Project 1 – GSS

Author(s):

Gabriel Jackson

Naad Kundu

Thao Nguyen

Kayla Nguyen

Halbert Nguyen

Ben Willoughby

Omar Zeineddine

DS 3001: Foundations of Machine Learning

March 12, 2024

SUMMARY

For this project, the research question was: "What are the impacts of age, religion, happiness, and education on political views between 2012 to 2022?" In order to answer this question, different methods were used, involving different types of graphs and visualizations. In order to have interpretable and un-skewed graphs, the data was first cleaned. This included only using data from the years 2012-2022 to better represent recent data and how it impacts political view. The column names were also changed to make the data more readable, as well as cleaning all of the important variables. The variables that were used in this analysis included age, degree, education, happiness, and political view. The variables were cleaned by converting the integer-typed variables to numeric, converting improper categorical variables into an ‘unknown’ category, and removing any unnecessary data to the analysis. To better represent the data, the variable names and categories were renamed as the General Social Survey used esoteric naming conventions. For the methodology of this experiment, the independent variables were age, degree, education, and happiness with the dependent variable being political view. These variables were plotted using bar and box-and-whisker plots. The results showed the General Social Survey revealed significant trends between politically-extreme individuals according to their age, happiness, and education. Overall, liberals tend to have higher levels of education, particularly at the graduate level, while conservatives tend to report higher levels of happiness. This demonstrates the role that education combined with political views can alter an individual’s level of happiness. Liberals also tend to be younger while conservatives are typically older. This demonstrates the impact that age (and related factors) have on political view. From this analysis, one potential further investigation would be to determine the role that age impacts political party

affiliation as an individual age. This would likely require a longitudinal study that would help determine the strengthening or weakening of an individual to a particular political party.

DATA

For the project, the data was originally very unclean and a lot of work had to be done to clean it and prepare it for the data visualization. First, in terms of the 'year' variable, it had to be converted from a string to numeric type. Also, the columns of the key variables were renamed to make them more readable to prepare for the visualization. Before any other key variables were cleaned-up, a challenge arose. The dataset stretched all the way back into the 20th century in terms of years, and older data should not skew the analysis's results. This is why it was decided to only include data from between the years 2012 to 2022. The reasoning is pretty simple; the most recent data possible is required to help explain more relevant political trends. Since 2022 was the most recent year available in the dataset, it only made sense to go back ten years from that year to 2012. This, therefore, allows for the most recent data to be shown and provides more of a substantial set of data points (compared to a shorter time period, like a 5 year range).

The next variable, 'age', involved some similar cleaning. Firstly, the age was converted to numeric-typed values (instead of string-typed). For age, it was decided to keep the NAs as it could be eventually graphed as a numeric-type. Another challenge that arose is that a lot of the age data points had a .0 at the end of their values (floating point type). This was fixed by converting them to the integer type (if they were not NAN). The next variable, degree, involved a similar process to cleaning the religion variable. The names were first capitalized (for example, bachelor's to Bachelor's). Then, any NAs were replaced with unknowns, as degree is a categorical variable. 'Education', a numerical variable, was the next key variable to clean. This

means that it was converted to a numeric-type, and left the NAs as is to allow for graphing later. The 'happy' variable, a categorical variable, was cleaned the same way as the other categorical variables- it was capitalized and any NAs found were replaced with "Unknown". The last key variable to clean was the 'political_view' variable. Like the other categorical variables, tt was capitalized and any NAs were replaced with 'Unknown'. Interestingly though, in order to best show this variable and represent it in an aesthetic and interesting manner, it was decided to create a new variable, 'political_view_id'. Since the 'political_view' variable had 7 values ranging from "Extremely liberal" to "Extremely Conservative", it was decided to make 'political_view_id' a numerical variable ranging from 1-7 that mapped 'political_view' to 'political_view_id'. This new variable used 1 to represent “Extremely Liberal” and 7 to represent “Extremely Conservative” with the other political views following this pattern of increasing conservatism with each higher number in “political_view_id”. Overall, this decision allowed to more easily graph and visualize different political affiliations with the other variables.

RESULTS

To determine the impact of age, religion, happiness, and education on political views, a comprehensive analysis began by utilizing various data wrangling and visualization techniques to convey the findings. Through the meticulous cleaning of the data from years 2012-2022, the comprehensibility of our data was ensured by focusing on key variables such as age, years of education, degree, happiness, and political views. As mentioned before in the summary, the newly created “political_view_id” variable is on a scale from 1 to 7, liberal to conservative respectfully.

The visualization efforts revealed a fascinating pattern regarding happiness and political views. Conservatives report the highest levels of happiness across all three levels; however, the amount of unhappy individuals are similar among both sides, suggesting an unclear pattern between political views and overall life satisfaction (Figure 1).

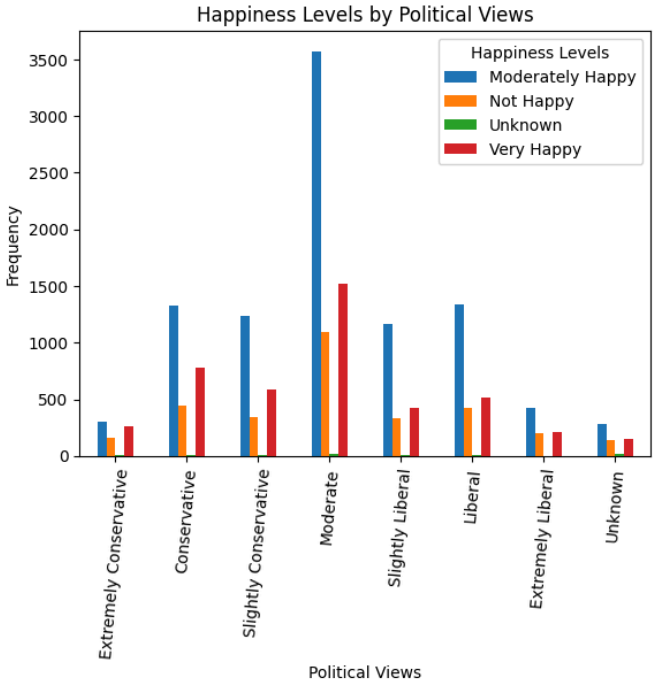

Figure 1: Happiness Levels by Political View

As the extremes were more considered, the individuals who identify themselves as “extremely conservative” tend to report as “very happy” in comparison to being “extremely liberal,” with most being only “moderately happy.” This disparity becomes less prominent when those who are only slightly forward with their political position are considered, signifying a convergence of happiness levels among less polarized individuals.

As for educational background, most of the surveyed individuals held up to a high school diploma, with the highest reaching up to graduate school (Figure 2). When the data was further examined, it can be seen that liberals tended to have higher levels of education, particularly

graduate school. This pattern highlights the potential influence that education can have on political views, suggesting that educational background can shape political views. By observing

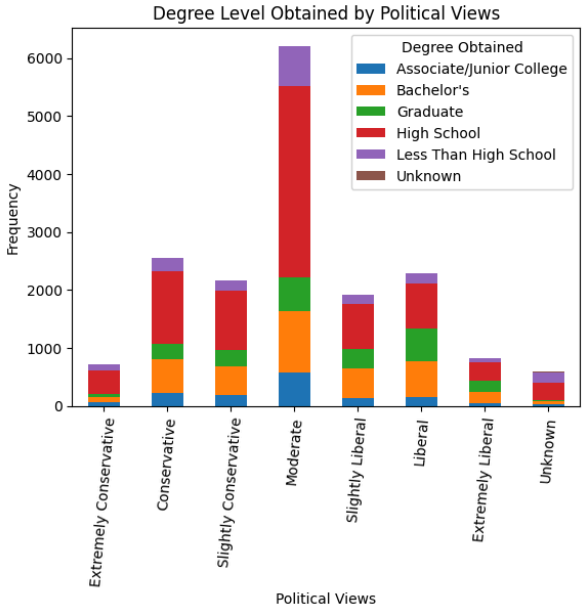

Figure 2: Degree Level Obtained by Political Views

the data, conclusions shouldn't be immediately founded, as correlation does not mean causation. Beyond the level of education, their fields of studies could also influence political views. For example, those with higher education might lean liberal due to the general environment in being exposed to liberal qualities such as inclusivity and equality.

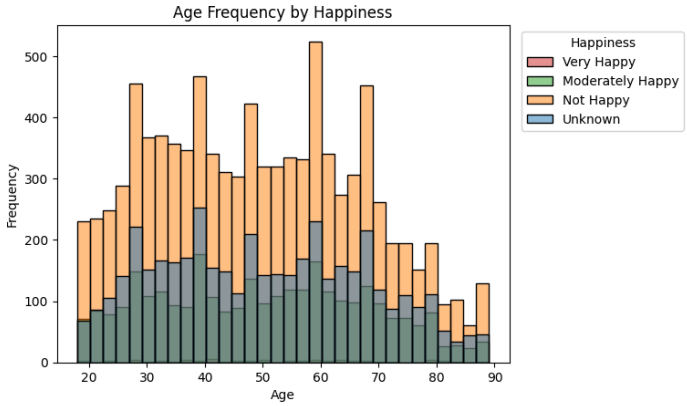

Figure 3: Age Frequency by Happiness

Ages from most of the surveyed lies between 30-70, with the amount of participants greatly decreasing above 70 years old (Figure 3). With the wide range of ages, the findings

indicate a relatively even distribution of happiness across the various age groups, suggesting that there are other factors than age that can impact an individual's happiness. As we can see on the graph above, the proportion of the orange bars (moderately happy) to green (not happy) are similar across all age groups. This means that no matter the age, humans are still in touch with their emotions. One consideration to think about is how education can affect happiness. Higher education allows for a higher ceiling in salary, employment opportunities, and social status. All of these can greatly improve the happiness as an individual as they move up in the workforce. On the other hand, the stressful lifestyles of being a student can adversely affect happiness. All of this suggests a complex relationship between education and happiness.

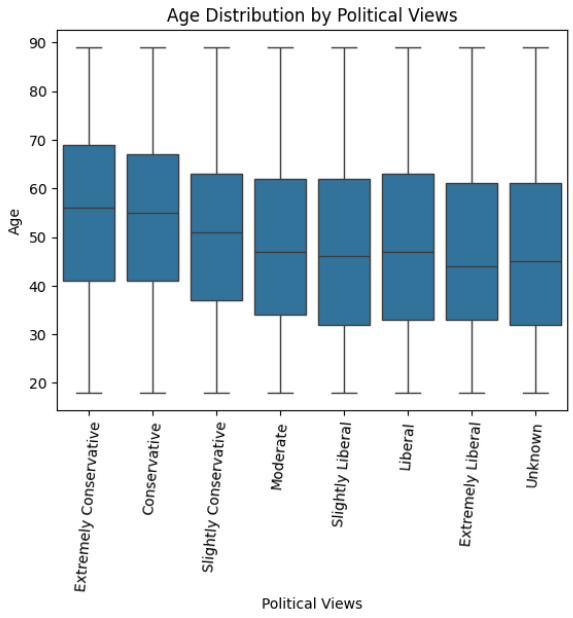

Figure 4: Age Distribution by Political Views

Although results from happiness vs. age were underwhelming, a more pronounced trend can be seen with political views and age (Figure 4). As can be seen from the box-and-whisker plot above, older individuals are more likely to lean conservative, where the liberal party contains individuals around 30 years old. This age related trend highlights the generational

differences in political ideas, as different generations have lived through different political climates.

One thing to mention was the overwhelming prevalence of the “moderate” category, as can be seen in the bar graphs above. This makes it difficult to interpret the data, going against the role of a data scientist whose goal is to be able to present data in a way that everyone can understand. This suggests that there are many neutral views among the surveyed population. Humans are naturally followers, this means that many will gravitate towards the moderate option in order to conform to the socially acceptable views. Additionally, there are individuals who don’t fully believe either side's views, so they probably chose “moderate” in order to find a way to most accurately capture their stance. By not totally adhering to a single stance but supporting certain ideas from both sides, these people aim to find a middle ground. The recurrence of selecting the “moderate” option could also signal a general discontent with the current political offerings. These individuals feel as if none of the sides fully convey their full beliefs, resulting in them to select “moderate” as a form of ambiguity. The prominence of the moderate category is a reflection of the evolving nature of political identity in society. It signifies a shift towards a more neutral and individualized approach to politics, where the current traditional labels no longer suffice to capture the complex and personalized view of these individuals. This trend in individualism challenges policy makers and political parties to rethink how they interact with the public, suggesting a need for change.

Through the comprehensive analysis of the General Social Survey, a better understanding between personal attributes and political views was attained. The bias towards selecting moderates across the dataset challenges the current political sides, advocating for a refined image and understanding of politics. As the relationships between certain variables were observed, it

can be seen that the subject of political orientation is composed from a diverse platter of life experiences, education, and morals.

CONCLUSION

In conclusion, the analysis of the General Social Survey revealed significant trends between voters belonging to more extreme sides of the political spectrum and their age, happiness, and education. Liberals tend to have higher levels of education, particularly at the graduate level, while conservatives tend to report higher levels of happiness. Liberals also tend to be younger while conservatives are typically older.

Ultimately, we believe that the findings should serve as a warning sign for conservative politicians. With higher education becoming more readily available and younger people opting to vote blue, the Democratic Party seems to possess an advantage in the demographic that will make up the future of American citizens. While higher happiness levels seem like a benefit conservatives could cite to attract voters to their cause, the self-reported nature of these scores make it difficult to determine whether those happiness rates were due to an actual better quality of life or if they were skewed upwards by a hidden, confounding variable such as religion which can positively affect worldview. Obviously, there are many other factors that determine a presidential election such as the candidates, the current state of the country, and campaign finances, however, voter demographics can serve as a solid indicator of which candidate's platform may be more appealing. Another important issue of using certain demographic qualities to predict political allegiance, and by extension an election, is the presence of an overwhelming moderate population. Just like this study shows, it is difficult to categorize these voters because they share both conservative and liberal characteristics. This concern, however, only serves to

reinforce the conclusion that liberals may have an advantage in elections in the future. If neither party can consistently rely on moderates during elections, then the party with more aligned voters should have the advantage.

One behavior the group would like to investigate further after conducting this analysis is if people that identify with a certain political party maintain those beliefs as they age. Political scientists often refer to the tendency for older people lean conservative and younger people to lean liberal as the “generation gap”. We would like to know whether that phenomenon is true over time and people switch from being liberal to conservative as they age or if older people are holding on to their beliefs from a significantly different political climate. If people do not have strong political party allegiance over time, the conclusion of Democrats having an advantage in demographics will not be as future-proof.

APPENDIX

No other resources were used within this paper, however additional graphs were considered and can be found within the visualizations ‘.ipynb’ file.