

Exploratory Data Analysis Lab- General Social Survey (GSS)

The variables I chose were wrkstat, marital, childs, educ, sex, race, born, parborn, granborn, earnrs, and income. These variables respectively measure: labor force status, respondent's self-employment or works for somebody, marital status, number of children, highest year of school completed, respondent's sex, race of respondent, was respondent born in this country, were respondent's parents born in this country, how many of respondent's grandparents born outside U.S., how many in family earned money, and total family income.

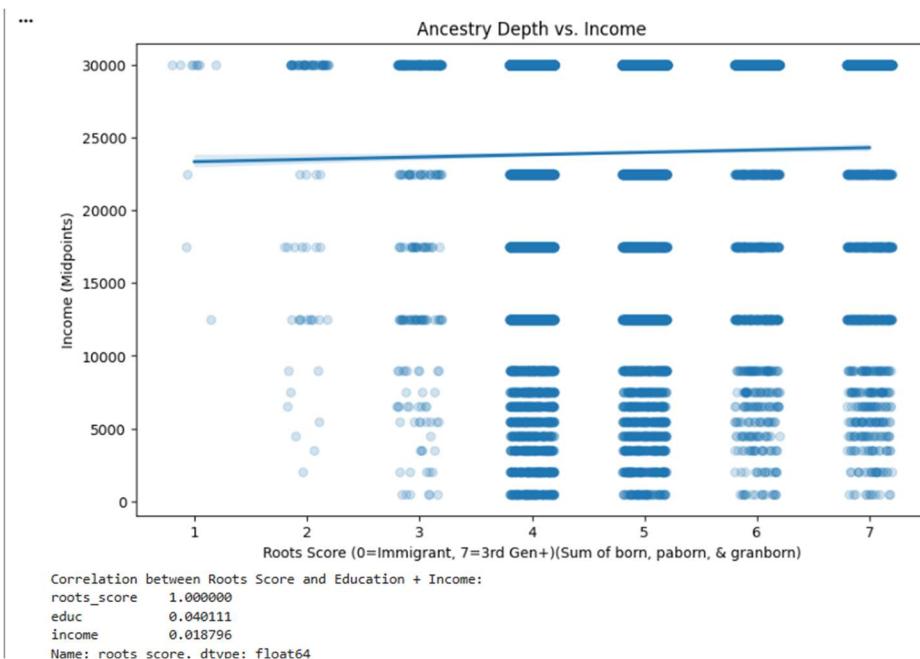
I decided on this mix of data to see how much of family income is driven by family structures and immigration status. By including variables like educ, wrkstat, and income, I can see the "standard" picture of someone's financial life. But the real reason I added born, parborn, and granborn is to look under the hood at the immigrant experience. I want to see if being a first, second, or third-generation American changes the trajectory of someone's education or how much their family earns. Adding in sex, race, and marital status helps me keep the analysis realistic; it lets me control for the different social pressures or hurdles people face that have nothing to do with their resume.

Three relationships I am interested in studying specifically, are:

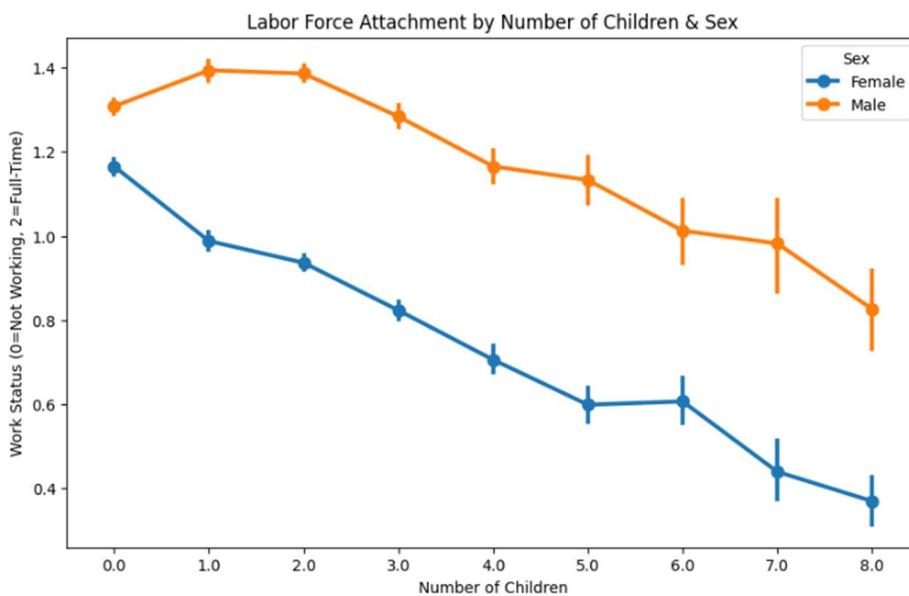
- 1) Does having more "roots" in the country (meaning your parents and grandparents were born here) actually lead to higher educ levels and income?
- 2) How much do marital status and the number of childs in a house impact a woman's wrkstat (labor force status) compared to a man's, and how does that affect the total family income?

- 3) If we look at people with the exact same educ level, how much do race and sex still influence their income and whether they are self-employed or working for someone else?

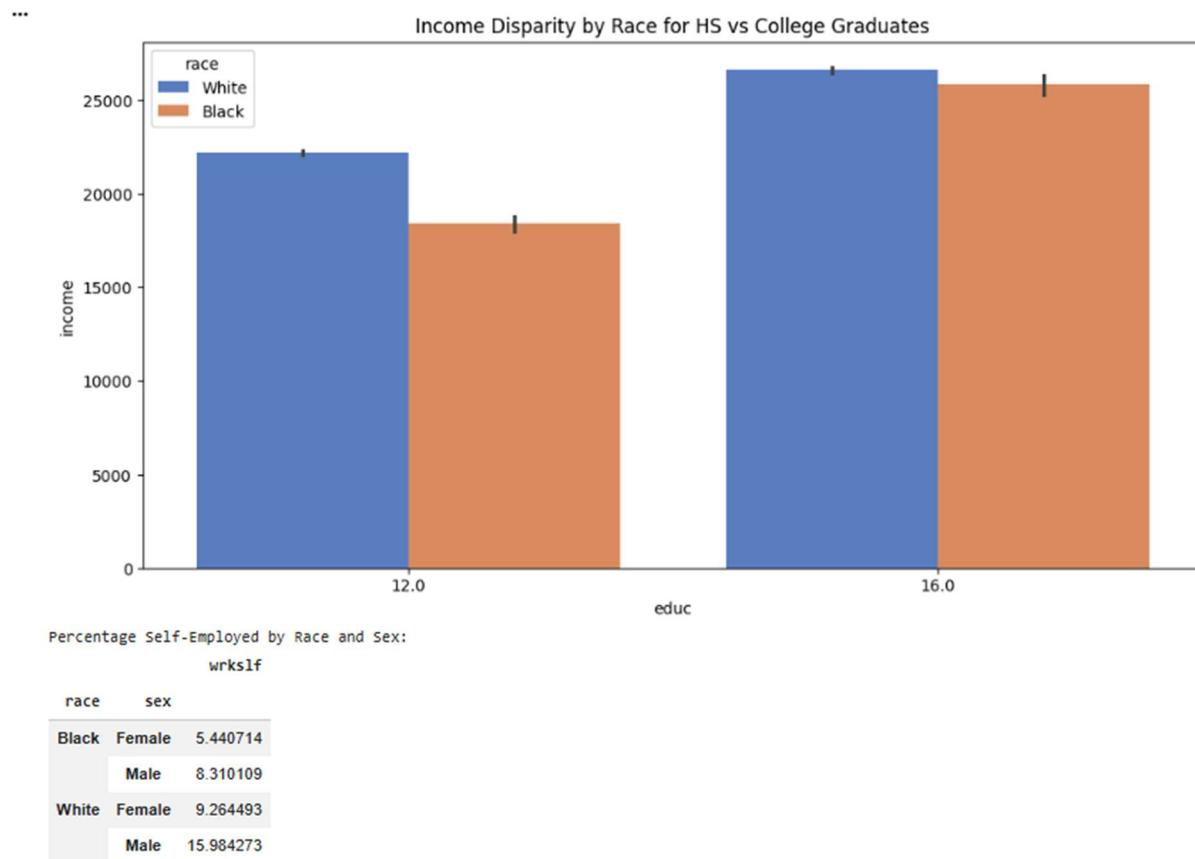
Question 1



Question 2



Question 3:



FINDINGS:

In investigating question 1, I found that there is little correlation between how long a person's family has been in the US (calculated by the sum of how many in the previous 2 generations and themselves were born in the US, meaning self + parents + grandparents), and their income and their educational attainment. There was a slight positive correlation, but it is mostly noise and should not be overinterpreted. I believe this is slightly biased, as the highest income bracket in this dataset is capped at \$25,000+. While that might have been a solid middle-class or even high-end salary back when this survey was first designed, in today's world, it doesn't tell us much. Since everyone from a school teacher to a CEO would be lumped into the

same "top" category, the data loses its ability to show the real differences in wealth. This bias likely smoothed over the exact correlations I was looking for, making the generational progress look flatter than it probably is. However, education is not subject to this bias, so it can serve as a decent proxy for income.

For question 2, I examined the number of children that affect a respondent's work status and total family income, and whether this varies by sex. As with my findings in question 1, the results for family income are likely to be heavily skewed or to show less variation than expected. Because the highest income bracket is capped at \$25k+, it is difficult to see the full impact of having more children or being married on a household's financial success. For example, a family with several children and a single earner might still hit that \$25k+ threshold, making them look identical in the data to a dual-income household with no children. This bias in the income variable means that any correlation found is probably underrepresented.

To get a more accurate result, I am focusing more on wrkstat to see whether there is a more noticeable difference in whether people are working full-time or part-time by gender and family size, since that variable doesn't have the same "ceiling" problem as the income data. In this, I found that men work more across the board, but labor force attachment decreases as the number of children increases. My best guess is that, as people get richer and older, they have more children.

Looking at my third question, I wanted to see the impact of race and education on earnings. As seen in the bar chart, white respondents earned more with 12 and 16 years of education (signaling graduating from high school or college). However, the gap is much smaller among college graduates than among high school graduates. This means that there is a much

smaller wage gap between Whites and Blacks when looking at college graduates as opposed to high-school graduates.

Another interesting thing I found was the rate of self-employment. In both Blacks and Whites, men had ~50% higher rate of self-employment, but Whites across both sexes had ~100% higher likelihood of working for themselves than Blacks.

While these models have evident flaws, such as the income cap, they still offer interesting insights into how a person's background relates to their current standing. They allow us to see how different life factors overlap, even if the data doesn't capture the full financial picture of today's world.