|  |
| --- |
| Oregon State University |
| Household Income vs. Housing Ownership |
| ST 599 Big Data Project 1 |

|  |
| --- |
| Sarah Guermond Andy Olstad Jasmine Pahukula Tim Skalland  4/21/2014 |

Overview

The American Community Survey (ACS) is distributed each year to a small sample of the population in each state including Puerto Rico. It provides updated information for communities to aid in planning efforts in between the main 10-year census survey. The ACS collects social, economic, housing and demographic data. This report will focus on the economic and housing information for each state.

*Question of Interest:*

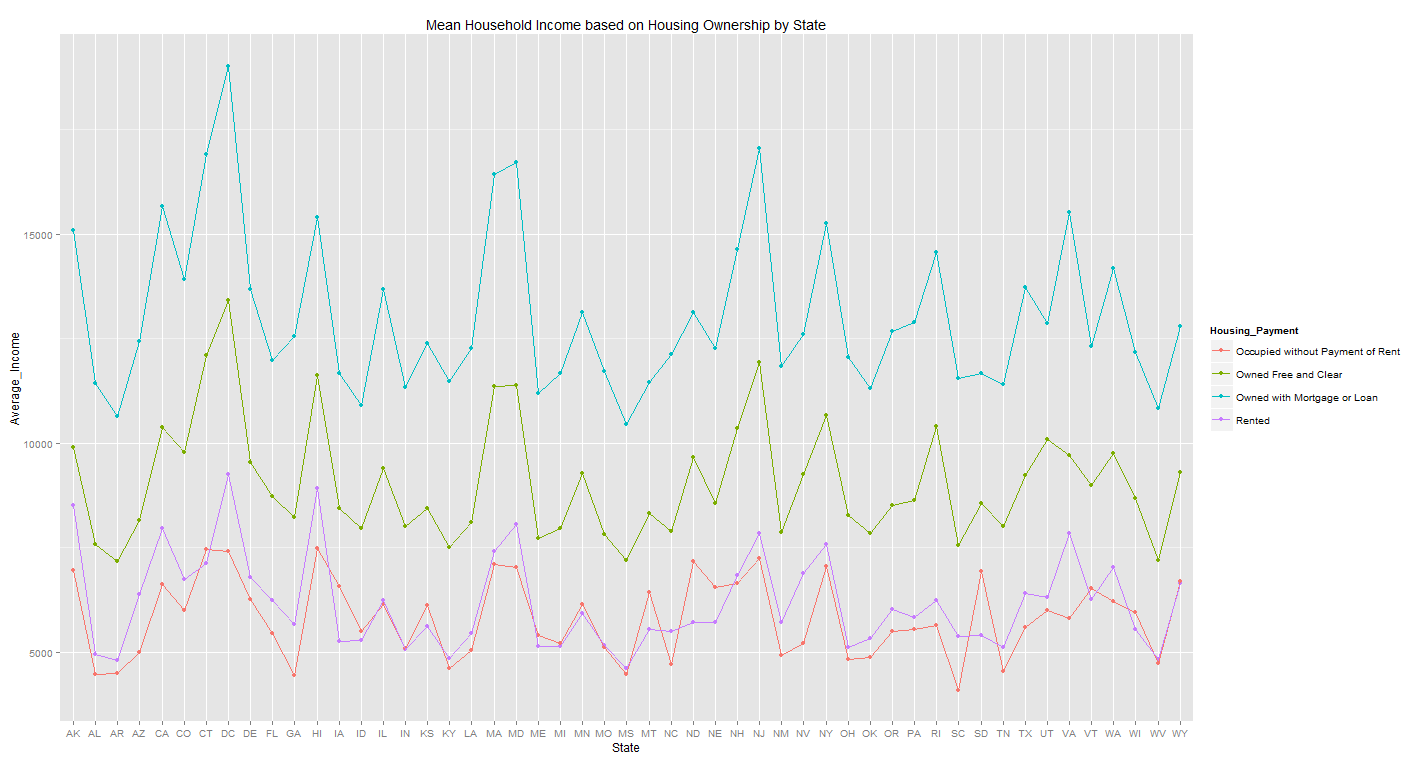
*How different are the mean household incomes for different household ownership (i.e. own with a mortgage or loan, own free and clear, rented, or occupied without payment of rent) -in the US? By state?*

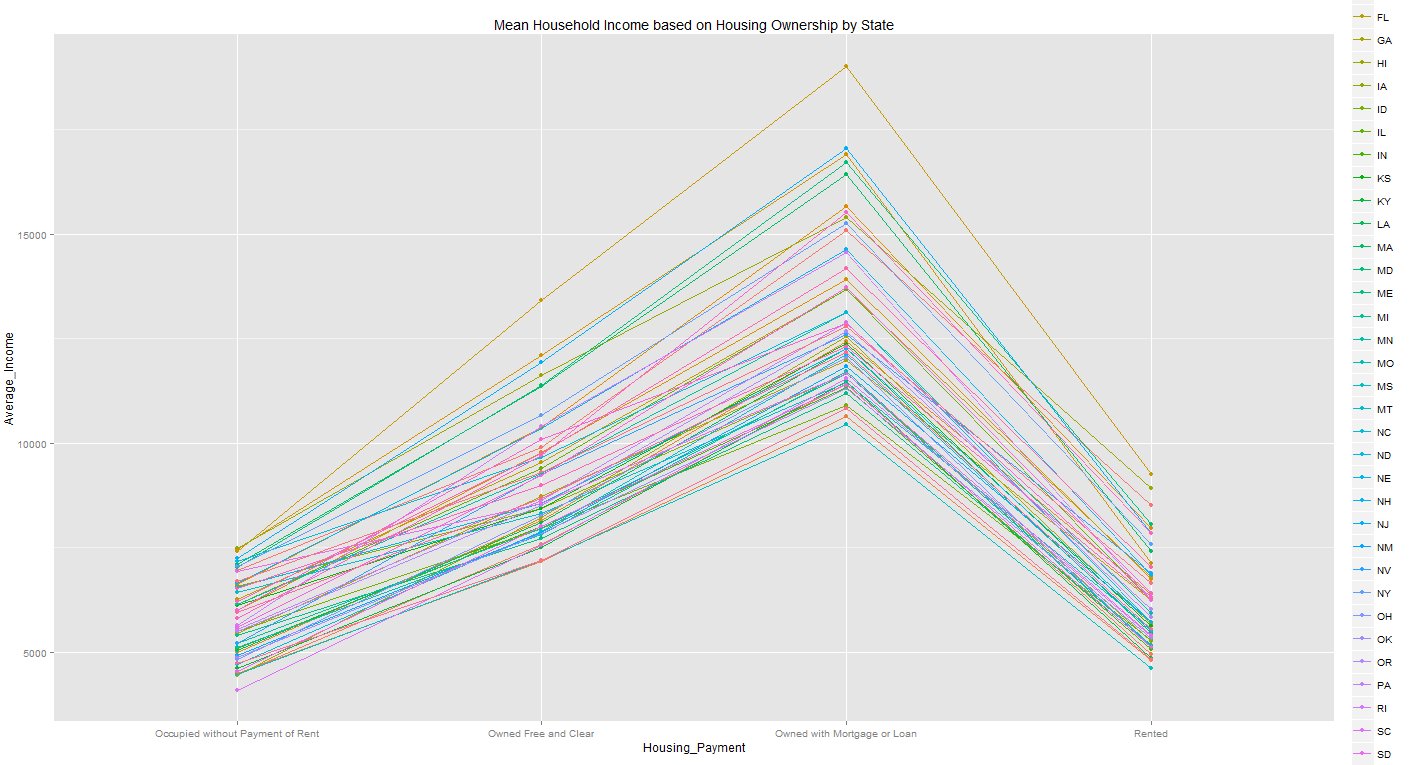
Findings

Respondents owning their house with a mortgage or other payment had a higher income than respondents owning their house free and clear in all states. (talk about possibility of retirement)

(Should do a statistical test between renting and rent-free, it doesn’t seem like they are statistically significant --Sarah)

\*I just took screen shots of these graphs, they still need a bit a polishing if we are going to include them. This is just to show you all what it will look like and if we even want to use them.





Obstacles and Solutions

* Git setup:
  + Problem: getting git to work and sync.
  + Solutions:
    - Delete local connection, create new version control project in Rstudio, paste changes to new file
    - Use github for Mac/Windows
* File downloads:
  + Problem: Automating the downloading process of each state from the web.
  + Solution: Download combined US household data.
  + Problem: Combined US household data comes in 4 files
  + Solution: Download all 4, use system() to merge in command line using cat.
* Data import:
  + Problem: ~200x4million dataset breaks R when you import it
  + Solution: Find columns of interest in the command line (put command here), cut, only import 4 columns
* Code compatibility:
  + Problem: 1 mac, 3 windows, windows cannot run command line data fixes
  + Solution: upload modified data file to git repo so it doesn’t have to be created every time
* Accidental branch diversion:
  + Problem: Code accidentally split into two versions, then merged, breaking code
  + Solution: Version control keeps track of code at different stages, allowing manual reversion to previous versions

Future Work

* Can get summaries per state based on year
* Differences between each of the 4 household ownership
  + Look at age: hypothesize that free & clear owners are retired, mortgage are still working
* Look at regions (East, South, West, Midwest)
* Take into consideration negative and null income