Tori Fredell, Kelsey Matsik, Ashleigh Curry, Caroline Wynne, Theresa Trinh
Professor Johnson
DS 3001
February 29, 2024

GSS Project

Summary

Our project posed the question: How does occupational prestige influence marital satisfaction among married individuals in the United States? To investigate this question we looked at the GSS data on factors that contribute to occupational prestige as well as marital satisfaction. For example, the ‘Respondent’s Job Prestige’ was combined with that of their spouse’s job prestige to create a measure of ‘combined occupational prestige’. This variable gave us a subjective measure of how respondents view their occupation and that of their spouse. This variable was cross examined with the respondent’s perceived marital happiness to examine any possible correlation between prestige (and the inferred costs and benefits) and marital happiness. We also examined respondent sex in relation to both variables individually as well in concert. Our analysis showed that for respondents who reported being “Very Happy” in their marriage, occupational prestige had a uniform distribution. This trend carried over even when we differentiated respondents by sex. Across our analysis we saw increased densities of “Not Too Happy” at perceived occupational prestige levels of approximately 38, 48, and 62. “Pretty Happy” saw similar increased densities at these ratings, but to a lesser extent. Overall our results indicate that people in happy marriages are happy regardless of their occupational prestige or sex.

Data

In the data wrangling phase of our study, we initially confronted data type warnings, which indicated the presence of mixed types in several columns. To address this, we changed all numerical data from objects to the float64 data type for the sake of computational efficiency. Then, to improve the clarity of the dataset, we revised column names wherever necessary. We also identified and discarded two columns that redundantly contained column names for each value, since they contributed no substantive information.

A key part of our wrangling involved comparing the 'respondent_prestige' column with the 'prestg105plus' column to examine any disparities. This analysis revealed that 'prestg105plus' did not offer any unique entries beyond what could already be found in the 'respondent_prestige' column, prompting us to consider only the latter column. This approach was similarly applied to the 'income16' and 'income' columns, as well as between 'prestige' and 'respondent_prestige', thereby eliminating unnecessary duplication and streamlining the dataset. In the case of ‘prestige’ and ‘respondent_prestige,’ the former contained many more NaN values, but we found that it contained some values for which ‘respondent_prestige,’ the significantly more populated column, had NaN. Therefore, for any rows where ‘respondent_prestige’ was empty but ‘prestige’ was not, we added in the ‘prestige’ values and created a new column, ‘combined_respondent_prestige.’

Our examination of the codebook indicated that missing values for the 'hapmar' variable were predominantly attributable to responses deemed 'not applicable' rather than outright 'no answer' responses or instances where the question was 'skipped on the web'. Therefore, we made the informed decision to exclude rows with NaN values for this variable.

Lastly, we removed several columns that were found to be too sparse for us to gain meaningful insights from or that were eventually deemed irrelevant when considered in the context of the study's objectives. This step was done in order to focus our analysis on only the most pertinent or helpful data. Because of all of these steps, the dataset is clean, coherent and ready for visualization and analysis. The data dictionary below provides the full list of variable names and their descriptions.

Data Dictionary

Each variable below corresponds to the GSS respondent's answer, unless indicated otherwise.

Variable	**Variable Type (Numeric/Categorical)**	**Description**
age	Numeric	Age
sex	Categorical	Sex (male, female)
educ	Numeric	Highest year of school completed (1-20 years)
marital_status	Categorical	Marital status (married, never married, divorced, widowed, separated)
work_status	Categorical	Work status
occupation	Categorical	Occupation title
combined_respondent_prestige	Numeric	Occupation title prestige, ranked by the respondent (standardized scale of 0-100)
respondent_hours_worked	Numeric	Number of hours worked last week
income	Numeric	Total family income
spouse_occupation	Categorical	Spouse's occupation title
spouse_prestige	Numeric	Spouse's occupation title prestige, ranked by the respondent (standardized scale of 0-100)
spouse_hours_worked	Numeric	Number of hours worked last week by respondent's spouse
marriage_happiness	Categorical	Perceived happiness in their marriage

Results

Summary statistics of our data were obtained to describe the overall dataset. Additional summary statistics were obtained for each of our variables of interest: combined respondent prestige, marriage happiness, and sex.

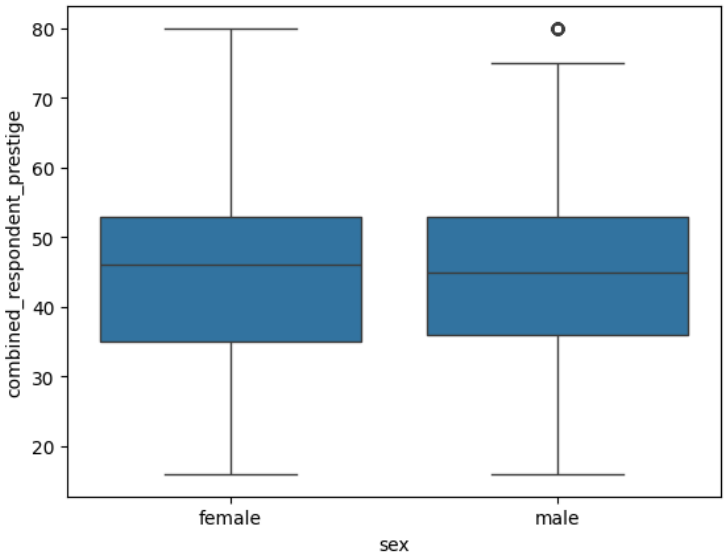
	age	educ	combined_respondent_prestige	respondent_hours_worked	income	spouse_prestige	spouse_hours_worked
count	14015.000000	14095.000000	14122.000000	14122.000000	13184.000000	14122.000000	14122.000000
mean	42.086621	13.884285	45.688996	41.161025	22984.412925	45.344640	41.077397
std	11.431608	2.877375	12.987408	13.966894	4602.611045	12.995856	12.953263
min	18.000000	0.000000	16.000000	0.000000	1000.000000	16.000000	0.000000
25%	33.000000	12.000000	35.000000	37.000000	25000.000000	35.000000	40.000000
50%	41.000000	14.000000	45.000000	40.000000	25000.000000	45.000000	40.000000
75%	51.000000	16.000000	53.000000	48.000000	25000.000000	53.000000	47.000000
max	89.000000	20.000000	80.000000	89.000000	25000.000000	80.000000	89.000000

Table 1: Summary Statistics of Dataset

Our sample pool was 52% women and 48% men. Of these, approximately 63% of respondents said they were ‘Very Happy’ in their marriage, 34% reported as “Pretty Happy”, with the remaining 3% ‘Unhappy’. ‘Combined Respondent Prestige’ ranged from 16 to 80, with the majority of ratings centering between 35 to 64. Rating with the highest occurrence was 35 followed by 47, at about .5% of the total allocated scores. Descriptive statistics were obtained for the variable ‘Combined Respondent Prestige’ grouped by marriage happiness and sex. From the table, conclusions regarding the relationship between occupational prestige and marital satisfaction can be drawn, as well as when stratified by sex. A boxplot was also utilized to visualize the distribution of combined respondent prestige between the two categories of sex (“female” and “male”) in order to observe the central tendency, spread, and possible outliers. From the boxplot, potential differences in combined respondent prestige across gender can be identified. As seen in Boxplot A below, men and women had self-rated them and their partners as similar levels of combined respondent prestige at a score of around 45. This was confirmed by our summary tables which produced a mean of 45.7. This plot indicates that one gender isn’t more likely than the other to inflate or deflate prestige scores.

		count	mean	std	min	25%	50%	75%	max
marriage_happiness	**sex**								
not too happy	**female**	252.0	43.115079	12.990598	16.0	32.0	45.0	50.00	80.0
	male	140.0	44.600000	13.201155	17.0	35.0	43.0	50.25	80.0
pretty happy	**female**	2572.0	44.821151	12.940995	16.0	35.0	46.0	53.00	80.0
	male	2257.0	44.802836	12.687936	16.0	35.0	44.0	51.00	80.0
very happy	**female**	4557.0	45.770682	12.845090	16.0	36.0	47.0	54.00	80.0
	male	4330.0	46.793303	13.204173	17.0	37.0	45.0	56.00	80.0

Table 2: Descriptive Statistics of Occupational Prestige grouped by Marital Satisfaction and Sex

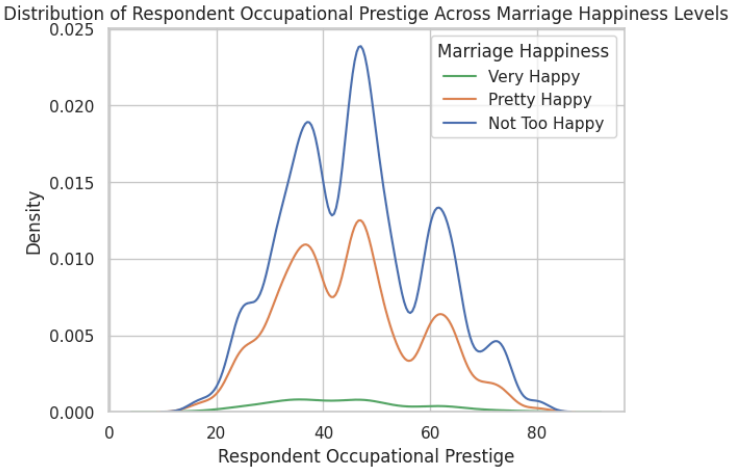


Boxplot A: Combined Respondent Prestige and Sex

In terms of visualizations, a grouped kernel density plot was decided in order to examine how the distribution occupational prestige varied amongst different sexes and levels of marital happiness. The first kernel density plot visualizes occupational prestige scores distributed within each category of marital happiness. This was done in order to see if individuals with higher levels of marital happiness tended to have occupations with higher prestige scores. The second kernel density plot visualizes occupational prestige scores distributed amongst sexes in order to see if there is any difference in distribution between male and females. Finally, a grouped kernel density plot was created, taking into account both marital happiness level and sex. This plot demonstrates how occupational prestige is distributed amongst the demographic factors of both marital happiness level and sex.

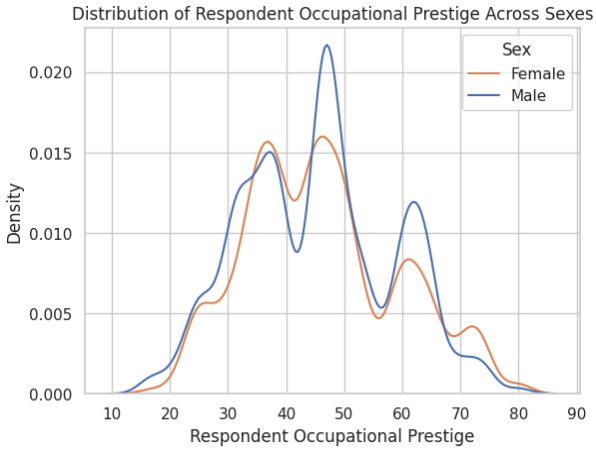
An examination of kernel density plots revealed consistent trends of happiness in relation to occupation prestige regardless of when gender was accounted for. Kernel Density Plot 1 (KDP1) below shows a uniform distribution of occupation prestige across people who reported being “Very Happy” in their marriage. Therefore, this data may indicate that people who are very happy in their marriage are so regardless of their job prestige. The happiness levels of “Pretty Happy” and “Not Too Happy” showed spikes in densities at similar levels of occupational prestige. For both variables, increased numbers of “Pretty Happy” or “Unhappy” marriages seemed to have occurred at perceived occupational prestige levels of approximately 38, 48. 38 and 48 concentrate around our mean prestige level of occupation (45.7), and the areas around them also have higher densities relative to the rest of the plot. Therefore, it appears that couples at below-average to average-levels of prestige are the least happy in their marriage. The graph steeply drops off after our peak around 48, until it peaks again at approximately 62. This would indicate that there is marital discord even at higher levels of occupational prestige. It’s likely that the sources of discontent at a combined prestige of 62 vs. 38 and 48 are quite different. The former likely stem more from time constraints and demand from the job while the latter may be

more likely to be financial in character. Break-down analysis of these variables is an avenue for further exploration to be acknowledged in our conclusion.



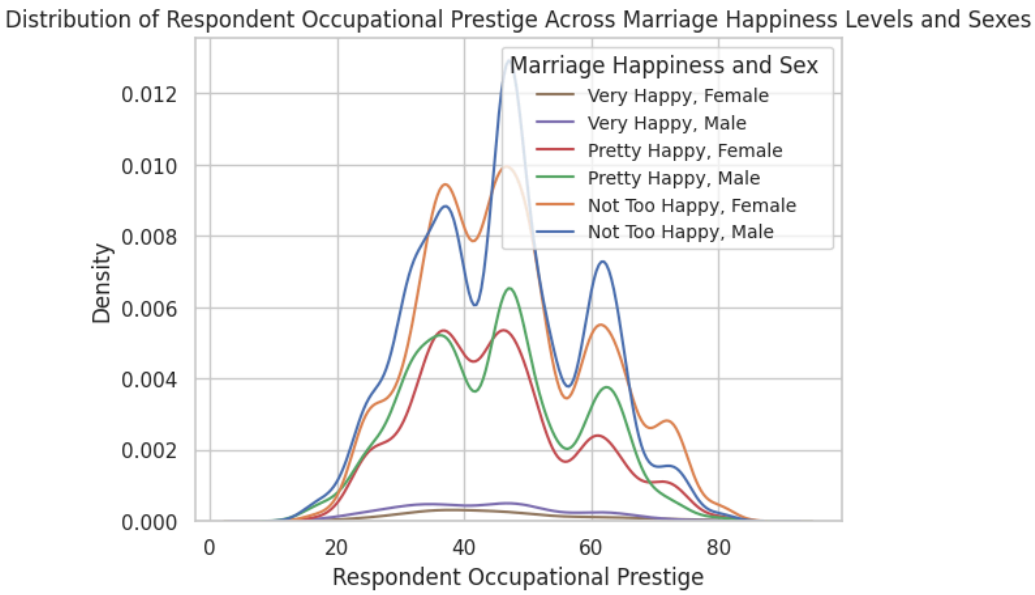
KDP1: Occupational Prestige and Marital Happiness

KDP2 shows us the relative densities of occupational prestige of men and women. The highest concentration of men (.025) have prestige of approximately 47. The highest concentration of women occur at prestige levels of .38 (.016) and .46 (.017). It's interesting to note that at the highest end of the prestige scale, women are found more frequently than men.



KDP2: Occupational Prestige Across Sexes

KDP3 compares marital happiness across occupational prestige levels and sex. When broken down this way we again see that people who were "very happy" in their marriage had a uniform distribution of occupational prestige regardless of respondent sex. Again, "Not Too Happy" in one's marriage had the highest densities, with peaks occurring around in the high 30s and low 40s. This pattern was seen irrespective of sex. The sex and happiness curves followed similar curves, with the biggest disparity seen between males and females in the "Not Too Happy Segment" at an occupational prestige level of approximately 43. Here, the difference in densities was .003. This would indicate that of men who are unhappy with their marriage, many of them are concentrated at this middle-level of relationship occupational prestige.



KDP3: Respondent Occupational Prestige across Happiness and Sexes

Conclusion

Overall our project reaches several probable conclusions. First is that people in happy marriages are likely happy irrespective of their occupational prestige. This would suggest that having a “high status” occupation is by no means a prerequisite for a successful relationship. It’s likely that their marriage serves as a buffer between their job and overall life satisfaction. A second conclusion is that marriage with average levels of combined job prestige have higher levels of unhappiness or moderate happiness then those on the tail ends of the spectrum. We also can infer that spikes in unhappiness that occur at the “more prestigious” end of the spectrum likely have different sources than those at the low-middle end. One caveat to our data and potential criticism is the sample size of the “Not Very Happy” pool. Only 3% of our samples self-selected as this. While potentially a positive reflection on the state of marriages in the United States, it also has the potential to skew our data, as outliers in prestige may have more pull than in the larger sample sets. This may explain why “Pretty Happy” and “Not Very Happy” have similar shapes, but the latter has higher peaks corresponding with higher density. A final conclusion comes from the role of sex. Across happiness levels, men tended to have higher variability than women, with the exception of the prestige level of 38 where women reported unhappiness at higher concentrations than men. This may occur because of the structural unpaid care work placed onto women. This burden is more pronounced at lower socio-economic levels that are likely associated with lower prestige.

Criticisms of our project likely would center on the subjectivity of our central variables. Definitions of marital happiness and occupational prestige vary by individual and are also

subject to the whims of the individual at the time of self-report (ie: had the couple just fought, gone on a date, etc) . There's also no way to account for if people gave honest ratings, or inflated or deflated their scores. People may want to inflate their prestige scores, or people in happier marriages may be more likely to inflate the prestige score of their spouse than those in unhappy marriages. People may give their prestige score based on how they perceive the general populace would see it instead of how they themselves see it or vice versa. However the fact that combined prestige levels were similar between the sexes points to increased reliability in terms of sex-differentiation.. As discussed above, the sample size of the "Not Happy" data set is also a potential criticism as this makes the data subject to more influence by outliers. In addition, most of the sample will concentrate in the middle of the occupational prestige level. This means we don't have the same quantity of data for the lower and upper ends, which again may overrepresent these densities similar to the "not happy" variable.

Further explorations pertaining to our project's research question could explore the contributing factors of "occupational prestige" and "marital satisfaction". Questions such as: 'Does Income Level or Hours Worked Matter More For Marital Happiness?' would help pinpoint more of a mechanism behind the role of occupation in marriage happiness. A further investigation could also look at the relation between prestige and divorce rates. Are "higher prestige" couples more or less likely to get a divorce than those at the lower end of the prestige spectrum? Does being "successful" make you more likely to get a divorce? Further research could also look at the couples on an individual basis. Is the "more prestigious" individual more or less unhappy in their marriage and how does this vary along the prestige spectrum?