Homeownership & Poverty in the United States

By Jace Henderson, Braden Swenson, and Jack Yelden

Introduction

- We knew that we wanted to investigate the rising rates of poverty in America
- The main point that we wanted to investigate was the impact that Home Ownership had on poverty
- We felt that providing easier ways to obtain housing would lead to lower poverty rates across the country

Introduction (contd.)

- The main statistic that we used was correlation
- We looked for a strong negative correlation
- We investigated data across 5 different races and a total value
- We selected data over the years 2005-2017

Data Preparation

- We were able to find 2 data sets that allowed for us to run our programs
    - Diversitydatakids.org
- Poverty & Homeownership by state yielded too few data points for us to make any meaningful conclusions
- These same rates sorted by county gave us over 7,000 data points once we cut out the empty areas

Data Preparation

geoid	name	year	total_est	aian_est	api_est	asian_est	black_est	hisp_est	nhisp_est	nhopi_est
05000US01001	Autauga County, Alabama	2005-2009	10.306	7.253	0	0	27.565	12.747	10.255	None

api_est	numeric	Estimate; Asian or Pacific Islander Alone

Data Preparation

- Once we exported both of the datasets as csv files we were able to combine them into one large, master database that allowed us to run our calculations more smoothly
- From here, we were able to clean it to get more accurate conclusions

Data Cleaning & Pre-Processing:

- Data cleaning was short
- Main concern was to remove blank data cells
- Used Excel's filter options to remove all the blank cells
    - This also helped our csv file become smaller so Python could run the necessary calculations
- Other main issue was deciding how many races to use
    - Started with 17 different columns, shortened to six
- This left a much more easily digestible data set

Data Analysis

- To compute the necessary correlation coefficient (r-value) we first converted the downloaded Excel file into a CSV file
- We then designed a Python Program Called Correlation Calculator that reads the CSV file and returns an r-value
- It would gather all the values of the Homeownership column, treat them as X values, and compare them to the corresponding Poverty rate (which represented Y values)
    - We made use of a function that could repeat the results whenever a new pair of columns were input meaning this program could easily be retooled to seek out correlations with other groups
- Figure 1 is the results

Figure 1

Correlation Coefficients	
Group	r-value
Total	-0.4244
African American	-0.2250
Hispanic	-0.2719
Asian and Pacific Islander	-0.3105
Non-Hispanic white	-0.3509
White	-0.3507

Data Analysis (cont)

- In order to show a strong negative correlation values should be less than -.75 but an r-value of -.5 would be significant
- As Figure 1 showed, while each of our variables had negative correlations, none of them reached a threshold of significance
- Our best correlation was in the “totals” column at -.4244 while the group that had the lowest correlation was African Americans at -.225

Data visualization

- We created scatter plots with Homeownership rates on the X axis and Poverty rates on the Y axis
- Figure 3 indicates that utilizing all the data points would render our models as one big blob
- Because creating scatterplots with a little under 8000 data points would be difficult to read, we instead decided to randomly select 800 rows of data and use Excel's built-in visualization tools to generate the scatter plots

Example of all data points

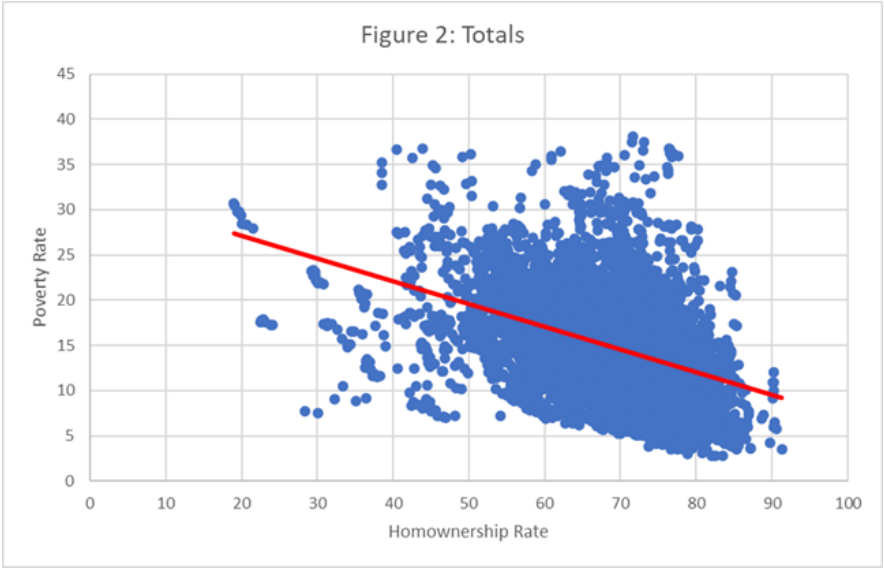
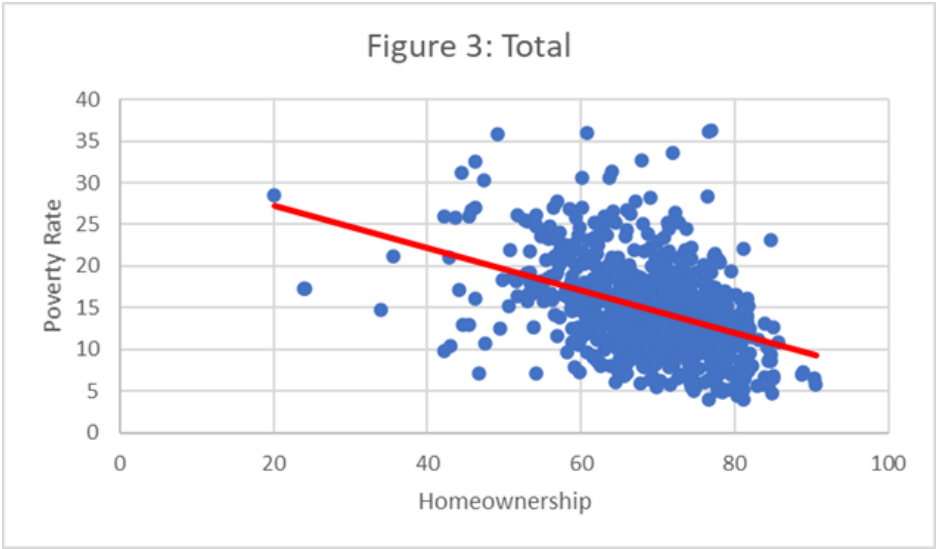
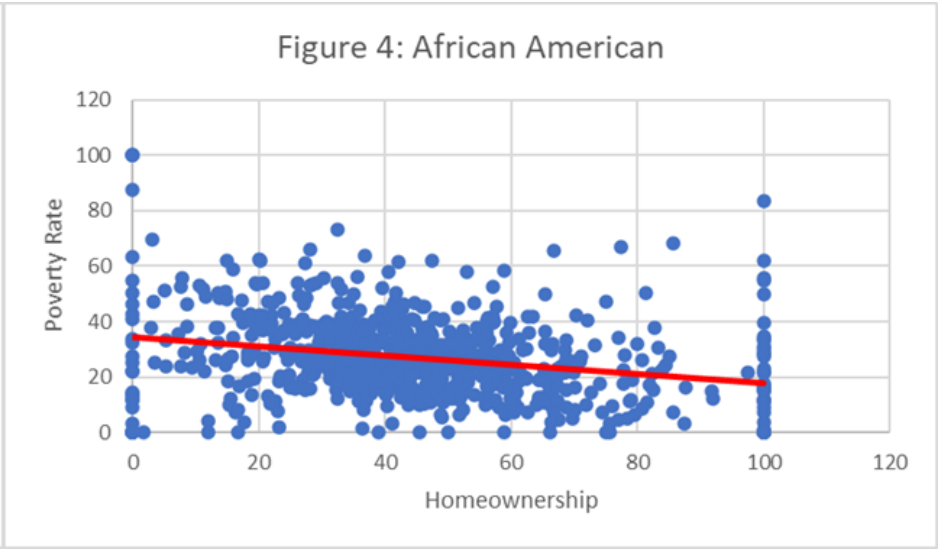


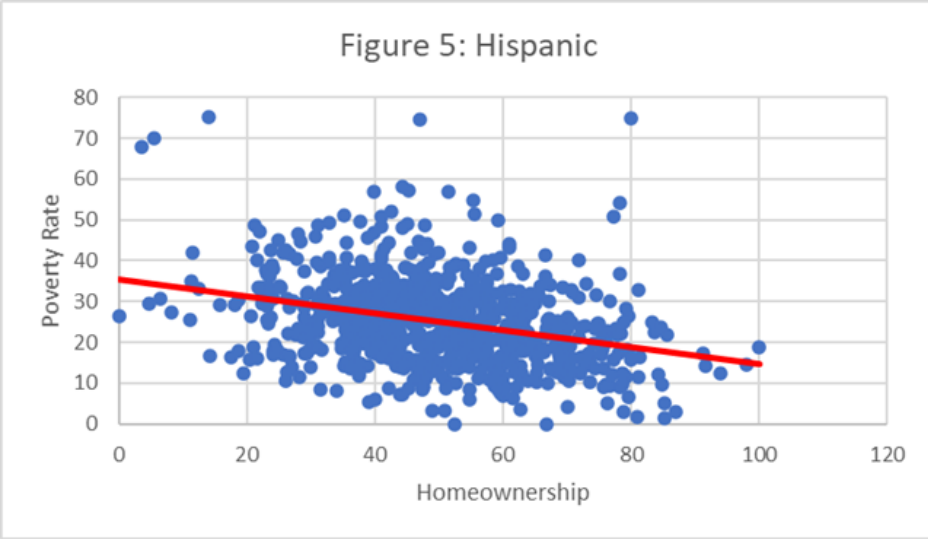
Figure 2 showcases how all the data points are laid out on a scatter plot. Each blue dot represents a data point from a specific county for a specific period. The x-axis (the independent variable) is the homeownership rate while the y axis (dependent variable) is the poverty rate. Unfortunately the mass of points makes it difficult to read



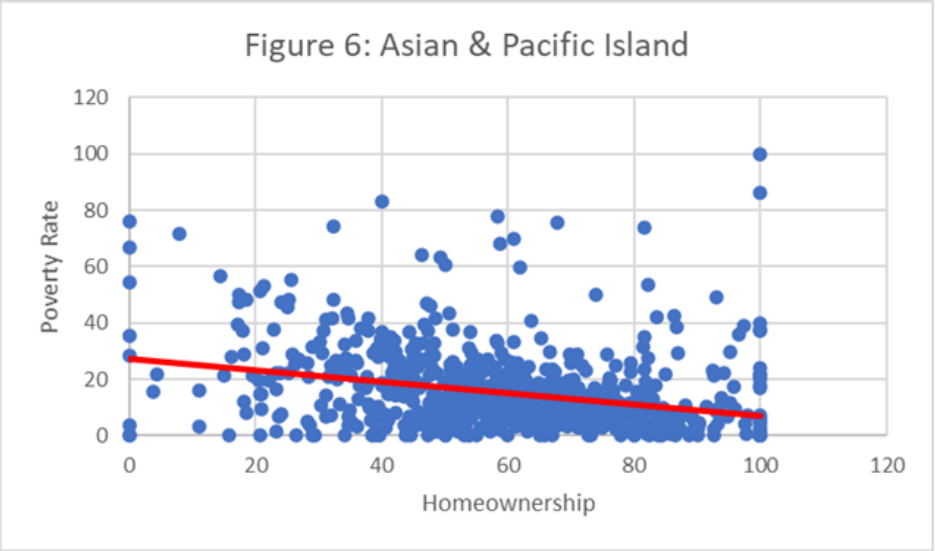
Sample size of 800 for the “Total” columns



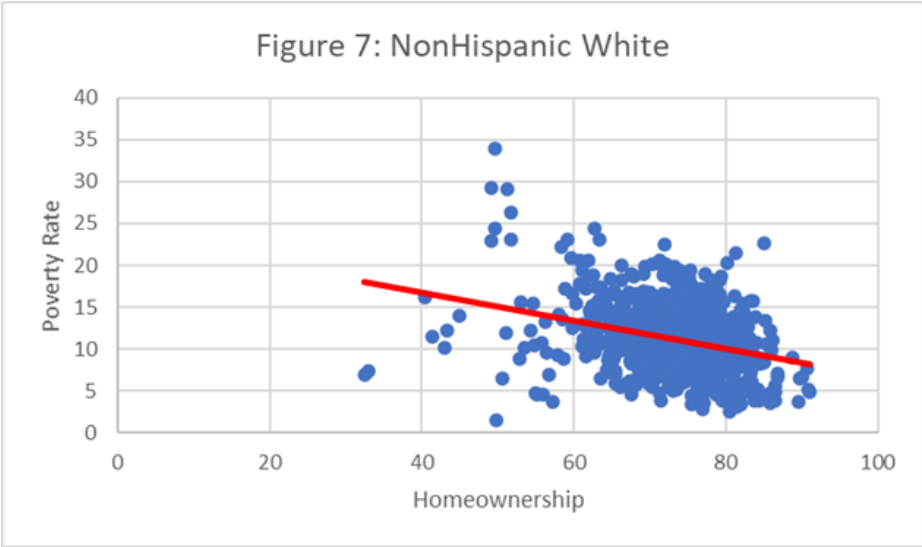
Sample for the “African American” columns



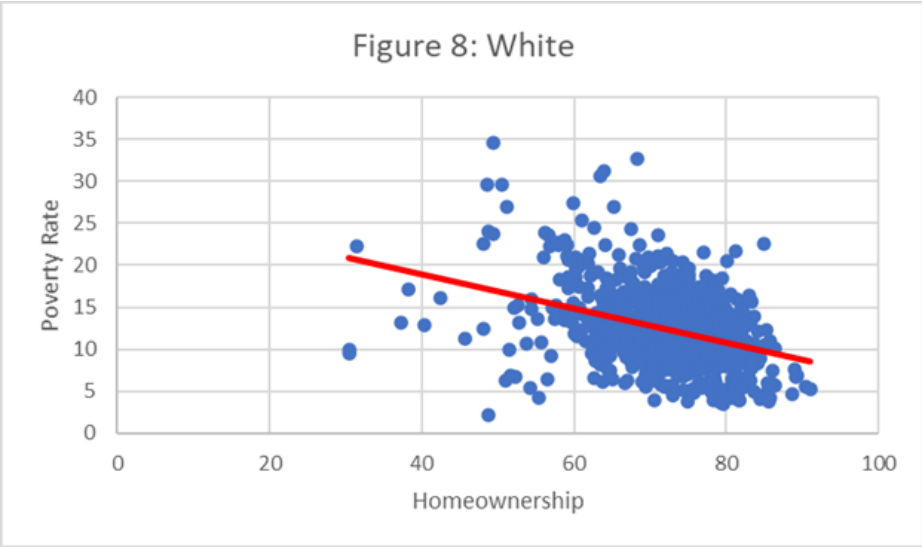
Sample for the “Hispanic” columns



Sample for the “Asian & Pacific Island” columns



Sample for the “NonHispanic White” columns



Sample for the “White” columns

Data Visualization (Cont)

- The scatter plots show a weak linear relationship between our two variables
- Very few data points touch the trendline, and outliers are common
- All trend lines are negative, showing that while the correlation coefficient might not be strong, it is still negative

Conclusion

- Poverty is a huge issue in the United States
- We believed there may be correlation between homeownership and poverty
- We conducted research and implemented programs to test our hypothesis
- We found negative correlation, but not enough to completely justify our claim
- Further research should be conducted as homeownership alone is not significant enough