Project Title:
Exploring the Correlation Between Race, Ethnicity, and Income Disparities in the United States

Authors:
Alex Alvarado, Sabrina Lee, Muad Abdi

Author Contributions:
- A.A, S.L., and M.A Designed the research
- M.A found or collected data
- A.A, S.L., and M.A. analyzed the data
- A.A. Made the figures
- S.L. and M.A. Wrote the report
- A.A. S.L., and M.A made the presentation

INTRODUCTION

Research Question:
How does race and ethnicity correlate with income disparities in the United States in the 21st century, and what are the primary socio-economic factors influencing this relationship?

Research Motivation and Summar:
Income inequality is one of the most pressing social and economic issues in the United States today. Studies have repeatedly demonstrated disparities in income levels among different racial and ethnic groups, yet the causes of these disparities remain a subject of debate. Structural and systemic factors, such as access to education, employment opportunities, discrimination, and historical economic disadvantages, may play a crucial role. This study aims to analyze income data to understand the correlation between race, ethnicity, and income disparities in the 21st century, identifying key socio-economic factors that influence these disparities.

METHODS

Data:
A dataset consisting of 17,000 individuals with demographic and income information, including White, Black, Hispanic, and Other races, and the other one.

Variable Operationalization:

This dataset is derived from the 1994 U.S. National Health Interview Survey and includes 17,870 observations. The variables below are operationalized as follows:

Earnings: Continuous variable representing annual labor income, adjusted to 2012 dollars. These were originally reported in income brackets and converted to specific dollar values using estimates from the Current Population Survey.

Race/Ethnicity: Categorical variable with four groups:

1 = Non-Hispanic White

2 = Non-Hispanic Black

3 = Hispanic

4 = Other

Data analysis procedure:
- Descriptive statistics to summarize income distribution among different racial and ethnic groups.
- Regression analysis to determine the impact of race and ethnicity on income levels, controlling for education, occupation, and other socio-economic factors.
- Clustering techniques to identify patterns within income groups.
- Data visualization using Matplotlib for histograms, scatter plots, and regression analysis.
- Statistical tools within Matplotlib to plot and analyze relationships.

RESULTS

Main Results:

This analysis focused on the relationship between race and earnings, revealing significant and persistent income disparities among racial groups. The data shows that White individuals (Group 1) earn the highest average income at $49,422.48. In contrast, Black individuals (Group 2) and Latinos (Group 3) earn substantially less on average—$36,162.11 and $37,529.10, respectively, despite similar median earnings. Individuals from other racial backgrounds (Group 4) have a slightly higher average income ($45,880.04) than Black and Latino individuals, yet still fall below White individuals.

These income gaps highlight the enduring impact of structural inequalities and systemic barriers that disproportionately affect minority communities. Even when accounting for the same earnings cap, these disparities suggest that equal opportunities remain out of reach for many individuals based solely on their racial or ethnic background. Future research should investigate the root causes of these differences, including discrimination in hiring and promotion practices, unequal access to education, and regional labor market inequalities. Addressing these issues is essential for advancing equity and closing the racial income gap in the U.S. economy.

Figure:

Histogram of Earnings by Race/Ethnicity
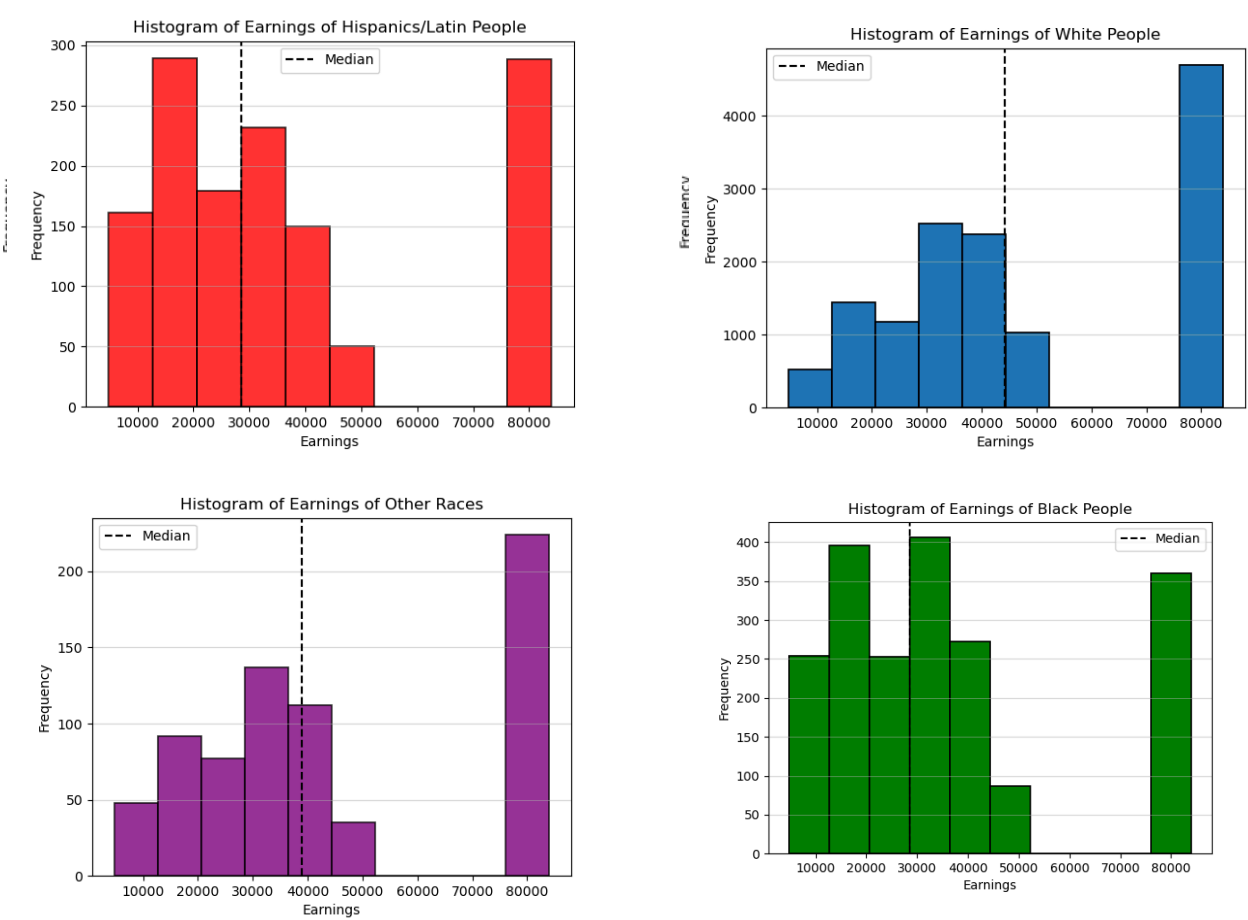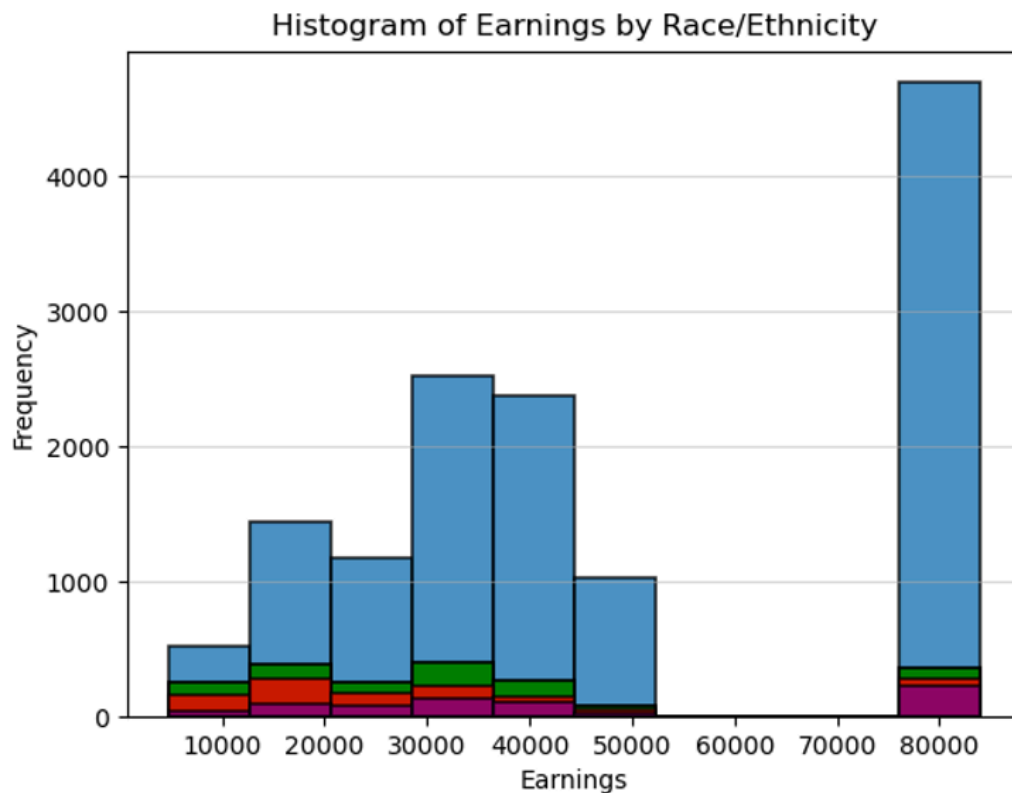
DISCUSSION

Limitations:

This research aimed to explore the relationship between race and earnings, but several limitations may have prevented it from fully addressing the research question. First, potential sampling bias exists since the dataset lacks information on how participants were selected, raising concerns about whether the sample accurately represents the broader population across different racial groups, as indicated by the varying numbers of participants from certain races. Measurement issues may also arise, particularly if race and earnings were self-reported or recorded inconsistently, leading to inaccuracies in representing earnings associated with specific racial groups. Additionally, the operationalization of earnings does not account for important distinctions such as hourly versus annual income or the number of hours worked, which could affect the results. The analysis overlooks significant unmeasured variables like personality traits, industry norms, and socioeconomic background, all of which could majorly influence earnings, introducing omitted variable bias. It also fails to account for individuals who have recently moved to the U.S. or hold specific immigration statuses while excluding certain racial groups, including Asians, Native Americans, and biracial individuals. Moreover, it neglects factors such

as heritage and regional residence. Lastly, without transparency in modeling decisions or consideration of alternative specifications, the analysis may suffer from researcher degrees of freedom, undermining the robustness and generalizability of its conclusion.

Future Directions:

To improve upon the findings of this study, several improvements could be made to the research design and the collection of data. The use of longitudinal data would provide a more accurate picture of racial and income disparities over time. Longitudinal data also enables the study of income mobility, life-cycle earnings, and the cumulative effects of discrimination or opportunity gaps. Future studies could also aim to be more precise with racial and ethnic categorization. The current broad categories do not fully capture the diversity and complexity within racial and ethnic groups.

Additional variables such as education level, industry of employment, and detailed geographic location could also be included in future research. These factors play an important role in shaping income and can help identify the structural and systemic forces contributing to inequality. Using more detailed data would allow researchers to explore how these factors interact with race and ethnicity. Applying machine learning methods like regression models or clustering could also help uncover deeper patterns in the data that may not be visible through traditional statistical analysis.

Future research could also benefit from adding qualitative insights. Surveys, interviews, or focus groups could provide context to the numbers and help explain the lived experiences behind income disparities. Combining these qualitative perspectives with quantitative analysis would give a more well-rounded understanding of systemic inequality. Studies could also look into the effects of public policy, like minimum wage laws, affirmative action, or tax credits, on income inequality across different racial and ethnic groups.

REFERENCE