Final Project

ECON 2250

Caroline Drury, Connor Prather, & Wendell Rogers

December 1, 2025

Introduction

Subject

The economic, educational, and social factors that predict income levels across the United States

Motivation

- Income inequality across U.S. states is large and persistent
- Several factors like cost of living, state tax policies, educational disparities, crime rate, demographics, etc. contribute to average income levels across states
- Our goal is to find which of these various factors contribute most strongly to average income.

Research Question

What economic, educational, and social factors best predict income levels across the United States?

Data Sets

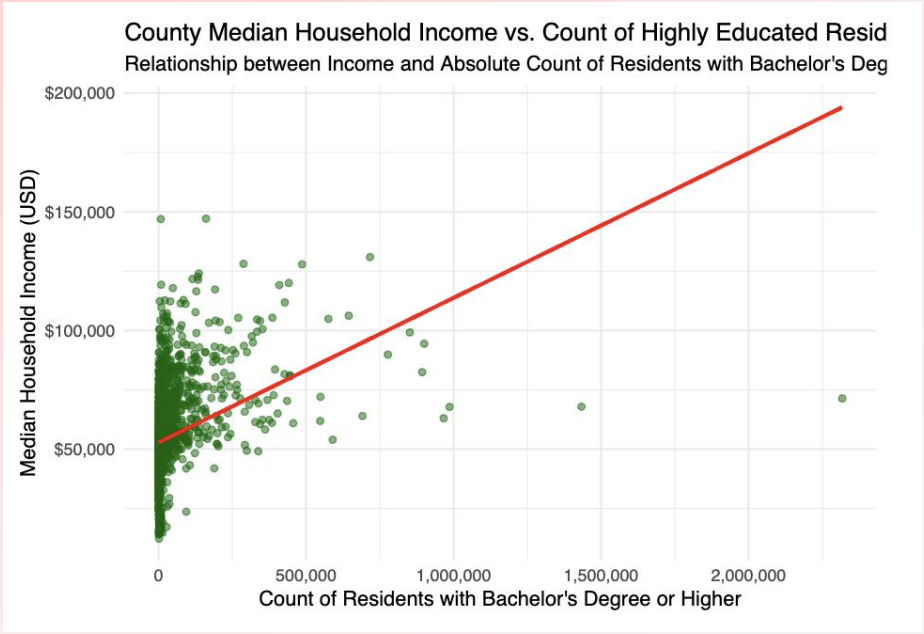
We pulled all our data from the US Census, filtering by county. The variables include...

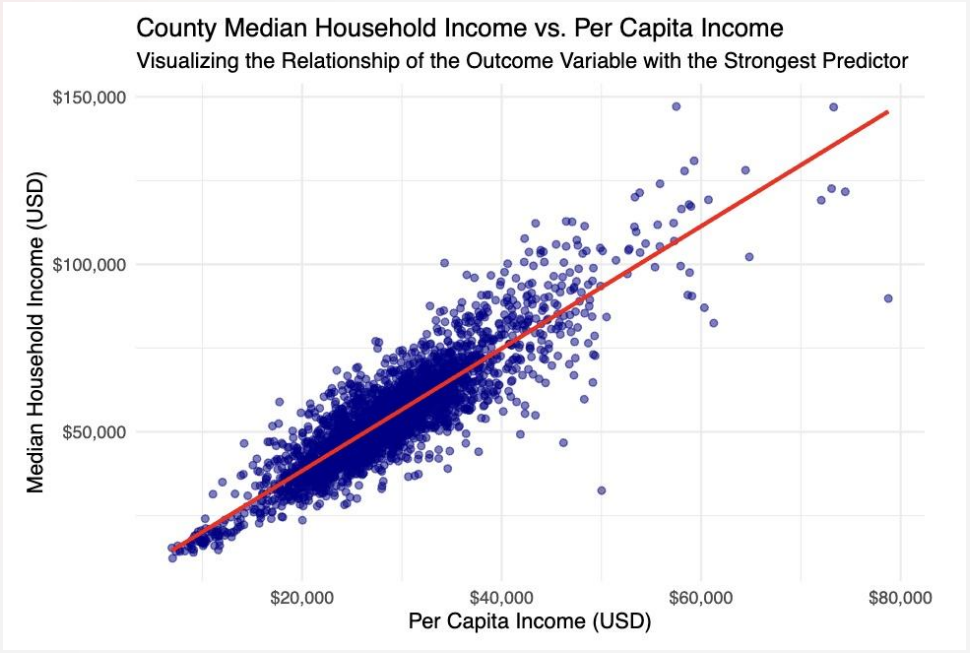
- **Median income:** This was our dependent variable, other variables were measured by their effect on this
- **Education level:** High School, Bachelors, Advanced degree etc.
- **Race**
- **Median Age**
- **Income per Capita**
- **County GDP**

How Our Model Works

- The model collected county-level US census data on demographic and economic factors
- Each variable was merged into one combined dataset
- The code then ran a linear regression that analyzed the factors we chose to see which had the strongest effects on income and how they could explain median household income

Descriptive Statistics

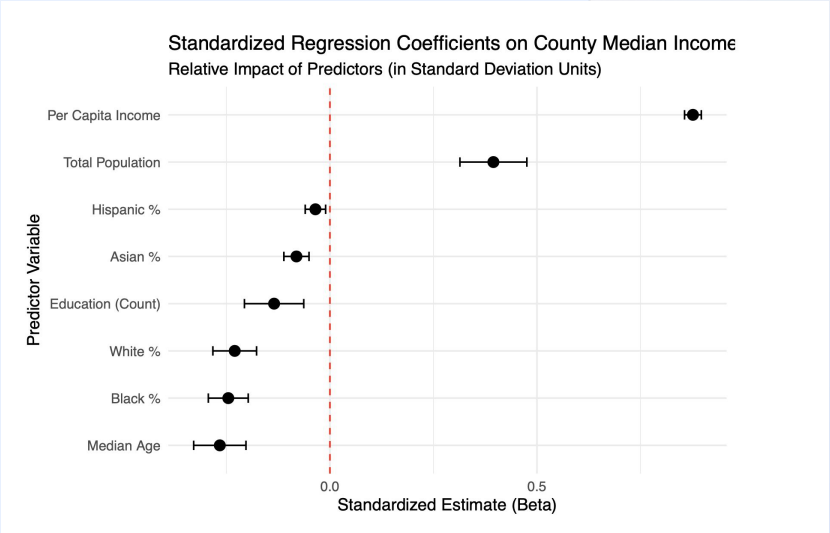




Results

Based on the standardized plot's visual arrangement:

- Per Capita Income has the most positive outcome. Essentially, this means that for each standard deviation increase in a county's PCI, median household income increases by the greatest amount (when in standard deviation terms) compared to all other variables.
- Racial Percentages have the most negative outcome. After accounting for all other factors, race has the strongest negative association with average household income.




```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65104  -3876   -277   3609  36021
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.205e+04  2.576e+03   8.561  < 2e-16 ***
## edu_bachelors_plus     -2.491e-02  6.748e-03  -3.691 0.000227 ***
## dp05_0038p             -2.621e+02  2.624e+01  -9.991  < 2e-16 ***
## dp05_0071p             -2.794e+01  1.003e+01  -2.787 0.005347 **
## dp05_0066p             -1.506e+02  2.889e+01  -5.215 1.95e-07 ***
## dp05_0037p             -2.052e+02  2.402e+01  -8.543  < 2e-16 ***
## median_age             -6.364e-01  7.711e-02  -8.254  < 2e-16 ***
## bea_personal_income     1.813e+00  2.139e-02  84.774  < 2e-16 ***
## gdp_county              1.862e-02  1.944e-03   9.580  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7004 on 3211 degrees of freedom
```

5

```
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.7958, Adjusted R-squared:  0.7953
## F-statistic:  1565 on 8 and 3211 DF,  p-value: < 2.2e-16
```

Black: dp05_0038p, **Hispanic:** dp05_0071p,
Asian: dp05_0066p, **White:** dp05_0037p

Next Steps

- Doing the project was a lot of fun and got us thinking of some other things we could do to expand on our research panel
- One test that would be interesting to investigate could be to multicollinearity, how the variables are related to each other. Investigating this can help determine whether our model could be improved by diversifying the variables we have
- Another next step could definitely be expanding the model. As it is the model has a good framework, with a few selected variables. It is set up to allow for easy additions such as geographical location and socio-economic background
- It could also be useful to run some additional tests to help reinforce the validity of the model. Some tests we could run are the

Thank you!

Any Questions?