Joint Returns as an Influence on Dependents in Massachusetts Zip Code Areas

Aditi Dass and Navraj Narula – CS591 K1 – Boston University

Motivation

According to the Internal Revenue Service (IRS), a dependent is a person who relies on someone financially. For our project, we have chosen to examine the number of dependents in congruence with the number of marriages per zip code area in the state of Massachusetts. Our assumption is that since married couples tend to signify more children, a discrepancy would signify an aging community. We would like to see how far this notion extends when considering different types of communities. More dependents in an area is generally not good for the economy, but more children result in greater future stability.

Project Goals

- Determine whether or not a correlation exists between the number of dependents and the number of joint returns.
- Identify zip codes of densely populated areas (cities, towns, etc.) vs. sparsely populated areas (rural, suburban, etc).
- Discover whether certain communities produce a higher number of dependents or number of joint returns than others and whether a community's ratio of dependents and joint returns makes it economically stable or not

Datasets

MA Zip Codes from the Internal Revenue Service

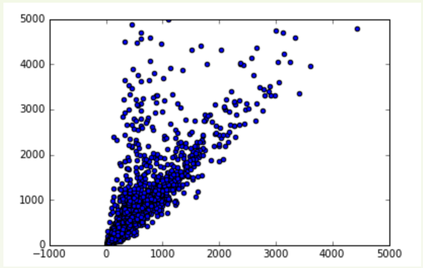
- 3870 rows, 112 columns
- Adjusted gross income, wages and salaries, number of returns, etc.

U.S. Zip Codes from ProximityOne

- 33,120 rows, 9 columns
- Population, land square meters, etc.

Dependents vs. Joint

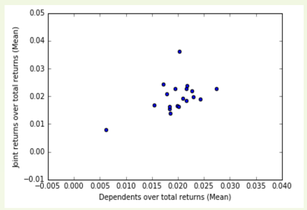
From our dataset of zip codes, we were able to extract the number of dependents and the number of joint returns. We found a correlation coefficient of 0.604 meaning that a slight correlation exists among these two attributes. The x-axis is the number of joint returns and the y-axis is the number of dependents.

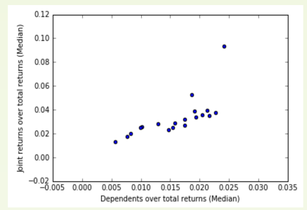


Community Analysis

Community Statistics were analyzed for both dependents and joint returns:

- **Mean** (dependents: 2870, joint returns: 2653)
- **Standard Deviation** (dependents: 1924, joint returns: 1689)
- **Median**(dependents: 3039, joint returns: 2892)
- When population size was accounted for, mean of dependents vs mean of joint returns for each community had a correlation coefficient of 0.253

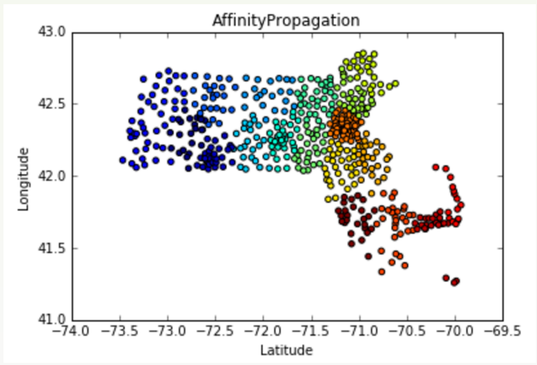




- Medians of both parameters provided a higher coefficient of 0.504 and clearer insight into the data
- Sparsely populated communities tended to have more dependents per joint return indicating there tended to be more aging populations there

Community Discovery

- Extracted longitudinal and latitudinal values of each zip code
- Used Affinity Propagation
- Number of clusters produced = 18
- Identified densely populated areas



Conclusion

There is a significant correlation between the number of dependents and the number of joint returns.

This correlation extends to different types of communities to a lesser but significant extent

Communities with a low number of joint returns and a high number of dependents correlate with aging areas in MA

The community made up of areas in Cape Cod Bay proved to be an outlier with a very high number of dependents and joint returns. This is because of the disproportionate number of retirement homes and beach houses along the Cape.

Future Work

- Use other means of discovering communities
- Find economically stable communities by deciding a benchmark for comparison
- Expand model to the entire USA

References

https://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-2013-ZIP-Code-Data-(SOI)

http://proximityone.com/cen2010_zcta_dp.htm

Funding Source:
Boston University
Computer Science Dept.