Relationships Between Demographics and Morality

Arik Gershman
March 24, 2025

I. Introduction

University of Maryland students in Professor Morawski's and Professor Alam's Fall 2023 data science courses participated in a survey asking for demographic information and opinions on moral dilemmas. A new set of students in Professor Morawski's class participated in similar surveys in Spring 2024 and Spring 2025. The surveys were almost identical except for some wording modifications. The surveys asked for the participant's school year, age, parental political ideology, personal political ideology, religiousness, and gender. This analysis aims to discover the relationship between these demographic variables and responses to moral dilemmas. After cleaning up the given data, dummy variables were created, hypothesis tests were performed, and visualizations were crafted. There were a few interesting results to consider. Out of all demographics, whether the participant was religious or not appeared to be the greatest indicator of their response to the moral dilemmas. However, this result and the others were not statistically significant at $\alpha = 0.05$. Therefore, these demographics did not significantly impact participants' responses to moral dilemmas.

II. Background

The surveys initially asked for demographic information, such as year (Freshman, Sophomore, Junior, Senior, and Graduate Student/Other), age, parental and personal political ideology ("Strongly liberal" to "Strongly conservative"), religiousness/spirituality, and gender ("Male", "Female", "Non-binary/other", and "Prefer not to say").

After collecting demographics, each survey detailed 14 situations where a person behaved in a morally questionable way. The 2023 surveys had identical scenarios. In the 2024 and 2025 surveys, people's genders in most moral questions were switched compared to the 2023 surveys. For instance, the following is a scenario from a 2023 survey: "My girlfriend is a doctor. Lately she's been complaining about pain in her right knee and constantly taking TONS of ibuprofen to treat (like 15-20 pills a day). However, it's only gotten worse. I've told her she needs to see someone about it but she told me it was fine. So I told her that if this is how she treats herself, I'd be concerned about being her patient. She was shocked and told me that was uncalled for. Am I a jerk?". The 2024 and 2025 surveys asked an identical question, but referring to "boyfriend" instead of "girlfriend". The participants were asked to select whether the person was "Not a jerk", "Mildly a jerk", or "Strongly a jerk".

Since none of the questions were required, participants could skip questions causing missing values and inconsistencies in the data. To list some of the discovered issues, there were 11 entries across all surveys where the participant skipped all the moral questions, some moral questions were intended to be identical across all surveys but contained mistakes making them appear as different questions, the 2023 surveys had entries with the participant's gender being "Famale" instead of Female, Sophomore was incorrectly spelled in all surveys as "Sophmore", the age entries had different types (ints, floats, and strings), Professor Alam's 2023 survey was the only survey asking the participants if they considered themselves compassionate, and finally, the 2023 surveys asked for participants to rate their "religiousness/spirituality", whereas the 2024 and 2025 surveys asked for participants to rate their "religiousness".

Before beginning the analysis, it was important to clean the data. The data was imported, and the surveys were stored in four data frames. To ensure the correct concatenation of the data frames into one, it was essential to fix wording mistakes in questions that were intended to be identical. For example, one otherwise identical question included a "\n" at the beginning in 2 of the 4 surveys. Another otherwise identical question asked the participants "Am I jerk?" instead of "Am I a jerk?" in 3 out of the 4 surveys. These columns were renamed to ensure consistency and correctness. The next step was to combine the four data frames into one, ensuring it included all the necessary information and no duplicate columns or rows. The columns in the data frame had lengthy names because they represented survey questions, so they were renamed to increase ease of use. For example, "How would you rate your religiousness" to "Religiousness", "What bests represents your gender?" to "Gender", "You could describe the adults you grew up with as..." to "Parental Political Ideology", and all the moral questions were renamed to "Q1", "Q2", up to "Q25" (11 questions with their gender-reversed counterparts, 22 total, and 3 questions that were identical across all surveys). After renaming, the gender variable mistake was fixed by converting all "Famale" entries to "Female". Additionally, the age variable consisted of one string, "50+". Since the analysis only considered whether the participant was at least 21 years old or under 21, the string "50+" was converted to an int, 50. The NaN entries for the age variable were replaced with 0s, and the rest were converted to ints. The NaN entries for the other demographic variables were filled with "No {variable} Entry". For example, for the "Self Political Ideology" variable, the NaNs were filled with "No Political Entry".

Two surveys asked "How would you rate your religiousness?" with options ranging from "Not religious at all" to "Strongly religious", whereas the other two used "How would you rate your religiousness / spirituality?", with options ranging from "Not spiritual at all" to "Strongly spiritual". These variables were combined into one because they collected identical information. To further prove this claim, it is beneficial to visualize the similarity in their distributions. The following visualization was crafted after modifying the entries in the "Spirituality" variable to match the ones in the "Religiousness" variable.

Comparison of Religiousness and Spiritulaity
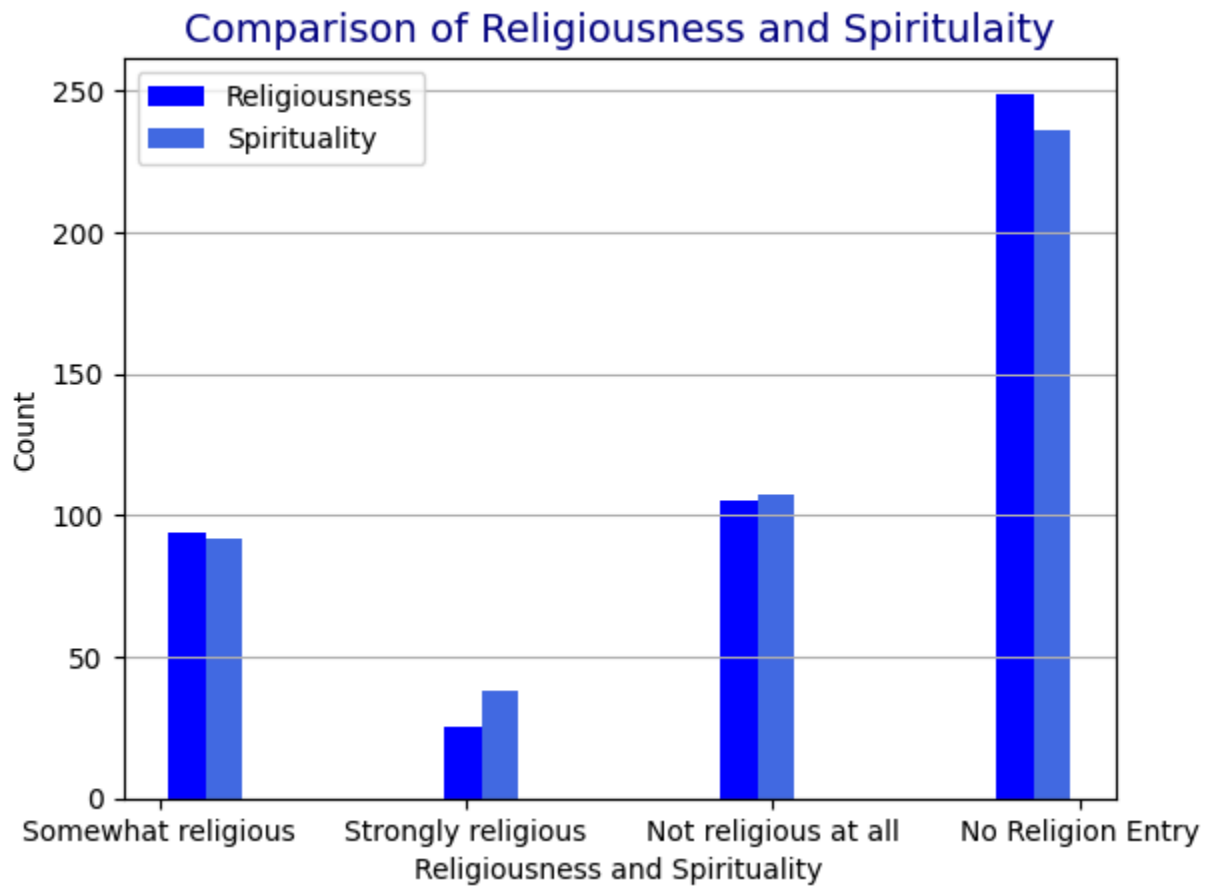
It was clear from both the visualization and common sense that these variables measured the same information. A 2-sample hypothesis test was performed to strengthen the claim further. A p-value of 0.994 showed no evidence that the two variables were different. To combine them, the "Religiousness_numeric" variable was created by assigning "No Religion Entry" a 0, "Somewhat religious" or "Strongly religious" a 1, and "Not religious at all" a 2. The same was done for the "Spirituality" variable. Then, a "Relig_numeric" variable was created by adding the "Religiousness_numeric" to "Spirituality_numeric". Since the questions appeared on different surveys and a 0 was assigned to missing entries, it made sense to add the two variables. The resulting "Relig_numeric" was ensured to only have entries between 0 and 2. "Relig_numeric" was then converted back to a categorical variable called "Relig", with the categories "Not religious at all", "At least somewhat religious", and "No religion entry".

III. Methodology

After cleaning the data, the exploration was initiated. The broad goal was to understand and compare the impacts of participants' demographics on their responses to the presented moral dilemmas. First, each variable was transformed into a binary variable. For "Self Political

Ideology", a "Liberal/Conservative" variable was created, including selections of "Strongly liberal" and "Mildly liberal" as "Liberal", "Strongly conservative" and "Mildly conservative" as "Conservative", and the rest as "Other". The same was done for the "Parental Political Ideology". The other demographic variables were also transformed. "Gender" was combined into "Male", "Female", and "Other", "Relig" was combined into "Not religious at all", "At least somewhat religious", and "Other", and "Age" was combined into "21+" (21 and older) and "21-" (less 21 and greater than 0).

Keeping our goal in mind, the next step was to create a quantitative variable representing participants' performance on the moral questions. A numeric variable was created for each moral question, converting all entries into numbers. After filling NaN entries of "Q{i}" (for $1 \leq i \leq 25$) with "No Entry", the "Q{i}_numeric" and "Q{i}_answered" variables were created. The "Q{i}_numeric" variable represented responses of "Not a jerk" with a 1, "Mildly a jerk" with a 2, "Strongly a jerk" with a 3, and "No Entry" with a 0. The "Q{i}_answered" represented "No Entry" with a 0 and the rest with a 1. Then, a column representing the participant's total score, "Total_jerk_score" was created by summing the values in "Q1_numeric" up to "Q25_numeric". The "num_questions_answered" column was created by summing the values in "Q1_answered" up to "Q25_answered". Since each participant may have answered a different number of questions, comparing total scores did not make sense. Before creating a variable for the average score, rows with fewer than half of the questions unanswered were removed to ensure a large enough sample (only rows with at least 7 answered questions were kept). Consequently, the "avg_jerk_score" variable was created by dividing the "Total_jerk_score" column by the "num_questions_answered" column. The "avg_jerk_score" was a float value ranging from 1 to 3. 1 meant the participant selected "Not a jerk" for all questions and 3 meant the participant only selected "Strongly a jerk". A visualization was created to understand this variable better.

Distribution of Average 'Jerk' Scores
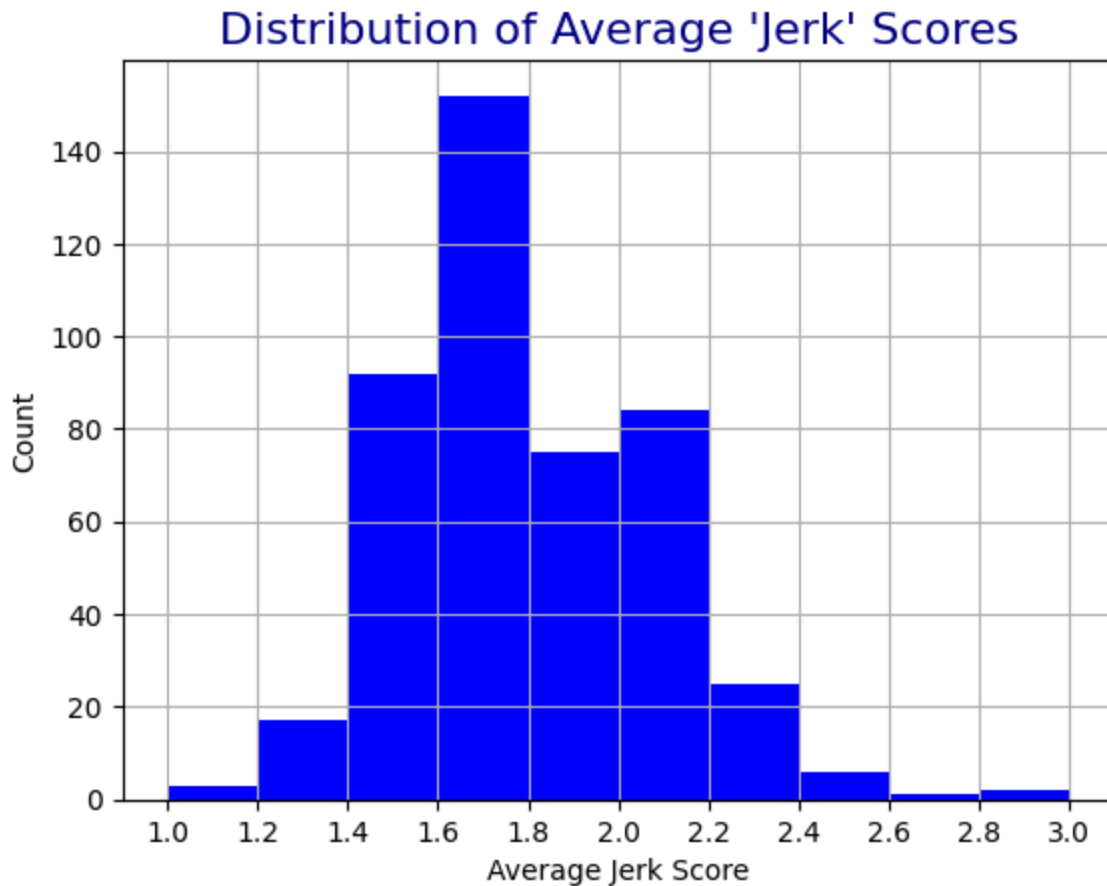
A two-sample t-test was used to investigate the relationship between a demographic variable and participant's responses to moral scenarios. For each demographic variable, the "avg_jerk_score" was grouped by the two categories of that variable, and t-tests were used to determine the likelihood of the two distributions being significantly different. For the "Male/Female" variable, the "avg_jerk_score" was grouped by "Male" or "Female" values and inputted into the "scipy.stats.ttest_ind" function. The function outputted a test statistic and p-value to show if the difference was statistically significant.

IV. Findings

Before running the hypothesis tests, it was interesting to examine the means of the "avg_jerk_score", grouped by the categories of each demographic variable.

Variable	Categories	avg_jerk_score	Difference in Average Score
Religiousness	At least somewhat religious	1.812059	0.040662
	Not at all religious	1.771397	
Self Political Ideology	Conservative	1.827567	0.033333
	Liberal	1.794234	
Gender	Male	1.793201	0.019714
	Female	1.773487	
Parental Political Ideology	Conservative	1.800047	0.007977
	Liberal	1.808024	
Year	Freshman/Sophomore	1.768250	0.036014
	Junior/Senior	1.804264	
Age	21 and older	1.799789	0.006313
	Younger than 21	1.793476	

It was clear from the table that the difference in the mean scores grouped by religion is the greatest, signifying a potential trend. From the table, at least somewhat religious participants tended to have higher average scores than not at all religious ones, indicating they were potentially more likely to consider certain actions as jerk behavior. However, nothing can be concluded before rigorously testing this hypothesis. A visualization was created to understand the difference in distributions better.

Average 'Jerk' Scores Grouped By Religiousness
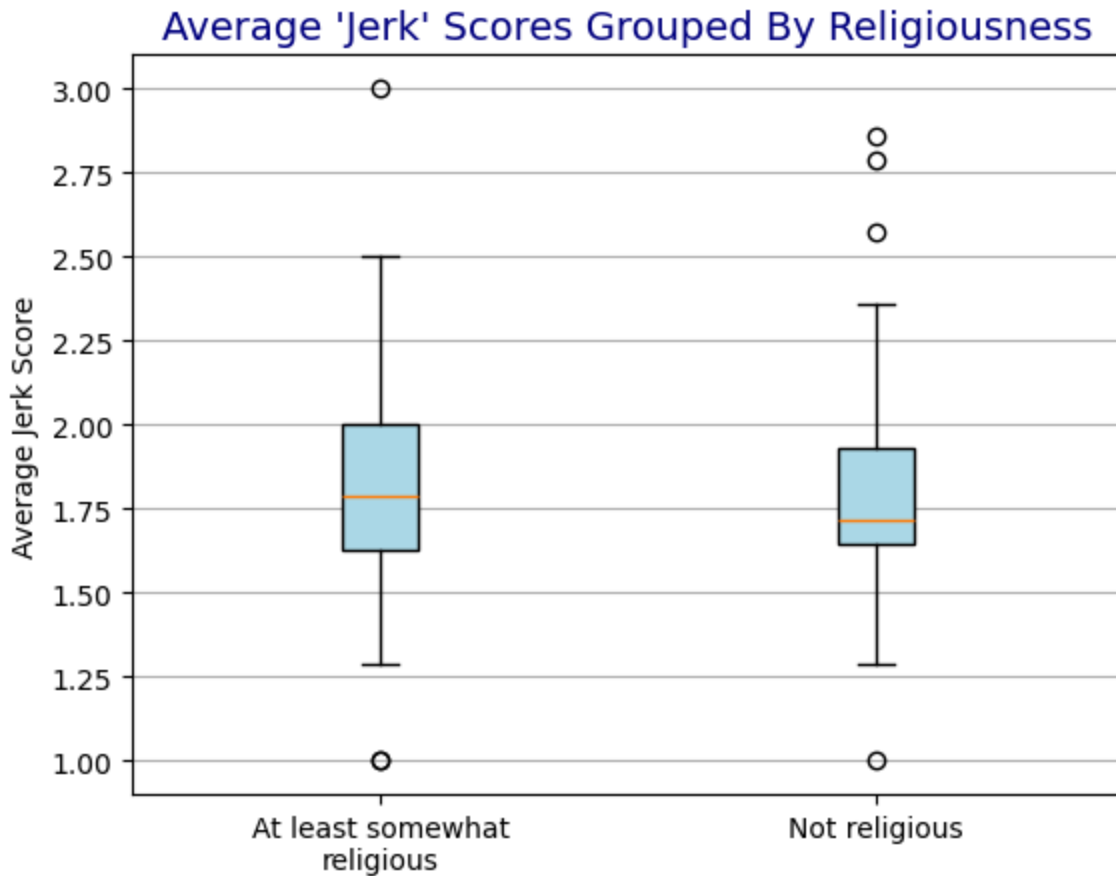
The null hypothesis was that the "average_jerk_score" for at least somewhat religious participants was equal to those not at all religious, and the alternative was that the "average_jerk_score" for at least somewhat religious participants was greater than those not at all religious. After setting $\alpha = 0.05$, the distribution of average scores for at least somewhat religious and not at all religious participants was compared through a one-tailed, two-sample t-test. The p-value for the one-tailed test was 0.052. Since the p-value was greater than the alpha, there wasn't enough evidence to reject the null hypothesis. This means there wasn't a significant difference between the average scores of at least somewhat religious participants and not at all religious ones. Therefore, it cannot be concluded that participants' religiousness was correlated with how they responded to the moral questions. However, it is important to note that if $\alpha = 0.1$, a common alpha level, the null hypothesis would be rejected, and there would be enough evidence to conclude participants' religiousness was correlated with how they responded to the moral questions.

The same procedure of comparing the distributions of "avg_jerk_score" grouped by the categories of each demographic variable was followed for the rest of the variables. Rigorous t-tests proved no significant differences between any "avg_jerk_score" distribution grouped by demographics. For personal political ideology, testing whether participants' "avg_jerk_score" was greater if they were conservative rather than liberal produced a p-value of 0.165. Testing

whether the "avg_jerk_score" of males was greater than that of females yielded a p-value of 0.247. Testing whether the "avg_jerk_score" of participants whose parents were conservative was lower than that of those with liberal parents produced a p-value of 0.395. Testing whether participants' "avg_jerk_score" was lower if they were freshmen or sophomores rather than juniors or seniors yielded a p-value of 0.093. Finally, testing whether participants' "avg_jerk_score" was greater if they were 21 or older rather than younger than 21 produced a p-value of 0.413. None of these results provide enough evidence to reject the null hypothesis with $\alpha = 0.05$. This means that demographics were correlated with how participants responded to the moral dilemmas.

V.   Conclusion

This study explored the relationship between demographic variables and responses to moral dilemmas across four similar surveys. After cleaning and examining the data, no statistically significant relationships between demographics and morality were discovered at $\alpha = 0.05$. The difference in responses to moral questions between non-religious and somewhat/strongly religious participants was notable, but not statistically significant. Demographics had no significant correlation to morality in this set of surveys.