Summary

Using data from the General Social Survey, this study delves into the relationships between an individual's education, their perceived social status, and their prestige score. We began by cleaning and organizing the data from 37 separate CSV files, focusing on those key variables such as degree, class, and prestige. We then used visual analytics to uncover insights revealing a strong correlation between educational attainment and prestige, especially when considering an individual's perceived social status.

The dataset was initially complex, which required merging, cleaning, and addressing missing data. Our visualizations, including plots, highlighted trends such as a prevalence of high school completion or lower education levels and a dominance of "working class" and "middle class" social statuses. Notably, those with advanced degrees tended to have higher prestige scores, and the same pattern was observed for individuals in higher social classes.

This project emphasizes the significant societal impact of education on an individual's perceived value and societal standing. While we had a robust methodology, it is important to recognize the potential limitations and biases inherent in survey-based datasets. We suggest future research opportunities, including the incorporation of additional socioeconomic variables, and more advanced statistical analyses to better understand causality in these relationships.

Data : 2 pages -

The data used for this analysis is sourced from the General Social Survey (GSS), containing information on education, prestige, and perceived social class. The dataset comprises a broad range of sociodemographic and attitudinal variables collected from a nationally representative sample of individuals. It replicates these questions yearly– providing a collection of data on social trends overtime.

This wealth of information includes details about individuals' educational backgrounds, their career and job-related aspects, and their subjective assessments of their own social status. The dataset is, therefore, a powerful tool for researchers and social scientists to explore and analyze various social trends and phenomena over time.

This analysis focuses on three variables: *degree*, which is the highest degree obtained by the respondent, *prestige*, which is a score assigned to professions based on a rating system developed by NORC (National Opinion Research Center). Using a nine-step ladder, the prestige scores in the GSS studies were formed by asking respondents what the social standing of various occupations were to them. The last variable involved is class; it is the respondent's self perceived social standings.

For *degree*, before any data cleaning or filtering, the original dataset consisted of a total of 72,230 records, with respondents providing their educational qualifications. Within this dataset, there are six unique options that respondents could select as their highest level of

education: "bachelor's," "less than high school," "high school," "graduate," "associate/junior college," and the presence of missing values ("nan"). Among these options, "high school" emerged as the most frequently chosen degree level. It was selected by 36,446 respondents, and the overall data indicates that most respondents did not continue school after getting a high school diploma.

Prestige level captures the perceived social prestige or status of the respondent's occupation. Before any data cleaning or filtering, this variable contained a total count of 24,303. In this dataset, the "prestige" variable employs a standardized prestige score, calculated as a straightforward mean value of the prestige ratings assigned to various occupation categories. These ratings are then converted to a scale ranging from 0 (representing the lowest prestige level) to 100 (representing the highest prestige level). The "prestige" variable exhibits diversity, encompassing 63 distinct numeric values assigned to the participants' jobs. However, the most prevalent prestige score among the respondents was 50.0. This score was selected by 1,913 individuals, signifying the prominence of this middle-of-the-road prestige level within the dataset. Additionally, it's worth noting that there were instances of missing data ("nan") in this variable.

The self-perceived social class is divided into six distinct categories, each offering a unique perspective on how individuals assess their societal status. These include “middle class,” “working class,” “"upper class,” “lower class, “no class,” and instances of missing data denoted as ("nan”). Prior to any data cleaning or filtering, the *class* variable encompassed a total of 68,894 records. Within these six social class categories, "middle class" stands out as the most prominent. To be specific, it was the most frequently chosen category, with 31,014 respondents identifying themselves as part of the middle class.

During the process of handling and preparing the data, several challenges were encountered. First, dealing with a large dataset required reading it in chunks, making it more manageable. To begin, the dataset's substantial size necessitated a chunked approach to reading the data. This method allowed for more efficient management of the data's scale. While this solved the issue of size, it brought about the challenge of consolidating these fragments into a coherent dataset. To address this, a new dataframe was created, enhancing the organization and overall cohesion of the data.

Another challenge emerged in the form of the "prestige" variable. The lack of comprehensive documentation regarding its meaning and calculation required a deeper dive into past documents and discussions to gain insight. This step was pivotal in fully comprehending the variable's significance within the dataset.

Initially, the consideration was given to replacing missing values with "Unknown" to maintain consistency across the dataset. However, upon further evaluation, a different approach was adopted. It was decided to drop rows with missing values and also exclude any instances where "prestige," "educ," and "class" were included as unique values. This choice was made because these variables were not essential for the analysis and were not part of the options

provided in the CSS codebook for their respective questions. This streamlined the dataset and ensured that only the relevant data points were retained for analysis.

Data consistency was also a vital concern. An inconsistency was detected in the "prestige" variable, where certain records included the term "prestige" as a category. To maintain uniformity, I standardized the variable by eliminating such anomalies

Lastly, to facilitate exploratory data analysis (EDA) and visualization, "no_na" versions of variables were created. These versions excluded records with missing values, yielding cleaner data for EDA, enabling a more focused analysis of patterns and relationships.

Results:

~~Our exploration focuses on the intricate relationships between an individual's educational background, their perceived social status, and their associated prestige score. Utilizing data from the General Social Survey, we embarked on a meticulous data cleaning process, subsequently employing descriptive statistics and visual analytics. Preliminary findings suggest a pronounced correlation between educational attainment and prestige, further nuanced by an individual's perceived social status.~~

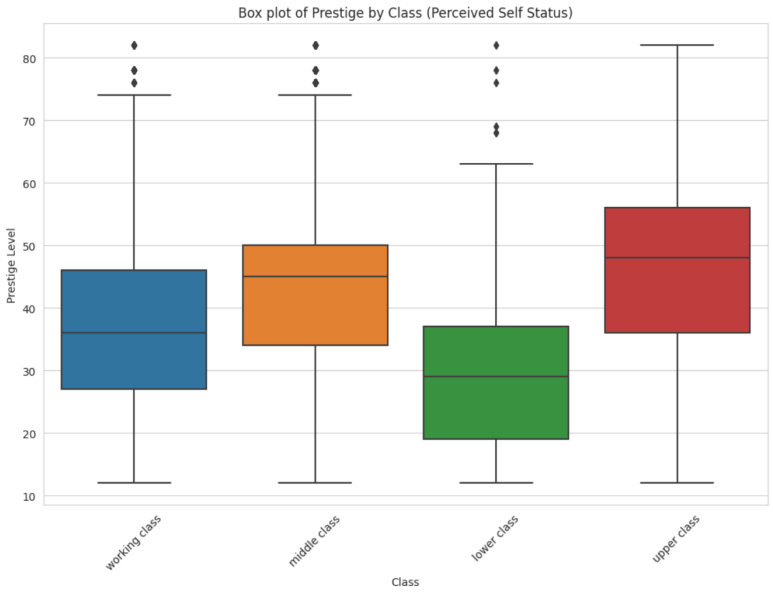
~~Originating from the General Social Survey, our dataset was initially fragmented across 37 discrete CSV files. Our attention converged on three pivotal variables: degree, class, and prestige. The process was twofold: firstly, we amalgamated these disparate files into a cohesive dataset. Next, we undertook a rigorous cleaning process. Columns were systematically renamed for clarity, while rows with incomplete or inconsistent data were addressed. A significant challenge lay in managing missing values and ensuring uniformity across textual entries.~~

Visualizations played a paramount role in our analytical process. Initial plots highlighted the educational distribution among participants, with a significant inclination towards high school completion or lesser education. In terms of perceived social status, the "working class" and "middle class" emerged as predominant. However, the most illuminating insights arose from box plots juxtaposing degree and class with prestige scores. An evident trend suggested individuals with advanced degrees often associated with higher prestige scores. This trend was paralleled in perceived social status, where those identifying as belonging to the upper echelons exhibited elevated prestige.

Our exploration underscores the profound societal implications of education on an individual's perceived value and societal standing. Although our methodologies were robust and meticulous, it's essential to approach survey-based datasets with caution, acknowledging potential inherent limitations or biases. Reflecting upon these findings, there's an evident horizon brimming with opportunities. Integrating more socioeconomic variables, adopting advanced

statistical analyses, and discerning causality beckon as tantalizing avenues for future research endeavors.

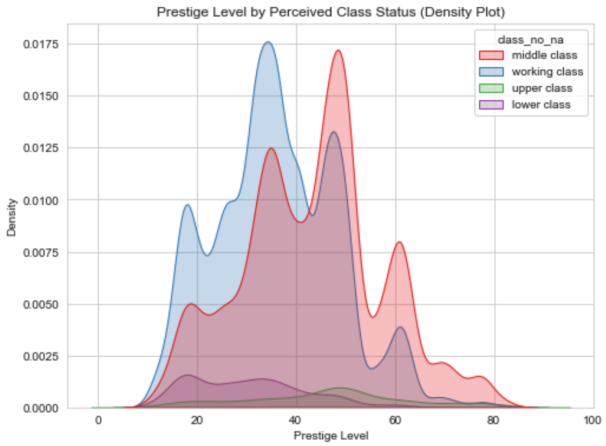
Visualization



The first figure shows the relationship between prestige score and perceived self status. Those who classify themselves as “upper class” have the highest prestige level, with a median of 48. Those who classify themselves as “working class” and “middle class” also have relatively high prestige levels, while those in the “lower class” have lower prestige levels, with a median of 29. While this trend may not be as pronounced as the one related to education levels, there is nonetheless a trend showing people who see themselves in a higher class having higher prestige scores.

There were distinct differences in the median for the various class levels established. For instance, consider the lower class and upper class. There is a significant difference in the medians where the upper class has approximately a prestige level of 50, while lower class has a prestige level of around 30, indicating that there may be a trend between one's socioeconomic status and the prestige level of their occupation.

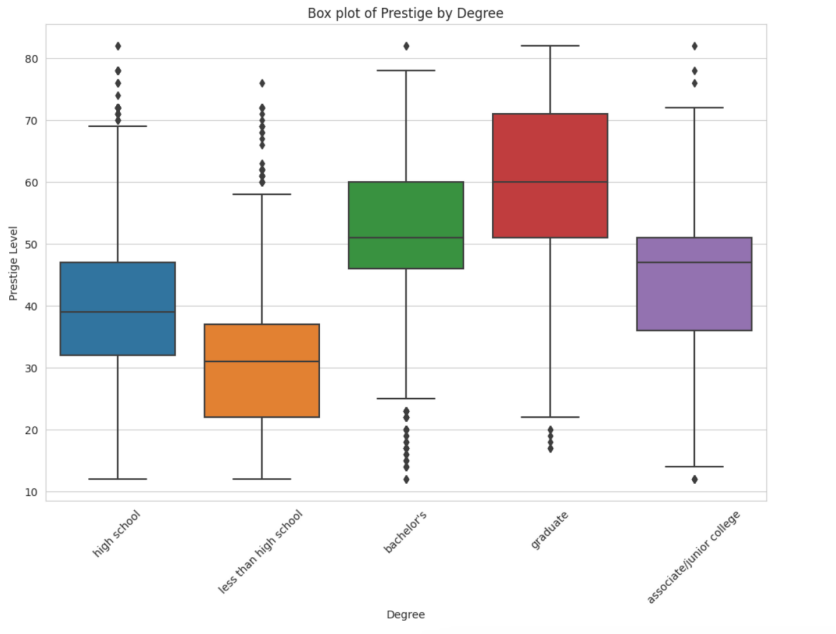
It is interesting to note that the prestige levels (medians) for the upper class and middle class are closely aligned, showcasing a similarity in the prestige of their respective occupations. In contrast, when we compare the maximum values, a more extensive range of prestige scores becomes apparent within the upper class category. The same can be said when comparing lower and working class.



This density plot backs up the claims of our other graphs findings. The highest density areas within each class align with higher prestige levels as the classes increase from lower class up to upper class.

This plot shows that our sample data might not be the best as there are very few data points for the lower and upper class, whereas for the middle and working class there are quite a few entries. This could however, could be attributed to the fact that in the population of the United States the majority of citizens lie within the middle or working class.

The low density specifically within the upper class could be the reason the interquartile range of the upper class in our first box plot was so large. Perhaps with better data or more data points to account for the data could change and give us a better understanding of what the true prestige levels of upper class citizens truly is.



The second figure shows the distribution of prestige based on the highest level of education achieved. There is a clear trend showing those with higher degrees of education having

higher prestige levels. Here, the median prestige level for those with a graduate degree is 60, while the median prestige level for those who did not complete high school is 31. There is a clear trend, from the lowest to highest levels of education, that higher degrees of education correlates with higher prestige levels.

From this graph it is clear that with each level of higher education completed prestige levels climb drastically in some cases. This is especially evident when comparing associate/junior college level to the highschool level. Despite having very similar ranges and interquartiles the average prestige level for those with an associate or junior college degrees is about 8 points higher than those with only a highschool diploma. This is the second highest gap between consecutive grade levels only being surpassed by the gap between a graduate degree and a masters degree. This shows how important any college level degree can be in terms of obtaining higher prestige in one's field of work.

Another interesting thing to note when analyzing the graph is that despite having quite different interquartile ranges the average prestige level between associate/junior degrees and a bachelor's degree were not extremely different. This could show that although level of education is important in determining prestige level, performance once in the real world or job market can definitely play a factor into prestige level as well. If someone with an associates degree were to outperform another person with a bachelor's degree their prestige level could very easily be greater than the person with a bachelor's degree.

Conclusion:

This project aimed to investigate the relationships between an individual's educational background, their perceived social status, and their associated prestige score using data from the General Social Survey (GSS). The dataset, which contained in-depth sociodemographic and attitudinal variables, allowed for a comprehensive exploration of these dynamics over time.

The study focused on three key variables: degree, prestige, and class. Before data cleaning and filtering, the dataset was substantial, consisting of over 70,000 records. This presented several challenges such as data size management, variable understanding, and missing data handling. After a thorough data cleaning, we were able to adequately prepare the dataset for analysis.

The preliminary findings of the analysis unveiled significant correlations between educational attainment, prestige, and perceived social status. Using visualizations, we were able to better highlight these relationships. Notably, individuals with advanced degrees tended to have higher prestige scores, reflecting the societal significance of education on one's perceived value and standing.

It is crucial to acknowledge that this analysis relies heavily on survey data, which inherently comes with limitations. Surveys are subject to response bias and the limitations of self-reporting. While every effort was made to clean and manage the data, there may still be residual issues that could impact the accuracy of the results. The analysis does not establish causation but rather highlights correlations. Additionally, the interpretation of "prestige" and "class" can be subjective, potentially influencing respondents' self-assessments.

Looking beyond the current project, there are several avenues for future research:

- Integration of More Socioeconomic Variables: Expanding the dataset to include additional socioeconomic variables could provide a more comprehensive understanding of how factors like income, race, or geographical location interact with education, prestige, and social status. This could offer insights into the intersectionality of these factors.
- Qualitative Analysis: Complementing quantitative data with qualitative insights, such as interviews or surveys with respondents, could provide a deeper understanding of the reasons behind the observed trends. It could also help explore the nuances of how individuals perceive their social status and the role education plays in this perception.
- Longitudinal Analysis: The GSS provides data over time. Conducting a longitudinal analysis to track changes in educational trends, social status, and prestige scores over the years could reveal evolving patterns and societal shifts.

In conclusion, this project sheds light on the complex relationships between education, prestige, and perceived social status. Despite its limitations, it serves as a valuable starting point for further exploration in the field of social sciences. The findings underscore the importance of education in shaping individuals' perceptions of their social status and emphasize the need for continued research in this domain to unravel the multifaceted interplay of sociodemographic factors.