Who Stays Single? Exploring the Factors Behind Non-Marriage in the United States*

Low Income and Education Levels Correlate with Higher Propensity to Remain Unmarried

Xizi Sun

November 28, 2024

This study examines the socio-demographic determinants of non-marriage in the United States using logistic regression analysis on individual-level data. The findings highlight that individuals with lower income and education levels, particularly below a bachelor's degree, are significantly more likely to remain unmarried. Additionally, gender and race also play critical roles, with men and certain racial groups, such as African Americans, showing higher unmarried rates. These results underscore structural inequities and socio-economic barriers shaping marital outcomes, offering insights into broader societal patterns and potential areas for intervention.

Table of contents

1 Introduction **2**

2 Data **3**

2.1 Data Tool . . . 3
2.2 Data Source And Measurement . . . 4
2.3 Variable Selection . . . 4
2.3.1 Outcome variables . . . 5
2.3.2 Predictor variables . . . 5
2.4 Data Selection . . . 6

*Code and data are available at: https://github.com/LilianS77/US_Marriage.

3 Model 6
3.1 Model Setup 6
3.1.1 Objective 6
3.1.2 Priors Explanation 8
3.2 Model justification 8
3.3 Model Validation 9
3.4 Results 9

4 Discussion 14
4.1 Determinants of Non-Marriage and Societal Implications 14
4.2 Insights into Education, Income, and Marriage Patterns 14
4.3 Gender and Race in Marital Trends 15
4.4 Weaknesses and Next Steps 15
4.5 Conclusion 16

A Appendix 17
A.1 How to Extract 2022 ACS Data from IPUMS 17
A.2 Model Assumptions 17

References 18

1 Introduction

Marriage is often regarded as a cornerstone of social and economic stability, shaping the lives of individuals and the broader dynamics of society. However, in recent decades, the United States has experienced a notable shift in marital patterns, with an increasing proportion of the population remaining unmarried. This trend raises essential questions about the underlying factors contributing to the decision not to marry and the societal implications of such changes. Addressing these questions is vital for understanding broader socio-economic inequalities and cultural dynamics.

This paper examines the determinants of non-marriage by analyzing a rich dataset from IPUMS USA (IPUMS 2024), which offers harmonized census and survey data spanning demographic, social, and economic variables. Prior studies have largely focused on individual aspects of marital behavior, such as income or education, but few have comprehensively analyzed the intersection of multiple socio-demographic factors. This study bridges that gap by exploring the combined effects of socio-demographic factors. Specifically, the estimand of this study is the reason of an individual remaining unmarried, conditioned on their socio-demographic attributes, including education level, income, race, gender, and age. This focus allows for a nuanced understanding of how these factors interact to influence marital decisions.

Using logistic regression analysis, the results reveal compelling patterns: individuals with higher education levels are more likely to marry, and economic stability, as indicated by higher income, is positively associated with marital status. Moreover, significant disparities are observed across racial and gender groups. African Americans exhibit higher non-marriage rates, while males show slightly elevated probabilities of remaining unmarried compared to females. These findings highlight persistent structural inequalities and cultural nuances shaping marriage in contemporary America.

Understanding the factors influencing non-marriage is crucial for addressing broader societal issues, including economic inequality, cultural shifts, and demographic changes. Marriage is often linked to financial stability, access to social networks, and overall well-being. The disparities uncovered in this study highlight systemic inequities that disproportionately affect certain groups, such as racial minorities and individuals with lower educational attainment. By identifying these disparities, policymakers and social institutions can better target interventions to reduce inequalities, promote inclusive opportunities for stable relationships, and adapt to the changing dynamics of family structures in modern society.

The remainder of this paper is organized as follows. Section 2 describes the dataset, variable selection, and measurement strategies used to examine marital status. Section 3 outlines the logistic regression approach, including the rationale behind model choice, prior specifications, and validation methods. Section 3.4 section presents key findings, supported by visualizations, and highlights the roles of education, income, race, gender, and age in predicting non-marriage. Finally, Section 4 explores the implications of these findings, acknowledges limitations, and proposes directions for future research.

2 Data

2.1 Data Tool

The dataset was analyzed using R (R Core Team 2023) and utilized several R packages for data manipulation, visualization, and analysis, including ggplot2 (Wickham 2016) for creating elegant graphics, dplyr (Wickham et al. 2023) for data wrangling, and here (Müller and Bryan 2023) for simplifying file management. The data was processed efficiently using Apache Arrow (Richardson, Dunnington, and Developers 2023), and visualizations were further enhanced with the scales package (Wickham, Seidel, et al. 2023). Bayesian modeling was implemented using rstanarm (Cepeda, Gabry, et al. 2023), while statistical results were tidied with broom (Robinson 2023) and broom.mixed (Bolker et al. 2023). Tables were constructed using kableExtra (Zhu 2023) for a polished presentation. Reproducibility was ensured with knitr (Xie 2014). Data was extracted from IPUMS USA (IPUMS 2024), and guidance on storytelling with data was drawn from Telling Stories with Data (Alexander 2023).

2.2 Data Source And Measurement

The data used for this study was sourced from **IPUMS USA (Integrated Public Use Microdata Series)** (IPUMS 2024), a comprehensive repository providing access to harmonized census and survey data from the United States. IPUMS USA is renowned for its meticulous process of standardizing variables across datasets, ensuring compatibility over time, and providing extensive metadata to support research. This harmonization process enables researchers to conduct longitudinal and cross-sectional studies on social, demographic, and economic trends.

The dataset consists of **individual-level microdata**, where each record represents a single person, numerically coded for all relevant characteristics. These records are organized into households, allowing for the study of individual behaviors and characteristics in the context of their family or co-residential settings. Unlike compiled statistics or pre-aggregated tables, this microdata structure provides researchers with unparalleled flexibility in exploring relationships between variables.

To address the diversity of record layouts, coding schemes, and documentation across the historical scope of the dataset, IPUMS implements a rigorous **harmonization process**. Variables are assigned **uniform codes**, ensuring consistency across census years (1850–2010) and the American Community Surveys (ACS) (2000–present). This standardization simplifies the analysis of long-term trends and facilitates comparisons across time and space.

2.3 Variable Selection

Using the IPUMS data extraction system, this study selected a focused subset of variables to examine the determinants of **non-marriage** in the United States. These include key demographic and socioeconomic characteristics, which are numerically coded for statistical analysis.

The table Table 1 shows the variables after data cleaning,

Table 1: Sample of cleaned data

Marital Status	Age	Gender	Race	Income	Education Level
Not_Married	73	Male	White	33900	Above_Bachelor
Married	43	Female	Asian	40000	High_School
Married	80	Female	White	13000	High_School
Married	66	Male	White	48000	Some_College
Married	52	Female	White	38100	Bachelor
Married	45	Female	American Indian	65000	High_School

2.3.1 Outcome variables

The primary outcome variable for this study is **Marital Status**, which categorizes individuals based on their marital state. The proportion of marital status categories is displayed in figure Figure 1. This variable allows for a comparison between individuals who have never married (`Not_Married`) and those who have (`Married`). For this study, **Not_Married**: Includes individuals who have never been married. **Married**: Includes individuals who are married, as well as those who are divorced, widowed, or separated.

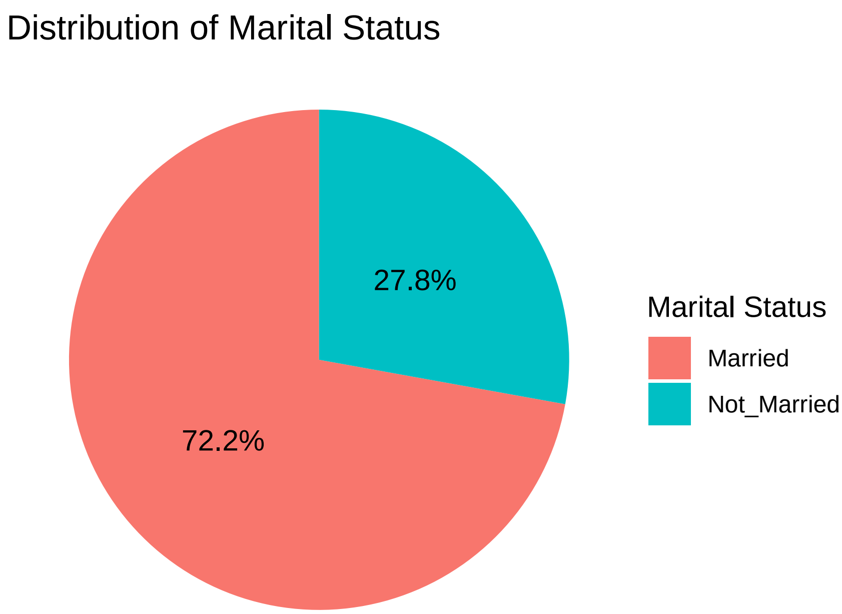


Figure 1: Proportion of Marital Status Categories

2.3.2 Predictor variables

This section focuses on the predictor variables included in the study. The distribution of predictor variables is displayed in Figure Figure 2. These variables capture demographic, socioeconomic, and personal characteristics, providing a comprehensive framework for analyzing factors associated with marital status. Below are the key predictor variables:

1. **Age**: A continuous variable representing the respondent's age in years.
2. **Gender**: A categorical variable indicating whether the respondent is male or female. Gender differences often play a role in marital patterns.

3. **Race**: A categorical variable categorized into White, Black, Asian, American Indian, and Other racial groups. This variable examines potential racial disparities in marital behavior.
4. **Education Level**: An ordinal variable indicating the highest level of education attained by the respondent. It is grouped into five categories: Below High School, High School, Some College, Bachelor's Degree, and Above Bachelor.
5. **Income**: A continuous variable measuring the respondent's annual income in dollars. Income reflects economic resources and may be associated with marital stability and decisions.

2.4 Data Selection

To understand the phenomenon of non-marriage in the United States, I selected the IPUMS USA dataset over alternatives such as IPUMS International. While IPUMS International provides harmonized census data from 104 countries with over 1 billion person records, its vast scope makes it less targeted for this study. My focus is specifically on the U.S. population, and IPUMS USA offers high-precision data from American Community Surveys (ACS) and federal censuses, making it a more suitable choice for studying societal patterns specific to the United States.

3 Model

For this analysis, I employed a Logistic Regression Model to assess the likelihood of an individual not being married (outcome variable) based on several demographic and socioeconomic predictors (predictor variables). This model was chosen due to the binary nature of the outcome variable, which distinguishes between individuals who are "Not Married" versus those who are married (including divorced, widowed, or separated).

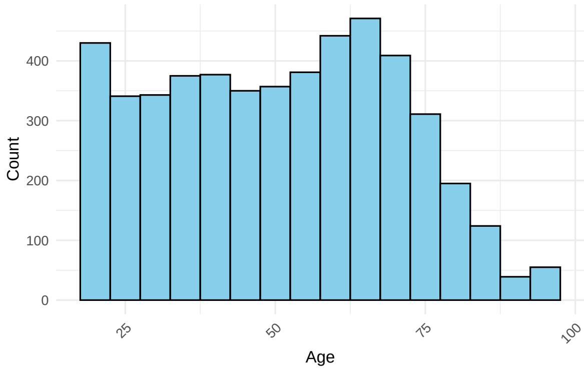
Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in **?@sec-model-details**.

3.1 Model Setup

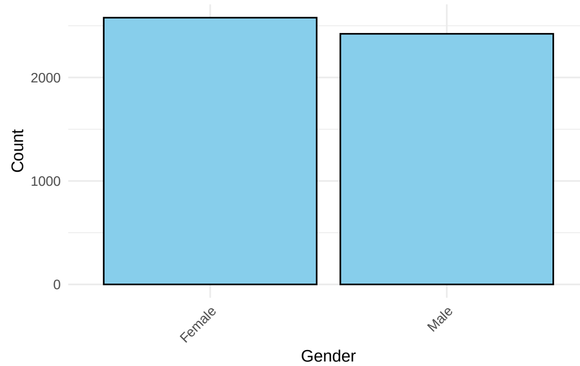
3.1.1 Objective

The primary objective of the model is to analyze and predict the factors associated with individuals' marital status, focusing on identifying key predictors for individuals who have never been married (`Not_Married`).

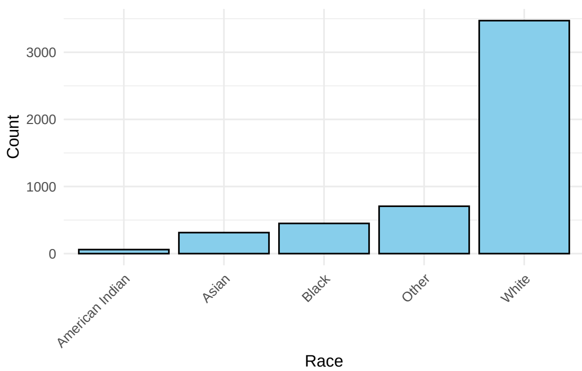
The logistic regression model used in this study is:



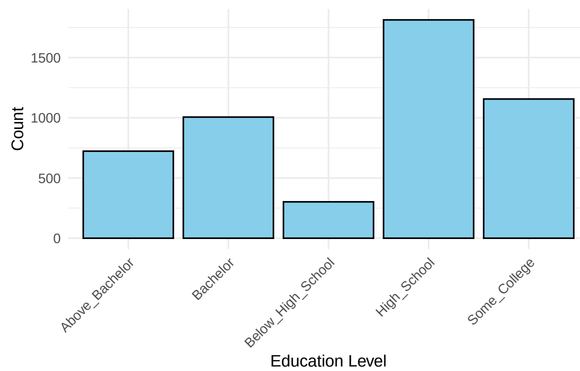
(a) Age Distribution



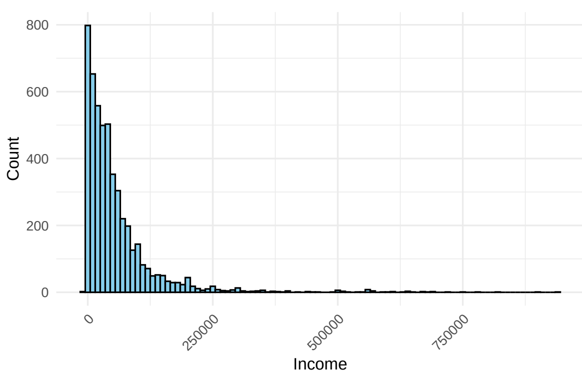
(b) Gender Distribution



(c) Race Distribution



(d) Education Level Distribution



(e) Income Distribution

Figure 2: Counts for Demographic Variables

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{gender} + \beta_3 \cdot \text{Race} + \beta_4 \cdot \text{Income} + \beta_5 \cdot \text{education\_level}$$

The priors for the coefficients are defined as:

$$\begin{aligned}
\beta_0 &\sim \text{Normal}(0, 2.5), \\
\beta_1 &\sim \text{Normal}(0, 2.5), \\
\beta_2 &\sim \text{Normal}(0, 2.5), \\
\beta_3 &\sim \text{Normal}(0, 2.5), \\
\beta_4 &\sim \text{Normal}(0, 2.5), \\
\beta_5 &\sim \text{Normal}(0, 2.5).
\end{aligned}$$

Where:

- $\hat{p}$: Represents the probability that an individual is classified as `Not_Married`.
- β_0: The intercept term, representing the log-odds of being `Not_Married` when all predictors are zero.
- β_1: The coefficient for the individual's age.
- β_2: The coefficient for gender.
- β_3: The coefficient for the race category.
- β_4: The coefficient for annual income.
- β_5: The coefficient for the education level.

3.1.2 Priors Explanation

In this model, normal priors with a mean of 0 and a standard deviation of 2.5 are assigned to all coefficients and the intercept. This choice reflects neutrality, implying no prior expectation of direction or magnitude for the coefficients.

The standard deviation of 2.5 reflects moderate uncertainty in prior beliefs, allowing the model to incorporate reasonable variability in predictor effects while avoiding overly restrictive priors.

3.2 Model justification

The logistic regression model was chosen for this study due to its suitability for predicting binary outcomes, in this case, whether an individual is categorized as "Not Married" or not. Logistic regression is widely used in social science research to understand relationships between a binary dependent variable and multiple independent variables, making it an ideal choice for analyzing marital status. This model allows for a straightforward interpretation of the effects of predictors, such as age, income, education, and gender, through odds ratios, providing insight into the likelihood of being “Not Married.” The dataset contains a mix of continuous and categorical variables, and logistic regression can handle this diversity effectively. Additionally, its theoretical simplicity ensures that the analysis remains interpretable while maintaining robustness. The choice of this model aligns with the study’s goal of identifying significant factors associated with non-marriage in the U.S., providing actionable insights while being grounded in established statistical methods. Despite its assumptions, such as linearity in the log-odds and independence of observations, logistic regression offers the flexibility to incorporate interaction terms and account for a range of predictor variables, making it a powerful tool for this research.

3.3 Model Validation

From Table 2, the validation indicates that the logistic regression model provides a robust understanding of the predictors of “Not Married” status. However, some predictors, such as race, show variability in their impact, necessitating cautious interpretation. The model’s good fit, coupled with well-defined confidence intervals and validation metrics, ensures that it provides reliable insights into the determinants of marital status.

Table 2: Summary of the model

term	estimate	std.error	conf.low	conf.high
(Intercept)	3.26	0.37	2.65	3.87
age	-0.08	0.00	-0.09	-0.08
genderMale	0.37	0.08	0.24	0.51
RaceAsian	-0.67	0.37	-1.26	-0.07
RaceBlack	0.20	0.34	-0.38	0.77
RaceOther	-0.67	0.35	-1.22	-0.10
RaceWhite	-0.67	0.33	-1.21	-0.11
Income	0.00	0.00	0.00	0.00
education_levelBachelor	0.30	0.14	0.07	0.54
education_levelBelow_High_School	0.48	0.21	0.15	0.83
education_levelHigh_School	0.21	0.14	-0.02	0.46
education_levelSome_College	0.18	0.15	-0.06	0.43

3.4 Results

The analysis integrates multiple visualizations to provide a comprehensive understanding of the demographic, economic, and social dynamics associated with marital status. Below, we summarize the key findings derived from each visualization.

The logistic regression coefficients, as shown in Figure 3, highlight the significant predictors of marital status. Race emerges as a crucial factor, with individuals identifying as "RaceBlack" or "RaceAsian" having substantial negative coefficients, indicating a reduced likelihood of being unmarried compared to other racial groups. Educational attainment also plays a major role; lower educational levels such as "Below High School" and "High School" are associated with a higher probability of not being married, as evidenced by their positive coefficients. Gender and income also contribute, with males marginally more likely to remain unmarried, while income has a smaller yet notable influence.

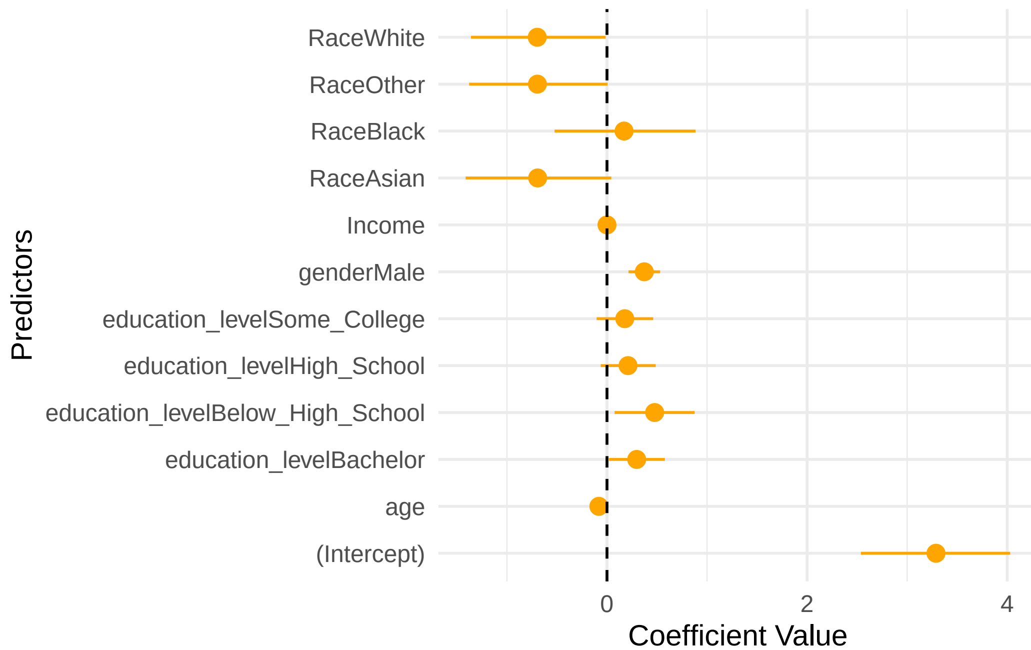


Figure 3: Coefficients of The Model

Figure 4 depicts the income distribution of unmarried individuals, revealing that the majority of unmarried individuals fall into lower income brackets. The distribution is heavily skewed, with a significant concentration of individuals earning below $50,000 annually. The density sharply declines for higher income levels, reflecting the economic constraints that may influence marital status. A small proportion of high-income individuals remain unmarried, as shown by the long right tail of the distribution.

Figure 5 shows the age distribution for unmarried individuals, which is dominated by younger age groups, particularly those in their 20s and early 30s. The density decreases substantially as age increases, illustrating that as individuals age, the likelihood of remaining unmarried diminishes. This trend aligns with societal norms where marriage becomes more prevalent in middle age, reflecting the intersection of age-related societal expectations and personal decision-making.

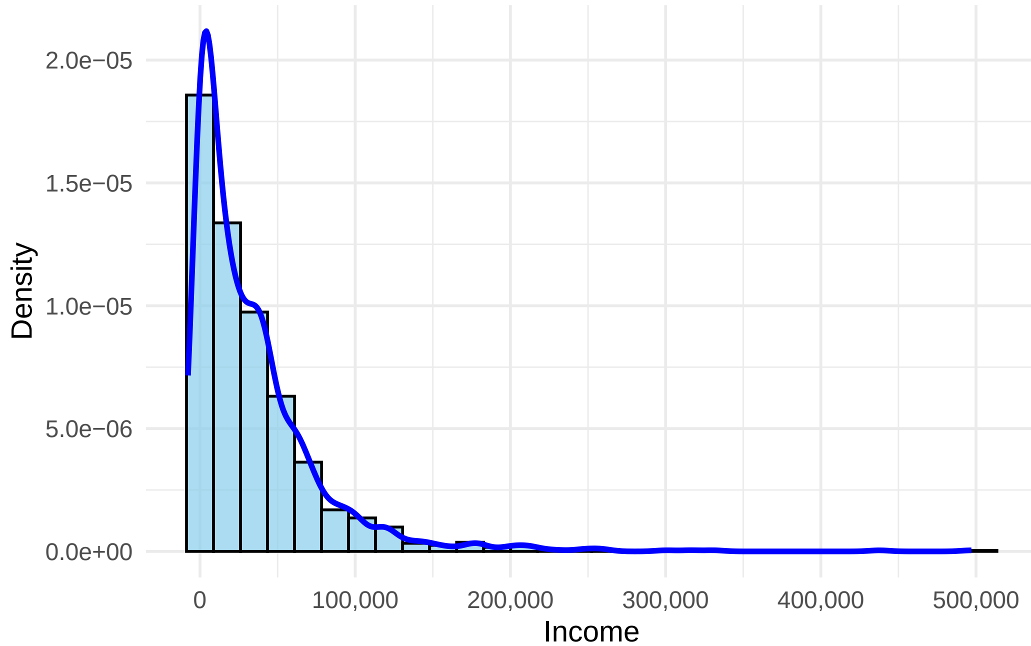


Figure 4: Income Distribution for Unmarried Individuals

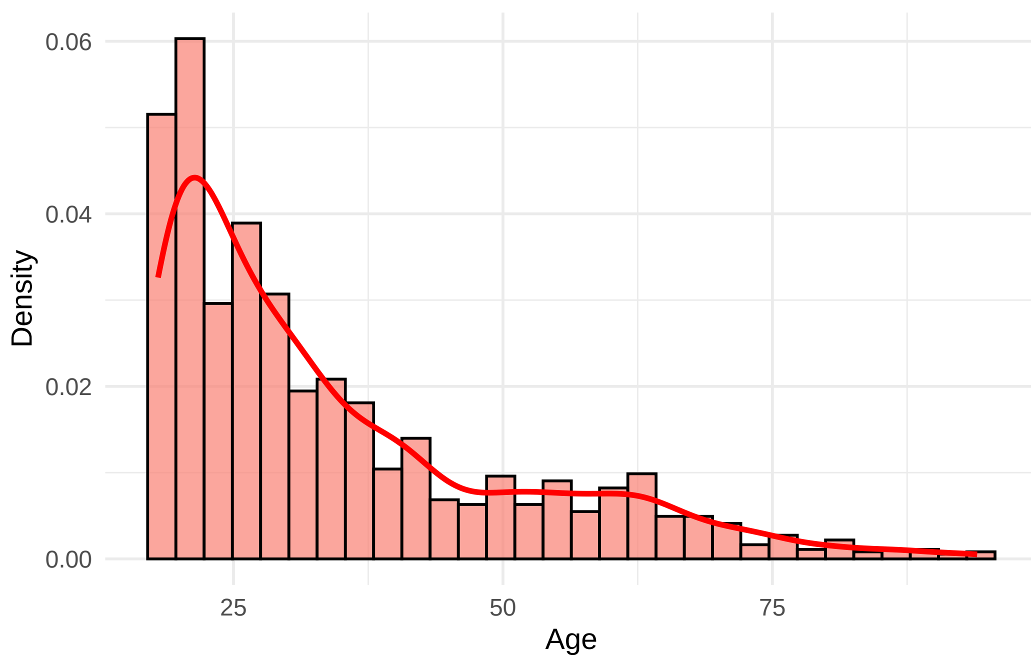


Figure 5: Age Distribution for Unmarried Individua

The proportions of marital status by education level are presented in Figure 6. Individuals with higher education levels, such as “Above Bachelor” and “Bachelor,” show significantly lower proportions of unmarried individuals, at 16% and 25%, respectively. Conversely, those with lower educational attainment, such as “Below High School” and “Some College,” exhibit higher proportions of unmarried individuals, with 32% in each group. This suggests that educational attainment serves as a strong predictor of marital outcomes, potentially due to its influence on economic stability and social networks.

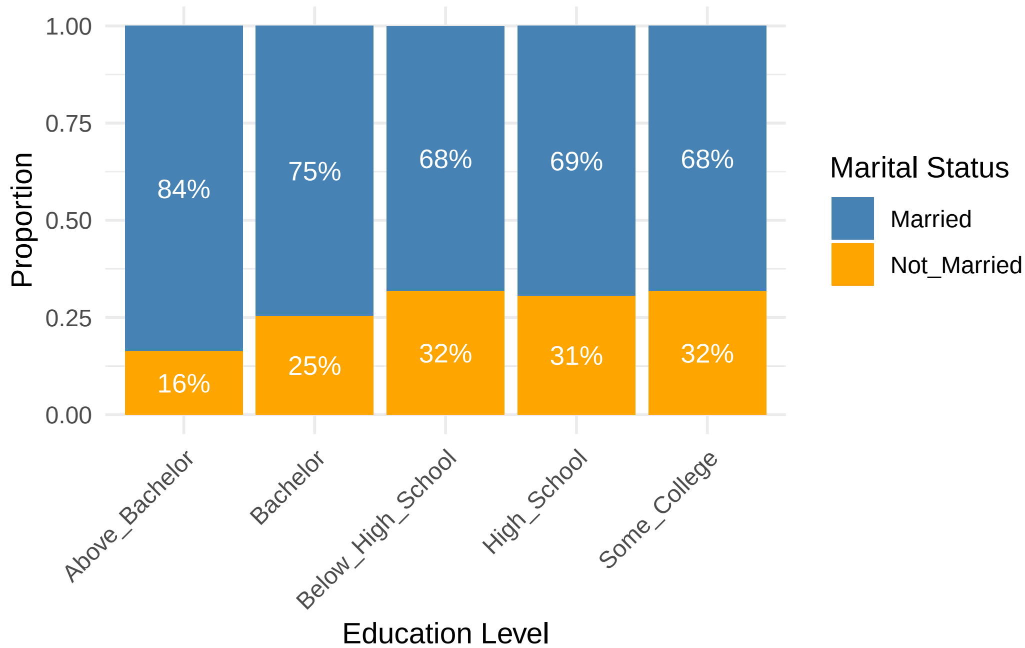


Figure 6: Proportions of Marital Status by Education Level

Figure 7 and Figure 8 highlight the proportions of marital status by race and gender. Race disparities are evident, with “American Indian” and “Black” populations showing higher proportions of unmarried individuals, at 37% and 43%, respectively. In contrast, “White” and “Asian” groups exhibit the lowest proportions of unmarried individuals, at 24% and 30%, respectively. Gender analysis reveals that males are more likely to remain unmarried than females, with 30% of males unmarried compared to 26% of females. These findings underscore the intersection of cultural, economic, and societal factors in shaping marital trends across both race and gender.

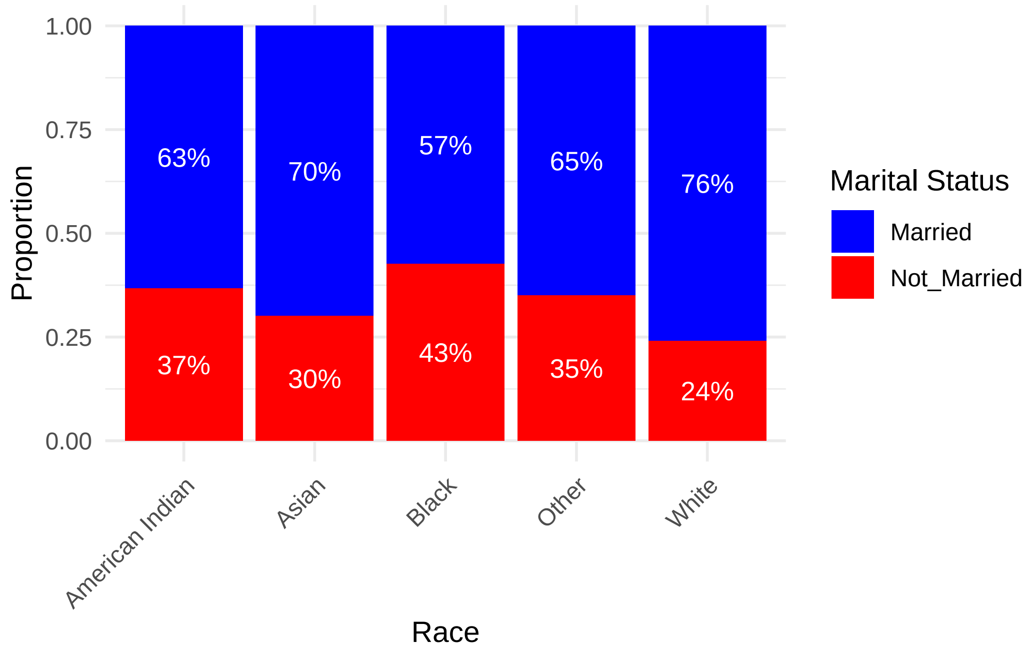


Figure 7: Proportions of Marital Status by Race

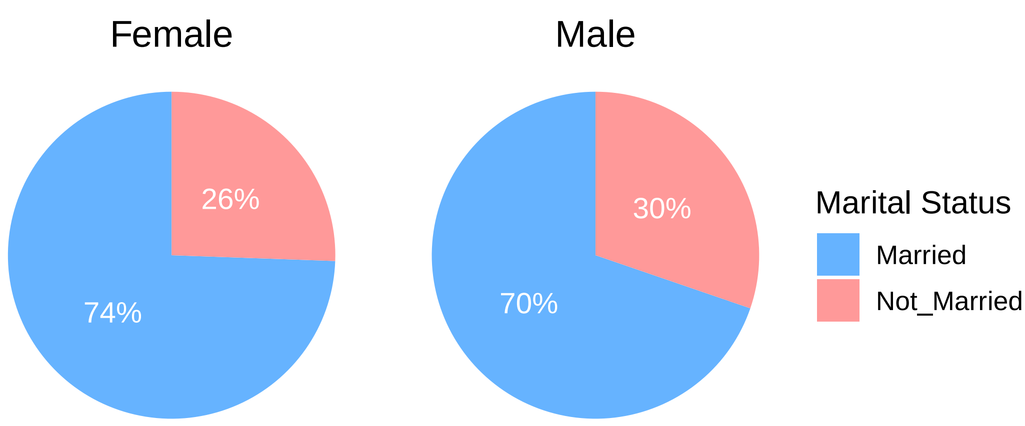


Figure 8: Marital Status Proportions by Gender

4 Discussion

4.1 Determinants of Non-Marriage and Societal Implications

This study delves into the socio-demographic factors that influence marital status, with a specific focus on the probability of remaining unmarried in the United States. By analyzing individual-level microdata through logistic regression, we identified significant predictors such as education, income, race, gender, and age.

For instance, the logistic regression model revealed that individuals with higher education levels, particularly those with a bachelor's degree or more, were significantly less likely to remain unmarried compared to those with lower education levels. Education, therefore, serves as a strong stabilizer in marital outcomes, providing individuals with economic resources and social capital that facilitate marriage. Simultaneously, the analysis underscored the heightened likelihood of remaining unmarried among African Americans, reflecting deeper structural inequities that impact marital patterns. Gender also emerged as a significant factor, with males exhibiting a slightly higher probability of being unmarried compared to females.

These findings, grounded in data and supported by statistical rigor, affirm prior literature while adding layers of nuance. They emphasize how individual socio-demographic traits intersect with broader societal structures to shape marital outcomes.

4.2 Insights into Education, Income, and Marriage Patterns

One of the most compelling findings of this study is the pronounced impact of education on marital status. Individuals with higher education levels—those with bachelor's degrees or above—demonstrated significantly lower rates of non-marriage. This is evident from the analysis showing that only 16% of individuals with education above a bachelor's degree were unmarried, compared to 32% of those with "Some College" education. The logistic regression coefficients also reinforced this relationship, with those holding higher degrees being far less likely to remain unmarried (e.g., coefficient for "Below High School" = 0.48).

This finding aligns with existing theories that higher education fosters economic stability, enhances social capital, and provides a greater sense of agency in life decisions, all of which are conducive to marriage. However, the results also highlighted that income, while less strongly predictive, plays a complementary role. The income distribution plot for unmarried individuals revealed clustering at lower income brackets, with the majority earning below $40,000 annually. This suggests that financial security continues to be an essential consideration in marital decisions, albeit to a lesser extent than education.

These patterns point to the interplay between socio-economic resources and marital outcomes, emphasizing how disparities in access to education and income perpetuate inequalities in family formation.

4.3 Gender and Race in Marital Trends

The analysis revealed striking disparities in marital trends across gender and racial lines. Gender-wise, men exhibited a higher likelihood of remaining unmarried compared to women, with 30% of men being unmarried versus 26% of women. This aligns with cultural norms and gendered expectations, where women are often perceived as more closely tied to family roles and marriage. Moreover, women may face greater societal pressure to marry, which could explain their lower rates of remaining unmarried.

Racial disparities in marital status were even more pronounced. For example, African Americans exhibited the highest rates of non-marriage (43%), significantly higher than those for Asian Americans (30%) and White individuals (24%). These patterns likely reflect a combination of structural inequalities, cultural factors, and historical contexts. Economic disparities, higher unemployment rates, and systemic barriers to wealth accumulation among African Americans may contribute to these higher non-marriage rates. Conversely, Asian Americans, who exhibited the lowest rates of remaining unmarried, may benefit from cultural norms that strongly prioritize marriage and family cohesion.

These findings underscore how structural inequities and cultural norms interact to shape marital trends. They also highlight the need for policies aimed at addressing systemic barriers and promoting equality in socio-economic outcomes to reduce disparities in family formation.

4.4 Weaknesses and Next Steps

While this study provides critical insights into the predictors of marital status, it is not without limitations. The reliance on cross-sectional data, for example, restricts the ability to establish causal relationships. While the findings suggest strong associations between education, income, race, gender, and marital outcomes, longitudinal data could better capture how these factors interact over time to influence marriage decisions.

Another limitation lies in the dataset itself, which does not include qualitative factors such as cultural attitudes, personal preferences, or psychological traits. These dimensions are critical for understanding the broader context of marital decisions. For instance, while income and education are robust predictors, societal expectations and individual life goals may play equally important roles that remain unquantified in this analysis.

Future research should aim to integrate mixed-method approaches, combining quantitative rigor with qualitative depth, to provide a holistic view of marital trends. Additionally, while this study focused exclusively on the United States, extending the analysis to include international data could uncover global patterns and cultural nuances. For example, exploring how marriage trends differ in countries with varying economic systems, religious practices, and gender norms would provide valuable comparative insights.

4.5 Conclusion

This study contributes to the understanding of non-marriage determinants by highlighting the interplay between education, income, race, gender, and age. Education emerged as the most significant predictor, with higher levels strongly associated with lower non-marriage rates. Income, while secondary, still played a notable role, emphasizing the importance of economic stability in marital decisions. Gendered expectations and racial disparities further revealed how societal structures and cultural norms influence marriage patterns.

The findings highlight the socio-economic inequities that shape family formation and underscore the importance of addressing structural barriers to promote marital stability. Future research should continue to explore these dynamics, leveraging longitudinal data and integrating qualitative perspectives to deepen our understanding of marriage and its evolving role in contemporary society.

A Appendix

A.1 How to Extract 2022 ACS Data from IPUMS

To obtain data from IPUMS, we start by navigating to the IPUMS USA section and clicking on Get Data. Next, we go to the Select Sample section, where we uncheck the "Default sample from each year" option and instead select 2023 ACS. After selecting our sample, we proceed to add variables of interest. For individual-level data, we might add variables from the Person section. For example, under Demographic, we could include variables like AGE, and under Person, we could add SEX, RACE, INCTOT (total personal income) and EDUC (education attainment). Once our variables are selected, we click View Cart, then proceed by clicking Create Data Extract. At this point, we review our selections, change the Data Format to CSV, and submit our extract for processing. Then we saved it locally as usa_00001.csv.

A.2 Model Assumptions

1. **Linearity**: The log odds of the outcome are linearly related to the predictors.
2. **Independence**: Observations are independent of each other.
3. **No Multicollinearity**: Predictors are not highly correlated.
4. **No Outliers or Influential Points**: Checked through residual analysis.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Bolker, Ben et al. 2023. *broom.mixed: Tidying Methods for Mixed Models*. https://bbolker.github.io/broom.mixed/.

Cepeda, Gabriel A., Jonah Gabry, et al. 2023. *rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm/.

IPUMS. 2024. "IPUMS USA." https://usa.ipums.org/usa/#:~:text=IPUMS%20USA%20collects,%20preserves%20and%20harmonizes.

Müller, Kirill, and Jennifer Bryan. 2023. *here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Dewey Dunnington, and Apache Arrow Developers. 2023. *arrow: Integration to Apache Arrow*. https://arrow.apache.org/.

Robinson, David. 2023. *broom: Convert Statistical Objects into Tidy Tibbles*. https://broom.tidyverse.org/.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. https://ggplot2.tidyverse.org/.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org/.

Wickham, Hadley, Dana Seidel, et al. 2023. *scales: Scale Functions for Visualization*. https://scales.r-lib.org/.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://haozhu233.github.io/kableExtra/.