

The Way to Happiness: Higher Socioeconomic Status?

Tina Qian, Yanlin Li

Introduction: Motivation and Exploratory Data Analysis

Happiness has long been one of people’s most important aspires, and the relationship between socioeconomic status (SES) and happiness has become a focal point of multidisciplinary research on individual well-being and societal cohesion. Socioeconomic status, encompassing elements such as income, education, and occupation stands as a pivotal determinant in shaping the quality of life for individuals. Understanding the complexities of this relationship holds significant implications for policymakers, educators, and social scientists alike. While conventional wisdom suggests that higher SES often translates to greater happiness, we are curious to confirm this assumption with data-backed evidence.

We have come across the dataset labeled “happiness,” provided by Wooldridge Data, which enables us to explore the correlation between various aspects of socioeconomic status and an individual’s self-assessed happiness.¹ This dataset, comprising 33 variables, contains demographic details of 17,137 individuals. Key variables include “happy” (indicating an individual’s overall happiness), “workstat” (employment status), “income” (total family income), “educ” (highest level of education attained), “female” (1 if female), “babies” (household members under 6 years old), “preteens” (household members aged 6 to 12), and “teens” (household members aged 13 to 17).

Our research question in this instance is whether individuals with higher socioeconomic standing tend to report greater levels of happiness. Our hypothesis is a positive correlation; more specifically, an employed male with high household income, high educational level and fewer young members at home is anticipated to be happier.

Concerning our response variable “happy,” we assigned numerical rankings of “1,” “2,” and “3” to its three categories—“not too happy,” “pretty happy,” and “very happy,” respectively—to facilitate analysis and model fitting.

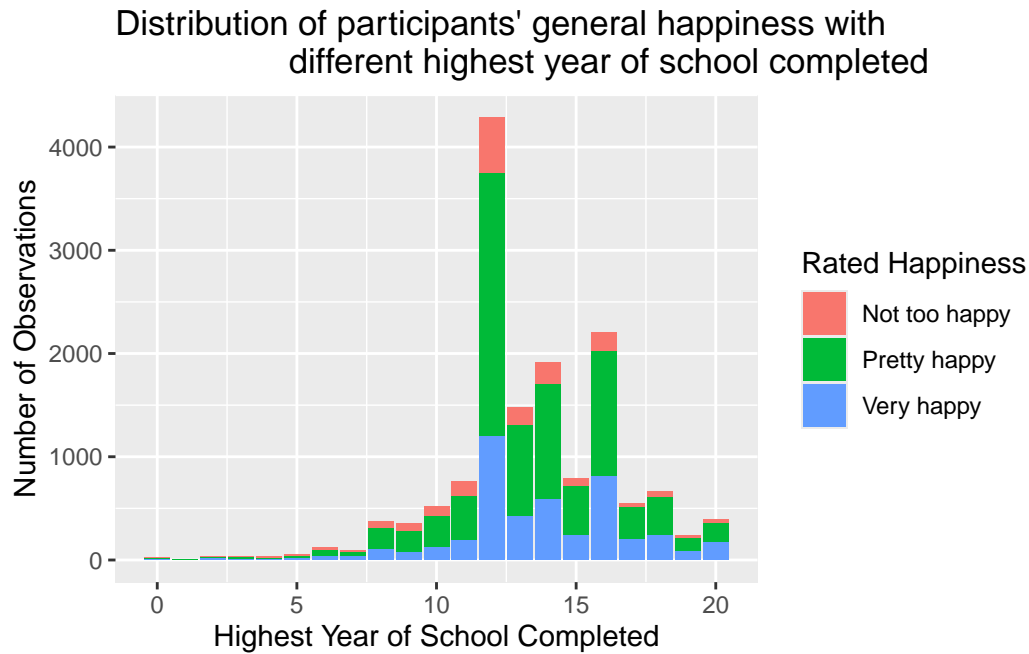
To create the predictor variable that records the number of household members under 17, which was not originally included in our dataset, we aggregated data from three existing variables:

¹<https://vincentarelbundock.github.io/Rdatasets/doc/wooldridge/happiness.html>

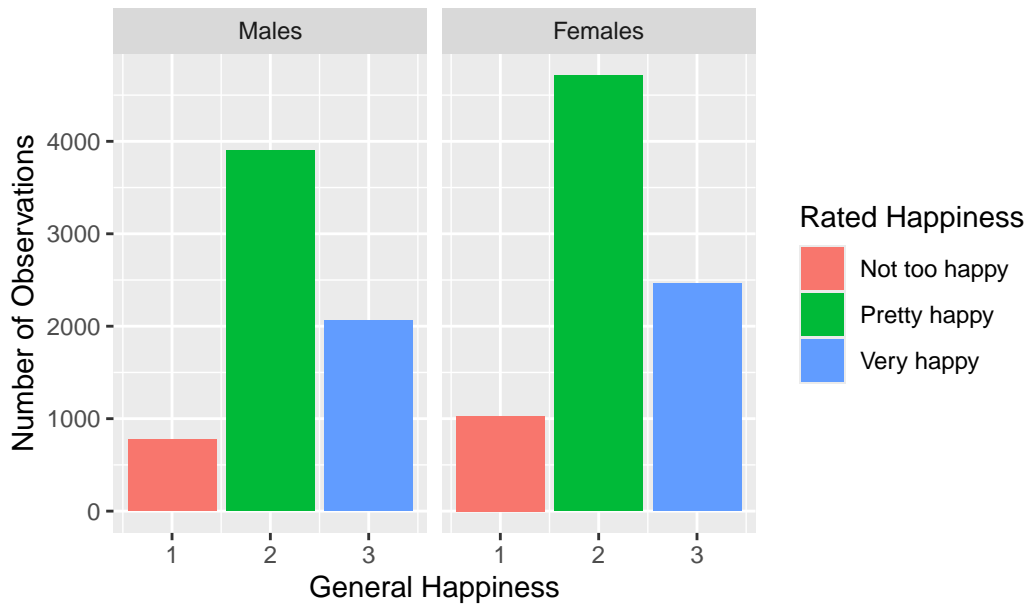
the number of babies, preteens, and teenagers in the household. We believe it's more logical to consider these age groups collectively since each requires financial support and additional care from older members of the household, given that they are not yet fully independent.

Given the N/As across “educ”, “workstat”, “income” and the “young_members” variable in total is approximately 1/3 of all observations, using imputation to generate missing data might not be the optimal choice in this context. We thus filtered the NAs across each column and generated data visualization and regression on the existing data.

We generated visualizations to explore the relationships between the predictor variables and the response variable. Since most of our variables, including the response variable, are categorical, we opted for bar plots for all our exploratory data analyses.

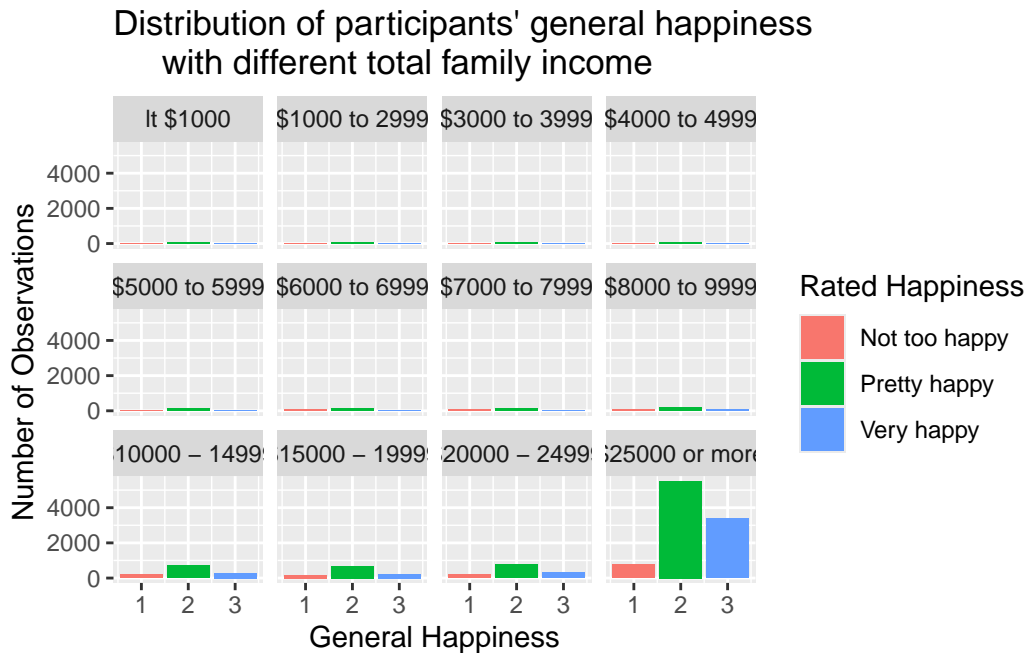


Distribution of participants' general happiness for both gender



Distribution of participants' general happiness with different work force status, adjusted for gender





The first bar plot examines the distribution of observations across different levels of highest year of school completed in the sample, accounting for their overall happiness. The majority of participants in the sample have completed 12-16 years of education, roughly equivalent to middle school and a high school diploma. Across almost every education level, more individ-

uals selected “pretty happy” than “very happy” or “not too happy,” possibly influenced by a tendency to prioritize these responses in general.

The second plot analyzes the distribution of observations for males and females separately, considering their overall happiness. Overall, participants tend to choose “pretty happy” more frequently than “very happy” or “not too happy,” and there are more female participants in the dataset than male.

The third plot examines the distribution of observations across different work force statuses, factoring in gender and overall happiness. Most participants in the dataset are employed full-time, with only a small proportion in school, temporarily unemployed, or not working. More female participants are homemakers or work part-time compared to male participants, while gender distribution is relatively even across other work force statuses. The overall pattern of happiness preference remains consistent across genders.

The fourth plot illustrates the distribution of observations across different household income levels, considering overall happiness. The majority of participants have a household income above \$10,000, especially those with incomes exceeding \$25,000, who strongly favor “pretty happy” and “very happy” over “not too happy.” The general pattern of happiness preference persists across all income levels.

The fifth plot displays the distribution of observations for participants with different numbers of young household members, adjusting for overall happiness. Most participants do not have anyone under 17 in their household, while those who do typically have one or two such members. The overall pattern of happiness preference remains consistent across these groups.

Finally, we decided on predicting whether people with higher socioeconomic status are happier with these predictors: work force status (“workstat”), total family income (“income”), highest year of school completed (“educ”), gender (“female”), and the number of household members under 17 (“young_members”).

Methodology: Interaction Assessment and Model Selection

Among our predictors, we examined two possible interactions: the interaction between work force status and gender, and that between household income and the count of household members under 17. Given the societal expectation for women to bear children and the stereotype of women primarily responsible for childcare, we anticipate a stronger correlation between being female and working part-time. Additionally, we anticipate that households with higher incomes will generally be able to support more members under 17, necessitating greater economic assistance from other family members.

Through an anova-f test, we conducted a comparison between the original model and an interaction model (involving both work force status and gender) to evaluate whether the interaction term, on average, performs better than alternatives, as opposed to the null hypothesis suggesting no interaction between work status and gender. This analysis aims to identify

discrepancies resulting from the interaction term between the initial additive model and the interaction model. A significant F-test for the interaction term(s) signifies that the model incorporating interaction effects offers a more suitable representation of the data compared to the model lacking interaction effects.

F test: (all coef associated with the interaction term are equal to 0).

H0: there is no interaction between working status and whether one is woman.

Ha: there is an interaction between working status and whether one is woman.

The p-value for the interaction term involving full-time work status and gender is 0.0003, indicating statistical significance as it falls below the conventional threshold of 0.05. This provides ample evidence to reject the null hypothesis. Consequently, we conclude that the association between happiness and full-time employment status varies based on the gender of the respondent.

We constructed a second model to assess additional interaction effects by incorporating another interaction term, this time between household income and the presence of young members, alongside the existing interaction model. Our findings revealed significant results for both F-tests. Specifically, the F-statistic value of 0.033 for the interaction term involving household income (measured in dollars per year) and the number of children (young_members) across all levels falls below the significance level of 0.05. This provides sufficient evidence to reject the null hypothesis. Consequently, we infer that the association between the general happiness score and household income (measured in dollars per year) is contingent upon the number of children within the household.

In our study, we contemplated employing both multinomial regression and ordinal regression models for the categorical response variable. Ultimately, we opted for the latter. The decision was straightforward, as the response variable “happy” exhibits an ordinal nature, with discernible gradations from “not too happy” to “pretty happy” to “very happy.”

After assessing and confirming all the above conditions, we went into fitting the ordinal regression model.

Results: Effects of Education, Gender-Adjusted Work Status, and Household Condition

| Working Statuses | Exponentiated coefficients | Gender Interactions | Exponentiated coefficients | Household Income | Exponentiated coefficients | Young Members Interactions | Exponentiated coefficients |
|-------------------------|----------------------------|----------------------------------|----------------------------|------------------|----------------------------|---------------------------------|----------------------------|
| Retired | 1.7024114 | Female | 1.418911 | \$1000 to 2999 | 1.0217438 | Young Members | 1.226093 |
| Working Full-time | 1.1200057 | Female & Unemployed | 1.0975054 | \$3000 to 3999 | 0.8965579 | Young Members & \$1000 to 2999 | 0.7860739 |
| School | 1.0344362 | Female & Working Part-time | 1.0761273 | \$4000 to 4999 | 1.1695977 | Young Members & \$3000 to 3999 | 0.7663593 |
| Working Part-time | 0.8146048 | Female & School | 0.9620414 | \$5000 to 5999 | 1.5157996 | Young Members & \$4000 to 4999 | 0.6559197 |
| Temporarily Not Working | 0.7772623 | Female & Temporarily Not Working | 0.9461765 | \$6000 to 6999 | 0.8439726 | Young Members & \$5000 to 5999 | 0.7104273 |
| Other | 0.5932338 | Female & Other | 0.8244014 | \$7000 to 7999 | 0.9211934 | Young Members & \$6000 to 6999 | 0.5911864 |
| Unemployed | 0.3977148 | Female & Working Full-time | 0.6489594 | \$8000 to 9999 | 0.9260782 | Young Members & \$7000 to 7999 | 0.6216896 |
| | | Female & Retired | 0.6095667 | \$10000 - 14999 | 1.3945867 | Young Members & \$8000 to 9999 | 0.8870423 |
| | | | | \$15000 - 19999 | 1.4007407 | Young Members & \$10000 - 14999 | 0.8236165 |
| | | | | \$20000 - 24999 | 1.7116136 | Young Members & \$15000 - 19999 | 0.8344152 |
| Other Factors | Exponentiated coefficients | | | \$25000 or more | 2.5982148 | Young Members & \$20000 - 24999 | 0.7486186 |
| Education | 1.0445367 | | | | | Young Members & \$25000 or more | 0.8509345 |

Education:

A person who has one more year of education is predicted to have 1.04 times the odds of being in the next higher score of happiness category compared to a person who has less years of education, while adjusting for other variables. Generally, having an additional year of education indicates a higher level of happiness.

Gender-Adjusted Work Status:

After adjusting for factors such as years of education, household income, number of children, and gender, individuals who are retired are expected to have 1.702 times higher odds of achieving a higher level of happiness compared to those who are homemakers. Similarly, full-time workers are anticipated to have 1.120 times higher odds of being in a higher happiness category compared to homemakers. Conversely, individuals in other employment statuses are predicted to have lower odds of reaching a higher happiness category than homemakers.

A woman is predicted to have 1.418 times the odds of being in the next higher score of happiness category compared to a man, while adjusting for all other variables. However, when accounting for the interaction between gender and employment status, it is found that the happiness levels of women, although diverging from the overall pattern, are still influenced by their work status.

When adjusting for education, income, and number of children, and compared to homemaking women, unemployed women have 1.098 times the odds of being in a higher happiness category, while part-time working women have 1.076 times higher odds. Conversely, women in other

employment statuses, including full-time work, temporary unemployment, being in school, or retirement, are predicted to have lower odds of achieving the next higher happiness level.

The interaction analysis reveals that retired and full-time working women do not necessarily experience higher happiness scores. This suggests that women who work often encounter challenges in both professional and domestic spheres, facing discrimination and microaggressions in the workplace while bearing additional domestic responsibilities at home. These factors may contribute to lower happiness scores compared to women who primarily focus on home-making.

Household Condition:

When considering various factors influencing an individual's happiness score, household income emerges as the primary determinant. Compared to individuals in households with a \$1000 total income, those with incomes exceeding \$10000 exhibit more than a 1-fold increase in the odds of reaching the next higher happiness category, after adjusting for other variables. Specifically, individuals in households with \$25000 or more in total income are predicted to have 2.598 times the odds of attaining the next higher happiness category.

However, individuals with household incomes ranging from \$1000 to \$2999 and \$4000 to \$5999 also demonstrate odds greater than 1 for reaching the next higher happiness category. This phenomenon may be attributed to class distinctions, which significantly influence perceptions of happiness within one's own socioeconomic stratum. Individuals on the cusp of upward mobility often face heightened challenges despite having higher household incomes relative to their peers.

While an additional child in the household generally corresponds to increased happiness, happiness scores tend to decline as the number of dependents rises across all income categories. The notable correlation between higher household income and larger family size offers a potential explanation.

Considering the interaction between the number of young dependents and household income, while accounting for education, gender, and employment status, individuals with household incomes above \$1000 are anticipated to experience reduced probabilities of being in the next higher happiness category for each additional young dependent, compared to those with a \$1000 income. This phenomenon suggests that as the number of young dependents increases alongside growing socioeconomic responsibilities, individuals with higher incomes also exhibit greater likelihood of achieving higher happiness scores.

Given that households with more children tend to have higher total incomes, the observed general positive correlation between happiness and increased young dependents in the household may be attributed to the influence of higher income rather than solely to the presence of additional children.

Discussion

In summary, the overarching conclusion is that employed females with higher household incomes, advanced educational attainment, and a greater number of children at home tend to report higher levels of happiness. However, two unexpected findings emerged: firstly, that females exhibit greater happiness compared to males, contrary to the assumption that males typically possess higher socioeconomic status and thus should be happier. This discrepancy may be attributed to societal and economic pressures placed on males, who are traditionally viewed as primarily responsible for household financial well-being. Secondly, the positive correlation between the presence of young household members and happiness may be explained by the emotional support and warmth provided by children to adults.

Some limitations in this research include: the potentially problematic strategy on dropping the N/A data (as some missing data may be due to lower SES status), the sampling bias caused by self-reported happiness, and potential missing interactions. In addition to addressing the limitations in our research, future studies in this field can concentrate on investigating the underlying reasons behind the two unexpected findings, evaluating the validity of the explanations we have proposed, and exploring additional factors that may influence the relationship between socioeconomic status and happiness.