Option #1: Capstone Project—Final Report: U.S. Organization

Henry Hopley

Colorado Status University – Global Campus

**ABSTRACT**

There are many considerations a political campaign must take when deciding what is important to the average voter. This paper attempts to take the three characteristics of race, religion and education and discover their ability to determine how a county in the United States will vote in a national Election. Knowing how a county typically votes can be crucial information to a campaign and can be used to further define strategy.

**RESEARCH HYPOTHESIS**

H0: There is no combination of characteristics that help predict which way a county will vote.

Ha: There is a combination of characteristics that help predict which way a county will vote.

**INTRODUCTION**

I selected the Democratic Party because I would love an opportunity to help more progressives elected. Democrats have always had policy that supports all those within the country, but focuses on providing a system of support for the least fortunate to have a stable and comfortable life. I believe there are inalienable rights that every human in the United States deserves to have without question and the Democratic Party is the organization that is best positioned to push those rights into law. Despite their best intentions however, I believe the size of the “Big Tent Party” is too large and old to easily adopt a platform that will attract the numbers they need to win elections. I believe that this country is headed down a dark path and without an infusion of new blood and ideas into the party, the Democratic Party is not long for this world.

There are a few primary methods that can help change the Democratic Party. The primary challengers in this election have pushed the party so hard that Joe Biden will have the most progressive platform in this country’s history. While it is amazing that he will be supporting so many progressive policies, he will need to be regularly pulled to the left in order to counter the moderates and the right. Data could be the key to provide information and actionable insight that gives the DNC the edge it needs to win the Presidential election as well as-as many down-ballot elections as possible.

**Services**

The various organizations of the Democratic Party provide several services related to political campaigns and party organization. There is established structure with paid employees across the country that provide operations support for voter outreach. These employees provide resources to grassroot efforts for supported candidates and policy. Those resources include massive data stores of all the information ever provided to them through surveys, polls, questionnaires, and public information such as voting history. This data is used by campaigns to find and reach out to potential voters in order to increase the chances they will support the Democratic candidate (About the Democratic Party – Democrats, 2019).

**Location**

There are Democrats across the United States but the headquarters is located in Washington D.C.

**Employee size**

There are approximately 200 members elected to positions across the United States for the DNC

**Revenue**

For the 2020 federal election, the Democratic Party has raised over $470 Million dollars. Historically, over a Billion dollars is raised for presidential election years. In 2017, it was revealed that the DNC was in a large amount of debt from the Obama campaign and the Hillary campaign had taken over the debt payments. While this was breaking news at the time, the DNC has since recovered with a healthy “war chest”. This money is collected through large fundraisers and grassroots donations.

**OBJECTIVES**

To determine if there is a correlation between various religions, races, and education levels in the results of national elections.

To analyze results of experimentation

To provide actionable insight on how results could be used in finding the right candidate and campaigning.

**OVERVIEW OF STUDY**

The first section of this report discusses how I concluded on the need to discover how a county might vote in order to help select the right candidate for a particular location. The second section discusses research methods and how this experiment will take place. I follow that with the third section explaining the findings from experimentation. In the fourth section I talk about the conclusions that can be made from my findings and in the fifth section, I recommend further investigation and action that can be taken from the conclusions made in this report.

**LITERATURE REVIEW**

As the partisan divisions grow in our country, there are fewer and fewer opportunities to improve voter turnout. As the NGP VAN field organizing guide describes, there are approximately one-third of total registered voters that are targeted the most during a campaign (Weiss, 2018). Knowing who to target and how many of those voters exist in a county can help determine how many resources are required there and how much effort a campaign should make. To further show this need, the special election California’s 25th Congressional district recently exemplified the need to have a candidate that inspires voters even in an environment like the Covid-19 pandemic (Sammon, 2020). From here it would be wise to discover what is important to people when they are choosing a candidate. It has been suggested and shown that there is a causal link between religion and politics, specifically between protestant evangelists and conservatism (Patrikios, 2008). Other arguments are made that race and education level are huge signifiers in determining support in an area, particularly white working-class voters, who vote for conservatives by a large majority (Thompson, 2019). With these statistics about a county known, would one be able to determine how a county will vote? Knowing that information could be vital in determining the amount of effort a campaign should put into a county? On top of that, if there are a large number of candidates, could knowing how a county will vote help party organizations choose which candidate to support?

**RESEARCH DESIGN**

With the historical datasets that I have including characteristics of voters within counties as well as the total number of votes in each county, I have a strong base to begin my research. Using predictive analytics will allow me to make connections within the data that may not be obvious from looking at spreadsheets. Regression will likely be the best predictive technique to determine correlation between variables and that can easily be conducted in SAS Studio. Multiple Linear Regression will be the best technique available that will easily tell if there is any correlation between the characteristics in my datasets and the results of the 2016 election.

There are three types of analytics that will be used in this project to provide insight. Descriptive analytics are used to determine historical trends and learn why something may have happened during that time. During this project, descriptive analytics will be used to analyze the data and determine what characteristics available correlated to the direction each county voted. Predictive analytics builds on those analytics and applies those trends to similar situations and can “predict future actions with a surprising degree of accuracy” (Relus Technologies, 2016). This technique would be used by taking the most effective predictors and applying them to current information to determine how counties might vote in 2020. Finally, prescriptive analytics helps determine the best course of action to take based on the available predicted outcomes. Prescriptive analytics would take the strongest predictors and advise how they could be used as well as exploring how they could be used to manipulate the results in November. All three are very important but arguably the most useful is predictive analytics.

To perform predictive analytics there are multiple techniques that can be used. To find the effect of the characteristics in my data on how a county votes, regression will be the best option. The other primary option are classification models which would help predict the outcome of each county. While that could be extremely useful for campaigning, the goal here is to find out what influences the results the most, which doesn’t fit into a neat category. Regression models on the other hand help predict the probability that a characteristic will have an effect in a county. This is far more important to improving get out the vote efforts as well as determining how to approach campaigning and which issues a campaign might want to focus on in a particular area.

By running significant variables through linear regression and multiple linear regression models, I should be able to reject my null hypothesis. I will start by running general statistical functions on all the data to find if anything immediately can be deemed worthy of investigation. After looking through the data and the statistical summary about it, I should have a fairly comprehensive list of variables that I believe will be worth further research. Those variables will put into a multiple linear regression model in SAS Studio as independent variables. This model will be run and based on the R-squared value and the p-value; I will be able to determine if there is any correlation with my target variable. The first regression I will perform will be a multiple linear regression on my target variable, the number of votes received by Barack Obama in 2012. I will remove variables that do not correlate in order to potentially improve the strength of the model. With those variables, I will then apply them to see if they also correlate with the turnout numbers for Hilary Clinton in 2016. I will then apply similar logic to the voters of Donald Trump and Mitt Romney by taking the variables that correlate with Trump and applying them to Romney. If I am able to create a strong correlation with all four candidates, then I should be able to reject my null hypothesis.

My project is quite lucky when it comes to security, privacy, and ethics. My data is publicly available and already aggregated into counties, thereby not showing any personal information about any one person. With public data, I don’t need to worry about the privacy of any individual as well. For the Texas Democratic Party, they use individual credentials to log in to their database and also have set tiers, to only allow certain levels of information to be visible based on the position of the user. My precinct chair has a higher level of permissions than I do in order for her to better coordinate outreach. Ethics will be slightly more challenging, but I have no problem stating my biases. I believe in science and because of that, do not want to give any reason for my findings to be doubted, due to my personal beliefs. Attempting to prove my biases wrong will be the best method to achieve unbiased results.

**FINDINGS**

There were 14 variables that were shown to be significant in finding a correlation with the target variable of voters per county for Barack Obama in 2012. 11 of them were different measurements of religious adherents while the others described education levels, the percent of the county that was considered rural, and the percentage of the county that was Hispanic. These variables were all < .05, ranging from <.0001 to <.0133. The R-Squared value for this model was .9283, indicating that 92.83% of the numbers for Obama in 2012 can be explained by the numbers in the independent variables. These independent variables were then applied in a multiple linear regression against the turnout numbers for Hilary Clinton, returning similar results. All but one of the variables used in the first model returned p-value of < .05. Muslims estimated per thousand returned a p-value of .0594 which is just over the significance level of .05. The R-Squared value for this model was .9420, slightly higher than that returned in the prior model.

The number of variables that accurately predicted the turnout for Donald Trump was smaller than the two Democratic candidates. There were only 8 variables that had significance level of < .05, ranging from <.0001 to <.0427. Three represented religious adherents, and the other five related to race and the percentage of rural voters. The R-Squared value for this model was .6462, indicating that 64.62% of the variance of Trump’s turnout numbers could be explained by the significant variables. This list of variables was applied to Mitt Romney’s turnout numbers and all variables returned p-values of <.05. The R-Squared value for this model was .6815.

**CONCLUSION**

Based on the results of the four models ran, there are significant variables that correlate with the numbers for all four presidential candidates. This leads me to reject my null hypothesis and accept my alternative hypothesis. It was important to test characteristics against the candidates individually in order to predict the outcome of a county for the democratic candidate. While there might be some benefit in finding a set of characteristics that could predict candidates on both sides of the aisle, the priority according to NGP VAN is to focus on undecided voters and low-engagement voters (Weiss, 2018). With that in mind, two sets of variables were able to show a strong correlation with the voter turnout in a county and the larger one occurred under the “Big Tent” party, further solidifying the idea that there are many different groups of people that are needed for a Democratic candidate to win the White House.

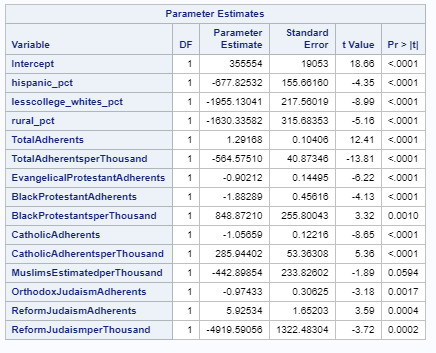


Figure 1. Independent Variables shown to Correlate with the Voter Turnout for Hilary Clinton

**RECOMMENDATIONS**

There are multiple actions that can be taken with this knowledge. Knowing there are specific characteristics that determine how a county can vote should lead to further research. This research could be done to perform a similar analysis based on a smaller sample size that possible represents individual counties or congressional districts. While it might be helpful to know important characteristics on a national scale like this report provided, it is only a stepping stone to far more detailed models. National information is rarely useful in down-ballot races due to the size of our country and the variety of people that live within it. Creating a template however that builds a model specific to a smaller area could be a key in learning the most important characteristics of the constituents that live there.

**REFERENCES**

BROWN JESSICA AUTUMN. (2016). The New “Southern Strategy” : Immigration, Race, and “Welfare Dependency” in Contemporary Us Republican Political Discourse. Geopolitics, History, and International Relations, 8(2), 22.

Davis, K., Patterson, D. (2012, September 13). *Ethics of Big Data.* O’Reilly Media Inc.

Flores, R. D. (2018). Can Elites Shape Public Attitudes Toward Immigrants?: Evidence from the 2016 US Presidential Election. Social Forces, 96(4), 1649–1690. <https://doi.org/10.1093/sf/soy001>

Thompson, J. (2019). A Review of the Popular and Scholarly Accounts of Donald Trump’s White Working-Class Support in the 2016 US Presidential Election. Societies, 9(2), 36. <https://doi.org/10.3390/soc9020036>

Jeakle, W. (2020, April 23). A New Poll Shows The Messaging Democrats Should Use To Defeat Trump. *Forbes*. <https://www.forbes.com/sites/williamjeakle/2020/04/23/democratic-poll/#6ed044bd1ac6>

‌NORRLOF, C. (2018). Hegemony and inequality: Trump and the liberal playbook. International Affairs, 94(1), 63–88. <https://doi.org/10.1093/ia/iix262>

Patrikios, S. (2008). American Republican Religion? Disentangling the Causal Link Between Religion and Politics in the US. Political Behavior, 30(3), 367–389. <https://doi.org/10.1007/s11109-008-9053-1>

Relus Technologies. (2016, May 13). *Descriptive, Predictive & Prescriptive Analytics - Relus Technologies - Medium*. Medium; Medium. <https://medium.com/@RelusTech/descriptive-predictive-prescriptive-analytics-d9aacb8ca5ce>

Sammon, A. (2020, May 18). How Democrats Lost a House Seat in California. The American Prospect. <https://prospect.org/politics/how-democrats-lost-a-house-seat-in-california/>

Vareca Ana. (2011). The influence of race, religion and class on election behavior in the United States of America (1930-2008). Sociologija, 53(1), 103–116. <https://doi.org/10.2298/SOC1101103V>

Weiss, G. (2018, August 17). *GUIDE: Field Organizing that Wins Campaigns*. Ngpvan.Com. <https://blog.ngpvan.com/guide-field-organizing-that-wins-campaigns>

‌