

Lab Assignment 10/13

Due Friday, October 16th at 5:59pm ET

You are welcome (and encouraged!) to consult your peers while working on your code. The goal of the lab is to prepare you for completing future homework assignments, so make sure you understand everything even if you are working with others.

Clustering the States

The goal of this exercise is to learn more about agglomerative clustering by completing a simple implementation in Python. The data set comprises state-by-state presidential election results 1988-2016. Your task is to fill in some of the function bodies, run the clustering code, and examine the output.

You'll be editing the code found in `state_election_clusterin.py`. The data can be found in `1988-2016.csv`. To complete the implementation, follow the instructions below.

0. Try running the code as is to get a feel for the output format. Note that the distance function used to calculate the distance between state is based on state name, so your clusters reflect that.
1. Fill in the `dist_euclidean()` function, and modify line 180 to use that function in your clustering algorithm.
2. The cluster distance function `single_link_dist()` uses single cluster linkage, which doesn't seem to work well for this problem (why not?). Fill in `avg_link_dist()` and `complete_link_dist()`, and change line 180 to use one of those.
3. Re-run the clustering algorithm -- does the output match your expectations?