Data Biography for Dataset Northeast Ohio Cities

The dataset *northeastohiocities.xls* contains four groups of four cities (16 in total) around northeast ohio.

The cities are “clustered” geographically to help testing the K-Means and SOM clustering.

The first worksheet contains four groups of three cities for training (12 in total) and the second worksheet contains a city from each group to use for testing of the predict functions.

To experiment in the testing, the population and elevation for each city is included along with the Latitude and Longitude.

To start, a tester can just try Lat and Lon. Then, see how much the prediction changes if you introduce the third and fourth dimension (variable) of elevation then population.

Is the clustering good enough to place the four guesses in their appropriate cluster? If not, why not?