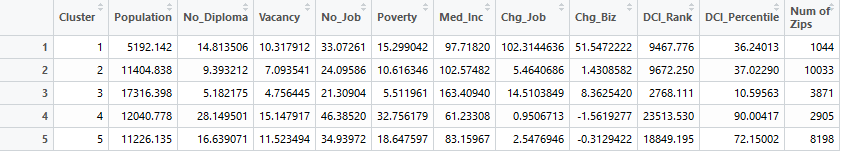
Cluster Analysis

This is a cluster analysis from the dci\_componenets\_v5 dataset. The original cluster analysis utilized all 26125 zip codes, but because of a small group of outliers in “Chg\_Job”, the desired results were not given. It was decided to remove values of “Chg\_Job” above 300, which resulted in 26051 (a loss of 74 zipcodes or less than 0.01%).

The subsetted data was called dci\_cluster\_26051. After running three methods to determine the best number of clusters (K’s) in a K-means clustering, it was determined that 5 clusters best represent our zip codes. The three methods were GAP analysis, the Silhouette method, and the Elbow method (within sum of squares); all three methods pointed to 5 being the best number of clusters. The picture below, is the summary statistics (averages) for each cluster for each DCI Component



Of the 5 zip codes, I have labeled them as followed:

* 1) Some Distress – Largest Chg\_Job and Chg\_Biz
* 2) Slightly Distress -Largest Cluster
* 3) Not Distressed – Most Affluent Cluster
* 4) Extremely Distressed – Cluster with the highest distress levels
* 5) Very Distressed – Second largest cluster, but high levels of distress

Cluster 1: This is the smallest cluster with 1044 zip codes falling in this category. This cluster does some levels of distress present in No\_Diploma, Vacancy, No\_ Job, Poverty, and a slightly lower Median Income. What makes this cluster special is the astronomically high Chg\_Job and Chg\_Biz, which is nearly 7 times higher in Change in Job and 6.16 higher inn Change in Business than the next highest cluster (Cluster 3). The average DCI Percentile is 36.24%. An interesting note is this cluster does show significant signs of distress, but has extremely high Chg\_Job and Chg\_Biz which is raising the DCI Rank and Percentile of these zip codes.

Cluster 2: This cluster is the largest cluster with 10033 zip codes and are only slightly. The zip codes in this cluster show very little signs of distress, very small deviations from Cluster 3 (the most affluent). However the main difference from Cluster 3, is a drop in around 37% of Median Income, and very small Chg\_Job and Chg\_Biz. The zip codes in this cluster have slight distress, but are the second wealthiest and have a stunted Chg\_Job and Chg\_Biz. This cluster has a DCI percentile of 37.02%

Cluster 3: This cluster is the most affluent, with the lowest levels of 5 DCi components: No\_Diploma, Vacancy, No\_ Job, Poverty, and the highest Median Income. This cluster has the second highest Chg\_job and Chg\_Biz. There are 3871 zip codes in this cluster, and have the highest DCI Percentile of 10.59%. The zip codes in this cluster have very little (if at all) distress compared to the other zip codes.

Cluster 4: This cluster contains the zip codes that are the most distressed. In components: No\_Diploma, Vacancy, No\_ Job, Poverty, all values are the highest. This cluster has almost no job growth, a negative change in business and the lowest median income. This cluster DCI percentile is 90% and contains 2905, these zip codes are in the most trouble.

Cluster 5: This cluster does have significant distress, but is better than Cluster 4. This is the second largest cluster with 8198 zip codes and a DCI percentile of around 72%. No\_Diploma, Vacancy, No\_ Job, Poverty are relatively high but it does have a lower Median Income. There is some (but relatively small) Chg\_Biz and only slight loss in Chg\_Biz.