Spring 2019

An Economic Regression Analysis of Homeless Rates in the United States

Jessica Silber
John Carroll University, jsilber19@jcu.edu

An Economic Regression Analysis of Homeless Rates in the United States

Capstone Project

May 2019

By: Jessica Silber

Abstract

During my time at John Carroll, I was a member and President of the Labre Project, a homeless outreach student organization founded fifteen years ago. I spent almost every Friday at John Carroll out on the streets of Cleveland delivering meals and chatting with people experiencing homelessness. Through this experience, I learned many facts and lessons about what it is like to be homeless, how people become homeless, and how the city of Cleveland is trying to solve problems surrounding the homeless population. Although I explored the topic of homelessness a vast amount, I never got the answer to two questions: 'What variables cause homelessness?' and 'What are the most effective ways to reduce homelessness?'. This paper is a search for the answers to these questions using economic research, statistics, and regression analysis.

Introduction

Shelter is one of the most basic needs of a human being. It provides safety, warmth, and protection. Yet, in the United States in 2018 there were 552,830 homeless people, which is 17 out of every 10,000 people. This number is down 15% from 2007, with the biggest decreases being seen in veteran, youth, and chronic homeless populations, while individual homeless rates, making up 67% of this population, remain high at 372,417 people. These statistics come from the National Alliance to End Homelessness, which defines "homeless" using the US Department of Housing and Urban Development's definition of homelessness. Their definition consists of anyone in one of the following four categories: Literally Homeless, Imminent Risk of Homelessness, Homeless under Federal Status, and Fleeing/Attempting to Flee Domestic Violence. "Literally Homeless" refers to those individuals who do not have a fixed or adequate nighttime residency. Those in the “Imminent Risk of Homelessness” category are individuals who are going to be evicted in the next 14 days. The “Homeless under Federal Status” are unaccompanied youth and various other individuals meeting a long list of requirements. Lastly, those under “Fleeing/Attempting to Flee Domestic Violence” are individuals who are running away or attempting to run away from domestic violence. Those under this category are usually recorded as either “Sheltered” or “Unsheltered.” . Sheltered homeless are individuals that are in emergency shelters, transitional housing, or safe havens, while unsheltered homeless are individuals who’s nightly home is in a public location or a private location not designated for sleeping. (HUD Exchange)

Literature Review:

The US Department of Housing and Urban Development published “The 2018 Annual Homeless Assessment Report” containing information on the current population of homelessness from point-in-time estimates of homelessness. In the report there are many intriguing statistics and clues to variables that might be crucial in my regression analysis. First, over 50% of the homeless population in the US are living in the large cities, and these cities shelter over 71% of people in sheltered homelessness. Conversely unsheltered homelessness is most common in largely rural areas with 40% of unsheltered homelessness living in these locations. Another trend in this data is that 60.2% of all homelessness are male, but only 55% of sheltered homelessness are male while 69.9% of all unsheltered homeless are male. (HUD Exchange)

The article “Childhood Homelessness and Adult Employment” was published by Springer-Verlag Berlin Heidelberg in 2017. This article focuses on childhood homelessness and the effects it has on the children when they grow up. In the paper they find a direct positive relationship between childhood homelessness and adult unemployment rates. Therefore for people experiencing homelessness as a child, they are more likely to be unemployed as an adult. The paper suggested this correlation is partially due to low high school graduation rates and high incarceration rates for this population. Although this research was conducted in Europe, I would like to see how unemployment, high school graduation rates, and homelessness are correlated in US States. (Cobb-Clark, 894)

In the economic research paper "When warm and cold don't mix: The implications of climate for the determinants of homelessness" conducted at George Mason University, the effect of climate on predicting homelessness rates are examined. It discusses the fact that in different climates, different variables are significant in predicting unsheltered homelessness. The paper found that warmer climates have higher homeless rates, but also more variation in these rates from city to city. The paper also found that housing prices, poverty rates, and religion were more correlated with unsheltered homelessness in cities with warm climates. The result of this paper was that climate, including temperature and precipitation, are very significant when discussing the variables that contribute to homeless rates in the US. (Corinth, 45)

In the research paper entitled "Do Local Economic Conditions Affect Homelessness?" the effects of rent rates, poverty rates, and local policies on homeless rates in the United States were investigated. They found that the median rent, the share of households in rental housing, and poverty rates have very strong positive correlations with homelessness. They also found that the poverty rate only has a positive correlation in cities where there are more shelters and beds available. The paper suggested that in cities with constraints on beds available, the economic condition of the city has little effect on homeless rates. (Hanratty, 640)

“The Impact of Federal Homelessness Funding on Homelessness” was written by a professor at George Mason University and was published in the Southern Economic Journal in 2017. This research paper explored the effects of federal spending on homeless rates. The paper found that an increase in government spending increase sheltered homelessness and is unrelated to unsheltered homelessness. The paper says that funding increases chronic homelessness, allowing people to stay in shelters for long periods of time. The paper suggests that government spending, which mostly goes toward increasing shelters and beds available, is not lowering homeless rates, but increasing them. I would like to explore if other variables the government could be investing money into, that would reduce homelessness. (Lucas, 545)

In the paper “Employment and Earnings Trajectories during Two Decades among Adults in New York City Homeless Shelters” the effects of employment, wages, and job stability on homelessness are researched. The paper discusses the negative stereotypes given to the population, and how it has contributed to the neglecting of employment opportunities as a predictor of homelessness. The paper found that over half of all those in homelessness work and that their average income was only $13,000. The paper suggests correlation between unemployment, low wages, and homelessness, although it could not tell which variables influenced the others. (Metraux, 652)

Data:

To conduct my research, I created a pooled dataset of both cross-sectional and time series data. The pooled dataset contains records from the 48 continuous United States (excludes Alaska and Hawaii) for the years 2007 to 2015. Each record contains information on 19 variables. Three of these variables will be considered independent variables: Total Homeless (TH), Unsheltered Homeless (UH), and Sheltered Homeless (SH). The remainder will be dependent variables: year, Housing Price Index (HPI) , average temperature (Temp), total precipitation (Precip), rural versus urban index (Rururb), male percentage (Male), poverty rate (POVRT), total debt outstanding (TDept), population (Pop), percentage of population who are high school graduates (HSG), percentage of population with a bachelor's degree or higher (BD), unemployment rate (Unemp), population density (PD), Census Bureau Gini of Income Inequality (CBG), median income (MI), and total year round beds in either Emergency Shelters, Temporary Housing, or Shelters divided by Total Population (BEDS).

This data was collected from a variety of sources and for a variety of reasons. Total Homeless, Unsheltered Homeless, and Sheltered Homeless statistics came from the US Department of Housing and Urban Development's Definition of Homeless. These variables are the same as those defined in the Introduction. I chose these three independent variables because I wanted to see what variables affect only sheltered or unsheltered homeless rates, and which variables affect total homelessness. I would like to explore the similarities and differences of these three statistical measures of homelessness.

The Housing Price Index data came from the Federal Housing Finance Agency. The Housing Price Index per state is the average change in price between sales of the same properties. This statistic is used to track housing price trends and is useful in measuring housing affordability. The higher the HPI gets, the less affordable housing is in a particular state, thus I believe that HPI will have a positive correlation on all three independent variables.

The data for average temperature and total precipitation per state came from the National Centers for Environmental Information. The average temperature is recorded in degrees

Fahrenheit and the total precipitation is measured in inches. Based off the research in the literature review, I hypothesize that average temperature will have a positive correlation and total precipitation will have a negative effect.

The rural versus urban index, male percentages, poverty rates, Census Bureau Gini for income inequality, and medium income came from the US Census Bureau. The rural versus urban index measures the rural population divided by the urban population per state. Thus when the statistic is greater than one, the state is more rural than urban, and when the statistic is less than one, it is more urban than rural. Due to the higher percent of homeless people in cities, I hypothesize that this index will have a negative effect on homeless rates. The male percentage is the percent of the state's population that is male. Since a majority of those experiencing homelessness are male, especially for unsheltered homelessness, I expect that this variable will have a positive effect on all of the independent variables with the largest effect being on unsheltered rates. The poverty rate per state is the percent of the state's population living below the designated federal poverty level per year. I predict this will have a positive correlation on homeless rates. The Census Bureau Gini for income inequality is a statistic representing the income distribution between different wealth classes. The closer the statistic is to 0, the more equal the income distribution, while the closer it gets to 1, the more unequal the income distribution. Since income inequality causes a growth in poverty rates and a greater distance between the rich and the poor, I expect the CB Gini for income inequality to have a positive effect on the dependent variables. Lastly, median income is the median of all incomes per state. I used the median instead of the mean due to the skewedness of income data from income inequality. The higher someone's income, the less likely someone is to become homeless, thus I expect this variable to have a negative correlation to homeless rates.

The data for Total Debt outstanding per state came from the Federal Reserve. The higher this statistic, the more debt the average person in a state might have. The more debt a person has, the harder it is to get out of poverty or homelessness. Thus I expect Total Debt to have a positive correlation to homeless rates.

The CQ Press's State Fact Finder Series provided the data for population, high school graduate rates, bachelor's degree or higher rates, and population density. I expect population rates to have a positive effect on homeless rates, since the higher the number of people in a state, the more likely the state is to have a high number of homeless people. For high school graduate and college graduate rates, I expect a negative correlation, since the more a person is educated, the more likely a person is to get a high paying job. Lastly, population density is the state's population divided by the total area of the state in square miles. Thus the more dense the state, the higher the value. I expect this to have a positive correlation, since the denser the state the more likely the state will have a lack of housing and a lack of affordable housing.

My last variable, Unemployment Rate, comes from the Bureau of Labor Statistics. This is a measure of how many individuals in the work force are unemployed, thus the higher this rate, the higher homeless rates are forecasted to be. Thus I predict a positive correlation.

Methodology:

The goal of my methods is to find which independent variables from my data influence the number of people experiencing both sheltered, unsheltered, and overall homelessness in various states, and explore how these variables interact with each other in predicting these population numbers. To conduct my research, I used R Studio. First, I ran basic linear models between each independent variable and dependent variables, and then graphed each model.

These models took on the form $Y = c_o + c_1 X$, where Y is the dependent variable and X is the independent variable. I then ran correlation tests on each set and recorded their correlation coefficients. Running the basic linear models allowed me to create graphs of the interaction between each independent and dependent variable, and the correlation coefficients allowed me to see the direction and degree of the correlation between each set of variables.

Next, I ran my regressions, by creating multilinear models for each dependent variable of the form:

$$Y = c_o + c_1 X_1 + c_2 X_2 + c_3 X_3 + c_4 X_4 + c_5 X_5 + c_6 X_6 + c_7 X_7 + c_8 X_8 + c_9 X_9 + c_{10} X_{10} + c_{11} X_{11} + c_{12} X_{12}$$
$$+ c_{13} X_{13} + c_{14} X_{14} + c_{15} X_{15} + c_{16} X_{16}$$

where each X_i represents a dependent variable, and each c_i is the correlation coefficient for that variable. I then reviewed each of the three models, looking at the adjusted R-squared value, the correlation coefficients for each variable, and the p-value of each independent. The adjusted R-squared value conveys how well each model predicted the specified dependent variable. The correlation coefficients specify how each independent variable influenced the dependent variable. Lastly, the p-value for each independent variable showed the significance of each independent variable on the model.

After reviewing each model, I chose to remove various variables from each model due to low significance levels, possible collinearity between variables, and unexpected degrees or directions of correlation coefficients. I then ran regressions of each new model, and reviewed them in the same method described for the first regression models.

To find out which regression model is superior, the full model or the reduced model, I ran an anova test on the set of models for each dependent variable. The anova test explains whether or not the variables removed were significant in the model. If the variables removed are not significant, then the reduced model is superior. I used a .001 significance level to determine this.

Based on the superior model for each dependent variable, I explored the effects of each independent variables in and between each model. I also explored the similarities and differences between each model. Lastly, I summarized these concluding results in various charts.

Results:

Attached are the graphs of each of the basic linear models between the dependent and independent variables, and below is a chart of the correlation coefficients for each.

	Total Homeless	Unsheltered	Sheltered
Year	-0.028	-0.048	-0.001
HPI	0.431	0.375	0.391
TEMP	0.236	0.316	0.1
PRECIP	-0.078	-0.15	0.012
RURURB	-0.364	-0.254	-0.393
POVRT	0.139	0.144	0.102
Male %	-0.01	0.114	-0.134
Tdebt	0.85	0.611	0.902
Population	0.898	0.791	0.804
HS Grad%	-0.329	-0.333	-0.251
BD%	0.1707	0.061	0.255
Unemployment	0.1962	0.233	0.114
Pop. Density	0.1426	0.025	0.23
CB Gini	0.426	0.258	0.501
Med Income	0.008	-0.045	0.059
Bed %	0.885	0.016	0.51

The graphs for the variable year depict a slight negative curve for unsheltered and total homeless, while sheltered homeless appears to be a flat line. The correlation coefficients concur with this showing the correlation coefficient for shelter versus year to be -.001. This evidence agrees with my research that overall homeless rates are falling, specifically for unsheltered populations.

The information for the housing price index, shows a strong positive correlation for all three homeless categories as predicted. The sheltered population is slightly more correlated than the unsheltered.

Temperature data shows the strong positive correlation I predicted. This correlation is weaker for sheltered homeless populations, and this can be seen by the large number of outliers in its graph. This could be due to the weather having a larger effect on those living outside of shelters, since they have an increased exposure to the elements.

The data on precipitation totals shows a negative correlation for unsheltered and total homeless populations as predicted. There is a slight positive correlation for sheltered homeless however, which could be due to the weather having less effect on those in sheltered populations or the weather driving more people into shelters, causing shelter rates to go up.

The graphs and correlation coefficients for rural versus urban index data show a negative correlation as predicted. There are a number of outliers between 0 and .5 and the plot resembles a logarithm curve, so a logarithmic model for rural vs. urban index may be more appropriate than a linear model.

The simple linear model for poverty rates depicts the negative relationship that was predicted. There is a slightly higher correlation with unsheltered than with sheltered.

The percentage of a population that is male has a positive correlation with unsheltered population and a negative correlation with sheltered populations. This might suggest that there is a higher number of men in the unsheltered population and a higher number of females in the sheltered homeless population.

Total debt and population per state both had a positive correlation with each of the homeless populations, specifically with sheltered homelessness.

High school graduate and college graduate rates were both predicted to have negative correlations. The high school graduate rate had the predicted negative correlation, but college degree had an unpredicted positive correlation. I am unsure of why this occurred; one theory is that many people come out of college with considerable debt which makes it difficult for people to make a living.

Unemployment, population density, and Census Bureau Gini of income inequality all had positive correlation values as predicted. Unemployment was significantly more highly correlated with unsheltered homeless populations than sheltered. This could be because unemployment causes sudden changes in income which may lead to people experiencing homelessness for the first time. People experiencing homelessness for the first time are more likely to end up unsheltered homeless, thus increasing unsheltered rates more than sheltered. Population Density had a very small correlation with unsheltered homelessness, suggesting it is not useful in predicting unsheltered homeless rates. The correlation coefficient for the Census Bureau for income inequality was significantly higher for sheltered homeless rates than unsheltered. This could be because the higher income inequality is, the harder it could be for a working individual to find housing, thus causing long term homelessness. Those experiencing long term homelessness are more likely to be sheltered homeless.

Medium income had a very small correlation with all three dependent variables and changing direction between sheltered and unsheltered homeless populations. These low values and unpredicted direction changes could be due to a number of reasons. One hypothesis is that when income inequality rises, it pulls the medium income higher, but there are still many people on the bottom of the income bracket.

The bed rate per state had an unpredicted, high correlation with sheltered and total homelessness and almost no correlation with unsheltered homelessness. The positive correlation could be because the more beds available, the more homeless people fall under the sheltered homeless category, thus the sheltered homeless population grows. The low correlation could be due to the number of outliers on the bed rate graph displaying a negative correlation. These hypotheses might suggest that increasing the number of beds decreases unsheltered homelessness but does not decrease overall homelessness.

Below are the regression outputs of the full and reduced models for total homeless:

Total Homeless Full:

```
Call:
lm(formula = TH ~ YEAR + HPI + TEMP + PRECIP + RURURB + POVRT +
    Male + Tdebt + Pop + HSG + BD + Unemployment + PD + CBG +
    MI + BEDS2, data = TH_Regression)

Residuals:
     Min       1Q   Median       3Q      Max
-28821.6  -3286.8     32.2   3668.0  30967.5

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.038e+06  3.329e+05  -3.117  0.00196 **
YEAR          5.007e+02  1.763e+02   2.840  0.00474 **
HPI           3.619e+01  3.776e+00   9.585  < 2e-16 ***
TEMP          2.581e+02  7.966e+01   3.240  0.00129 **
PRECIP       -4.510e+01  3.271e+01  -1.379  0.16866
RURURB        1.211e+03  6.008e+02   2.015  0.04455 *
POVRT        -6.418e+02  2.036e+02  -3.151  0.00174 **
Male          1.481e+03  7.655e+02   1.934  0.05379 .
Tdebt         2.453e-01  2.881e-02   8.514 3.11e-16 ***
Pop           1.641e+00  1.081e-01  15.170  < 2e-16 ***
HSG          -4.554e+02  2.094e+02  -2.175  0.03021 *
BD           -5.012e+01  1.337e+02  -0.375  0.70801
Unemployment  4.605e+02  1.445e+02   3.187  0.00155 **
PD           -8.632e+00  2.044e+00  -4.223 2.97e-05 ***
CBG          -3.018e+04  3.122e+04  -0.967  0.33430
MI           -3.614e-01  8.000e-02  -4.518 8.16e-06 ***
BEDS2         4.047e+09  7.723e+08   5.240 2.57e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6041 on 415 degrees of freedom
Multiple R-squared:  0.9181,	Adjusted R-squared:  0.9149
F-statistic: 290.7 on 16 and 415 DF,  p-value: < 2.2e-16
```

Total Homeless Reduced:

```
Call:
lm(formula = TH ~ YEAR + HPI + TEMP + RURURB + POVRT + Male +
    Tdebt + Pop + HSG + Unemployment + PD + CBG + MI + BEDS2,
    data = TH_Regression)

Residuals:
     Min       1Q   Median       3Q      Max
-29268.2  -3259.0     84.7   3648.5  31219.6

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.027e+06  3.234e+05  -3.175 0.001608 **
YEAR          4.805e+02  1.694e+02   2.837 0.004769 **
HPI           3.558e+01  3.567e+00   9.976  < 2e-16 ***
TEMP          2.491e+02  7.867e+01   3.166 0.001659 **
RURURB        1.157e+03  5.990e+02   1.931 0.054156 .
POVRT        -5.732e+02  1.964e+02  -2.919 0.003698 **
Male          2.197e+03  5.676e+02   3.870 0.000126 ***
Tdebt         2.478e-01  2.871e-02   8.631  < 2e-16 ***
Pop           1.638e+00  1.074e-01  15.245  < 2e-16 ***
HSG          -4.994e+02  1.780e+02  -2.806 0.005243 **
Unemployment  4.272e+02  1.425e+02   2.998 0.002878 **
PD           -8.523e+00  2.023e+00  -4.213 3.09e-05 ***
CBG          -4.015e+04  2.883e+04  -1.392 0.164520
MI           -3.604e-01  6.996e-02  -5.151 4.00e-07 ***
BEDS2         4.109e+09  7.703e+08   5.334 1.58e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6040 on 417 degrees of freedom
Multiple R-squared:  0.9177,	Adjusted R-squared:  0.9149
F-statistic: 332.1 on 14 and 417 DF,  p-value: < 2.2e-16
```

The full model for the total homeless population had a high adjusted R-squared of .9149, meaning that 91.29% of the response variable’s variation can be explained by this multi-linear model. The p-values for the dependent variables precipitation, bachelor’s degree, and Census Bureau Gini of income inequality all appeared to be insignificant in this model. Along with precipitation and bachelor’s degree appearing insignificant, I also suspected collinearity between precipitation and temperature; and high school graduates and bachelor degree graduates. I ran correlation tests on these sets, and I found the correlation coefficient for the first was .3288997 and the second was .4837658. These are significant correlations, and thus I excluded these variables from all of the reduced models.

For the reduced model for total homelessness, I removed precipitation and bachelor’s degree. The adjusted R-squared remained the same at .9149. The Census Bureau Gini remained insignificant, so if I ran a third model, I would have considered removing that variable. I than ran an anova test to see which model was superior:

```
Analysis of Variance Table

Model 1: TH ~ YEAR + HPI + TEMP + RURURB + POVRT + Male + Tdebt + Pop +
    HSG + Unemployment + PD + CBG + MI + BEDS2
Model 2: TH ~ YEAR + HPI + TEMP + PRECIP + RURURB + POVRT + Male + Tdebt +
    Pop + HSG + BD + Unemployment + PD + CBG + MI + BEDS2
  Res.Df        RSS Df Sum of Sq      F Pr(>F)
1    417 1.5215e+10
2    415 1.5143e+10  2  71294110 0.9769 0.3773
```

The test shows an insignificant F statistic, thus the reduced model is a better fit, meaning that precipitation and bachelor degree are insignificant in predicting total homelessness.

Below are the regression outputs for the full and reduced models of unsheltered homelessness:

Unsheltered Homeless Full:

```
Call:
lm(formula = UH ~ YEAR + HPI + TEMP + PRECIP + RURURB + POVRT +
    Male + Tdebt + Pop + HSG + BD + Unemployment + PD + CBG +
    MI + BEDS2, data = UH_Regression)

Residuals:
     Min       1Q   Median       3Q      Max
-22162.3  -3244.5   -115.4   2882.6  28128.3

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.085e+06  2.899e+05  -3.744 0.000207 ***
YEAR          5.157e+02  1.535e+02   3.359 0.000853 ***
HPI           4.295e+01  3.288e+00  13.065  < 2e-16 ***
TEMP          3.535e+02  6.936e+01   5.096 5.28e-07 ***
PRECIP       -5.330e+01  2.848e+01  -1.871 0.061986 .
RURURB        2.045e+03  5.231e+02   3.910 0.000108 ***
POVRT        -5.806e+02  1.773e+02  -3.275 0.001147 **
Male          1.362e+03  6.665e+02   2.044 0.041593 *
Tdebt         3.377e-02  2.508e-02   1.346 0.178908
Pop           1.106e+00  9.416e-02  11.746  < 2e-16 ***
HSG          -1.134e+02  1.823e+02  -0.622 0.534397
BD           -2.149e+02  1.164e+02  -1.846 0.065603 .
Unemployment  7.192e+02  1.258e+02   5.715 2.10e-08 ***
PD           -5.912e+00  1.780e+00  -3.321 0.000975 ***
CBG          -5.333e+04  2.718e+04  -1.962 0.050471 .
MI           -2.411e-01  6.966e-02  -3.461 0.000593 ***
BEDS2        -3.132e+09  6.725e+08  -4.658 4.30e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5260 on 415 degrees of freedom
Multiple R-squared:  0.8079,    Adjusted R-squared:  0.8005
F-statistic: 109.1 on 16 and 415 DF,  p-value: < 2.2e-16
```

Unsheltered Homeless Reduced

```
Call:
lm(formula = UH ~ YEAR + HPI + TEMP + RURURB + Pop + POVRT +
    Male + Unemployment + PD + CBG + MI + BEDS2, data = UH_Regression)

Residuals:
     Min       1Q   Median       3Q      Max
-23095.7  -3014.2    157.4   2823.6  29034.9

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -9.034e+05  2.522e+05  -3.582 0.000380 ***
YEAR          3.989e+02  1.290e+02   3.093 0.002116 **
HPI           4.156e+01  3.101e+00  13.405  < 2e-16 ***
TEMP          3.418e+02  6.097e+01   5.606 3.77e-08 ***
RURURB        2.031e+03  5.265e+02   3.859 0.000132 ***
Pop           1.223e+00  5.095e-02  24.007  < 2e-16 ***
POVRT        -3.516e+02  1.552e+02  -2.265 0.024034 *
Male          2.101e+03  4.904e+02   4.285 2.27e-05 ***
Unemployment  6.999e+02  1.250e+02   5.599 3.91e-08 ***
PD           -5.488e+00  1.727e+00  -3.177 0.001597 **
CBG          -5.686e+04  2.386e+04  -2.382 0.017641 *
MI           -2.632e-01  6.086e-02  -4.325 1.91e-05 ***
BEDS2        -2.836e+09  6.344e+08  -4.470 1.01e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5311 on 419 degrees of freedom
Multiple R-squared:  0.8023,    Adjusted R-squared:  0.7966
F-statistic: 141.7 on 12 and 419 DF,  p-value: < 2.2e-16
```

The full model for unsheltered homlesssness has a lower adjusted R-squared of .8005. The variables precipitation and bachelor degree still appeared insignificant, along with total debt and high school graduates. For the reduced model, I removed all four of the insignificant variables from the previous model. The adjusted R-squared was slghtly lower at .7966, but all the variables in the second mdel are significant at a .01 level. Below is the anova test of the two models:

```
Analysis of Variance Table

Model 1: UH ~ YEAR + HPI + TEMP + RURURB + Pop + POVRT + Male + Unemployment +
    PD + CBG + MI + BEDS2
Model 2: UH ~ YEAR + HPI + TEMP + PRECIP + RURURB + POVRT + Male + Tdebt +
    Pop + HSG + BD + Unemployment + PD + CBG + MI + BEDS2
  Res.Df        RSS Df Sum of Sq      F  Pr(>F)
1    419 1.1820e+10
2    415 1.1481e+10  4 339295036 3.0662 0.01651 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F statistic is insignificant at a .001 level, thus we can assume the reduced model is superior. precipitation, bachelor degree, high school, and total debt are insignificant in predicting unsheltered homelessness.

The full and reduced models for the sheltered homeless population are below:

Sheltered Homeless Full:

```
Call:
lm(formula = SH ~ YEAR + HPI + TEMP + PRECIP + RURURB + POVRT +
    Male + Tdebt + Pop + HSG + BD + Unemployment + PD + CBG +
    MI + BEDS2, data = SH_Regression)

Residuals:
     Min       1Q   Median       3Q      Max
-10644.0  -1325.7    195.2   1538.2  20237.7

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.754e+04  1.754e+05   0.271 0.786533
YEAR         -1.505e+01  9.290e+01  -0.162 0.871393
HPI          -6.760e+00  1.990e+00  -3.398 0.000745 ***
TEMP         -9.540e+01  4.198e+01  -2.273 0.023554 *
PRECIP        8.196e+00  1.724e+01   0.476 0.634680
RURURB       -8.347e+02  3.166e+02  -2.637 0.008684 **
POVRT        -6.110e+01  1.073e+02  -0.569 0.569396
Male          1.182e+02  4.034e+02   0.293 0.769662
Tdebt         2.115e-01  1.518e-02  13.933  < 2e-16 ***
Pop           5.345e-01  5.699e-02   9.379  < 2e-16 ***
HSG          -3.420e+02  1.103e+02  -3.100 0.002070 **
BD            1.648e+02  7.047e+01   2.339 0.019809 *
Unemployment -2.586e+02  7.616e+01  -3.396 0.000749 ***
PD           -2.720e+00  1.077e+00  -2.525 0.011938 *
CBG           2.315e+04  1.645e+04   1.407 0.160179
MI           -1.203e-01  4.216e-02  -2.854 0.004531 **
BEDS2         7.179e+09  4.070e+08  17.640  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3183 on 415 degrees of freedom
Multiple R-squared:  0.9267,	Adjusted R-squared:  0.9238
F-statistic: 327.8 on 16 and 415 DF,  p-value: < 2.2e-16
```

Sheltered Homeless Reduced

```
Call:
lm(formula = SH ~ HPI + TEMP + RURURB + Tdebt + Pop + HSG + Unemployment +
    MI + BEDS2, data = SH_Regression)

Residuals:
     Min       1Q   Median       3Q      Max
-10813.0  -1461.5    236.6   1673.1  20403.6

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.619e+04  7.953e+03   3.294 0.001072 **
HPI          -6.826e+00  1.628e+00  -4.193 3.36e-05 ***
TEMP         -6.816e+01  3.816e+01  -1.786 0.074805 .
RURURB       -6.272e+02  2.950e+02  -2.127 0.034034 *
Tdebt         2.094e-01  1.233e-02  16.983  < 2e-16 ***
Pop           5.814e-01  4.707e-02  12.352  < 2e-16 ***
HSG          -2.667e+02  7.653e+01  -3.484 0.000545 ***
Unemployment -2.886e+02  7.159e+01  -4.031 6.58e-05 ***
MI           -7.579e-02  3.025e-02  -2.506 0.012594 *
BEDS2         7.523e+09  3.859e+08  19.493  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3213 on 422 degrees of freedom
Multiple R-squared:  0.924,	Adjusted R-squared:  0.9224
F-statistic: 570.4 on 9 and 422 DF,  p-value: < 2.2e-16
```

The adjusted R-squared for this full model is the highest at .9238, meaning this model was the best at predicting the variation in its specified population. This model also had the most insignificant variables: year, precipitation, poverty rates, male percentage, bachelor degree, population density, and the Census Bureau Gini. For my reduced model, I removed all seven insignificant variables. The adjusted R-squared for the reduced model was only marginally lower at .9224, and all the variables are significant at a .05 significance level. The anova test for the last two models is shown below:

```
Analysis of Variance Table

Model 1: SH ~ HPI + TEMP + RURURB + Tdebt + Pop + HSG + Unemployment +
    MI + BEDS2
Model 2: SH ~ YEAR + HPI + TEMP + PRECIP + RURURB + POVRT + Male + Tdebt +
    Pop + HSG + BD + Unemployment + PD + CBG + MI + BEDS2
  Res.Df        RSS Df Sum of Sq      F  Pr(>F)
1    422 4356046278
2    415 4204870306  7 151175972 2.1315 0.03939 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F statistic is insignificant at a .001 level, and thus the reduced model is superior. Thus all seven variables removed are insignificant in predicting unsheltered homelessness.

Below is a table summarizing the regression results for all three independent variables:

	TH Significant	TH Insignificant	UH Significant	UH Insignificant	SH Significant	SH Insignificant
Year	+		+			x
HPI	+		+		-	
TEMP	+		+		-	
PRECIP		x		x		x
RURURB	+		+		-	
POVRT	-		-			x
Male %	+		+			x
Tdebt	+			x	+	
Population	+		+		+	
HS Grad%	-			x	-	
BD%		x		x		x
Unemployment	+		+		-	
Pop. Density	-		-			x
CB Gini	-		-			x
Med Income	-		-		-	
Bed %	+		-		+	

The variables that remained significant across all three categories of homelessness were the Housing Price Index, temperature, rural versus urban index, population, unemployment rate, median income, and beds available rate. The variables that were significant for unsheltered but not sheltered were year, poverty rates, male percentage, population density, and the Census Bureau Gini. The variables that were significant for sheltered, but not unsheltered were total debt and high school Graduates Rates.

Conclusion:

The goal of this paper was to discover what causes homelessness and which variables might decrease homelessness if altered. The variables that were found to be significant in causing homelessness were: Housing Price Index, temperature, rural versus urban index, population, unemployment rate, median income, and beds available rate. A number of these variables are difficult or impossible to change such as temperature, rural versus urban index, and population. The other variables can all be manipulated or affected by people and policies, and could be used to help lower homeless rates in some way.

The housing price index could be lowered by creating more affordable housing in cities with high housing prices. This would cause a decrease in the housing price index, and as my research shows, it would lead to a decrease in total homelessness. Unemployment rates can be changed by the government creating or subsidizing more jobs in the US or by companies choosing to hire more people. This would cause a decrease in unemployment and a decrease in homelessness. Median income can be increased by companies paying their lower employees more and their higher employees less. This would draw the median income up, and cause homeless rates to decrease. Lastly, increasing the number of beds in shelters would decrease unsheltered homelessness, however it would increase sheltered homeless.

Along with these conclusions, there were a few other pieces of information I discovered in my data. Firstly, year appeared insignificant in predicting sheltered homelessness, which suggests that for the time period my data came from, 2007 to 2015, this rate was stagnant, neither increasing nor decreasing as time progressed. Next, the regression for the sheltered population had a much higher adjusted R-squared than the unsheltered regression. This suggests that the regression of unsheltered homelessness was missing some dependent variables. Lastly, sheltered homelessness and unsheltered homelessness are two very different forms of homelessness that exist in the United States, and when the government or organizations are looking to reduce homelessness, they should take into account how different variables will affect both populations along with the total population.

In the future, this research could be continued in several ways. The regression could be improved by adding more variables such as substance abuse, number of veterans in a state, or mental illness rates. This research could also be done at a city or country level to explore homelessness in a broader or more focused approach. The regressions could also be explored further by using non-linear models or changing the variables in the multi-linear models. Lastly, this research could also be furthered by exploring how these independent variables affect other categories of homelessness, such as chronic homelessness or veteran homelessness.

Bibliography

Board of Governors of the Federal Reserve System. (n.d.). Retrieved from https://www.federalreserve.gov/

Cobb-Clark, D. A., & Zhu, A. (2017). Childhood Homelessness and Adult Employment: The Role of Education, Incarceration, and Welfare Receipt. Journal of Population Economics, 30(3), 893–924. https://doi.org/https://link.springer.com/journal/volumesAndIssues/148

Corinth, K., & Lucas, D. S. (2018). When warm and cold don't mix: The implications of climate for the determinants of homelessness. Journal of Housing Economics, 41, 45-56. doi:10.1016/J.JHE.2018.01.001

Hanratty, M. (2017). Do Local Economic Conditions Affect Homelessness? Impact of Area Housing Market Factors, Unemployment, and Poverty on Community Homeless Rates. Housing Policy Debate, 27(4), 640–655. https://doi.org/http://www.tandfonline.com/loi/rhpd20

HUD Exchange. (n.d.). Retrieved from https://www.hudexchange.info/

Lucas, D. S. (2017). The Impact of Federal Homelessness Funding on Homelessness. Southern Economic Journal, 84(2), 548–576. Retrieved from http://jcu.ohionet.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&dbecn&AN=1687514&site=ehost-live

Metraux, S., Fargo, J. D., Eng, N., & Culhane, D. P. (2018). Employment and Earnings Trajectories during Two Decades among Adults in New York City Homeless Shelters. Cityscape: A Journal of Policy Development and Research, 20(2), 173–202. Retrieved from http://jcu.ohionet.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&dbecn&AN=1737394&site=ehost-live

Morgan, K. O., & Morgan, S. (2018). *State rankings 2018: A statistical view of America*. Thousand Oaks, CA: Sage/CQ Press.

National Alliance to End Homelessness. (n.d.). Retrieved from https://endhomelessness.org/

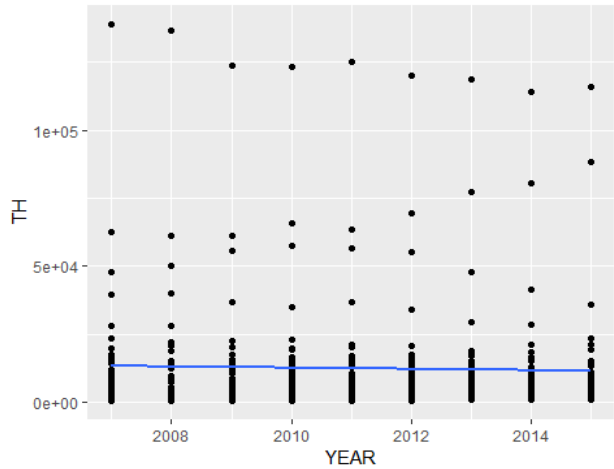
United States Interagency Council on Homelessness (USICH). (n.d.). Retrieved from https://www.usich.gov/

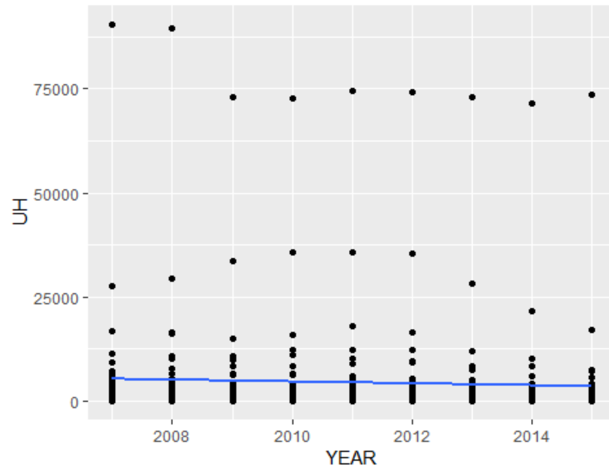
US Bureau of Labor Statistics. (n.d.). Retrieved from https://www.bls.gov/home.htm

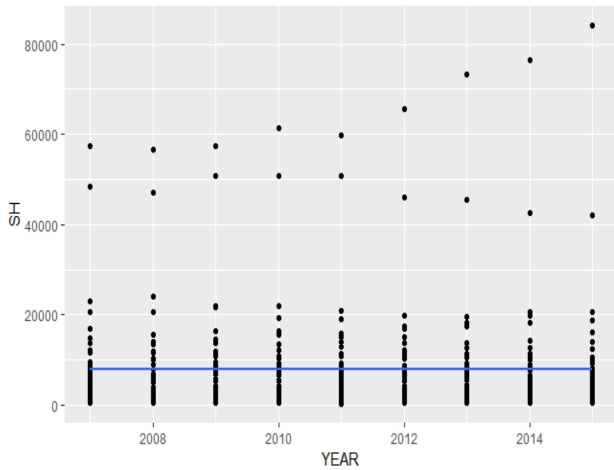
US Census Bureau. (n.d.). Census.gov. Retrieved from https://www.census.gov/en.html

Basic Linear Graphs

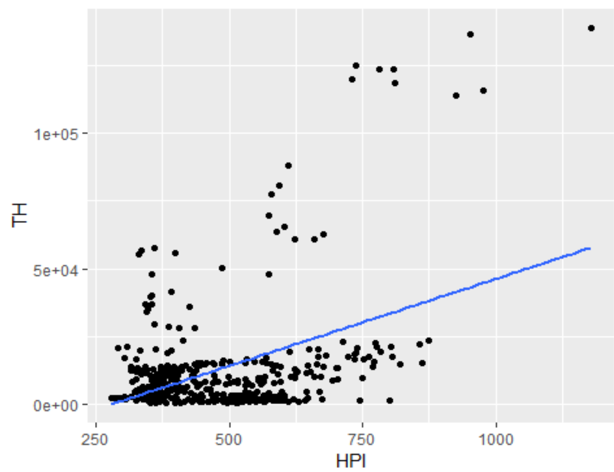
Year:

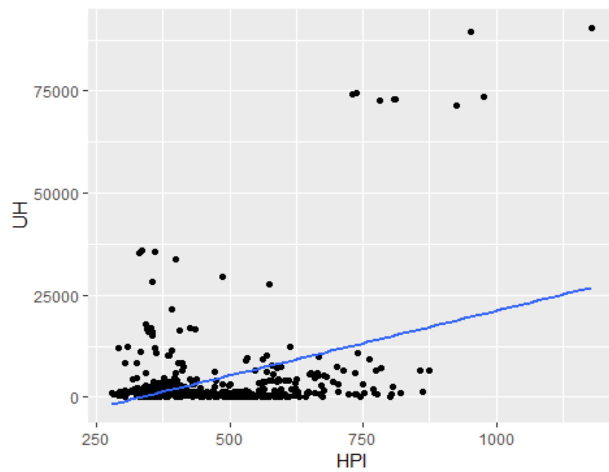


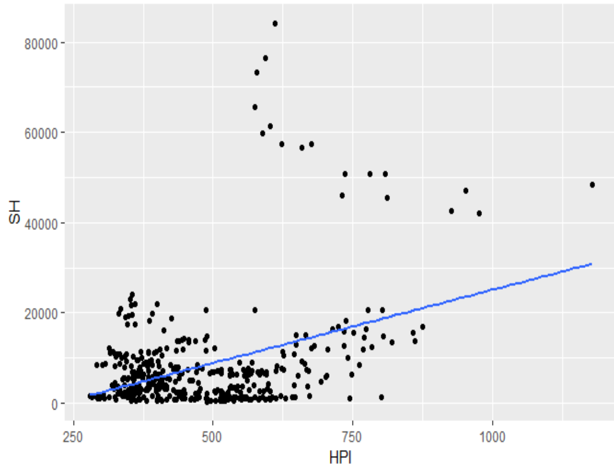




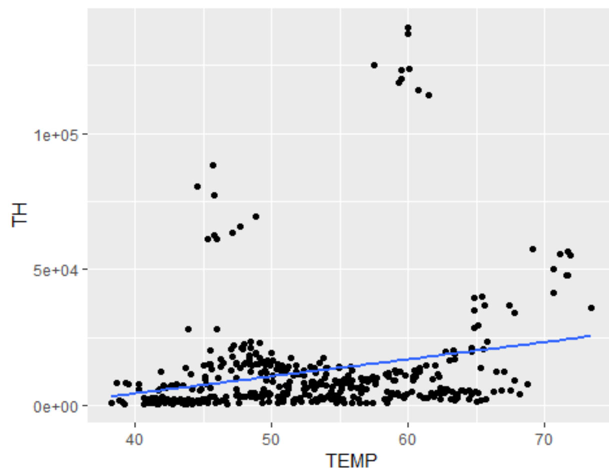
Housing Price Index:

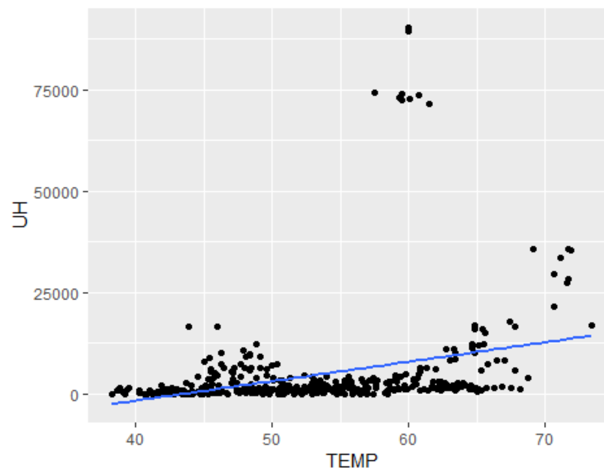


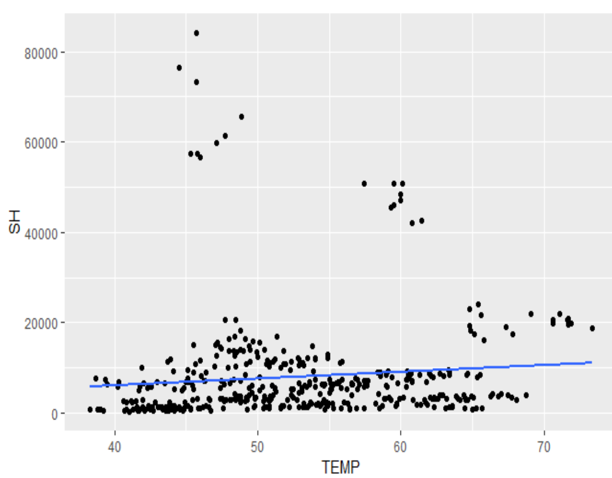




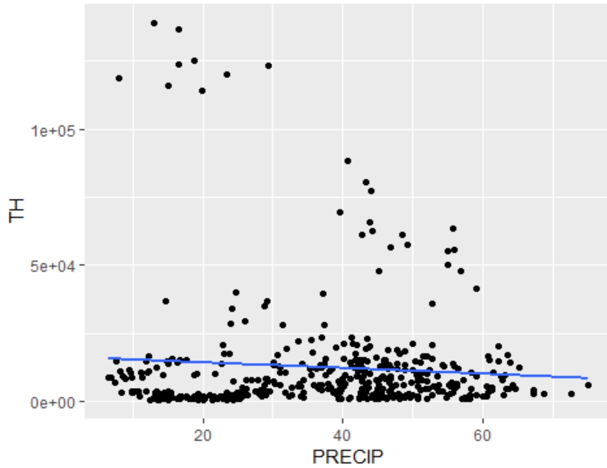
Temperature:

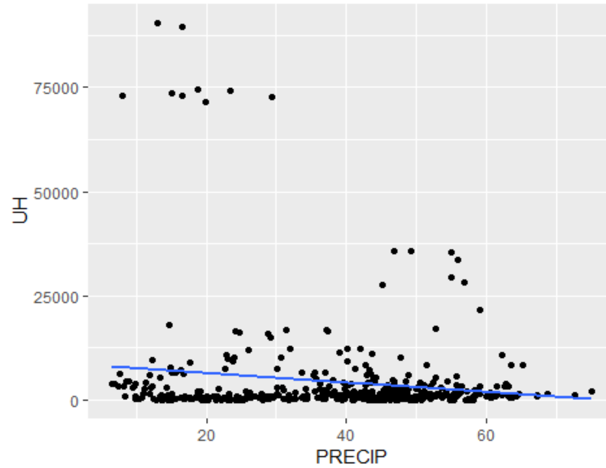


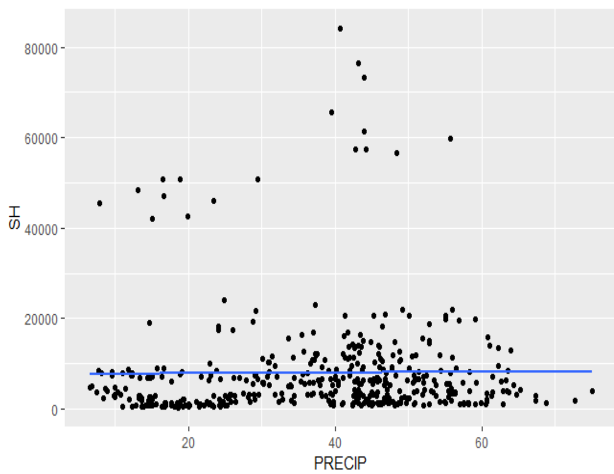




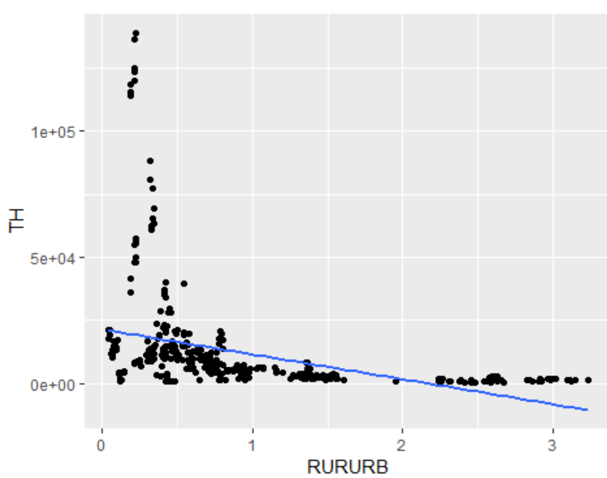
Precipitation:

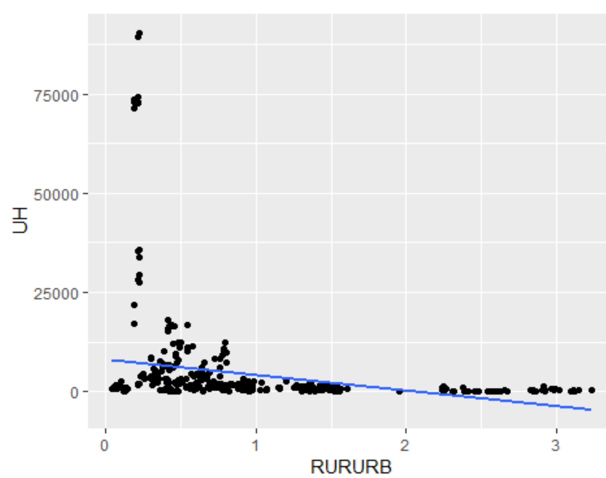


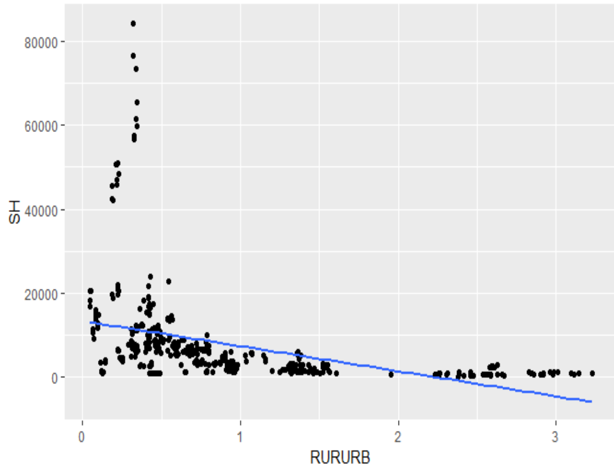




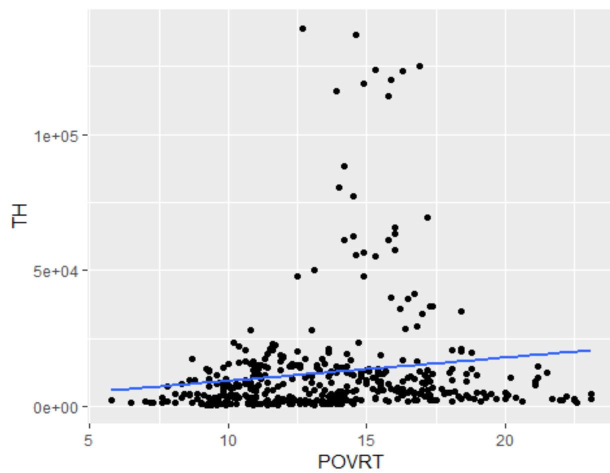
Rural versus Urban Index:

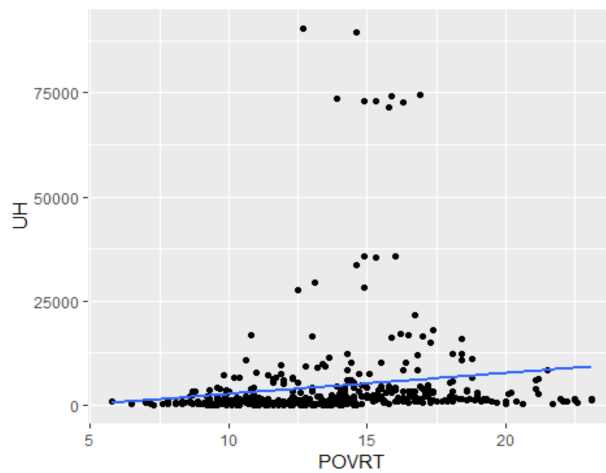


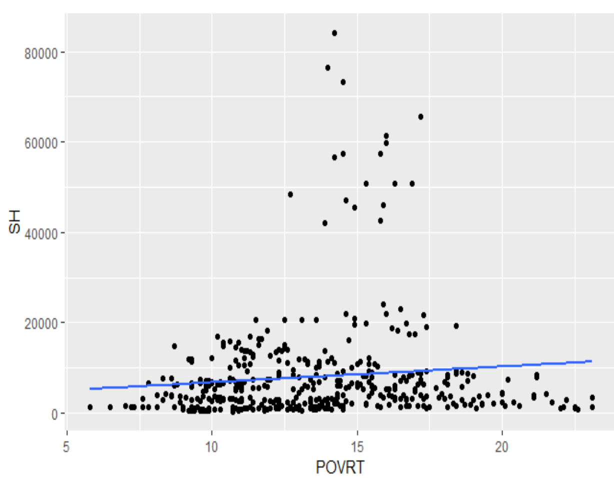




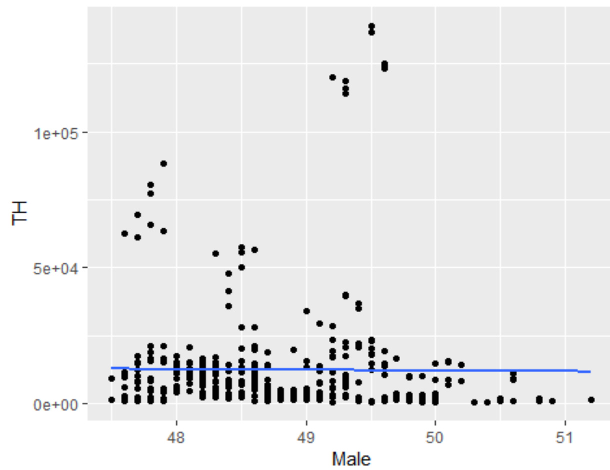
Poverty Rate:

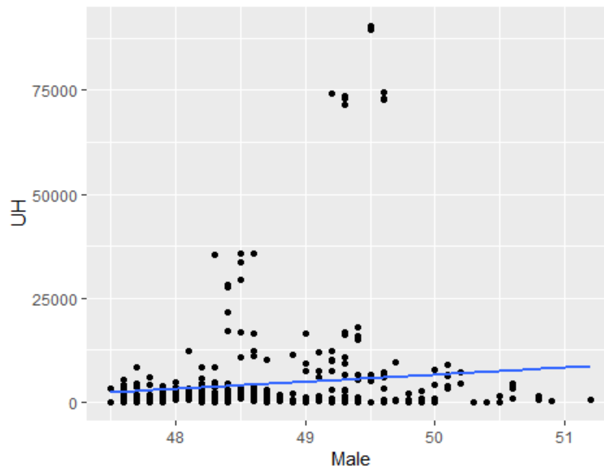


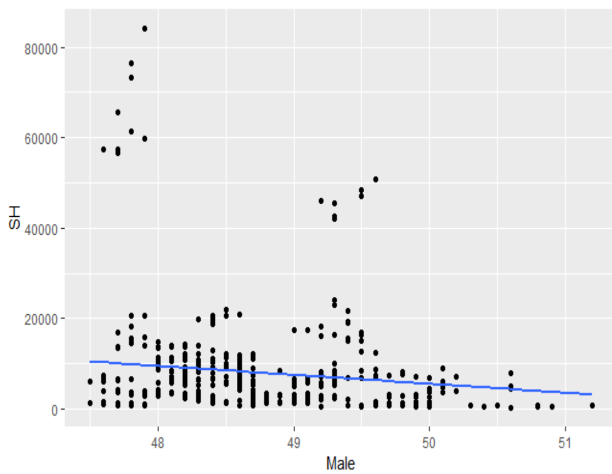




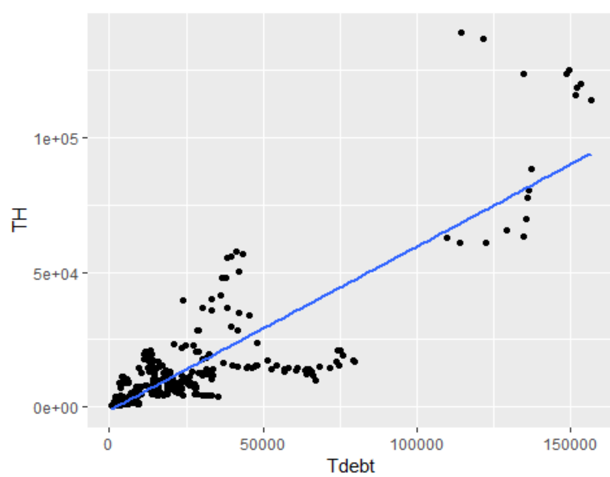
Male:

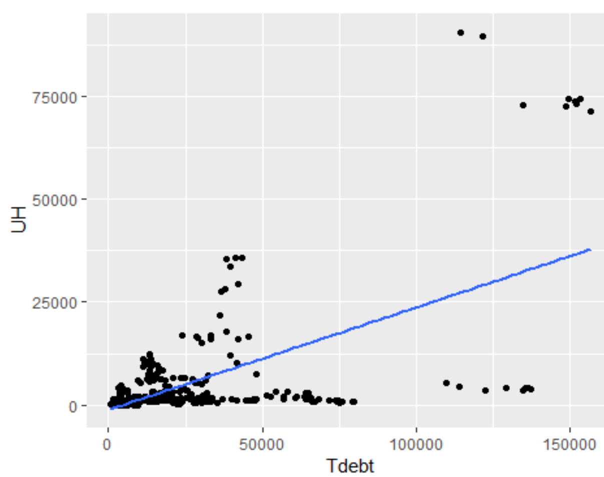


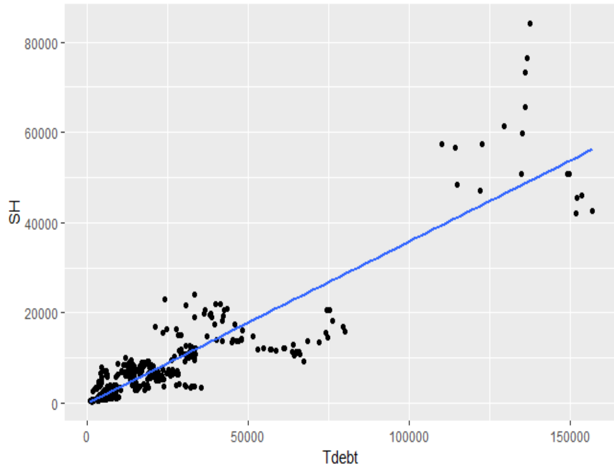




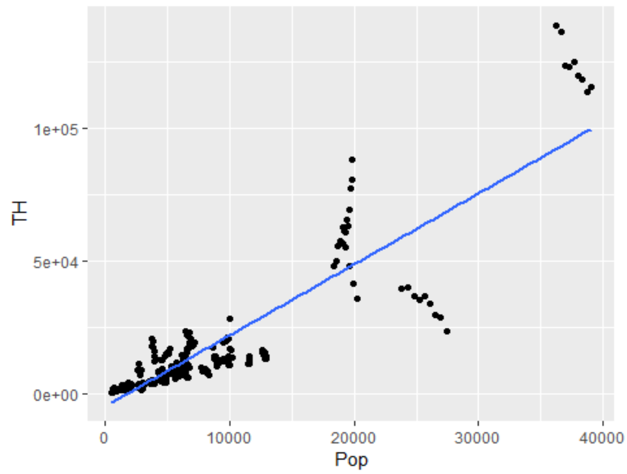
Total Debt:

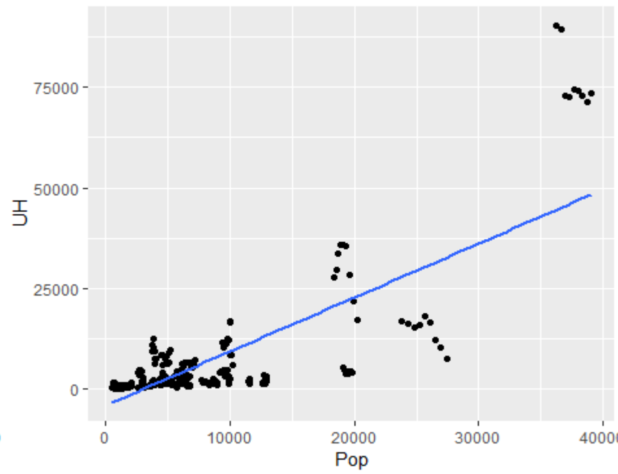


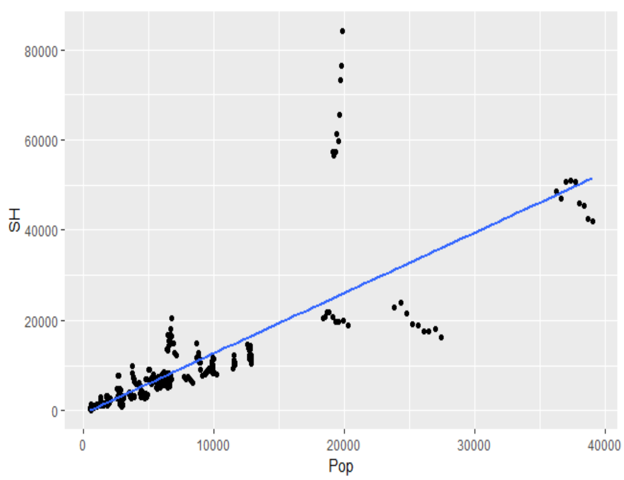



SH
Tdebt
80000
60000
40000
20000
0
0
50000
100000
150000

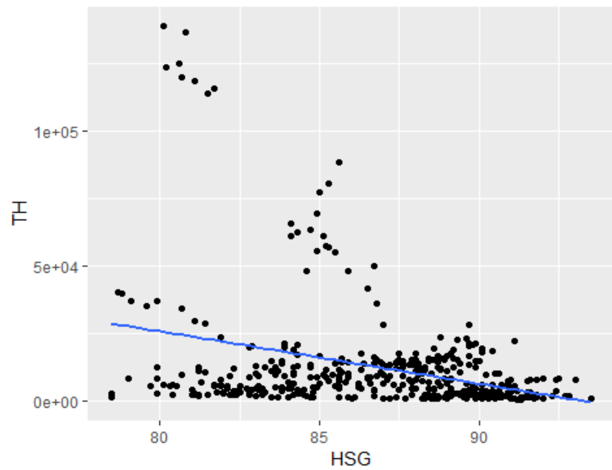
Population:

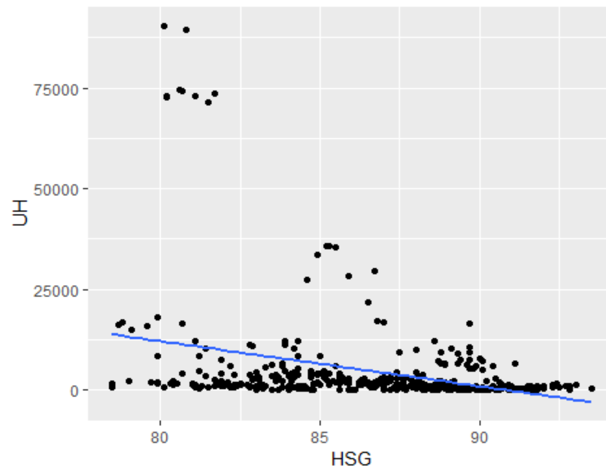

TH
Pop
1e+05
5e+04
0e+00
0
10000
20000
30000
40000

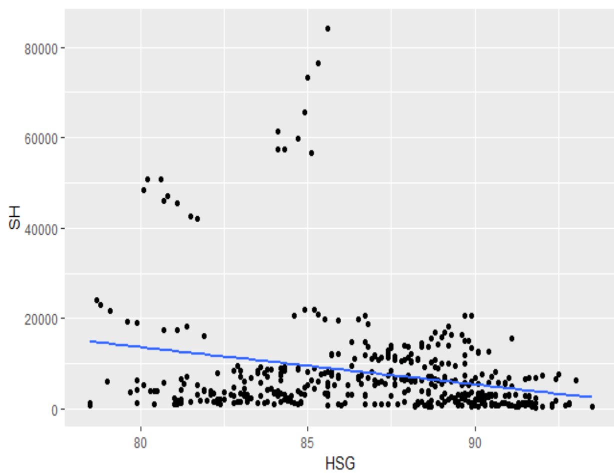

UH
Pop
75000
50000
25000
0
0
10000
20000
30000
40000


SH
Pop
80000
60000
40000
20000
0
0
10000
20000
30000
40000

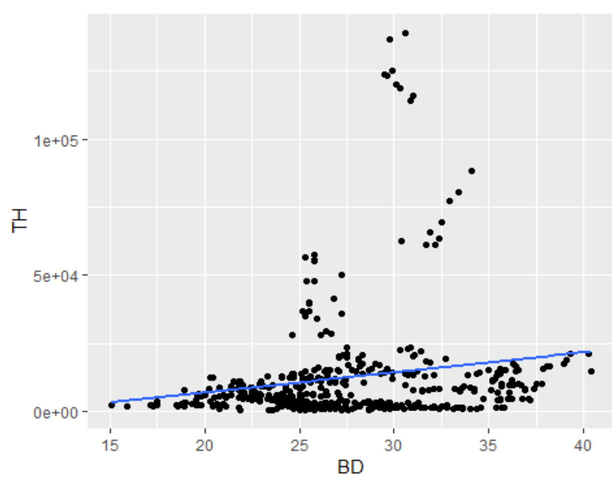
High School Graduate Rate:

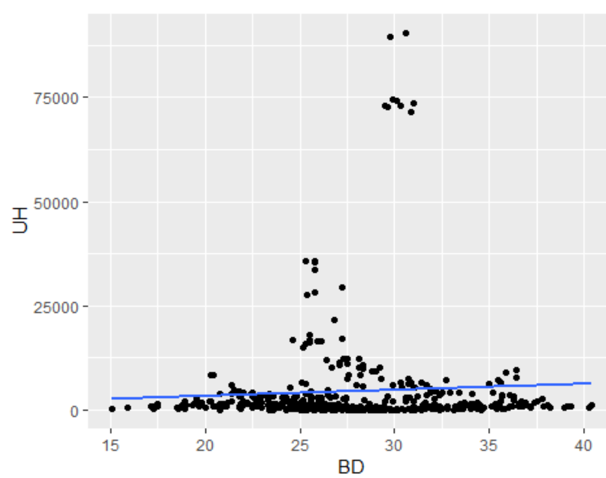


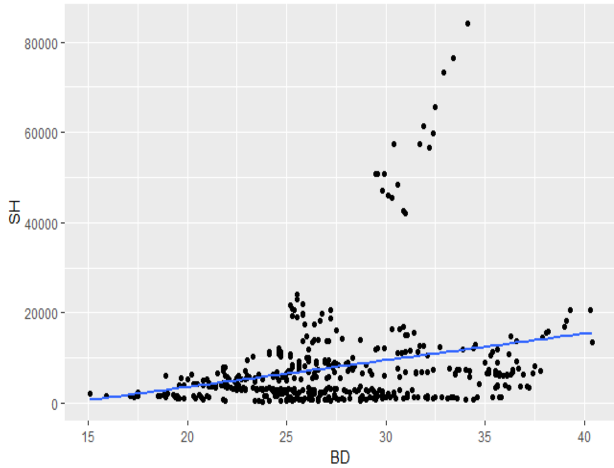




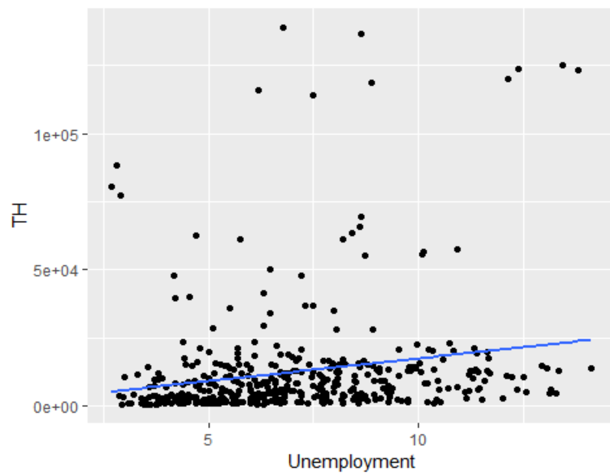
Bachelor Degree Rate:

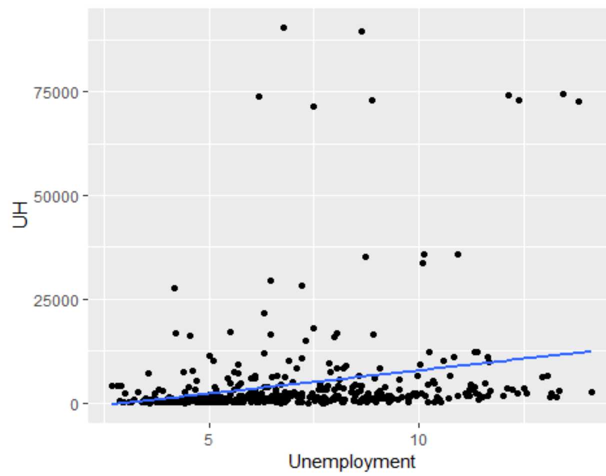


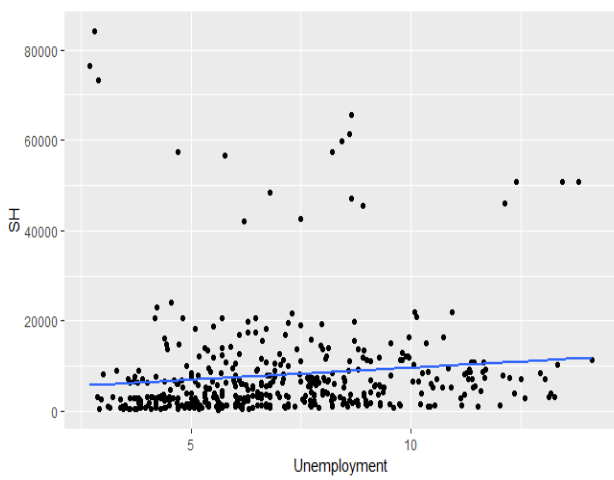




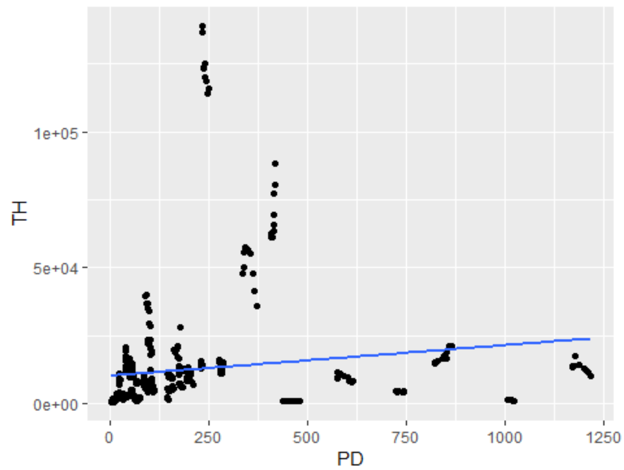
Unemployment:

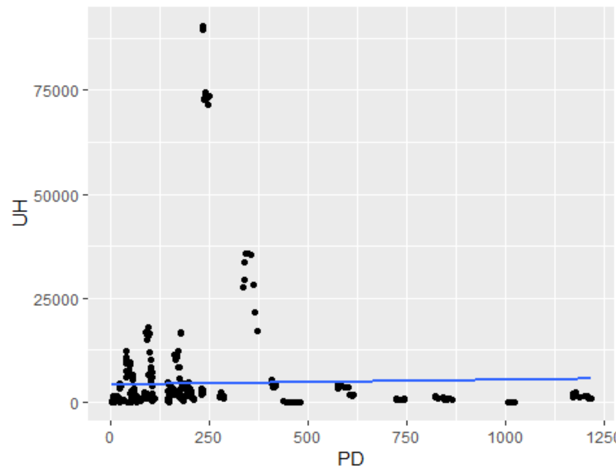


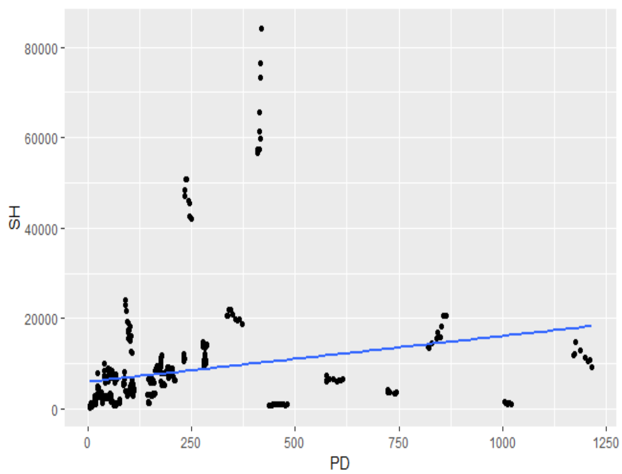




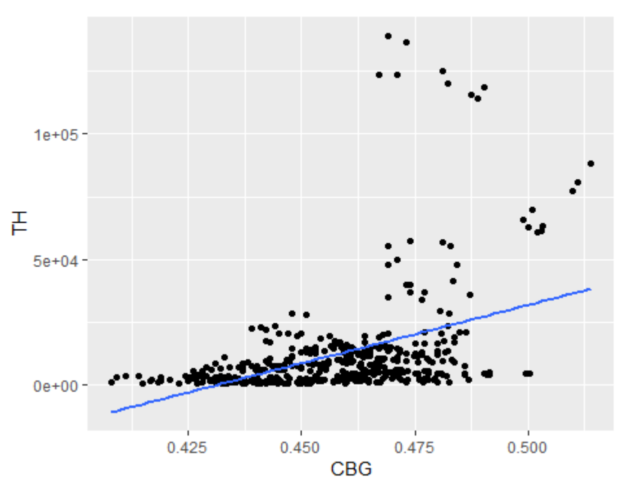
Population Density:

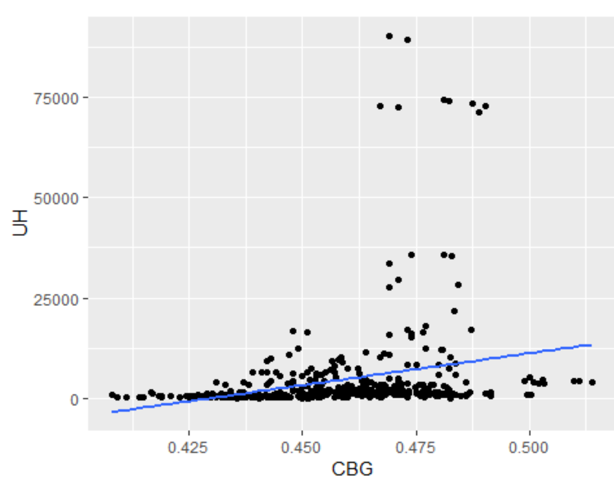


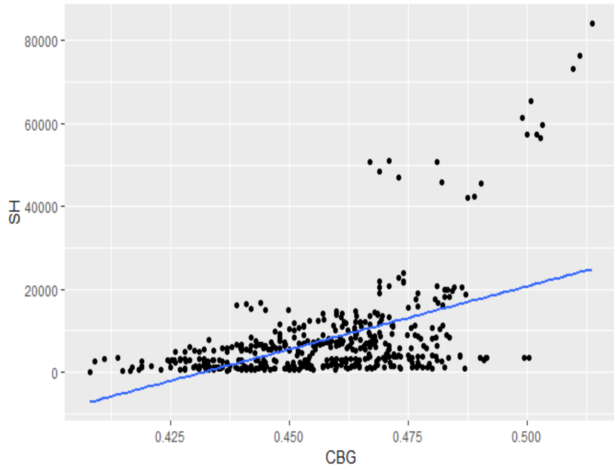


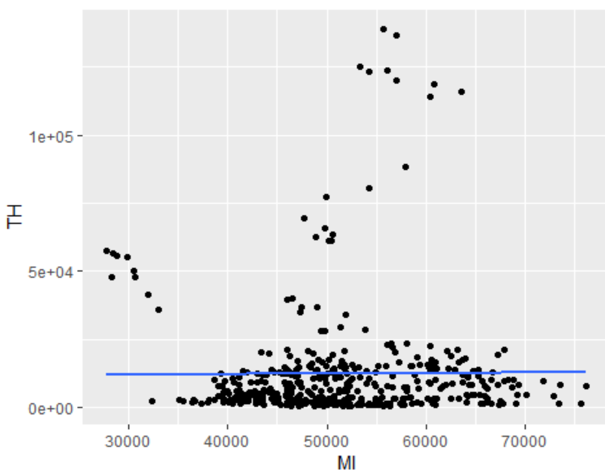


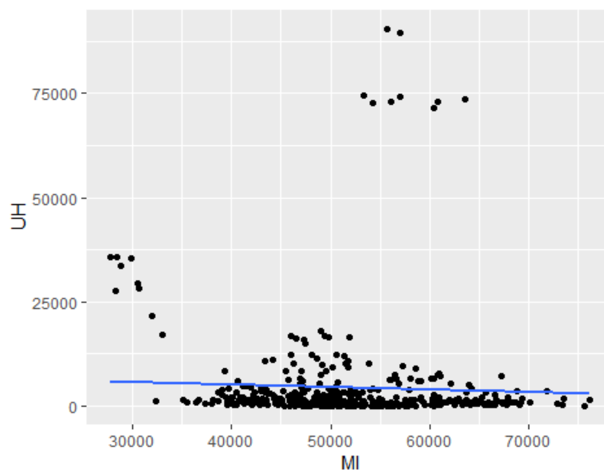
Census Bureau Gini for Income Inequality:

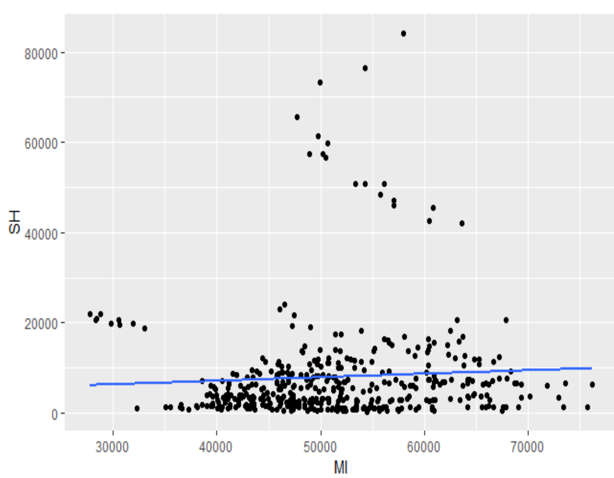




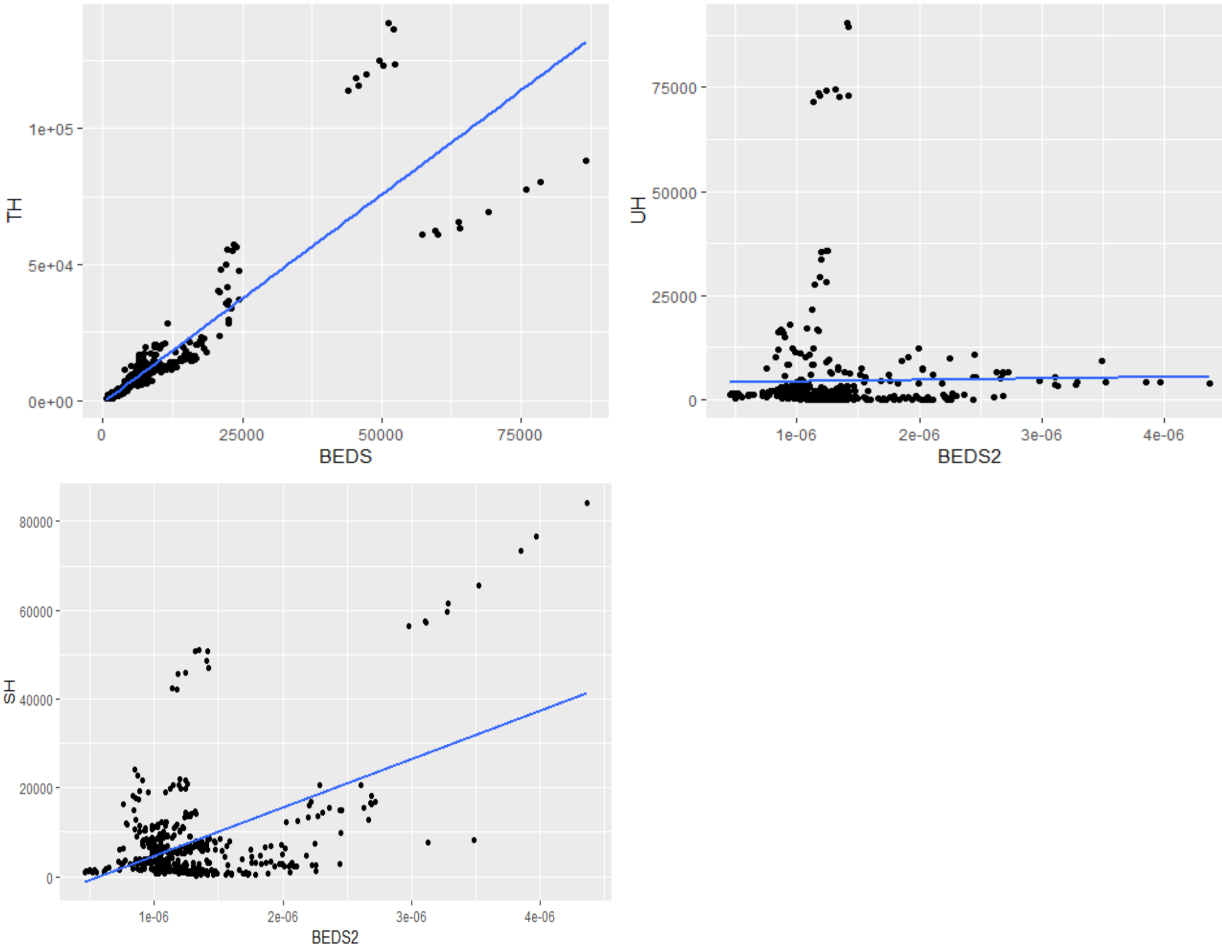

80000
60000
40000
20000
0
SH
0.425
0.450
0.475
0.500
CBG


1e+05
5e+04
0e+00
TH
30000
40000
50000
60000
70000
MI


75000
50000
25000
0
UH
30000
40000
50000
60000
70000
MI


80000
60000
40000
20000
0
SH
30000
40000
50000
60000
70000
MI

Bed Rates:


TH
1e+05
5e+04
0e+00
0
25000
50000
75000
BEDS
UH
75000
50000
25000
0
1e-06
2e-06
3e-06
4e-06
BEDS2
SH
80000
60000
40000
20000
0
1e-06
2e-06
3e-06
4e-06
BEDS2