Flavien Moise, Tenzin Nargee, Charlie Payne, Duncan Tanner

03.12.2024

Professor Johnson

DS 3001

GSS Project Findings

Summary:

The project addresses the relationship between three primary data points: "RLOOKS," a numerical variable which evaluates the surveyor-reported physical attractiveness of the surveyee, "OWNGUN," a binary variable which evaluates whether or not the surveyee owns a gun a or not, and "PRESTG10," a categorical variable which reports the surveyee's self reported job prestige. Our approach to analyzing relationships involved a series of three graphs after cleaning the data cleaning: "Distribution of Job Prestige by Physical Attractive," "Distribution of Job Prestige by Gun Ownership," and "Distribution of Physical Attractiveness by Gun Ownership." The largest correlative effects we observed existed between physical attractiveness and job prestige, and between physical attractiveness and gun ownership. We also observed a similar distribution of kernel density between job prestige and gun ownership.

Data:

The first variable we looked at in this analysis was "RLOOKS." The surveyor of the variable answered the question "How physically attractive is the respondent?" on a 1-5 scale from "very

unattractive" to "very attractive." This variable presented the most challenges for two reasons. First being the nature of the variable required the survey to be conducted face-to-face rather than via web self-administered questionnaire. This eliminated almost half of the total respondents from being evaluated for this variable, only 1,672 out of 3,544, with "not applicable" accounting for 1,864 surveyees and "don't know" accounting for the remainder, significantly limiting the variable's sample size. The second issue with the "RLOOKS" variable is the subjective nature of the evaluation, which presents potential inconsistencies for its analysis.

The second variable in question was "OWNGUN," which assessed the question "Do you happen to have in your home (or garage) any guns or revolvers?" on a scale from 1-3 with 1 being "YES," 2 being "NO", and 3 being "REFUSED." The survey further evaluates what type of gun the surveyee owned, but we chose to focus on the broader evaluation. The main challenge with this variable involved converting the variable from numeric to a boolean true or false, to aid in graphical representation and digestibility, which we accomplished via the "map" function. Like "RLOOKS," there was a considerable amount of null values that had to be dropped.

The last variable was "PRESTG10" which self-assessed the prestige of the respondents occupation. Responses to the variable were measured on a scale from 10-89, and were collapsed and presented in value counts for 10-point bins starting with 10-19. Only 6.2% of responses resulted in null values.

We followed similar steps in the cleaning process for each of the three variables. We first used "head" and "describe" functions to get a general sense of the data set and relevant available data. We used "value_counts" to observe numbers of each recurring observation and to achieve a general idea of

how potential graphs would look. For dealing with missing values we used a combination of ".isnull" and ".sum" functions to determine the number of total missing values in each variable. We also used ".dropna" to drop missing values from our analysis. It's important to recognize that doing this removed observations that were otherwise complete for the prestige variable. This process, although necessary due to the fact we couldn't substitute missing values in looks any other way, could have removed valuable information about prestige. Cleaning, overall, was not intensive. For the most part, surveys were almost entirely completed for the variables we look at. There was a more than insignificant count of missing data for "RLOOKS" and "OWNGUN."

Results:

The first method of analysis involved comparing "RLOOKS" with "PRESTG10." We did so using a series of vertical box plots with "RLOOKS" on the x-axis and "PRESTG10" on the y-axis, shown in *Figure 1*: "Distribution of Job Prestige by Physical Attractiveness." After data cleaning, the graph was created on the "agdf" dataframe. The distribution of the boxplot describing job prestige in the lowest physical attractiveness category demonstrates an extreme positive skew, with the lowest median, second-highest upper quartile, and highest maximum observed. By contrast, the distribution of job prestige for the highest physical attractiveness category shows the highest median with an almost perfect distribution. The strongest trend observed in the data is a strong positive correlation between median self-reported job prestige and survey-reported physical attractiveness, with a difference of medians of almost 10 points between the highest and lowest physical attractiveness categories. The

conclusion we are able to draw from this observation is that as attractiveness increases, so does job prestige.

This conclusion comes with a few caveats. The first being the amount of missing data for the "RLOOKS" variable (1,872) in comparison to the missing data for "PRESTG10" (218).

The second caveat that exists is the question of consistent levels of subjectivity across levels of physical attractiveness. It could be the case that respondents who are viewed objectively attractive have a corresponding subjective agreement with this valuation, which research has shown correlates to a higher sense of entitlement (Teng et. al.). A higher sense of entitlement can be extrapolated and applied to how the surveyee self-reports their own job prestige, resulting in the trend observed. For our analysis, however, it is purposeful to assume normalized levels of subjectiveness when analyzing the variables in question, resulting in our conclusion being purely correlative. However, this conclusion is relatively weak due to significant overlaps between the boxplots for different levels of attractiveness.

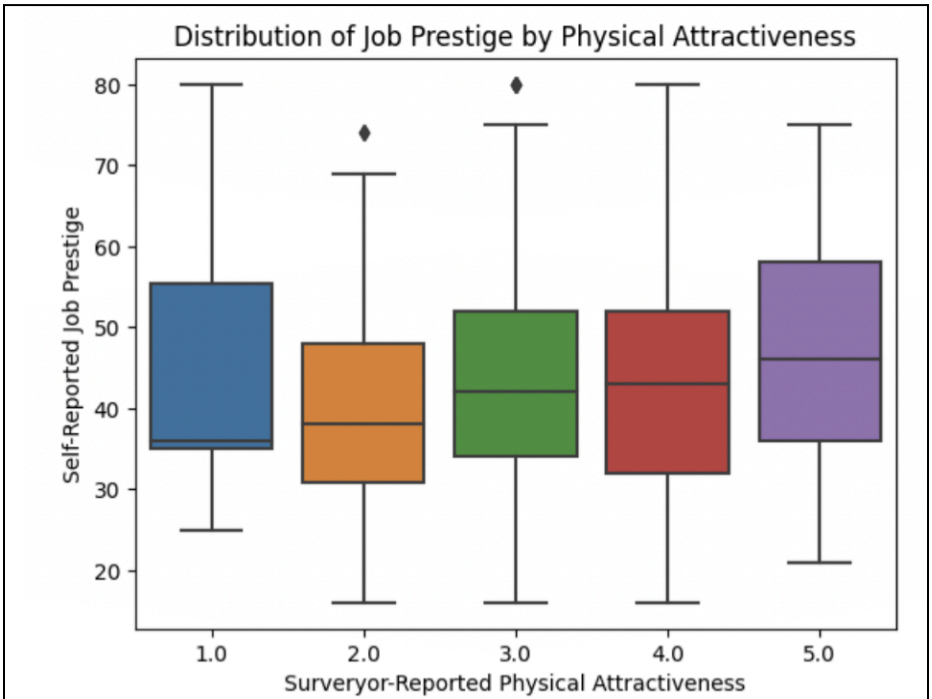

Figure 1: Distribution of Job Prestige by Physical Attractiveness

Our second area of analysis dealt with correlations between "PRESTG10" (x-axis) and "OWNGUN" (y-axis) shown in *Figure 2*. We approached the analysis using a combined overlapping histogram and kernel density line plot. We can see very similar shapes in the kernel density distribution, with overlying peaks and valleys in frequency of job prestige for both gun owners and non gun owners. Another important observation is the relative smoothness of gun owners when compared to non gun owners across all levels of self-reported job prestige. Gun ownership has peaks, such as a mode between levels 30 and 40 of job prestige and the next most frequent level of gun ownership at 50 job prestige, but overall is not as drastic as non-gun ownership. This may be attributed to a lack of reporting for gun ownership, or another causal relationship that is undetermined based on the available data. Self-reported job prestige is skewed left, possibly indicating that people's perceived prestige is lower than it actually may be. That said, there doesn't appear to be a correlation between prestige and owning a gun.

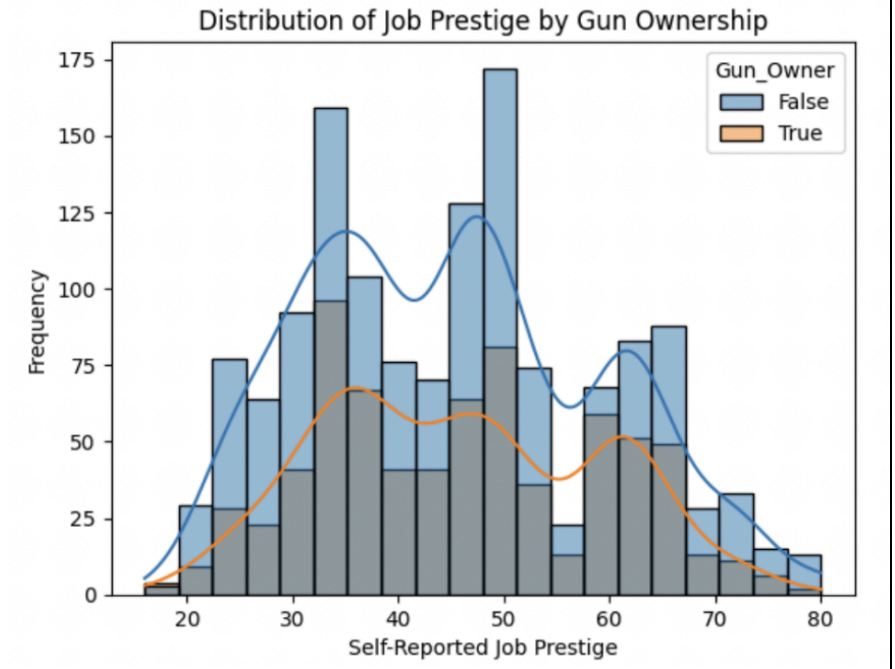

Figure 2: Distribution of Job Prestige by Gun Ownership

The last area of analysis addressed the Distribution of Physical Attractiveness by Gun Ownership, shown in *Figure 3*. The histogram shows percentages of total non-gun and gun owners separated into five categories of attractiveness, from "very unattractive" (1) to "very attractive (5). The most important initial observation to note is the distribution for all data is significantly skewed right, with 90.89% of surveyees receiving scores of 3 or higher on the attractiveness scale, and 36.23% of surveyees receiving a 4 or 5 compared to just 9.11% receiving a 1 or 2. This potentially skews the following observations of correlativity. However, an important correlation of the two variables can still be derived. While both more non-gun and gun owners are perceived as more attractive, a higher percentage of total gun owners received attractiveness scores of 4 and 5 than percentage of total non-gun owners did. From this available data, we can conclude that there is a stronger positive correlation between owning a gun and perceived attractiveness than there is between not owning a gun and perceived attractiveness. Of course, this is a correlative relationship and not causal.

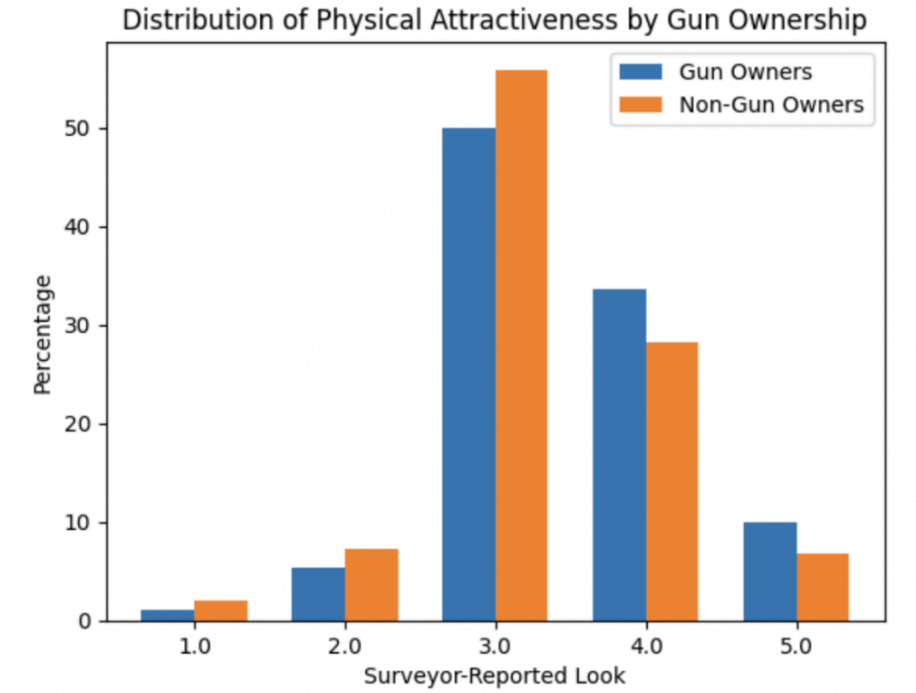

Figure 3: Distribution of Physical Attractiveness by Gun Ownership

It is also important to remember the significant amount of missing values for both "RLOOKS" and "OWNGUN." The attractiveness variable was only able to measure 47.2% of total surveyees, and the gun-ownership variable failed to record a "YES" or "NO" answer from 35.9% of total surveyees due to a lot of missing and blank data. Both numbers are significant, and thus potentially had nullifying effects on our conclusion.

Conclusion:

Overall, this analysis explored the relationships between three variables: surveyor reported attractiveness of the surveyee, gun ownership, and self-reported job prestige of the surveyee. The analysis was organized into three comparisons: attractiveness and prestige, prestige and gun ownership, and gun ownership and attractiveness. There were two key observations. First, higher levels of attractiveness correlated positively with higher self-reported job prestige. The conclusion we can draw from this suggests people who perceive themselves as more successful have higher median levels of perceived attractiveness by others. Second, a higher percentage of gun-owners received attractiveness scores of 4 and 5 compared to non gun owners. From this correlative relationship, we conclude that if you own a gun, you are more likely to also be perceived as attractive than non-gun owners. The comparison between prestige and gun ownership yielded no correlative conclusion.

While we did find correlations, it is necessary to recognize that there was a significant amount of missing values for both physical attractiveness and gun ownership variables due to the nature of the survey type (for attractiveness) and survey question (for gun ownership). Missing values in these

variables forced us to ignore a chunk of data for self-reported job prestige variable where values did not overlap, reducing confidence in conclusions we drew. Additionally, the subjective nature of reporting physical attractiveness and self-reporting prestige introduces potential bias in the study. The surveyor could have underlying preferences that affect these ratings, resulting in correlations that could be misrepresentative and the individual could have a subjective perception of themselves, which may not represent their real-world prestige.

Future research should aim to address missing data issues in the "OWNGUN" and "RLOOKS" variables, mitigate potential bias of the surveyor, and analyze possible causal relationships. For example, as mentioned previously, there may be a possible causal relationship between objective attractiveness and subjective attractiveness, and thus between objective attractiveness and self-reported levels of prestige.

References:

Fei Teng, et. al., "Mirror, Mirror on the Wall, I Deserve More Than All: Perceived Attractiveness and Self-Interested Behavior", Evolution and Human Behavior, Volume 43, Issue 6, 2022, Pages 536-547, ISSN 1090-5138, https://doi.org/10.1016/j.evolhumbehav.2022.09.005.

Contributions:

Flavien: outlined project, imported data, set up the data sets, completed first graph

Tenzin: outlined project, completed second graph, and formatted .ipynb

Charlie: outlined project, set up and completed second graph, assisted with paper conclusion

Duncan: outlined project, completed summary, data, results sections of paper, assisted with conclusion