Student-Generated Political Beliefs Dataset Analysis

Chris Wu

Introduction

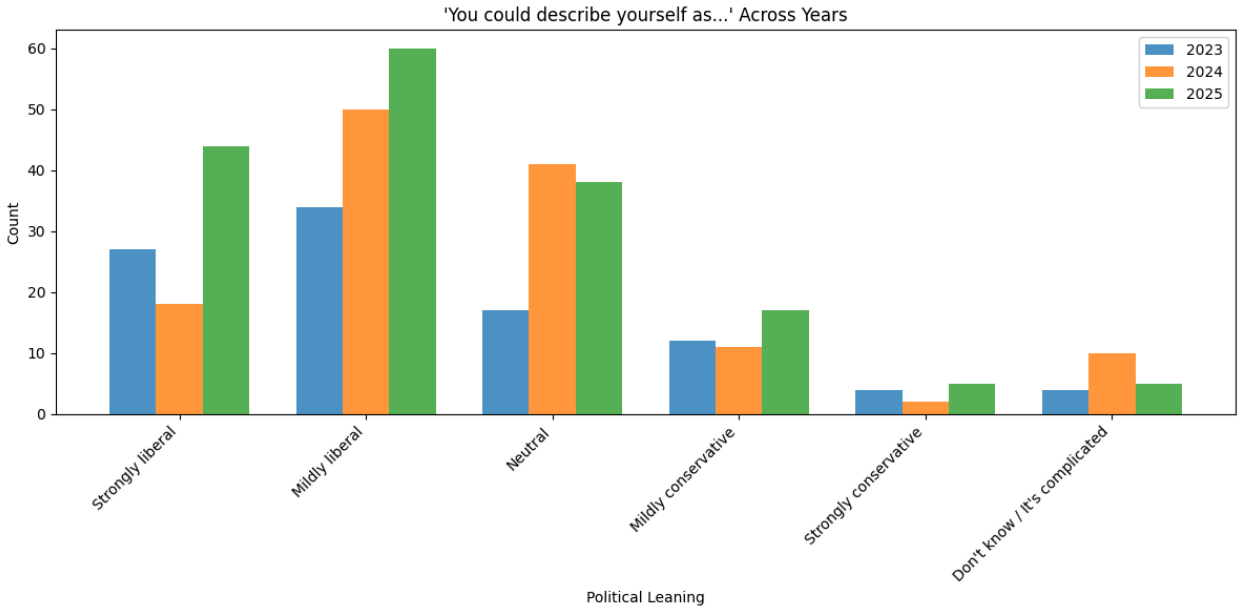
The related graphs, code, and datasets can be found in the following repository: https://github.com/cheesuschris/320Homeworks/tree/main/Homework2. To check this homework out yourself, clone the repo and run the script.

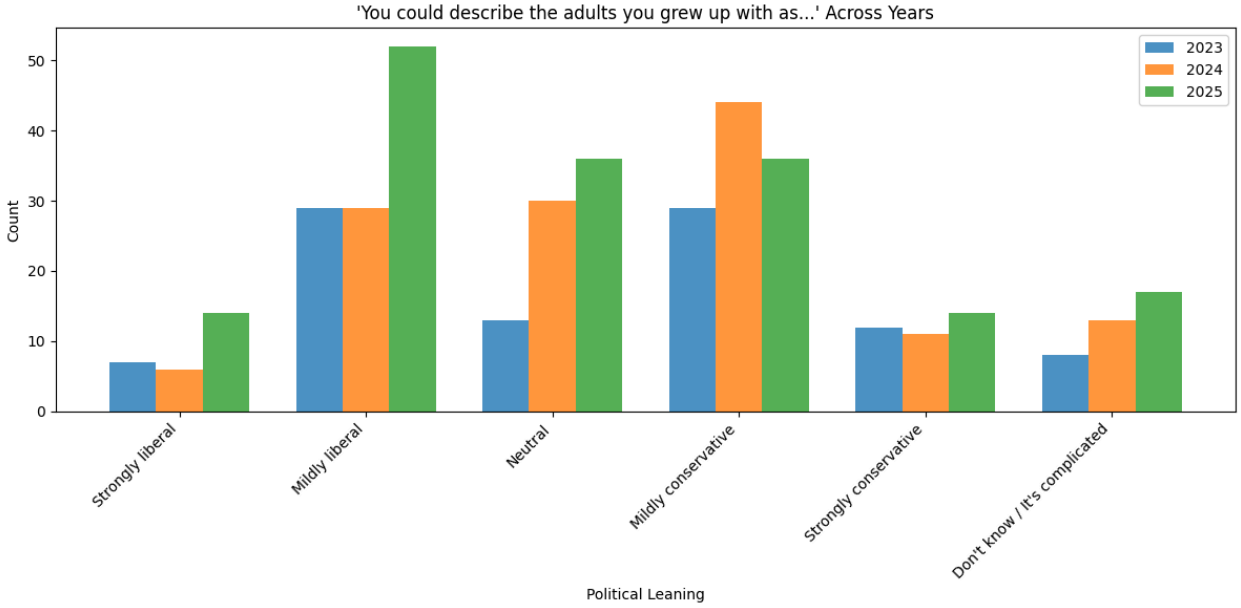
Report summary

Between the recent years of 2023-2025, the CMSC320 (Intro to Data Science) class's professors at the University of Maryland - College Park have conducted google form surveys upon its students regarding their political beliefs. These student answers, in addition to some of their qualities such as age and gender, were then turned into data points to study how many different factors of a student's life influenced their political views. Within this report contains a study for the Fall class of 2025's responses as well as their relation to previous years' datasets that slightly differ; an introductory exploratory data analysis and visualization, followed by more specific column correlation, priming, what-if, gender-swapping, and data-over-time tests (and their graphs) are also offered within the related script.

Interesting findings

After data cleaning the class of 2025's dataset – and besides the obvious correlation observations that could be made by just plain assumptions from looking at the dataset's column names – there were also some interesting insights hidden within the data. For one, it turns out that a student's gender impacts their political beliefs and decision-making much more than that student's church attendance or sport watching. Another question within the study that was primed and intended to introduce bias also revealed priming DOESN'T introduce as much bias to students' responses as much as one would think. Surprisingly, a household's self-described political views impact different political decisions than an individual student's self-described political views. Similarly, across the years students have also leaned towards being neutral and mildly liberal (and even strongly liberal) more than they have leaned towards staying conservative – however, the students' perspective of their parents' political beliefs have remained more stable and slightly less left-leaning between 2023 and 2025.





This analysis provides analysis of the political view datasets generated by the students over the years, as well as insights and potential explanations similar to the ones mentioned within this paragraph.

Background

Explaining the dataset

The analysis comes from the class of 2023, class of 2024, and class of 2025's student political views survey data. Each dataset roughly includes:

- Student Response Timestamp

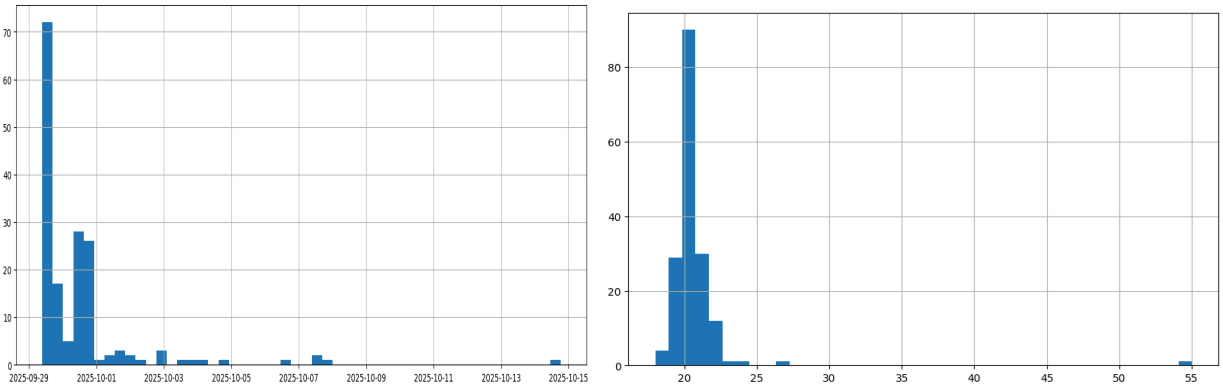
- School Year of Student
- Age of Student
- Student's Political Description of the Adults they grew up with
- Student's Political Description of themselves
- Church attendance / Self-Description of Spirituality / Self-Description of Religiousness
- Sports Watching & Parents' Sports Watching (only available in Fall 2025)
- Gender of Student
- A series of "Am I A Jerk?" questions, some questions being intentionally primed, and some datasets having opposite gender people (in relation to other datasets) within the scenario being described to the reader

The Fall Class of 2025 had a total of 169 responses to the survey.

Data Cleaning Issues

Upon exploration of the Fall Class of 2025, there weren't any irrelevant data columns that should've been merged or dropped. No duplicate responses or rows appeared in the dataset, and no rows had ALL of its columns missing – however, the first response in the dataset contained SOME columns missing, starting from the "AIAJ" question on column 16 and on. Hot deck imputation couldn't be performed as no (even slightly) similar rows were found; as a result the N/As were replaced with the mode of the data, and the risk of a biased mode was taken on (since the column responses were ordinal the N/A replacement wasn't that bad of a choice). Additionally, suspicions were raised that since the only missing responses come from row 0, either one of the TAs or Professor Morawski himself inputted this response; this would make sense since the students were required to respond to the questions, but perhaps the TAs or professor had access to skip some questions.

Response time and age found a couple of acceptable outliers that were ultimately NOT removed to form a "typical age or response time" dataset (all responses should be fair) – this introduced a right skew to those distributions:



The left graph describes date of response, while the right graph describes the age of responder.

Investigating trends over time introduced the Fall 2024 and Fall 2023 datasets, which also came with many more N/A rows than the Fall 2025 dataset – these questions were probably not required for those classes. For Fall 2023, the "Famale" response option had to be replaced with "Female", and for all dataframes the "Sophmore" response option was replaced with

“Sophomore”. N/As for “What year are you?” were dropped, and N/As for “How old are you?” were dropped rather than replaced with the median/mode or hot-deck imputation because their year level in school was also recorded but seniors also had varying ages among themselves, as did juniors and sophomores as well; the responses for “How old are you?” were also converted to integers.

Findings

Some obvious findings were proven trivial through the computation of a chi-squared test:

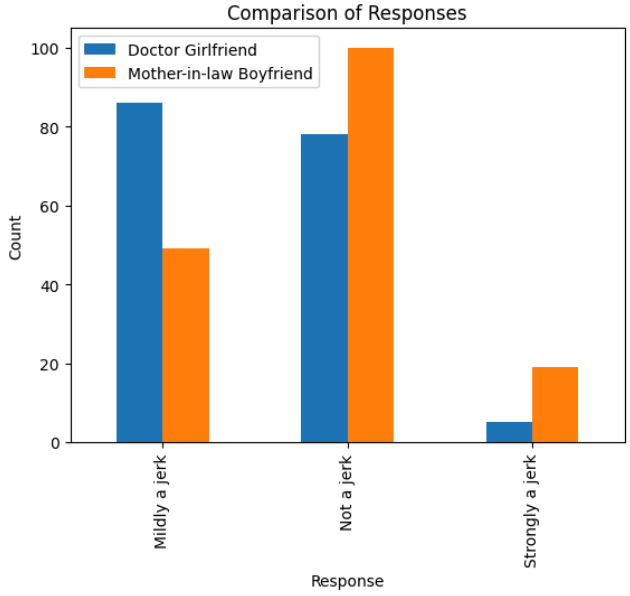
- A student’s political description of their parents heavily relates to their own political description of themself, with a p-value of 7.33x10^-13.
- A student’s attending church heavily relates to their own attending church, with a p-value of 2.79x10^-20.
- A student’s parents watching sports heavily relates (but slightly less) to their own watching of sports, with a p-value of 8.58x10^-6.

Some theories about correlations within the dataset were also tested to reveal less immediately obvious insights as well:

- A chi-squared test relating gender to a student’s political description of themself gave a p-value of 8.62x10^-5, meaning that we reject the null hypothesis that these columns are not related. However, the same test relating church and sports to the description resulted in much higher p-values > 0.05, meaning that we fail to reject the null hypothesis and the church and sports factors of a student’s life aren’t significantly related to their political beliefs/views.
- A student’s gender was found to be significantly correlated to the responses recorded in some of the AIJA questions as well: the refusal to turn in the lost cat question, conflict with mother-in-law’s boyfriend question, relatives refusing to support an LGBTQ wedding question, and refusal to let sister be a bridesmaid at wedding due to dyed hair questions achieved significant p-values. These questions were more geared towards personal freedom and respect.
- A student’s age was found to be significantly correlated to some AIJA questions: the doctor girlfriend’s ibuprofen intake question, trust fund kid splitting finances question, and the child with a negligent non-paying mother questions achieved significant p-values. These questions were more geared towards both hard-boundaries in a person’s life, serious interpersonal relationships, and setting up/thinking ahead for the future.
- A student’s year in school, as well as their attendance in church, both were found to NOT be significantly correlated to any AIJA questions.
- A student’s watching of sports was found to be significantly related to the negligent non-paying mother AIJA question – one interpretation of this result could be sports playing a big role in students’ lives who DO watch sports (and vice versa), so much so that they are more sensitive (or not) to a semi-maternal role in their life.
- A student’s political self-description was also found to be significantly related to the relatives refusing to support the LGBTQ wedding question. This can be explained due to simply the popularity of LGBTQ being a liberal topic – students tend to correlate positives and negatives to this topic when they think of themselves as being liberal or conservative.

Theories about Primed questions were also investigated:

- The girlfriend's ibuprofen question was primed with the narrator's own view clearly being included in the details of the anecdote (e.g. "a TON of ibuprofen"). However, from the multiple bar chart below we are able to see that, when compared to a non-primed question, even the doctor girlfriend question reached a more "mellow" result – students tended to respond to the doctor girlfriend question with mildly a jerk more so than reach a strong viewpoint of not a jerk or strongly a jerk, while they responded to the non-primed question more strongly. The effect of priming this question is not obvious and wasn't good at introducing bias to the question.



Theories about a household's political views as compared to a student's individual political views were also taken looked into:

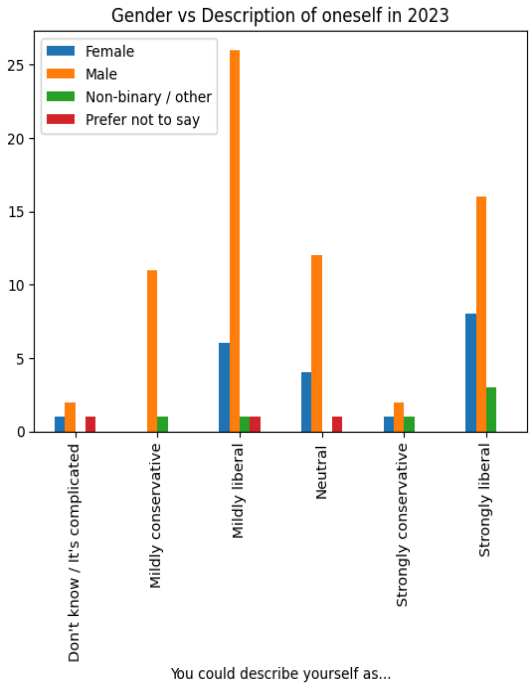
- A chi-squared test revealed significant correlation between a household's political views (combined view of parent and student) and the AIJA questions related to the wife's non-allowance of drinking due to pregnancy, as well as to the relatives refusing to attend the LGBTQ wedding. These topics are more familial and are heavy inter-family relationship questions, which make sense.

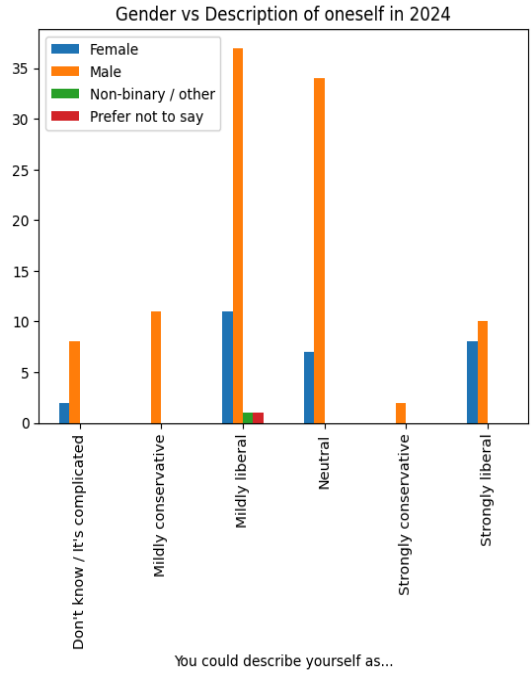
Gender swapping was also taken into account with the Spring of 2025's dataset – their dataset had male participants within the AIJA questions asked rather than the female participants included in the Fall of 2025's dataset:

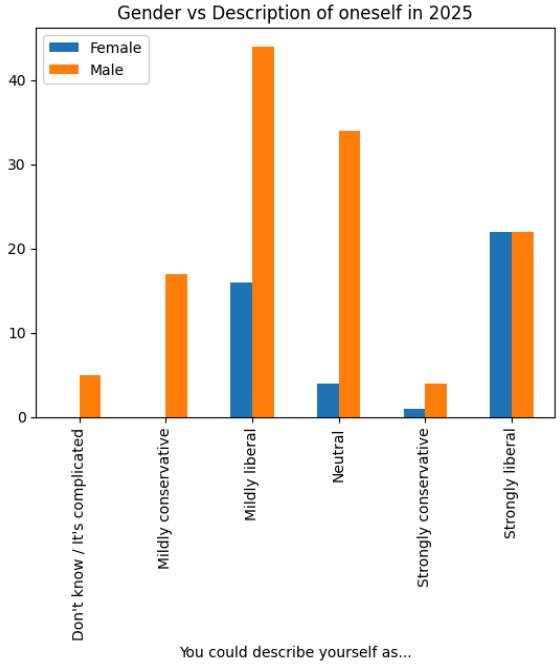
- Chi-squared test revealed the BOYFRIEND as a doctor taking ibuprofen was significantly related to student gender, the BROTHER dropping off a complaintive son was significantly related to student gender, and abandoning the BROTHER because of a seat upgrade was also significantly related to student gender. One view of this result as compared to the female question results could be the confounding conservative view that males are naturally stronger and therefore tougher mentally than females when enduring hardship.

Student Statistics over Time were also recorded and visualized in the graphs below:

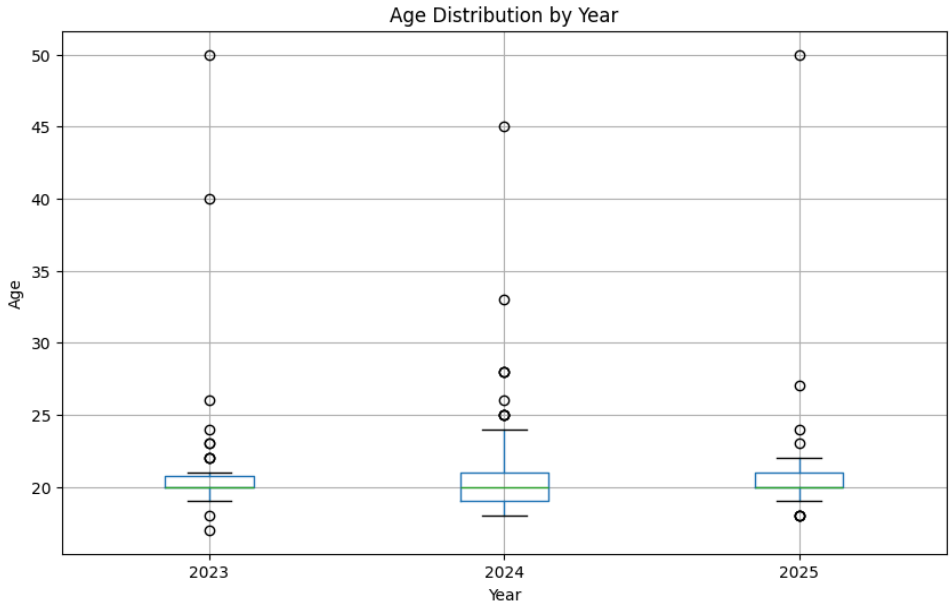
- Gender vs. Descriptions of oneself over time are shown below. Over time, there are clearly non-binary/other and prefer not to say responses (one reason that stands out for this could be the election that happened in 2024); students also clearly favor mildly liberal and neutral responses and playing safe, continuing to do so and even grow these response categories over time:



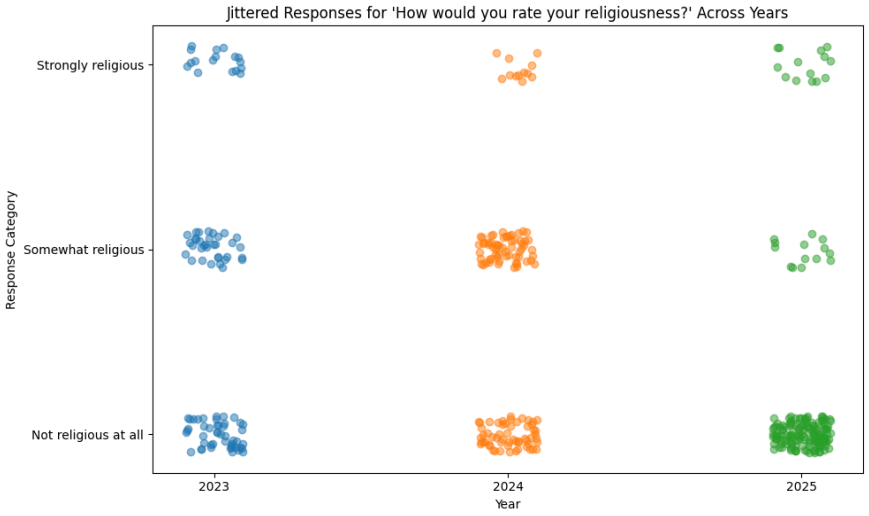




- Class Age and Year representation over time are shown below. “50+” responses were replaced with “50” for the sake of visualization – the most common student demographic is juniors, although there are some sophomores and seniors each year. There is a freshman outlier in 2024 as well as other outliers in 2023 and 2025:



- Class Religiousness representation over time is shown below. Students initially were spread relatively even between the different religious categories, but slowly shifted towards not religious over the years and as more responses were filled out (there are less data rows in previous years than in later years):



Conclusions

This report supports the following theories.

Theory 1 – Gender affects political and moral views.
Students' gender was strongly linked to how they described their political beliefs and judged moral situations. This suggests that gender roles and experiences shape how people think about fairness, respect, and social issues.

Theory 2 – Family beliefs influence but don't control personal beliefs.
Students often shared similar political views with their parents but also showed independent opinions on topics like LGBTQ rights. This means students learn from their families but develop their own beliefs as they grow older.

Theory 3 – Priming has little effect on moral judgment.
Changing the wording or framing of a question didn't strongly change how students answered. This shows that most students rely on their own moral sense rather than being easily influenced by how a question is written.

Quantitative and qualitative analysis on student's environments as well as political beliefs over time offer valuable insights into how young teens/early 20's think and respond politically according to outside factors in their life.