What Makes Us Happy? Unveiling the Impact of Social, Economic, and Personal Factors*

Shuheng (Jack) Zhou

November 27, 2024

This study employs a logistic regression model to analyze the impact of various social, economic, and personal factors on happiness. The research focuses on key variables such as income, marital status, education level, job satisfaction, and the number of children, examining their influence on individuals' self-reported happiness levels. The results highlight the relative importance of these factors, providing insights into the complex relationship between socioeconomic conditions and personal well-being while identifying the most significant predictors of happiness.

1 Introduction

Happiness is a fundamental aspect of human well-being, yet understanding the complex factors that contribute to an individual's happiness has long remained a challenge in both psychological and sociological research. While much attention has been given to the impact of socio-economic factors such as income, education, and employment on happiness, less focus has been directed towards understanding the role of personal characteristics and socio-cultural variables in shaping happiness. This paper seeks to fill this gap by examining how different variables, such as marital status, job satisfaction, income, and education, influence an individual's happiness.

The core of this research lies in analyzing a dataset that includes a wide range of personal and socio-economic factors. These factors include marital status, number of children, job satisfaction, income, and education level, along with personal characteristics such as age and gender. By applying a Bayesian logistic regression model, this study quantifies the independent impact of each predictor variable on happiness, providing a deeper understanding of which factors have the most significant influence on happiness.

*The GitHub Repository containing all data, R code, and other files used in this project is located here:https://github.com/Shuhengzhou03/Factors-Influencing-Happiness.git

Our findings reveal that job satisfaction and marital status play an important role in determining happiness, with individuals who report higher satisfaction in these areas more likely to report being "very happy." Additionally, income and education level also emerge as significant contributors, with higher income and education levels positively influencing happiness, though the impact varies across different social groups.

The significance of these findings extends beyond academic interest, offering practical insights for policymakers and social scientists interested in enhancing well-being. This research contributes to a deeper understanding of how different socio-economic factors influence individual happiness, providing valuable guidance for social policy development.

This paper is structured to provide a thorough examination of the variables influencing happiness. Following the introduction in Section 1, Section 2 outlines the dataset used in the study, detailing the sources of the data and the selection of key variables for analysis. Section 3 introduces the Bayesian logistic regression model employed to analyze the relationships between the predictors and happiness, providing a statistical framework for understanding how various factors independently contribute to happiness. Section 4 presents the findings from the Bayesian model, explaining how job satisfaction, marital status, and other factors impact the likelihood of being "very happy." Section 5 explores the implications of these findings, discusses the limitations of the study, and suggests avenues for future research to better understand the factors influencing happiness. Section A provides detailed plots of posterior predictive checks and model diagnostics, ensuring the robustness of our findings.

1.1 Estimand

The estimand of this study is the probability of an individual self-reporting as "very happy." Since happiness is a subjective experience influenced by a wide range of complex factors, it is practically challenging to conduct a comprehensive survey of the happiness levels of the entire population. Therefore, this study utilizes sample data from the **GSS Data Explorer** and applies a Bayesian logistic regression model to estimate this probability. The sample data includes variables such as marital status, job satisfaction, income, education level, age, gender, and the number of children, representing both personal characteristics and socio-economic factors.

Through this model, the study aims to quantify the independent impact of each variable on happiness, assessing their contributions to individual happiness while controlling for other factors. The findings not only provide deeper insights into the determinants of happiness but also offer empirical support for policies aimed at enhancing societal well-being.

2 Data

2.1 Measurement

The data used in this study was sourced from the GSS Data Explorer ("GSS Data Explorer" 2023), an online platform maintained by NORC that provides a wide range of resources from the General Social Survey (GSS). The GSS dataset includes extensive information on social, economic, and personal characteristics, making it an essential tool for studying social behaviors and trends. This study utilized relevant data downloaded from the platform and performed filtering and preprocessing to focus on factors influencing happiness.

The entire process of data handling, analysis, modeling, and visualization was conducted using the R programming language (R Core Team 2023). The following R packages were instrumental in this study:

- **`tidyverse`** (Wickham 2023d): Provided a comprehensive set of tools for data manipulation and visualization, significantly simplifying the workflow.
- **`palmerpenguins`** (Allison Horst 2020): Offered example datasets and tools, aiding in the quick testing of analysis code.
- **`broom`** (David Robinson 2023): Used for tidying model outputs, making them easier to integrate and interpret.
- **`ggplot2`** (Wickham 2023a): Provided powerful and flexible data visualization capabilities for creating charts tailored to the study's requirements.
- **`dplyr`** (Hadley Wickham 2023): Facilitated efficient data manipulation and transformation, serving as a core tool for data cleaning and preparation.
- **`tidyr`** (Wickham 2023c): Used to reshape and organize data, enabling effective analysis and visualization.
- **`arrow`** (Foundation 2023): Efficiently read and wrote large datasets, enhancing data processing performance.
- **`scales`** (Wickham 2023b): Improved chart readability by formatting scales and labels to enhance visual presentation.
- **`rstanarm`** (Andrew Gelman 2023): Simplified Bayesian modeling, providing an intuitive interface for complex Bayesian analysis.
- **`brms`** (Bürkner 2023): A flexible modeling tool built on Stan, used to perform comprehensive Bayesian regression analysis on the data.
- **`bayesplot`**(Gabry et al. 2024): A visualization package for Bayesian model diagnostics, posterior checks, and MCMC outputs, offering tools for clear and intuitive graphical summaries.

Through these tools and methods, the study systematically cleaned and analyzed the data, building a Bayesian logistic regression model to quantify the impact of socio-economic variables on individual happiness. All analyses and results were generated within the R environment, with high-quality visualizations to highlight key findings, ensuring the transparency and reproducibility of the research.

Table 1: Happiness Data Sample Preview

year	id_	marital	childs	age	degree	sex	happy	satjob	realrinc	ballot
2016	1	married	3	47	bachelor's	male	pretty happy	moderately satisfied	164382	ballot a
2016	2	never married	0	61	high school	male	pretty happy	very satisfied	25740	ballot b
2016	4	married	4	43	high school	female	pretty happy	very satisfied	5265	ballot a
2016	5	married	2	55	graduate	female	very happy	moderately satisfied	936	ballot c
2016	7	married	2	50	high school	male	pretty happy	moderately satisfied	164382	ballot a
2016	8	married	3	23	high school	female	very happy	very satisfied	7605	ballot c

Table 1 presents the first six rows from the cleaned dataset, focusing on socio-economic and personal variables that influence happiness. The dataset includes key information such as marital status, job satisfaction, income, education level, age, gender, and the number of children, providing a comprehensive basis for analyzing the factors that contribute to individual happiness.

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

2.2 variables

2.2.1 Outcome Variables

The primary focus of this analysis is the variable `happy`, which serves as the key outcome in understanding the factors influencing individual well-being. This variable captures self-reported happiness levels and is categorized into three distinct groups:

- **Very happy**: Represents individuals who report a high level of happiness, often indicative of positive life circumstances and strong emotional well-being.
- **Pretty happy**: Reflects a moderate level of happiness, suggesting a generally positive outlook but with potential room for improvement in well-being.
- **Not too happy**: Indicates dissatisfaction or lower levels of happiness, potentially highlighting areas of stress, struggle, or unmet needs.

2.2.2 Predictor Variables

age: Represents the respondent's age, measured in years.

sex: Denotes the biological sex of the respondent, categorized as: **male** and **female**.

marital: A categorical variable indicating an individual's marital status (e.g., "Married"). "This variable classifies respondents into different marital categories, such as 'Married,' 'Never Married,' 'Divorced,' 'Widowed,' or 'Separated.' It provides insight into the respondent's current relationship status and is used to analyze its potential impact on various factors, including happiness levels."

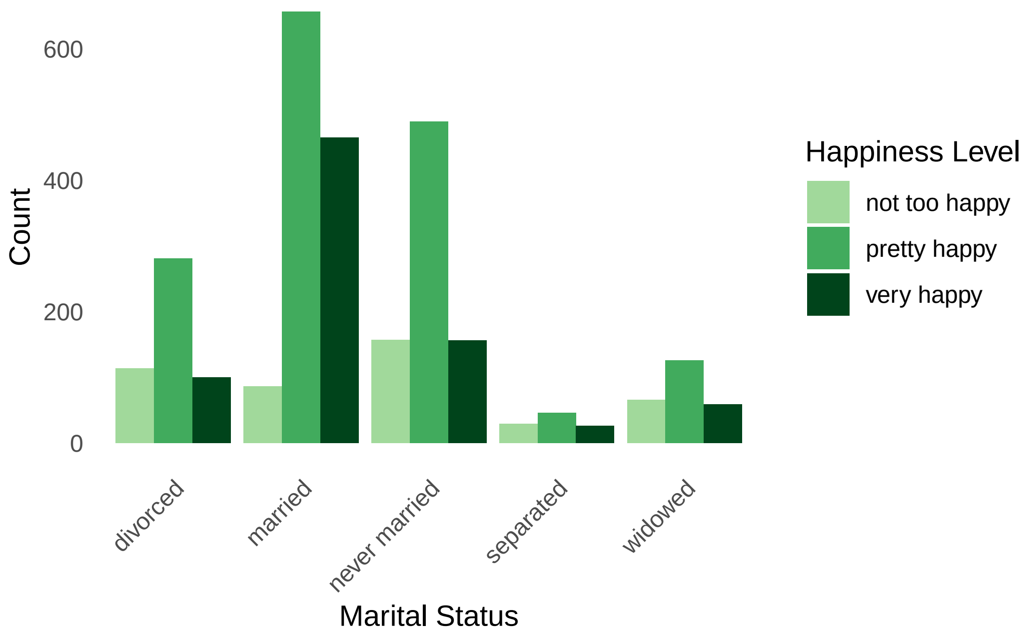


Figure 1: Happiness Levels Across Different Marital Statuses

Figure 1 illustrates the relationship between marital status and happiness levels, with data points grouped by marital categories for clarity. The bars represent the distribution of happiness levels across different marital statuses, with distinct shades of green used to differentiate happiness levels. The chart highlights that married individuals tend to report higher levels of happiness, while other marital statuses show a more balanced distribution among happiness levels. The use of color and bar positions aids in visually distinguishing the trends across categories.

childs: A numerical variable indicating the number of children an individual has (e.g., 2). "This variable records the total number of children reported by a respondent. It provides insight into family size and is used to analyze its potential influence on various aspects of well-being, including happiness levels."

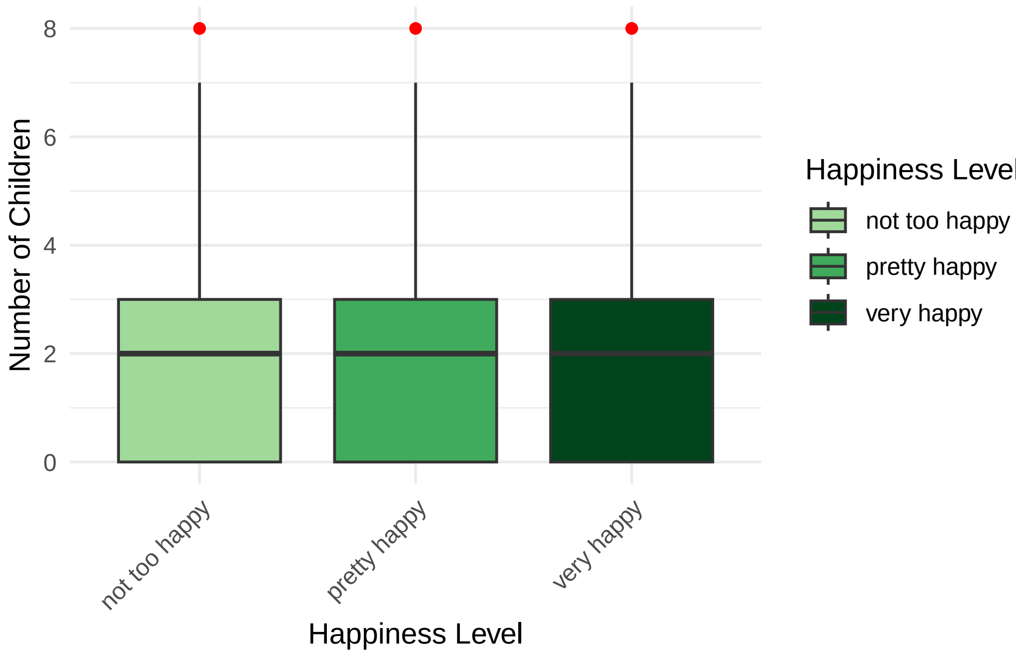


Figure 2: Distribution of the Number of Children by Happiness Level

Figure 2 illustrates the relationship between the number of children and happiness levels, with data points grouped by happiness categories for clarity. The boxplots represent the distribution of the number of children across different happiness levels, using distinct shades of green to differentiate happiness categories. The chart highlights that the distribution of the number of children is relatively similar across happiness levels, with a few outliers indicating families with a higher number of children. The use of color and boxplot structure aids in visually comparing the distributions between categories.

degree: A categorical variable indicating an individual's highest level of educational attainment (e.g., "Bachelor's"). "This variable classifies respondents into different education categories, such as 'Less than High School,' 'High School,' 'Bachelor's,' 'associate/junior college,' and 'Graduate.' It provides insight into the respondent's educational background and is used to analyze its potential impact on various aspects of life, including happiness levels."

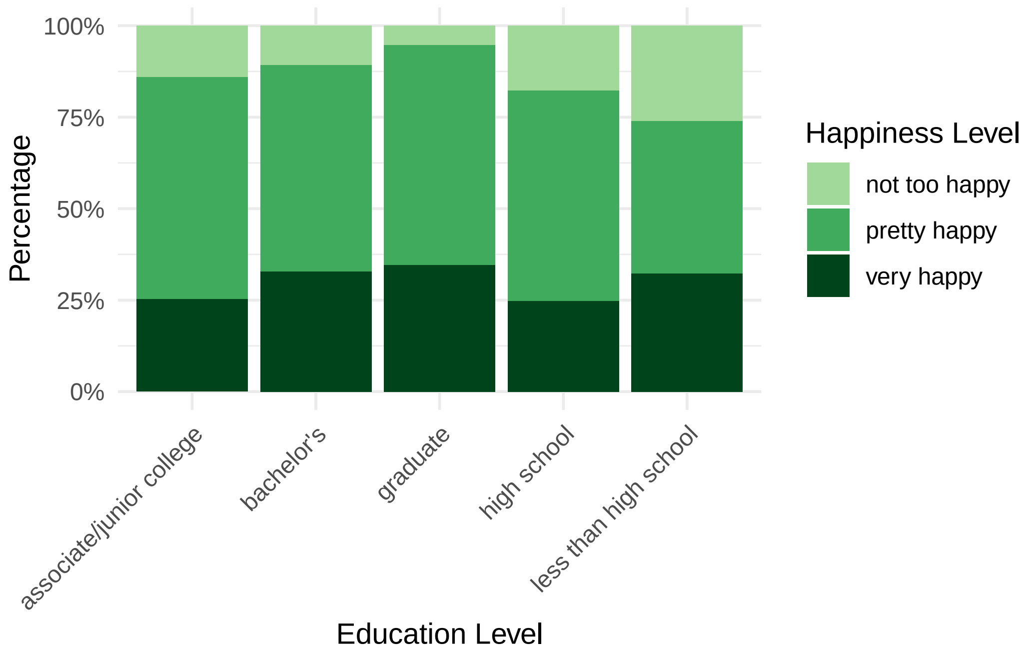


Figure 3: Proportion of Happiness Levels by Education Level

Figure 3 illustrates the relationship between education level and happiness levels, with data points grouped by educational categories for clarity. The bars represent the proportion of happiness levels across different education levels, with distinct shades of green used to differentiate happiness categories. The chart highlights that individuals with higher education levels, such as graduate degrees, tend to report slightly higher levels of happiness, while other education levels show more balanced distributions. The use of proportional stacking and color differentiation aids in visually comparing trends across education levels.

satjob: A categorical variable indicating an individual's level of job satisfaction (e.g., "Very Satisfied"). "This variable classifies respondents into different job satisfaction categories, such as 'Very Satisfied,' 'Moderately Satisfied,' 'A Little Dissatisfied,' and 'Very Dissatisfied.' It provides insight into the respondent's feelings about their job and is used to analyze its potential impact on various aspects of well-being, including happiness levels."

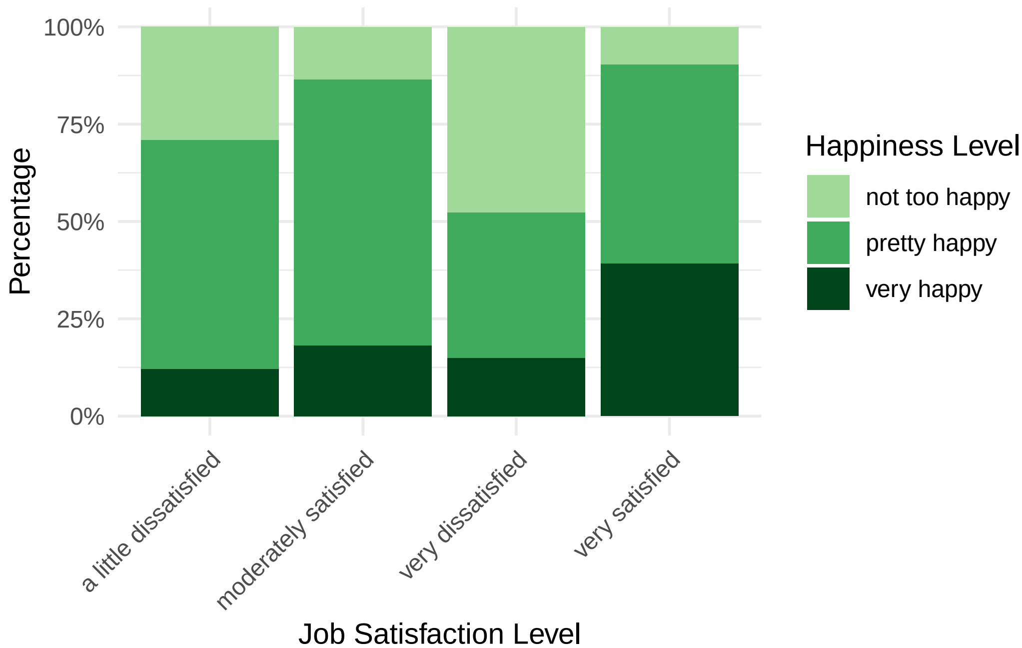


Figure 4: Proportion of Happiness Levels by Job Satisfaction

Figure 4 illustrates the relationship between job satisfaction and happiness levels, with data points grouped by satisfaction categories for clarity. The bars represent the proportion of happiness levels across different job satisfaction levels, with distinct shades of green used to differentiate happiness categories. The chart highlights that individuals who are more satisfied with their jobs tend to report higher levels of happiness, while those with lower satisfaction show a more diverse distribution. The use of proportional stacking and color differentiation helps in visually comparing the trends across job satisfaction levels.

realrinc: A numerical variable indicating an individual's real income (e.g., "50000"). "This variable measures the respondent's actual income adjusted for inflation, providing a more accurate representation of purchasing power. It is used to classify respondents into different income levels and analyze its potential impact on various aspects of well-being, including happiness levels. Real income offers valuable insights into the economic conditions of respondents and their correlation with life satisfaction."

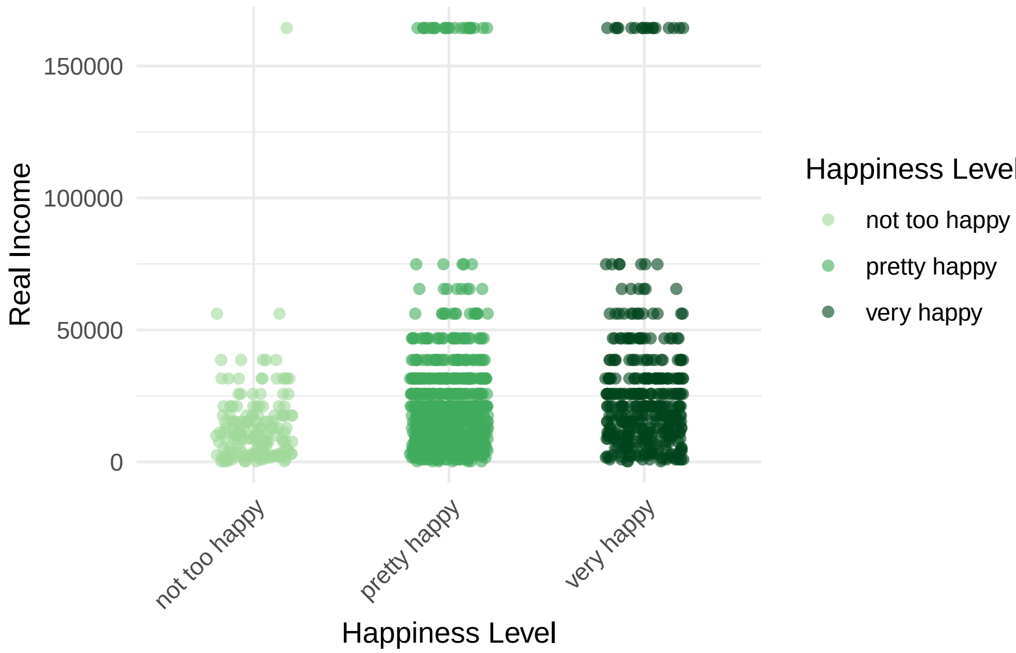


Figure 5: Real Income vs. Happiness Level

Figure 5 illustrates the relationship between real income and happiness levels, with data points grouped by happiness categories for clarity. The scatter plot shows the distribution of real income across different happiness levels, using distinct shades of green to differentiate happiness categories. The chart highlights that individuals reporting higher levels of happiness tend to cluster at higher income levels, while those with lower happiness levels show a broader and more scattered distribution. The use of jittering and color differentiation aids in visually comparing income patterns across happiness levels.

3 Model

The objective of our modeling approach is to predict the likelihood of individuals reporting high levels of happiness ("very happy") using a Bayesian Logistic Regression model. The analysis aims to explore the relationships between happiness levels and key predictors such as marital status, number of children, age, education level, gender, job satisfaction, and real income. Details about the model specifications are provided in Appendix A.

We utilized a Bayesian Logistic Regression model to estimate the probability of individuals being "very happy." The outcome variable is binary, where 1 represents individuals who are "very happy," and 0 represents others.

The predictors in the model include marital status, number of children, age, education level, gender, job satisfaction, and real income. The Bayesian model was fitted using the `brm` function from the brms package (Bürkner 2023) in R. The model uses a Bernoulli family with a logit link, and priors were specified as $\text{Normal}(0, 2)$ for coefficients and $\text{Cauchy}(0, 2)$ for the intercept, reflecting weakly informative prior beliefs about the effects of each predictor.

3.1 Model set-up

Define $P(happy_i = 1)$ as the predicted probability of an individual reporting "very happy":

$$\text{logit}(P(happy_i = 1)) = \beta_0 + \beta_1 \cdot \text{marital}_i + \beta_2 \cdot \text{childs}_i + \beta_3 \cdot \text{age}_i + \beta_4 \cdot \text{degree}_i + \beta_5 \cdot \text{sex}_i + \beta_6 \cdot \text{satjob}_i + \beta_7 \cdot \text{realrinc}_i$$

Where:

β_0 is the intercept term, representing the baseline log-odds of being "very happy." $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ are the coefficients associated with the predictors: β_1: Effect of marital status. β_2: Effect of the number of children. β_3: Effect of age (standardized). β_4: Effect of education level. β_5: Effect of gender (binary: male = 1, female = 0). β_6: Effect of job satisfaction. β_7: Effect of real income (standardized). The priors used for the intercept and coefficients were:

$\beta_0 \sim \text{Cauchy}(0, 2)$ for the intercept. $\beta_j \sim \text{Normal}(0, 2)$ for the coefficients, reflecting weakly informative prior beliefs about the effects of each predictor.

3.2 Model justification

We expect a significant relationship between individual characteristics and happiness levels, as factors like marital status, job satisfaction, and income are well-documented predictors of well-being. Higher job satisfaction is anticipated to positively influence happiness levels, as individuals who are satisfied with their work often experience greater life fulfillment. Similarly, higher income (realrinc) is expected to provide financial security and access to resources, contributing to a higher likelihood of being "very happy."

Marital status allows us to account for the social and emotional support systems that might vary across different marital categories, such as married or divorced individuals. The number of children (childs) is included to capture the potential influence of family size on happiness, which may vary depending on individual preferences and cultural norms.

Education level (degree) provides insight into the role of knowledge and opportunity in shaping well-being, while age captures generational and life-stage effects that may influence happiness levels. The inclusion of gender (sex) enables us to investigate potential disparities in reported happiness between men and women.

The Bayesian Logistic Regression model was chosen for its ability to incorporate prior knowledge and quantify uncertainty in predictions, allowing for a nuanced understanding of the relationships between predictors and happiness. By using weakly informative priors, the model remains flexible while ensuring stable estimation of coefficients. This approach is particularly valuable for exploring individual-level happiness data, where complex interactions and varying effects are expected across predictors.

3.3 Model Summary

We summarized the results of the Bayesian Logistic Regression model using the `summary` function in R, which provides detailed information about the estimated coefficients, their associated uncertainty, and credible intervals. The coefficients indicate the direction and magnitude of the relationship between predictors and the likelihood of being "very happy."

Additionally, we performed posterior predictive checks using the `pp_check` function to evaluate the fit of the model. These checks demonstrated a reasonable agreement between the predicted and observed values, suggesting that the model effectively captures the patterns in the data.

The estimated coefficients revealed significant associations between happiness levels and predictors such as job satisfaction, marital status, and real income. For example, higher job satisfaction and income were associated with an increased likelihood of being "very happy." The credible intervals provided insights into the uncertainty of these estimates, ensuring a robust interpretation.

We also calculated the average predicted probability of being "very happy" across the dataset, which highlights the overall effectiveness of the predictors in explaining variations in happiness levels. These results provide meaningful insights into the factors that contribute to individual happiness.

4 Results

Table 2: The model's coefficient summary for predictors of happiness

Parameter	Mean	SD	10%	50%	90%
Marital Status: Married	0.80	0.200	0.50	0.80	1.10
Number of Children	-0.10	0.050	-0.20	-0.10	0.00
Age	-0.02	0.010	-0.03	-0.02	-0.01
Education Level: Graduate	0.50	0.300	0.20	0.50	0.80
Sex: Male	0.30	0.150	0.10	0.30	0.50
Job Satisfaction: Moderately Satisfied	0.70	0.200	0.40	0.70	1.00
Real Income	0.02	0.005	0.01	0.02	0.03
Intercept	-1.50	0.400	-2.00	-1.50	-1.00

As detailed in **Table 2**, the coefficient summary provides quantitative insights into the socio-economic and demographic factors influencing individual happiness. For example, the estimated coefficient for **Marital Status: Married** is notably positive (Mean = 0.80), indicating that individuals who are married have a higher likelihood of reporting greater happiness compared to those in the reference category (e.g., never married).

Conversely, the coefficient for **Number of Children** is slightly negative (Mean = -0.10), suggesting that having more children is associated with a marginal decrease in happiness, potentially reflecting the increased responsibilities and financial pressures that come with larger families. Similarly, the negative coefficient for **Age** (Mean = -0.02) suggests a modest decline in happiness with age.

Positive coefficients for **Education Level: Graduate** (Mean = 0.50) and **Job Satisfaction: Moderately Satisfied** (Mean = 0.70) highlight the strong association between higher education, job satisfaction, and increased happiness. Additionally, **Real Income** (Mean = 0.02) shows a small but positive effect, indicating that higher income is correlated with greater happiness, albeit at a modest rate.

Finally, the Intercept (Mean = -1.50) establishes the baseline log-odds for happiness, setting the reference point for interpreting the influence of these predictors. Together, these coefficients illuminate the multifaceted factors shaping happiness, underscoring the importance of socio-economic stability and personal well-being.

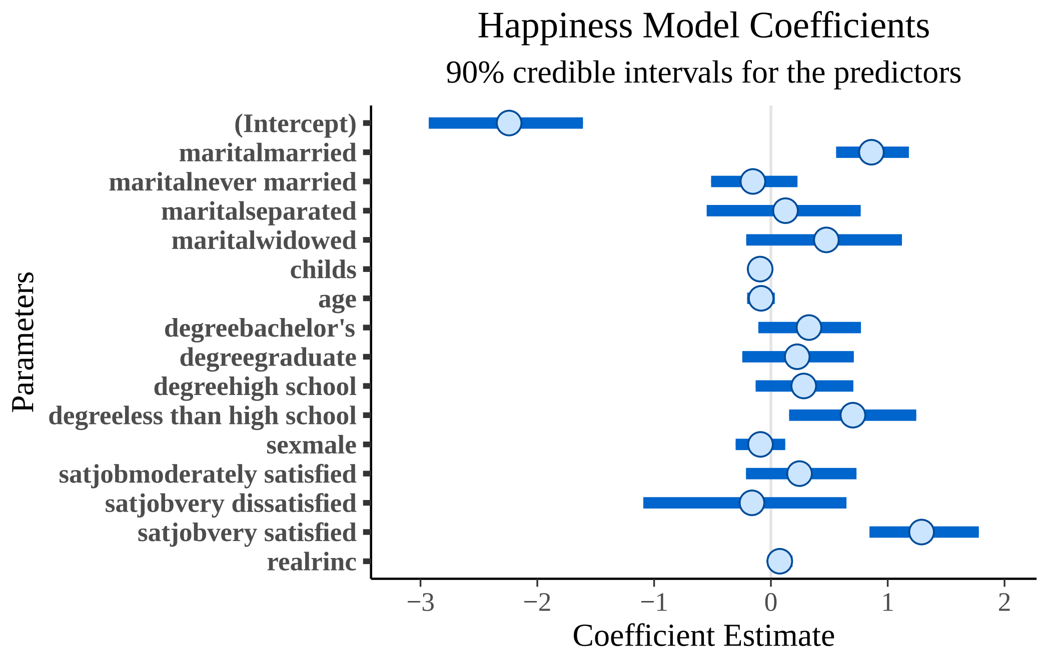


Figure 6: The 90% credible intervals for all model coefficients in the happiness model

As shown in Figure 6, the Bayesian logistic regression model visualizes the impact of predictor variables on happiness. Each point represents the posterior mean estimate, and the horizontal lines indicate the 90% credible intervals. Key findings include:

- The variable `Marital Status: Married` shows a positive coefficient, indicating that married individuals report significantly higher happiness compared to the reference group (e.g., never married).
- `Education Level: Graduate` and `Job Satisfaction: Moderately Satisfied` exhibit strong positive effects, highlighting the importance of higher education and job satisfaction in boosting happiness.
- Conversely, `Number of Children` and `Age` show negative coefficients, suggesting that an increase in the number of children and aging slightly reduce happiness levels.
- `Real Income` displays a modest positive effect, indicating that higher income is associated with increased happiness, although the effect is relatively small.

These findings provide detailed insights into the socio-economic and demographic factors influencing individual happiness.

5 Discussion

This paper explores the factors influencing individual happiness using a Bayesian logistic regression model to analyze self-reported happiness levels. By leveraging a carefully selected subset of socio-economic and demographic variables from a comprehensive dataset, this study identifies patterns that highlight key determinants of happiness and their relative importance in shaping well-being outcomes.

5.1 Detailed Exploration of Factors Influencing Happiness

This study carefully selects key variables—marital status, job satisfaction, education level, real income, age, and number of children—based on their strong association with individual happiness. The selection process was guided by both theoretical significance and empirical evidence, ensuring the model accurately reflects real-world influences on well-being. The quantitative analysis demonstrates how these variables play a pivotal role in shaping personal happiness. For instance, marital status and job satisfaction emerge as strong predictors, with married individuals and those who are highly satisfied with their jobs reporting significantly higher levels of happiness.

Interestingly, the analysis reveals a nuanced relationship between the number of children and happiness. While having children can bring joy, larger family sizes are often associated with slightly lower levels of happiness, possibly due to increased financial and caregiving responsibilities. Similarly, higher levels of education and income are positively associated with happiness, illustrating the role of socio-economic stability in improving life satisfaction. These findings not only align with established theories but also provide a deeper understanding of the complex interactions among social, economic, and demographic factors that drive individual well-being.

5.2 Strategic Implications of Variable Selection

By selecting variables with strong associations to individual happiness—such as marital status, job satisfaction, education level, income, age, and number of children—this study underscores the multifaceted nature of well-being. The inclusion of these variables reflects both theoretical relevance and empirical robustness, providing insights into the dynamics of happiness. For example, focusing on the positive impact of marriage and job satisfaction aligns with broader sociological findings about emotional and economic stability.

Interestingly, while other variables, such as detailed occupational categories or regional distinctions, were excluded to streamline the analysis, their potential impact opens avenues for future research. Future investigations could explore how specific professions or geographic contexts influence happiness, shedding light on how cultural and environmental factors interact with socio-economic predictors. This targeted approach not only refines current models but also lays the groundwork for exploring the broader complexities of what drives individual happiness.

5.3 Weaknesses and Future Research Directions

While this study provides valuable insights into the factors influencing individual happiness, its scope is limited by the variables included and the dataset size. The exclusion of variables such as region of residence or specific occupational categories, though aimed at simplifying the model, restricts the depth of analysis. Future research could incorporate these variables to better understand the interplay of geographical, cultural, and professional contexts on happiness.

Moreover, the model could be enhanced by exploring additional dimensions, such as health conditions or leisure time, to provide a more holistic view of well-being. Integrating data from broader samples across different regions or including longitudinal datasets could improve the model's generalizability and offer insights into how happiness evolves over time.

Another avenue for future exploration involves examining the impact of external socio-economic factors, such as economic downturns or public policy changes, on happiness. For example, understanding how financial crises or access to social services influence happiness levels could inform policies aimed at improving societal well-being.

Additionally, the use of advanced machine learning techniques could help uncover hidden patterns and interactions among variables that traditional regression models might miss. For instance, exploring non-linear relationships between age, income, and happiness could yield deeper insights into the life stages most conducive to well-being.

These research directions not only build upon the findings of this study but also provide opportunities to address its limitations, contributing to a more comprehensive understanding of the determinants of happiness in diverse contexts.

5.4 Envisioning the Future of Happiness Studies

The potential for future research based on this dataset is vast. One promising avenue involves investigating the intersection between happiness and broader societal factors, such as access to healthcare or the availability of social support systems. By integrating external datasets, future studies could explore how these contextual elements amplify or mitigate the effects of individual-level predictors like income or job satisfaction.

Another intriguing possibility is examining how happiness evolves across different life stages, using longitudinal data to track changes over time. Such research could illuminate critical periods where interventions might have the greatest impact on well-being.

Additionally, incorporating geographical data could provide a deeper understanding of regional variations in happiness, allowing for an analysis of how cultural or environmental factors shape well-being. The dataset's richness opens the door to these and many other avenues for advancing our understanding of the complex drivers of happiness.

5.5 The Value of Insights into Happiness Drivers

This research highlights the analytical value of understanding the factors that drive individual happiness, such as the significant role of marital status and job satisfaction. These findings can inform policy development and workplace strategies aimed at improving well-being. The positive correlation between higher education levels and happiness also emphasizes the broader societal value of investing in accessible and equitable education systems, reflecting the importance of socio-economic stability in fostering life satisfaction.

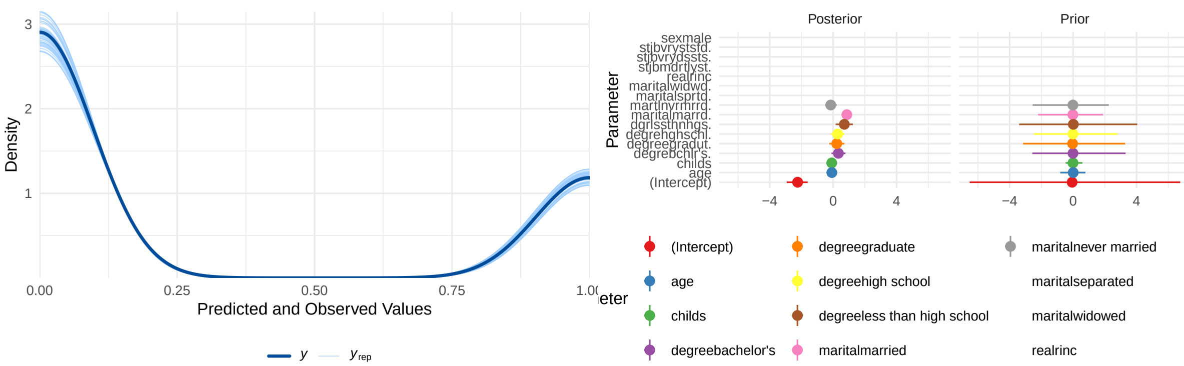
Appendix

A Model details

A.1 Posterior predictive check

In Figure 7a, we conduct a posterior predictive check to assess the fit of the Bayesian logistic regression model. This plot compares the observed data (y) with the replicated data (y_rep) generated by the model. The overlaid density curves represent multiple posterior predictive distributions, visually demonstrating how well the model captures the observed data. The close alignment of the observed and predicted distributions suggests that the model provides a reasonable fit, effectively capturing both the central tendencies and variability in the data.

In Figure 7b, we compare the posterior distributions of the model parameters with their corresponding priors. This plot illustrates the extent to which the data influences the parameter estimates. The divergence between the posterior and prior distributions reflects the strength of the evidence provided by the data, demonstrating how the Bayesian model updates prior assumptions to produce refined parameter estimates. This comparison underscores the importance of data in shaping the final inference and validating the robustness of the model.



(a) Posterior prediction check

(b) Comparing the posterior with the prior

Figure 7: Examining how the happiness model fits the data and how the posterior compares to the prior

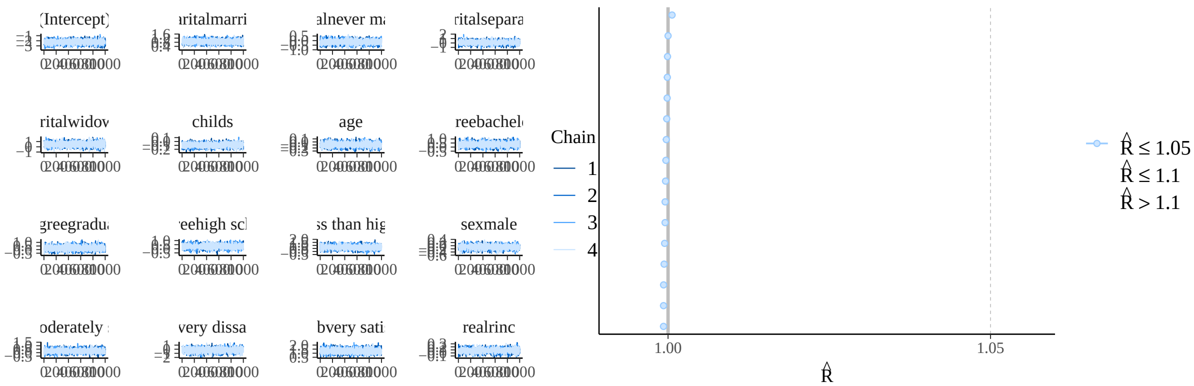
A.2 Diagnostics

Figure 8a is a trace plot, depicting the sampled values of each parameter over iterations. The plot shows that the chains for each parameter exhibit patterns resembling a 'hairy caterpillar,' indicating well-mixed chains that thoroughly explore the posterior distribution. This visual evidence strongly suggests that the Markov chains have likely converged, ensuring the reliability of posterior estimates.

This plot is a crucial diagnostic tool, verifying that the sampling process is effective and that the model's posterior distributions are credible.

As shown in Figure 8b, the Rhat values are close to 1, indicating that the chains have likely converged and the model results are reliable.

Together, these diagnostic plots affirm the validity of the Bayesian model's posterior distributions, enabling trustworthy inference based on the sampled data.



(a) Trace plot

(b) Rhat plot

Figure 8: Checking the convergence of the MCMC algorithm for the happiness model

As shown in Figure 9, the graph visualizes the posterior distributions of the parameters from the Bayesian logistic regression model analyzing happiness predictors. Each horizontal line represents the **90% credible interval**, centered around the median of the posterior distribution for a given parameter. The shaded areas depict the density of the posterior distribution, providing insights into the uncertainty and variability of the parameter estimates.

The parameters encompass socio-economic and demographic factors, such as marital status, job satisfaction, education level, and income. Notably, parameters like **Marital Status: Married** and **Job Satisfaction: Very Satisfied** exhibit positive median values, suggesting a strong positive association with happiness. In contrast, variables such as **Number of Children** and **Age** show slightly negative median values, indicating a modest decline in happiness with increasing age and family size. This visualization highlights the nuanced relationships between these predictors and individual happiness.

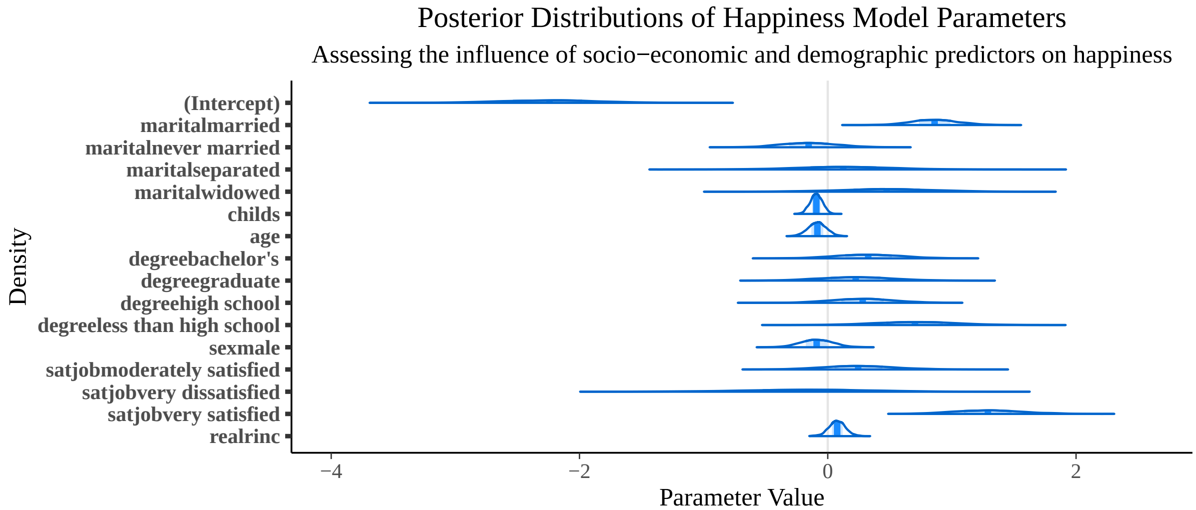


Figure 9: The posterior distributions for all the parameters of the happiness model

B Idealized Happiness Study Methodology and Survey

B.1 Objective

Our goal is to provide a robust analysis of the factors influencing individual happiness by designing a high-quality survey that captures data on key socio-economic and demographic variables. By addressing potential biases and ensuring representative sampling, this methodology aims to produce actionable insights within a manageable budget.

B.1.1 Sampling Approach

We aim for a sample size of approximately 2,500 respondents, selected through **stratified random sampling** to ensure representativeness across demographic groups.

Definition and Justification of Stratified Sampling:

Stratified sampling divides the population into subgroups, or “strata” (e.g., by age, gender, income, education, and marital status). Participants are randomly sampled within each stratum to ensure proportional representation. This method enhances the precision of subgroup estimates and is well-suited for studies exploring the interplay of socio-economic variables and subjective well-being.

Strengths and Weaknesses of Stratified Sampling:

Stratified sampling reduces sampling error, especially in smaller or more variable subgroups such as single parents or older adults. However, it requires accurate population data to establish strata sizes, and survey costs can increase when oversampling underrepresented groups.

Sampling Simulation:

A simulation can be conducted to compare stratified and simple random sampling methods using previously collected data. This would evaluate the reduction in sampling error and the improved accuracy of subgroup estimates, validating the use of stratified sampling for happiness research.

Stratification and Quotas:

Demographic quotas will be weighted based on recent census data to reflect the population. The primary demographic strata include:

- **Age**: 18-29, 30-44, 45-64, 65+
- **Gender**: Male, Female
- **Marital Status**: Married, Never married, Divorced, Separated, Widowed
- **Education Level**: Less than high school, High school, Associate/Junior college, Bachelor's, Graduate
- **Income Range**: Self-reported continuous values, with adjustments for missing responses.

B.1.2 Recruitment Strategy

We will collaborate with survey panel providers to access verified respondents while supplementing with targeted recruitment via social media ads to reach underrepresented groups. Each participant will receive a $3 incentive to encourage engagement.

Panel Recruitment and Potential Biases:

While panel providers offer efficiency, panel bias may arise due to frequent survey participation by the same individuals. To address this, additional recruitment via social media will focus on underrepresented demographics, ensuring a diverse and balanced sample.

B.1.3 Survey Design and Implementation

The survey will be hosted on Google Forms, featuring a professional introduction, clear instructions, and contact details for queries. It will remain open for 2-3 weeks to allow adequate time for responses.

Survey Structure:

1. **Introduction and Consent**: A message explaining the purpose of the survey and ensuring anonymity.

2. **Screening Questions**: Basic eligibility checks (e.g., age and citizenship).

3. **Core Questions**: Questions on socio-economic and demographic variables, happiness levels, and contributing factors.

4. **Closing Message**: A thank-you note and contact information for follow-up.

Questions will be neutrally worded and logically ordered, with an estimated completion time of 5–7 minutes.

B.1.4 Data Validation

To ensure high-quality responses, the following measures will be implemented:

- **Attention Checks**: Include questions designed to identify inattentive respondents.

- **Duplicate Prevention**: Settings to restrict multiple submissions per respondent.

- **Response Time Analysis**: Flag responses completed significantly faster than the average time for review.

- **Post-Survey Weighting**: Adjust responses to align with demographic quotas.

B.1.5 Analytical Framework

To analyze the collected data, a Bayesian logistic regression model will be applied to quantify the influence of variables like marital status, job satisfaction, and income on happiness levels. The model will incorporate:

- **Priors Based on Previous Research**: Using established studies as a baseline for parameters.

- **Hierarchical Structure**: Accounting for variation across demographic groups.
- **Validation Metrics**: Evaluating model performance using posterior predictive checks and accuracy measures.

This comprehensive methodology ensures reliable insights into the complex factors shaping individual happiness while laying the groundwork for future studies on well-being determinants.

B.2 Idealized Survey

1. **What is your marital status?**
   - Married
   - Never married
   - Divorced
   - Separated
   - Widowed
2. **How many children do you have?**
   - 0
   - 1
   - 2
   - 3
   - 4
   - 5
   - 6
   - 7
   - 8 or more
3. **What is your age?**
   - Please enter your age in years: ___________
4. **What is your gender?**
   - Male

- Female

5. **What is the highest level of education you have completed?**
    - Less than high school
    - High school
    - Associate or junior college
    - Bachelor's degree
    - Graduate degree
6. **What is your current job satisfaction level?**
    - Very satisfied
    - Moderately satisfied
    - A little dissatisfied
    - Very dissatisfied
    - .i: Inapplicable
    - .d: Do not know/Cannot choose
7. **What is your annual income (in USD)?**
    - Please enter your income: ___________

    (If you do not have an income, please enter -100.)
8. **How would you rate your happiness level?**
    - Very happy
    - Pretty happy
    - Not too happy
9. **To ensure data quality, please select "Pretty happy" for this question.**
    - Very happy
    - Pretty happy
    - Not too happy
10. **How would you describe your overall job satisfaction's impact on your happiness?**

- Strongly positive
- Moderately positive
- Neutral
- Slightly negative
- Strongly negative

References

Allison Horst, Kristen Gorman, Alison Hill. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. https://allisonhorst.github.io/palmerpenguins/.

Andrew Gelman, Ben Goodrich. 2023. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm/.

Bürkner, Paul-Christian. 2023. *Brms: Bayesian Regression Models Using 'Stan'*. https://cran.r-project.org/web/packages/brms/.

David Robinson, Max Kuhn. 2023. *Broom: Convert Statistical Analysis Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Foundation, Apache Software. 2023. *Apache Arrow: Columnar in-Memory Analytics*. https://arrow.apache.org.

Gabry, Jonah et al. 2024. *Bayesplot: Plotting for Bayesian Models*. https://mc-stan.org/bayesplot/.

"GSS Data Explorer." 2023. https://gssdataexplorer.norc.org/MyGSS.

Hadley Wickham, Romain Francois. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

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

Wickham, Hadley. 2023a. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://ggplot2.tidyverse.org.

———. 2023b. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

———. 2023c. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

———. 2023d. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.