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:

---

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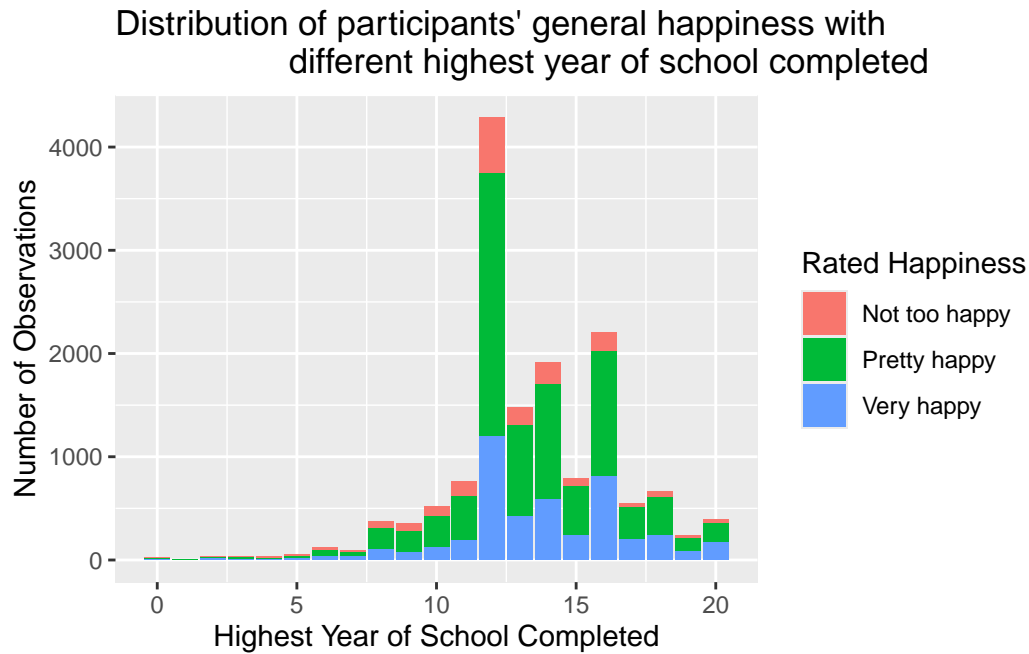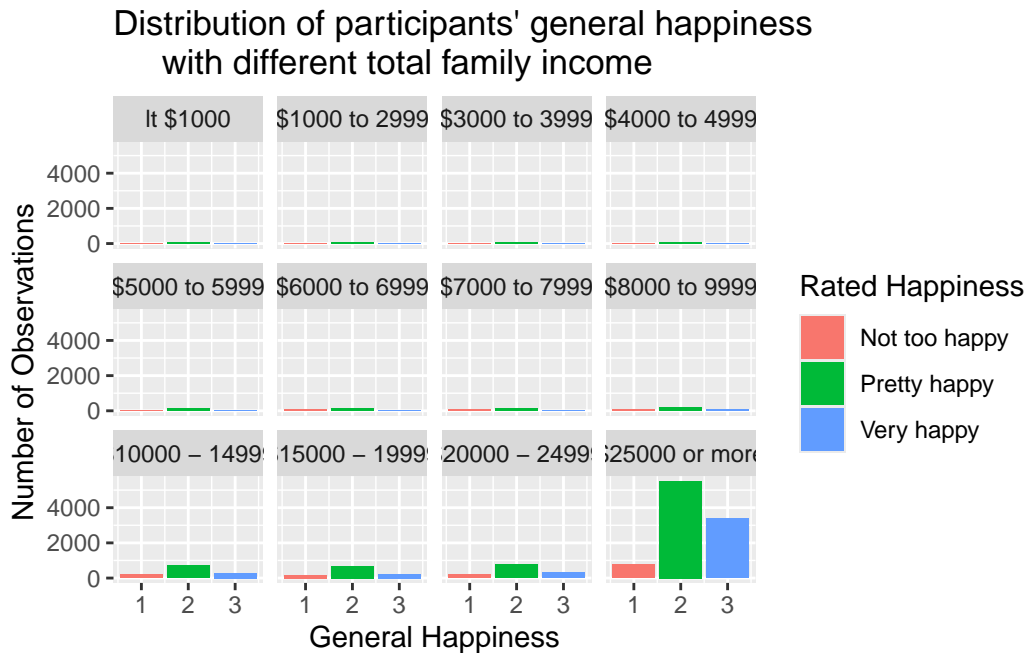
Distribution of participants' general happiness for both gender
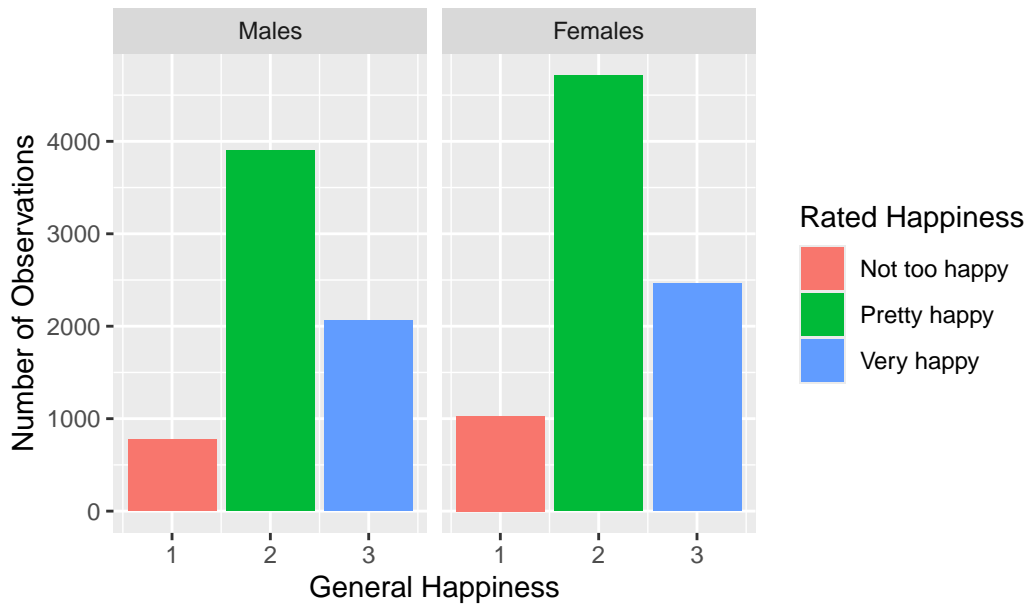
Distribution of participants' general happiness with different work force status, adjusted for gender

Distribution of participants' general happiness with different total family income

Distribution of participants' general happiness with different number of household members under 17

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 fest, we compared the original model with an interaction model (interaction between work force status and gender) to assess whether the interaction term on average is superior to others versus the null hypothesis that there is no interaction between workstat and woman. This test is performed to detect the difference caused by the interaction term between

the original additive model and the interaction model. A significant F-test for the interaction term(s) indicates that the model including interaction effects provides a better fit to the data than the model without 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 interaction term between full-time working status and female is 0.0003, which is lower than a significance level of 0.05 and statistically significant. We have sufficient evidence to reject the null hypothesis, and thus the relationship between happiness and whether one is working full-time depends on whether the respondent is female.

We created a second model comparing another interaction (between household income and young members) on top of the first interaction model to detect any potential interaction effects. Our results showed that both f tests are significant, The f statistic value 0.033 for interaction term between household income (dollars per year) and number of children (young_members) across all levels are lower than the significance level 0.05. We have sufficient evidence to reject the null hypothesis, and thus the relationship between the score of general happiness and household income (dollars per year) depends on the number of children in 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

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

Based on the outcomes derived from the regression model, we have formulated interpretations corresponding to each of the predictor variables.

workstat:
According to the ordinal regression model, a person who retired is predicted to have 1.213 times the odds of being in the next higher score of happiness category compared to a person who keeps house, while adjusting for years of education, household income, the number of children they have, and whether being a woman. In the work force status categories, people with all other status are predicted to have lower possibilities of being in the next higher score of happiness category compared to those who keep houses, while adjusting for years of education, household income, the number of children in their household, and whether being a woman.

educ:
A person who has one more year of education is predicted to have 1.05 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 work force status, household income, the number of children in their household, and whether being a woman.

young_members:
A person who has one more number of children in their household is predicted to have 1.021 times the odds of being in the next higher score of happiness category compared to a person who has less numbers of children in their household, while adjusting for work force status, household income, years of education, and whether being a woman.

female:
A person who is a woman is predicted to have 1.007 times the odds of being in the next higher

score of happiness category compared to a person who is a man, while adjusting for years of education, household income, and number of children in their household.

income:
According to the ordinal regression model, a person in a household with \$25000+ total income is predicted to have 2.709 times the odds of being in the next higher score of happiness category compared to a person in a household with \$1000-2999 total income, while adjusting for years of education, work force status, the number of children they have, and whether being a woman. In the total family income categories, people with all other status are predicted to have lower possibilities of being in the next higher score of happiness category compared to those with \$1000-2999 total income, which is the lowest of all, while adjusting for other variables.

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.