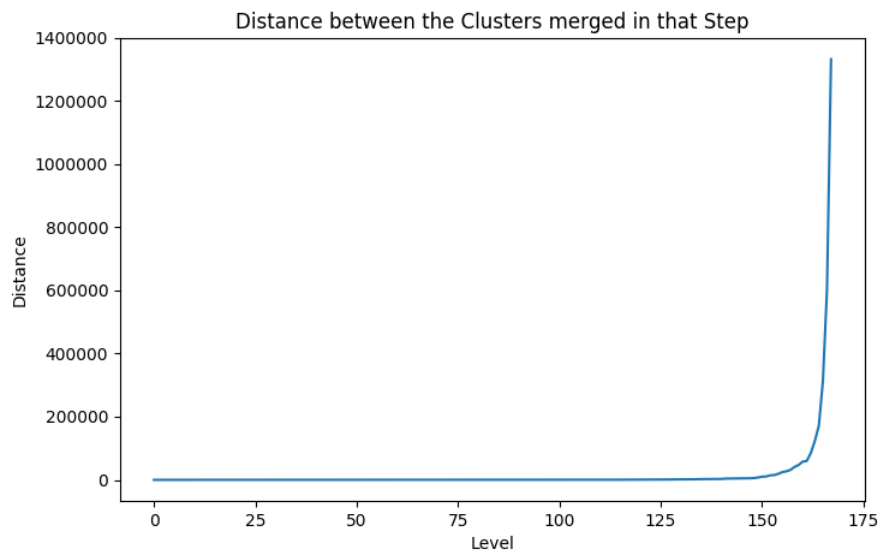
For hierarchical clustering and the plot of the dendrogram we use the package `scipy.cluster.hierarchy`. Function `linkage` provides the hierarchical clustering algorithm for different methods. We use `method='centroid'` in part (a) and `method='complete'` in part (b). Function `dendrogram` is used to plot the dendrogram.

(a) merge clusters with the closest centroids



The Dendrogram shows that we would get 4 clusters if we decide to stop clustering at a distance of 0.1m and 2 clusters if we decide to stop at 0.5m.

(b) merge clusters so that the new diameter is the smallest among all options



The Dendrogram shows that with this method we would get 5 clusters if we decide to stop clustering at a distance of 0.1m and 2 clusters if we decide to stop at 0.5m.