**Data Science with R- Project 3**

**Problem Statement**: Analyze the monthly seasonally adjusted unemployment rates for the given employment data of the U.S. “unemp.csv” file, covering the period January 1976 through August 2010 for the 50 U.S. states(*n* = 50).

The requirement is to cluster the states into groups that are alike. Here, each state is characterized by a feature vector of a very large dimension (*p* = 416). Its components represent 416 monthly observations.

For the purpose of an illustration, assume that New York and California form a cluster.

You will need to calculate 416 monthly averages (of two observations each). This vector of averages is called the centroid for that cluster.

Note that the sum of the squared distances from the centroid of this cluster expresses the within-cluster sum of squares. To explain it better, in the mentioned example, for New York and California, there are 2*(*416*)* = 832 such distances.