

<Finding Peers by Age Bracket: Patterns in US Population>

Richard Deslauriers

Analysis Technique

In order to locate large groups of a certain age bracket, census data will be used. Collected information of the ages and percentages of each age bracket in each state is the data that will be focused on. By using a self-organizing map patterns in the typical locations of each age bracket will emerge. In this way, one would be able to locate one's peer group and potentially move to a large cluster to be closer to their peers. Current data will be used to allow for time to locate a new home and move there, before population trends and patterns change.

A self-organizing map will be used to organize and group states based on the percentages of age brackets found in each of the United States. The self-organizing map (SOM) is a subtype of artificial neural networks. It is trained using unsupervised learning to produce low dimensional representation of the training samples while preserving the topological properties of the input space. In the training phase weights of the whole neighborhood are moved in the same direction, similar items tend to excite adjacent neurons. Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar apart.

The use of the SOM is to find relationships between states and their population densities of certain age groups. States with similar percentages of each age group will map together. With the states mapped together, I hypothesize that east and west coast states will offer the highest ratio of young adults to all other age brackets. If not that exact, the SOM should still help in locating peer groups somewhere in the United States. From the map, it may be possible to draw conclusions of age bracket clustering based on weather patterns or even political state affiliation. It is possible that an outside influence other than the randomness of birthplace affects the residence of U.S. citizens.

The data set used has been created by the U.S. Census Bureau. It has 50 entries for each of the United States. For each state there are four age groups; children, senior citizens, young adults and older adults. Each state contains a raw figure for the number of each the census gathered, as well as a percentage for each bracket in each state. The dataset includes information over time and current trends. The data set will need to be transformed to include only current trends and their percentages. Raw numbers aren't as important as the percentages each age bracket represents in a particular state. If inconclusive results arise from a SOM using percentages instead of raw data, the data set can be readjusted using raw numbers and the SOM can be run and trained again. In this way, one data set becomes two data sets if necessary.

Assumptions

- Living with one's peers is an important part to having a happy life
- State percentages are used, rather than city; age clustering is consider even throughout each state
- Certain states will contain higher percentages of certain age brackets
- The Census captures enough information; illegal immigrants who are unaccounted for will not impact the results
- There is an appreciable difference between population clusters in each state

Appendices

1. <http://www.census.gov/>
2. http://en.wikipedia.org/wiki/Self_organizing_map