**Cluster Analysis, 1990-2017**

We are interested in identifying census tracts with marked and noticeable change in low-income residents in recent decades. To this end, census tracts in the study area are divided into XXX-TBD tract typologies based on the 1990 percentage of low-income residents (“baseline”) and the percentage change in low-income residents from 1990 to 2017 (“change”). We analyze baseline and change simultaneously because change in itself can be misleading.

Tract typologies are created using model-based clustering along a Gaussian finite mixture distribution; see the R package [mclust](https://cran.r-project.org/web/packages/mclust/index.html). Because the 2017 data comes from the ACS and includes sample error, we run 1,000 cluster analyses with simulated baseline and change values. In each iteration, the baseline and change values vary based on 2017 estimates and MOEs. Each census tract is assigned a new baseline value, where the mean is the census tract’s estimate, the standard deviation is the tract’s MOE divided by 1.645, and the range of possible new values is normally distributed. The new baseline value is used as an input in creating a new change value.

**Table 1** below shows the results of the cluster analysis.

See cluster.csv for cluster assignments for all observations. The archive folder contains clusters.R, which was the script used to run this analysis.

**Table 1: Cluster Analysis Results**

|  |  |  |
| --- | --- | --- |
| Model Name | Number of Groups | *n* |
| 120 |  |  |
| 121 |  |  |
| TOTAL | | 1000 |