|  |
| --- |
| 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 and Puerto Rico?*

Findings

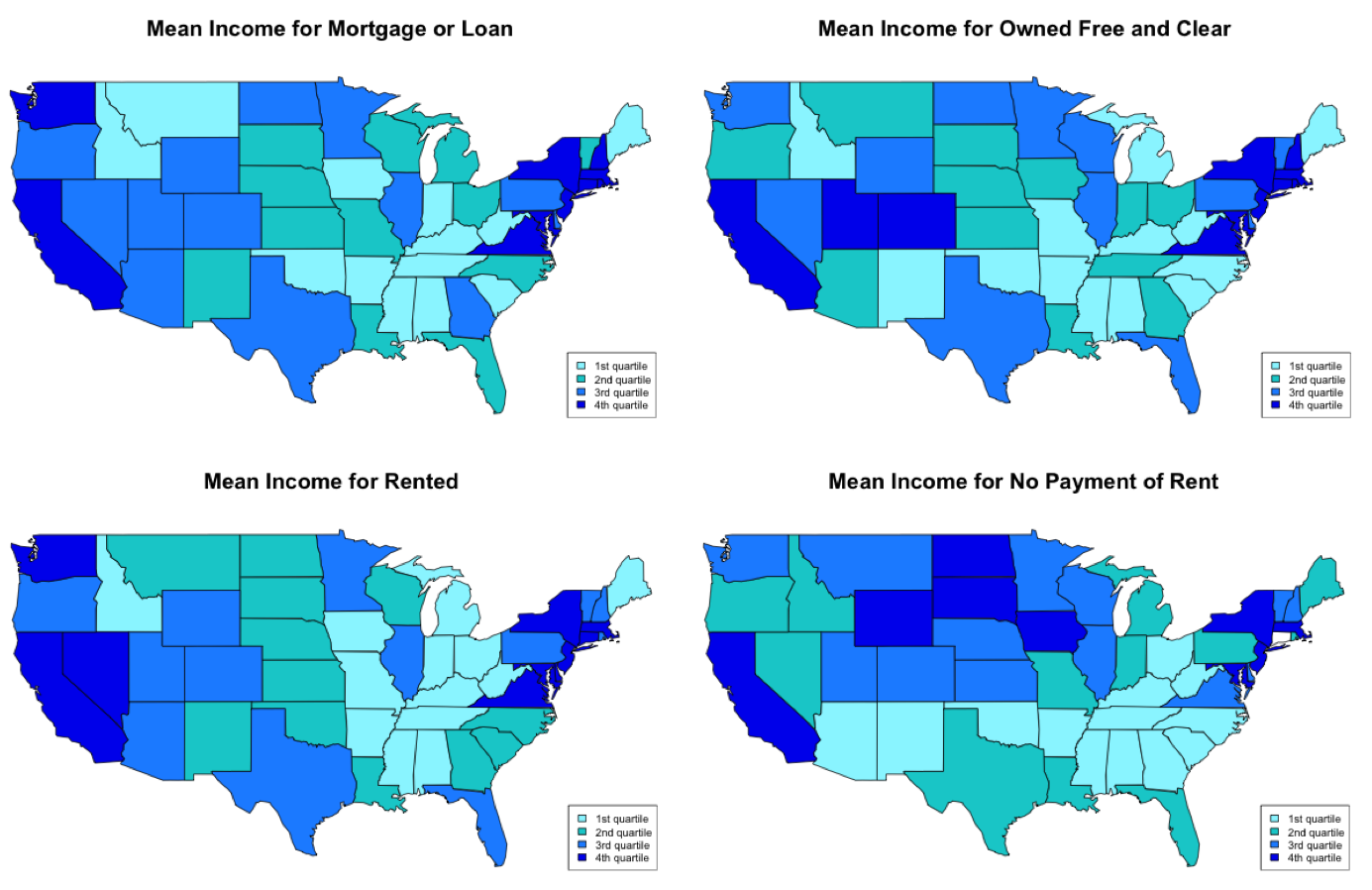
We find that mean household income for those having a mortgage or loan is highest in Washington D.C. ($154,465), followed by Connecticut ($126,760), New Jersey ($124,364), Maryland ($119,358) and Massachusetts ($118,119).

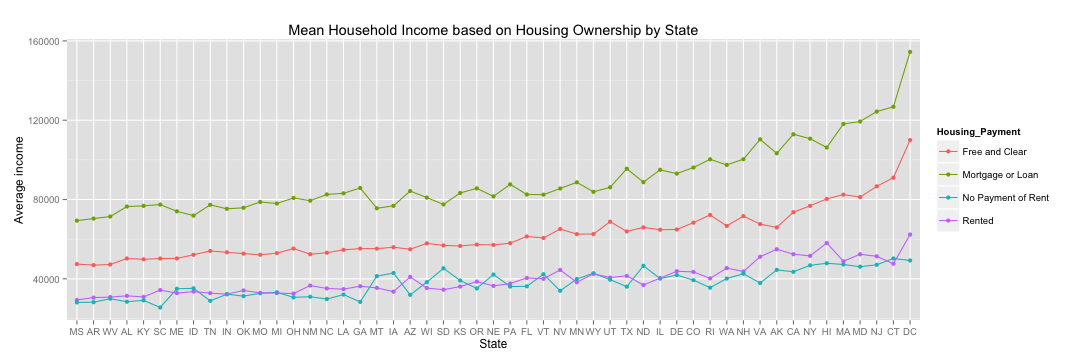
We find that mean household income for those owning free and clear follows similarly to above.

However, for those renting the mean household income is highest in Washington D.C. ($62,337), followed by Hawaii ($58,077), Alaska ($54,918), California ($52,423) and Maryland ($52,404).

For those having no payment of rent, the mean household income is highest in Connecticut ($50,264), followed by Washington D.C. ($49,266), Hawaii ($47,860), Massachusetts ($47,224) and New Jersey ($47,081).

It was interesting to see that Mississippi, Alabama and Arkansas were near the lowest mean household incomes across all ownership types, as seen in the graphs below.



 We also notice that households owning with a mortgage or loan had higher mean income than owning free and clear across all states. A possible explanation may be due to retirees (ie: smaller income but likely to have paid off the house). We see that renting and occupying without rent were fairly similar within all states.

More interesting trends: Midwest has high income/occupy with no rent and Florida has higher income for owned free and clear than surrounding region.

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: ~200 x 4 million dataset breaks R when you import it
  + Solution: Find columns of interest in the command line (head -n 1 ss12hus.csv | tr ',' '\n' | nl | grep -w 'TEN\|HINCP\|ST\|ADJHSG' | less -S), use cut -d, -f7,8,40,54 ss12hus.csv to 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
* Examine differences between each of the 4 household ownership types
  + Hypothesize that free & clear owners are retired, mortgage payers are still working
* Examine by region (East, South, West, Midwest)
* Take into consideration negative and null income
* Examine counts – what proportion of people fall into what category in each state?
* Include weighting: this analysis pretended the ACS data was from a simple random sample.