Siddhida Pandya, Haven Biddix, Saurish Reddy

Section: 12:30

POLI 281, Fall 2023, Final Project

I. Introduction

We decided to utilize the dependent variable "Who did you vote for?" and the independent variable regarding religious participation to investigate how the religious status of voters impacts how they voted in the 2016 election. Specifically, we would look into whether more religious people voted for Donald Trump than less religious people. Furthermore, this will prove if religious voters have a positive correlation to voting for Donald Trump, or if non-religious or less religious voters have a positive correlation to voting for Hillary Clinton or another candidate. This may be the case due to voter bias geared towards candidates with elements of their own identity and representation, specifically religion. We observed that the Republican Party, of which Donald Trump is a part, tends to value religion, particularly Christianity, which is the major religion in the United States, more than the Democratic Party.

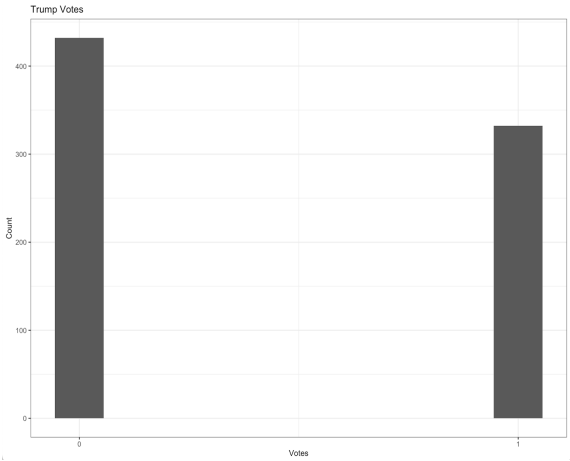
II. Introducing your independent variable and dependent variable

"Who did you vote for?" is a question that can produce important observations specific to the voting behavior of the 2016 election. This is because, by surprise, Trump ended up achieving triumph over Hillary and winning this election, all due to the votes of the citizens of the United States. Each citizen, however, comes from a variety of different religious, ethnic, and income backgrounds that shape their vote for their specified candidate. Thus, those candidates that people see upholding their ideals are more likely to get their votes. Who these citizens voted for is vital to figuring out the relationship between the independent variable to the result of the election, and everything in between. Thus, by using the dependent variable that analyzes who each vote went to, we can tie together the independent variable of religion to see how religious beliefs impact voting choices.

Religious beliefs have played a significant role in shaping the political preferences of many individuals, influencing their decision to support candidates like Donald Trump. One key factor is the

alignment of Trump's policies with conservative values commonly held within certain religious communities, particularly evangelical Christians. Trump's stance on issues such as abortion, traditional marriage, and family values resonated with religious conservatives who prioritize these values based on their faith. Many religious voters also perceived Trump as a defender of religious freedom, viewing his administration as a bulwark against perceived encroachments on their ability to practice their faith freely. Additionally, the intersection of religious identity with Trump's policies created a narrative of moral alignment and cultural preservation, contributing significantly to the decision of religious individuals to support him.

III. Prep work: the dependent variable

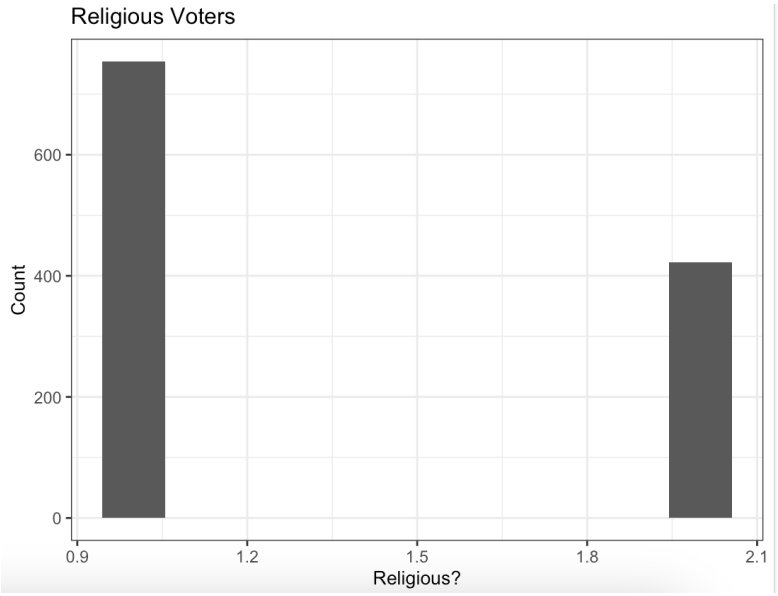


The figure in part A shows us the number of people who did and did not vote for Trump. These respond to the question "Did you vote for Trump?" where 1 is yes and 0 is no. From the figure in A, we can see that fewer people voted for Trump than people who did not vote for Trump. The negative numbers in the dataset represent no data, so these numbers were coded as NA. We are not worried about item nonresponse, because it would not significantly change the figure, as we already know that Trump did not win the popular vote, and 1 in this figure would remain the minority.

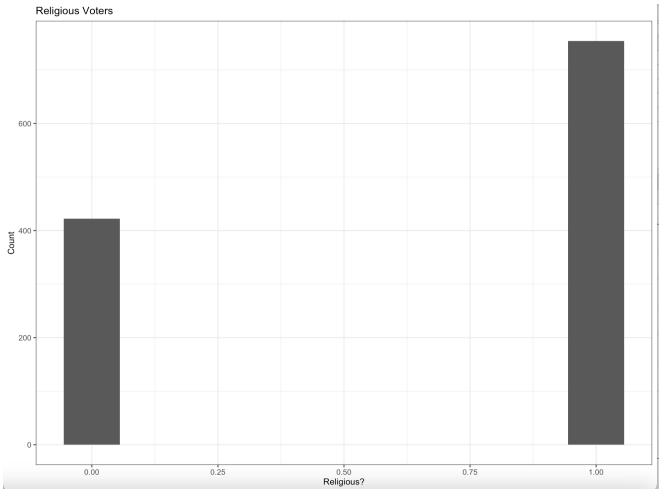
We coded all responses less than 0 as NA, as the negative numbers represent no data. We felt that, because the amount of "no data" responses was a minority, the inclusion of these numbers did not

advance our dependent variable. Further, the dependent variable has a binary response that does not include values that represent a lack of data.

IV. Prep work: the independent variable

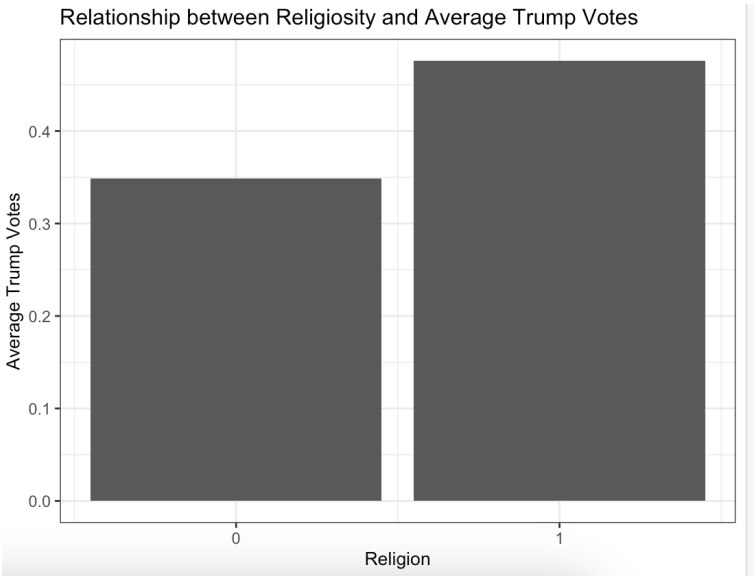


The bar graph illustrates the binary voting preferences, with values of 1 representing those who practiced religion by attending religious events and values of 0 representing those who did not practice religion (did not attend religious events).. The analysis of this graph reveals insightful patterns in the respondents' voting behaviors. Approximately 425 individuals indicated a decision not to practice religion, as reflected by the height of the corresponding bar representing the binary value of 0. Conversely, the second bar, showing a binary value of 1, indicates that around 750 voters expressed their interest in religion by attending these religious events. This stark contrast in the number of respondents favoring and opposing religiosity suggests a notable divergence in the electorate's sentiments, in this case, the sentiment of religion. We are not too concerned with item nonresponse here because we coded in a way where any numeric values under 0, are nullified and assigned NA. These are the people who decided they did not want to answer the question. We are not too worried about these respondents as they are not meaningful to our overall study/analysis.



We decided to recode the variables by assigning any numeric value below 0, NA. We decided that because the negative numbers were a minority value in comparison to 0 and 1, there was no need to analyze those values. These values also denoted whether a voter refused or was unsure, which is not significant/relevant to our data and statistics. Moreover, we were focused on those who answered Yes (1) or No (0). So it was acceptable to nullify the numbers ranging in the negative numeric values. We also recoded our code so that the numeric value 2 equals 0. Because we are working with binary values, rather than 2 meaning “No”, it was easier for us to visualize the value 0 as it aligns with the binary system.

V. Initial analysis: data visualization



The x-axis of this figure represents religion, wherein 0 is non-religious and 1 is religious. The y-axis simply represents average voters who voted for Trump. From initial analysis, it is clear that a majority (nearly half) of Trump voters are religious, meaning we can logically correlate the level of religiosity and voting for Trump. This figure alone supports the idea of a positive correlation between being religious and voting for Trump due to the average Trump votes that correlate to each version of religiosity. Thus, from this initial analysis, we can see that religiosity does seem to have an impact on the dependent variable. There seems to be a relationship between those who are religious having a higher amount of average Trump votes than those who are not religious. This is because, as we predicted, it seems that more people who tended to express religiosity had an increased amount of average Trump votes and those who did not had a decreased amount of Trump votes. Thus, this does match our initial expectations of those that who expressed more religiosity voted for Trump more, on average.

VI. Initial regression analysis

	Voted for Trump
	Fit1

Slope	0.128
	(0.038)
Intercept	0.348
	(0.032)
N	763
R^2	0.0144

The slope is positive, meaning that there is a positive relationship between being religious and voting for Trump. Because religiosity is coded as a binary variable, wherein 0 is non-religious, the intercept (0.348) represents the average Trump vote for non-religious voters. The R^2 shows that 1.4% of the dependent variable, voted for Trump, can be correlated with religiosity. The number of observations, N, is 763, which is less than the total number of rows in the dataset, but this is to be expected because we restricted this linear regression model to only individuals who voted for Trump and whether or not they were religious.

VII. Introducing: confounders

We are highlighting four confounders in our analysis: geographical location (including specific regions and the distinction between urban and rural settings), socioeconomic status, race/ethnicity, and media influence. When diving into confounders and their specific roles the first thing that must be looked at is whether the confounding variable is associated with the independent and dependent variables. The second is whether the confounding variable distorts the association with the independent and dependent variable

In examining the relationship between religious affiliation and voting for Trump, it's crucial to address potential confounding variables. Geographical location, including specific regions and urban-rural

distinctions, emerges as a critical factor influencing both religious beliefs and voting behavior. Certain religious affiliations may be more prevalent in distinct geographic areas, introducing a potential confounding influence. Recognizing this, researchers can implement strategies to control for geographical variations and ensure a more accurate understanding of the connection between religiosity and voting behavior. Additionally, socioeconomic status adds complexity by influencing both religious beliefs and political preferences. Varying economic circumstances shape values tied to religious affiliations and the likelihood of supporting Trump, allowing researchers to navigate confounding influences and uncover the unique contribution of religious identity to voting decisions. Race and ethnicity is another confounding variable that can skew the relation between being religious and voting for Trump. Depending on the race or ethnicity you are and your likelihood to vote for Trump, some races may be more inclined to favor Trump due to certain policies promised to be implemented. On the flip side, some ethnic groups and races may be affiliated with a certain religion that does not accept certain things that Trump has said or done. These are all potential examples and factors to consider when investigating how the confounding variable can distort the relationship between our independent and dependent variables. Furthermore, media influence serves as a crucial confounding variable, shaping both religious beliefs and political opinions. Acknowledging this role is vital for analysis, enabling researchers to account for external factors impacting religious identity and voting preferences. Certain media outlets push specific political affiliations which can be a red flag, especially in distorting the variables.

Gender, education level, age, and economic opinion may act as confounding variables in the relationship between religious affiliation (independent variable) and voting for Trump (dependent variable) due to their potential associations with both factors. Gender-specific patterns in religious beliefs and voting behavior, the influence of education on worldviews, generational variations in religious affiliation, and the intertwining of economic opinions with political preferences can independently impact the observed associations. Failing to account for these variables poses the risk of misattributing the relationship between religiosity and voting for Trump, emphasizing the need to consider these factors for a more accurate and nuanced understanding.

VIII: Prep work: confounders

- Gender:
    - First, we added gender to the anes dataset that we created, and renamed its original label “V161002” to “gender”.
    - We then changed the female variable, 2, to 0 in the dataset so that it would be a binary variable where the female is 0 and the male is 1.
- Age:
    - We again added the V161267 variable to the anes dataset, and recoded it as “age”.
    - All of the variables under 0 were useless to the dataset, so we recoded those values as NA.
    - Finally, we rescaled the variables so that the ages ranging from 18 to 90 were placed on a scale from 0 to 1.
- Economic:
    - We added the V161140 variable to the anes dataset, and renamed it to be “economic.”
    - We recorded all of the data that says “don’t know” or “stayed the same” to be NA, as it seemed to hinder the point of the data, not enhance it.
    - Finally, we changed the variable 3, which is “Gotten worse” to 0, which means that if one thinks the economy has gotten better, they are 1, and if it has gotten worse, it is 0.
- Education:
    - We added the V161270 variable to the anes dataset as “education”.
    - Then, we recoded the numbers above 16 as NA, and the numbers below 0 as NA, as they do not contain important data for our analysis.

- Finally, we rescaled the remaining numbers to be on a scale from 0 to 1, so we rescaled it for simplicity to show the overall range of education levels.

	Min	Max	Mean	Median
Gender	0	1	0.4746	0
Age	0	1	0.4425	0.4444
Economic	0	1	0.4964	0
Education	0	1	0.6623	0.6667

Correlation Matrice

	Religion	Trump_votes
Gender	-0.0702	0.0727
Age	0.0974	0.1676
Economic	-0.1367	-0.6786
Education	0.0507	-0.0945

Based on these matrices, we think that the variable that is most likely to be a confounder is Education. When looking at the matrix that shows the correlations, we can see that the independent variable that has the highest correlation to both our independent variable, religiosity, and the dependent variable, Trump votes, is the economic variable. It shows that there is a strong correlation between whether people think the economy got better or worse and whether they voted for Trump. Notably, since it is negative for Trump votes, we can see that when they thought the economy got worse under a Democrat President, Obama, they tended to vote for Trump. Hence, the negative coefficient in the correlation. It does not change the discussion we wrote due to the fact that we stated that those independent variables, including the economy, do seem to have an impact on how people would vote,

which could potentially outweigh the impact religiosity has on Trump's votes. It was established as a confounding variable in our opinion, as well as in the analytical confounding tables.

IX. Multivariate regression

	Voted for Trump?	Voted for Trump?	Voted for Trump?
	fit1	fit2	fit3
Religiosity	0.128	0.035	0.035
	(0.038)	(0.037)	(0.038)
Economic opinion	–	-0.665	-0.645
	–	(0.035)	(0.037)
Education	–	–	-0.102486
	–	–	(0.111)
Age	–	–	0.166
	–	–	(0.074)
Gender	–	–	-0.007
	–	–	(0.035)
Intercept	0.348	0.756	0.741

	(0.032)	(0.039)	(0.092)
N	763	440	432
R^2	0.014	0.461	0.466

To summarize this regression table, it focuses on three different types of linear regression models that were executed. First, it was simply the bivariate regression done with the independent variable we chose, religiosity, and the dependent variable, whether Trump was voted for or not. Then the next one introduced economic opinion as another independent variable, and the third one introduced gender, age, and education on top of religiosity and economic opinion. These individual columns were created to show the change in the coefficient of each independent variable as more confounding variables were added. It also shows the progression of the change in the intercept, the R-squared, and the number of observations.

First and foremost, while analyzing the change in intercept while moving across columns, there seems to be a significant increase in the value of the intercept as more independent variables are added to the analysis. Thus, when the independent values are 0, there seems to be an increase in the value of the dependent variable. This can be analyzed by recognizing that as more confounding variables are added to the regression analysis, the value of the dependent variable when the independent variables are 0 tends to increase.

Additionally, our main independent variable, religiosity, was noted to significantly decrease across the three columns. The initial analysis was a coefficient of 0.128, but as more independent variables were added, the coefficient decreased to 0.035. This is an interesting result because it shows that the effect of religiosity on voting for Trump significantly decreased when economic opinion was added to the analysis, which means that it is a very impactful confounding

variable. Because the impact of religiosity decreases as the impact of economic opinion increases, it shows that religiosity does not have as much of an impact on whether or not a person will vote for Trump as much as whether they think the economy is better or worse will. Additionally, the fact that the impact of religiosity did not decrease or increase in the third column, but rather stayed the same shows that the other independent variables that were added to the analysis do not have as much of an impact on the dependent variable as economic opinion does. It shows that economic opinion decreased the impact of religiosity on Trump's votes, but the other independent variables did not manage to decrease the coefficient of religiosity on Trump's votes. Therefore, I do not believe that they were serious confounders to the analysis since their impact on the independent variable that we chose to analyze, religiosity, was minimal and not statistically significant.

These changes truly impact our perception of the independent variable that we chose to analyze. The fact that the coefficient for the slope of religiosity in the regression model significantly decreases with the addition of a highly likely confounding variable displays that there is not as strong of a correlation between religiosity and Trump votes as there is between economic opinion and Trump votes. This shows that confounding variables do have a very strong impact in our initial comparison which shows that religiosity is not a reliable or strong comparison or explanatory analysis to utilize in explaining why Trump won the election. Thus, when the analysis was adjusted for potential confounders it showed that our main independent variable, religiosity, did not have as much of an impact on Trump votes as it did when no confounders were involved, which changed our perception of how truly impactful the variable we chose is.

X. Conclusion

In our hypothesis, we predicted that being religious and voting for Trump would have a strong positive correlation. If this were true, religion and Trump votes would have a correlation coefficient at or near 1. However, once we coded the linear regression model for these two variables, we found that the correlation coefficient was closer to 0 than to 1, which signifies a weak correlation. The main reason why we predicted a strong correlation between these two variables was because of everyday observations of how Republican candidates are often more outwardly Christian than Democrats, and Christianity is the majority religion in the United States. However, after finding the weak correlation, we inquired that religiosity either doesn't have as strong of an impact as other potential independent variables and that the values of various denominations within Christianity and non-Christian religions may have influenced some of these voters to vote for another candidate.

When analyzing other independent variables, we found that gender and Trump votes had the weakest correlation. We found that opinion on the economy during the last year had the strongest correlation with whether or not someone voted for Trump. We hypothesized that people who felt that the economy had improved over the past year were likely in a higher income bracket. In contrast, people who felt that the economy had gotten worse were likely in a lower income bracket. The dataset could be further improved by including the economic income of voters. Furthermore, our policy recommendation focuses on taxation, more specifically raising the tax on the top 5% tax bracket, which we feel would cause a more even distribution of voting results.

Overall, I would conclude that our analysis is thorough. It tests our research question of whether there is a correlation between being religious or not and the likelihood of voting for President Trump. We used numerous methods to strengthen our analysis. This includes linear

regression, histograms/bar charts, and multivariate regression. Our code demonstrates these methods when you run it. When running it, you can see the various formulas and processes we take to strengthen and support our research question. Based on this, we feel confident in concluding that we have a very strong analysis. However, despite investigating four potential confounders, we are still concerned about alternate confounders. Referring to part VII, we can see that other confounders include geographic location, socioeconomic status, race/ethnicity, and media influence. Although these exist and can MAYBE distort our results, it is not something to be overarchingly worried about. For a research question on the influence of religious beliefs on voting for Trump, concerns about confounders like geographic location, socioeconomic status, race/ethnicity, and media influence can be mitigated. Given the specificity of the question, these factors are less likely to distort the relationship substantially. Focusing on the direct connection between religious affiliation and voting behavior allows for a more targeted analysis, reducing the need to overly worry about extraneous variables that may have a minimal impact in this context.

Moreover, we can see that this is the case in this analysis where we would much rather hyperfocus on the correlation between the independent and dependent variables, to counteract getting distracted with confounders that CAN POTENTIALLY be distorting (which is not likely).

Despite these confounders, we believe that our measures for our independent and dependent variables are sufficient. For starters, we have a very reliable sampling population. It is fixated on those who voted in the Presidential election and nullified (NA) respondents who responded “did not say”, “would rather not say”, “did not register”, etc. By nullifying those

respondents, we were able to focus just on those who voted for Trump or did not. This gives us very strong measures to analyze how the independent and dependent variables are related. Moreover, our sampling size is also very large which increases the statistical power of a study, enhancing the reliability and precision of findings. It helps capture the diversity within a population, reduces the margin of error, and allows for more robust generalizations. A larger sample size also improves the likelihood of detecting genuine effects, making research results more credible. Another consideration to make when looking at our measures and how they are sufficient is that we have performed a very thorough data analysis. Multivariate regression enhances the analysis by accounting for potential confounders, ensuring a more precise assessment of the relationship between being religious and voting for Trump. Charts visually illustrate patterns and aid in identifying trends or outliers, offering a clearer understanding of the data. Linear regression provides a quantitative measure of the strength and direction of the association, allowing for precise predictions. Together, these methods strengthen the investigation, providing a comprehensive and robust analysis of the influence of religious affiliation on voting behavior. There are no further caveats to be made because we are fairly confident about our analysis and the association between our independent and dependent variables.

Overall, I can say our group did learn a lot about the 2016 election, and what drove people to vote for Trump. We know now that though religiosity may have some correlation, economic opinion seems to be the independent variable with the most correlation to Trump votes. However, overall, I would say the biggest takeaway that we gained from this is how many factors go into political elections. There is religion, as we analyzed, but there are also academics, economics, age, gender, and many other factors that are not even included in this dataset that

may play a role in how people vote. There is much to be said about how political elections work from the local to the national scale; however, this project gave us an insight into how different aspects of voters' lives can impact the turn of an election in a very insightful manner.

XI: Individual contributions

In this project, all group members wanted to make sure that each part came together in an organized, understandable manner. To do this, we made sure to either work on the project together in person or work over Zoom and be in unanimous agreement on how to divide a task. Because there is no way to make a collaborative file on R Studio, Siddhida wrote a majority of the code on her computer after collectively agreeing on how we wanted to code the variables. To split the difference, Rish and Haven did a majority of the writing on the document. Furthermore, we all looked over each part of the final project as a means of offering feedback and ensuring equal collaboration. We worked on it in part in order, as a majority of the assignment builds upon previous sections, and it's nearly impossible to skip steps. To do this, we started both the R script and the document about a week in advance to effectively answer each part and to plan the division of labor.