

Exploring Socioeconomic Patterns in the 2021 Household Census in England

Oluwadamilola Oreofe Ajakaiye | Student ID: 2439749 | MSc Data Science | 7CS034 Data Science

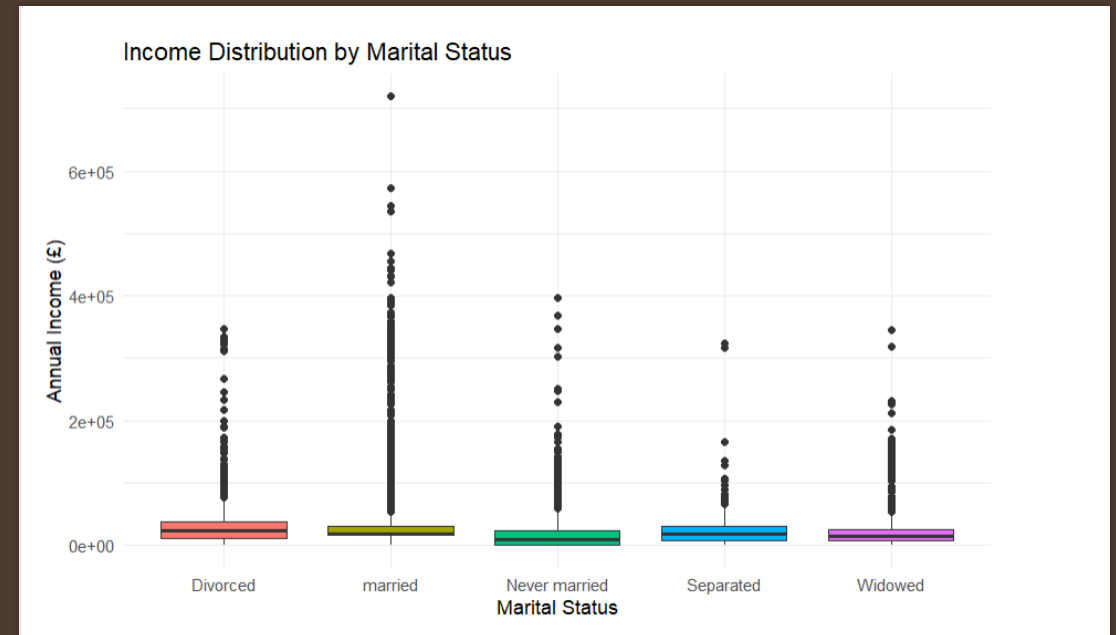
Objective

- This report explores the relationships between demographic factors (age, marital status, ethnicity and gender) and socioeconomic variables (income and education level) within a 2021 household England's census dataset.

Demographic Factors	Socioeconomic Variables
Age, Marital Status, Ethnicity and Gender	Income, Highest study level.

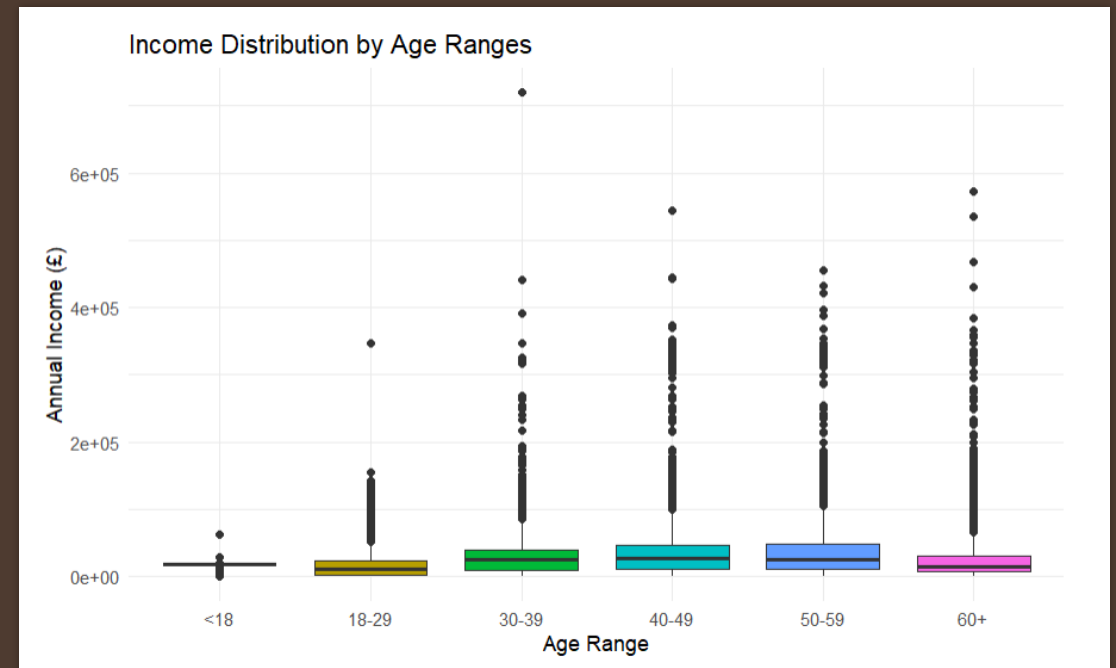
Income by Marital Status

- Married individuals shows a higher median income compared to those with other marital statuses, suggesting that marital status may influence financial stability.



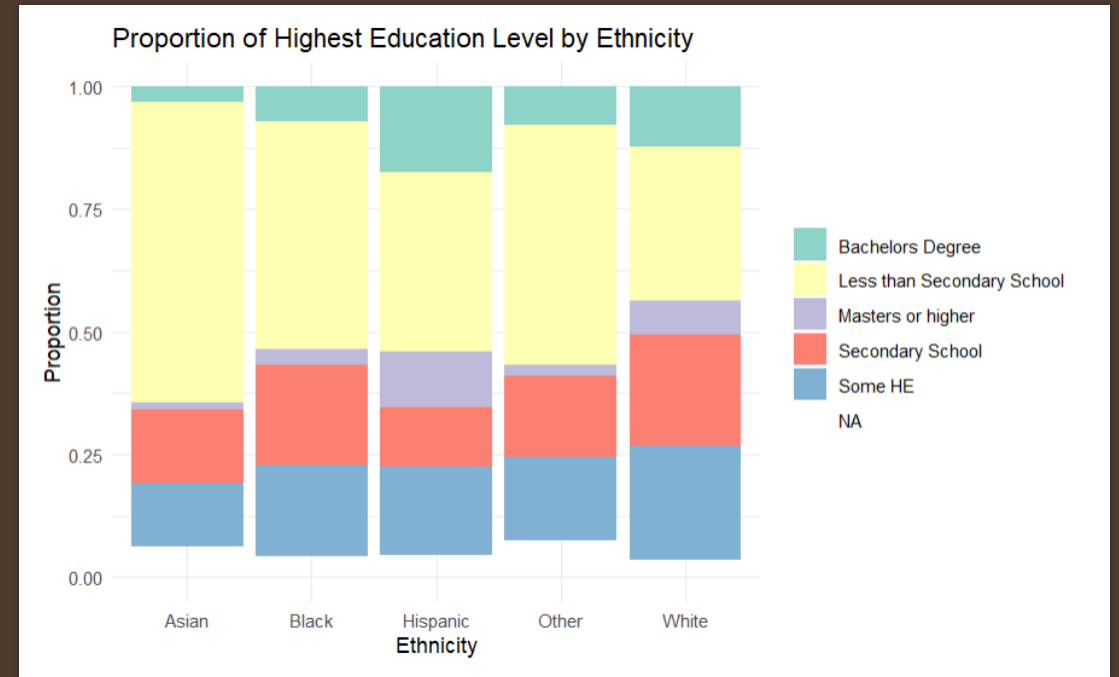
Income by Age Ranges

- The age range of 30 to 39 appears to have the highest average income. However, individuals aged 60 and above tend to have more income earners within that group. In contrast, the 18 and over age group has the fewest income earners and records the lowest income levels.



Highest Education Level by Ethnicity

- The plot indicates that individuals of Hispanic descent have the highest proportion of higher education attainment



Dataset Limitations

1. **Sample Size:** While the dataset provides a broad overview of various demographic and socioeconomic factors, the sample size may not fully represent the entire population. Specific demographic groups, such as ethnic minorities, could be underrepresented, leading to potential biases in the analysis of income distribution, education attainment, or marital status. This limits the generalizability of the results across a wider population.
2. **Missing Data:** There may have been missing or incomplete data points, particularly in sensitive categories like income or education level. Missing values in these key variables can skew the analysis, especially if certain demographic groups are more likely to have incomplete data. Although cleaning techniques were applied, the absence of significant amounts of data could impact the accuracy of the findings.
3. **Additional Variables:** The inclusion of additional variables, such as regional or industry-specific income data, would greatly improve the depth of analysis. Income levels can vary significantly based on geographic location or the sector of employment, but these factors were not accounted for. Similarly, incorporating data on employment status or family size, would provide a more nuanced understanding of how socioeconomic factors interact with demographic traits in the population.

Thank you