

Ryan Brady
Jacob Rabin

"The codes and results derived by using these codes constitute my own work. I have consulted the following resources regarding this assignment:" (Rselenium Github, R Stack Exchange,data.gov,politco)

Group Member's Responsibilities

While group projects are difficult to properly split up time and effort, all members of our group worked hard and contributed what they could to the project. We both aided in the development of ideas and discussion on how to execute our project. Downloading and figuring out how to use RSelenium was headed by Ryan Brady who then taught everyone how to use the program in order to webscrape on Politico. Jacob and Ryan found the data sets on data.gov on poverty and education. We all contributed to making the histograms and boxplots. Ryan constructed the election graph and education graph across the United states map. We both contributed equally in our analysis and report.

Goal of the Project

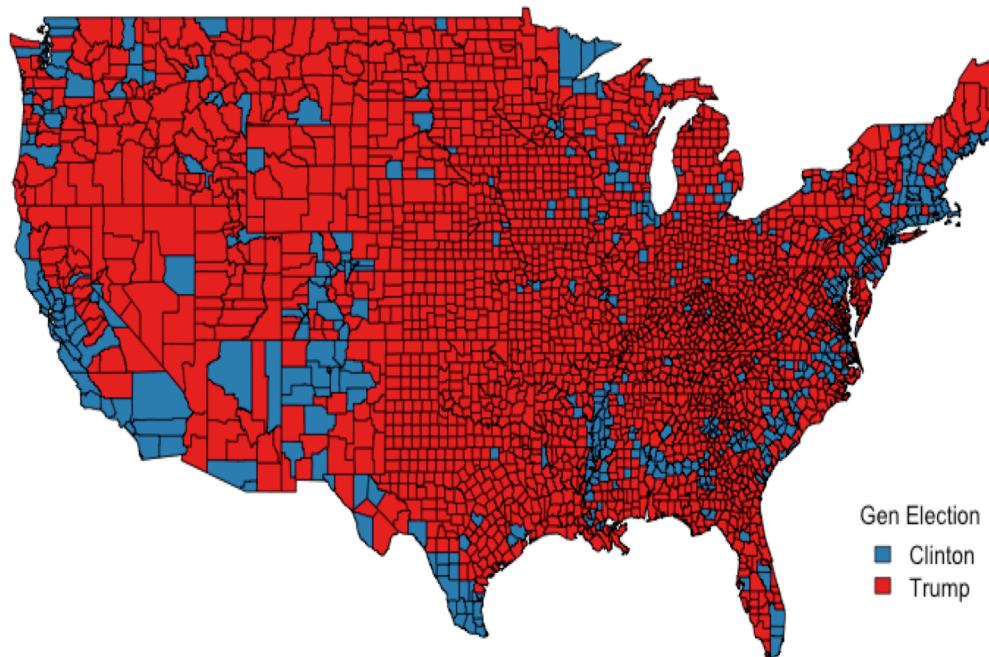
Our main goal of the project is to see the education rates and poverty rates across individual counties, and compare them to whom the majority of the constituents in those counties voted for in the 2016 General election. We attempt to answer the following questions: does the education rate of counties predict whether the county voted for Hillary Clinton or Donald Trump? How does poverty rates in a county influence how the people in that county vote? Answering these questions will help us develop more informed conclusions about the past presidential election. To answer these questions we utilize different histograms, box plots, and maps of the United States in an attempt to find a relationship between voting results, education rates and poverty rates throughout the counties of the United States.

Gathering the Data

For this project we used two different data sets, one came from data.gov which contained the poverty and education rates of each county in the United States. The other dataset was web-scraped from Politico. In order to get the data from Politico we installed the package R Selenium. Politico's website contains a javascript which loads all of the data slowly onto the page. This means we could not simply use the packages, rvest, XML, or xml2, since they only read the page source. R Selenium allows us to read the data after it loads on the page by using a program called Selenium which allows us to run a firefox browser through R. It actively takes elements from the page and reads them into R. After the data loads into R, we placed the elections results into a data frame of each of the counties and compared the amount of votes received for Clinton and Trump. After, we cleaned the county names taken from politico and matched them into the list of counties provided by the maps function in R.

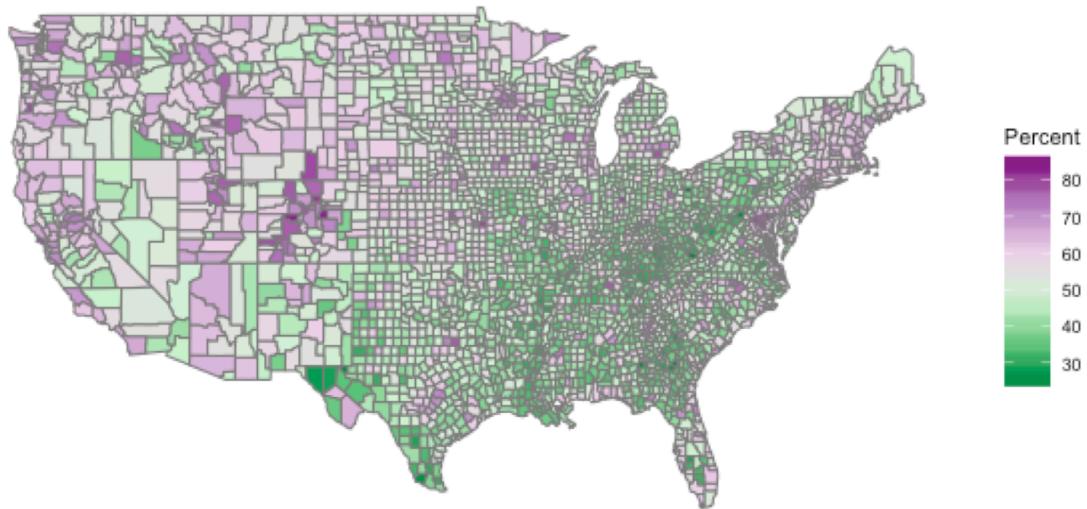
Report:

United States 2016 General Election Results

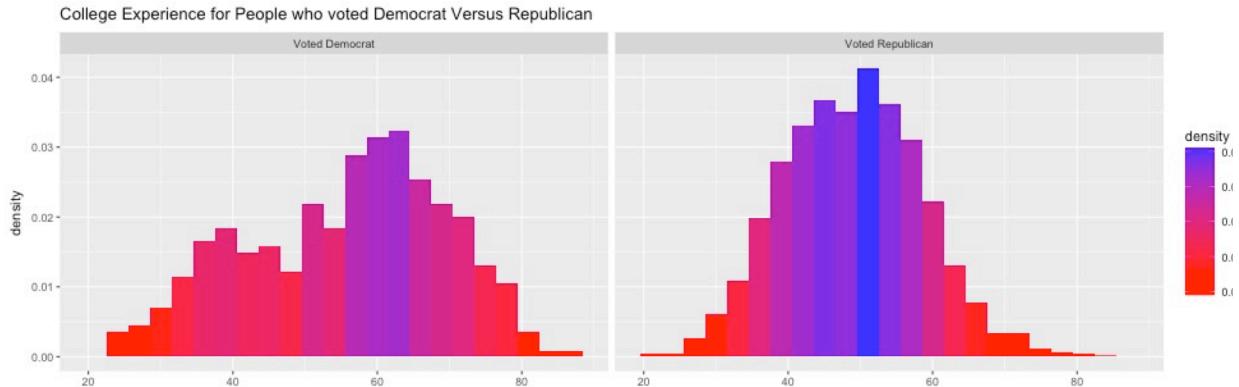


When observing overall election map we see that many of the counties with a majority of voters for Clinton were either in the North East, West Coast, and Border States. There are two major exceptions, a group of counties in Mississippi and Alabama as well as a group in Colorado and New Mexico. The blue counties in New Mexico and Colorado make sense due to the historic liberal voting tendencies such as Colorado decriminalizing the possession and recreational use of cannabis in 2014. The counties in Mississippi and Alabama seem to be all located in similar parts of their respective states. Most of the counties Donald Trump won tend to be inland and in the south, as well as counties that did not contain major cities. This illustrates how most of Donald Trump's voter base tends to live in areas where agriculture is more important and have a lower population density.

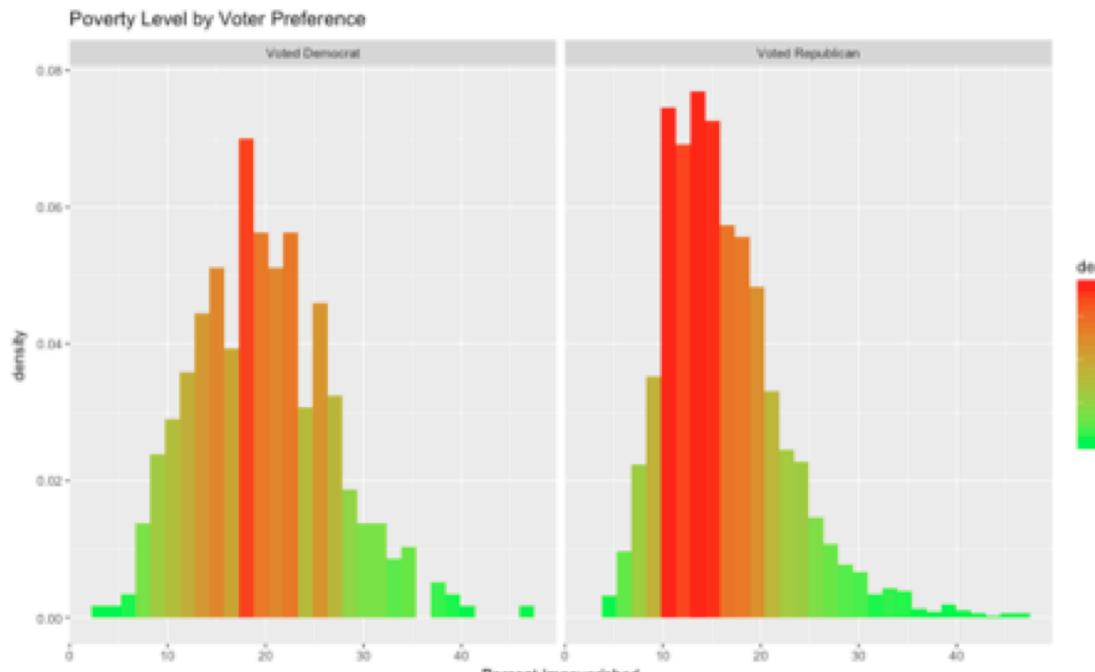
Percent of Citizens
With Associate Degree
or Higher



This map of the United States plots each county's education level (the percent of citizens who have an associate degree or higher). Comparing this graph to the results of the election, we observe that counties with an overall higher percentage of educated citizens generally voted for Hillary Clinton. This is easiest to see when looking at counties Clinton won in the far West, Midwest, and Northeast. Although there are some exceptions such as uneducated southern areas that voted for Clinton, this is most likely due to a higher percentage of uneducated immigrants that favor Clinton's policies. When looking at the uneducated populations that voted for Clinton in Alabama and Mississippi it would be useful to see the ethnic makeup of those regions. Also of note, the areas where Clinton won in Colorado and New Mexico have higher education rates than the rest of surrounding Middle-America. It is also interesting that there is at least one area in every state that has a much higher education rate than the rest of the state. For instance when looking at Minnehaha county in South Dakota it is a much lighter shade of orange than the other surrounding counties. Sioux Falls is contained within this county and contain one of the few higher education universities in the state of South Dakota. Thus while counties which voted for Clinton tend to have a higher education rate, there are still many counties dispersed throughout the United States that have high education rates and voted for Donald Trump.

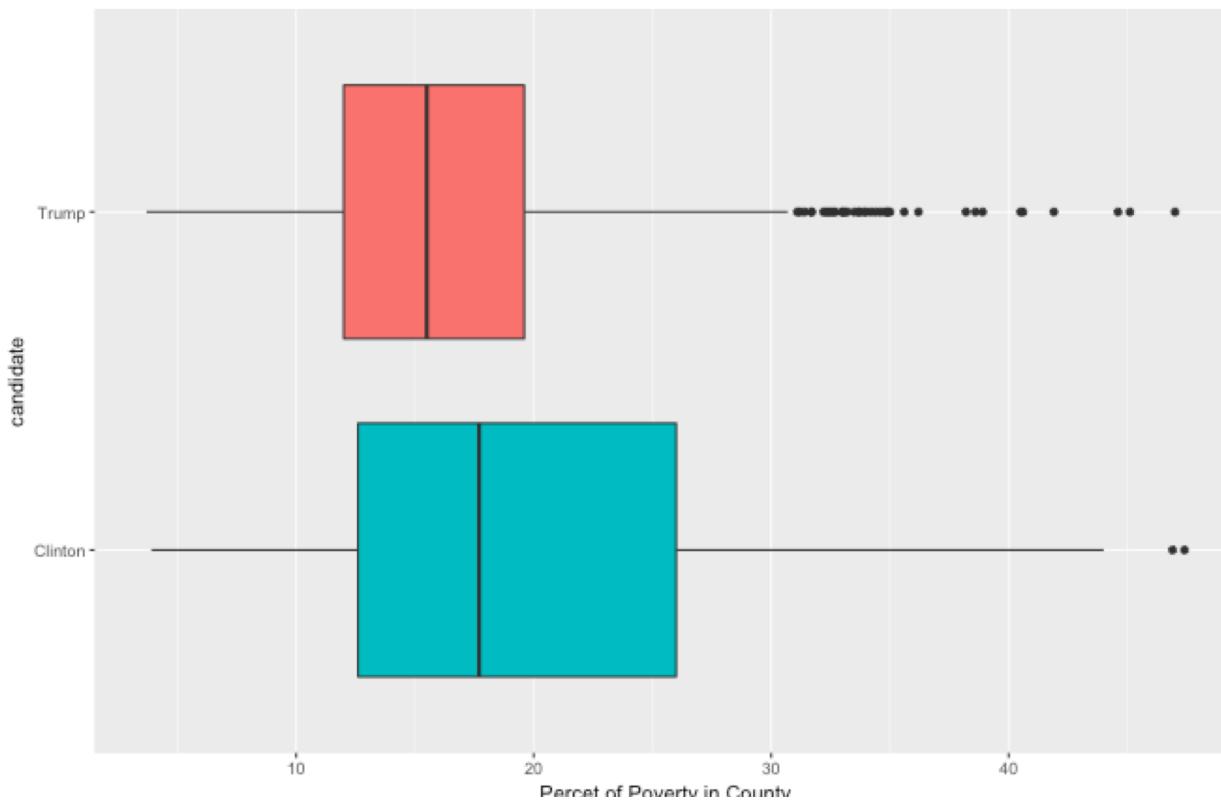


The graph above plots the percent of counties with an Associate's Degree or higher education against the county's voter preference. On the left plot, we see a bimodal distribution in counties that voted for Hillary Clinton. One minor peak is between 38 and 41% while the absolute peak is at 62 to 65%. For counties Trump won, the maximum peak is between 50 and 53% and has a fairly normal distribution. The overall average is 58.4%, so majority of counties that voted for Trump had an education level well below the overall average. These graphs also convey a larger range of education levels for voters in Clinton supporting counties compared to Donald Trump supporting counties.



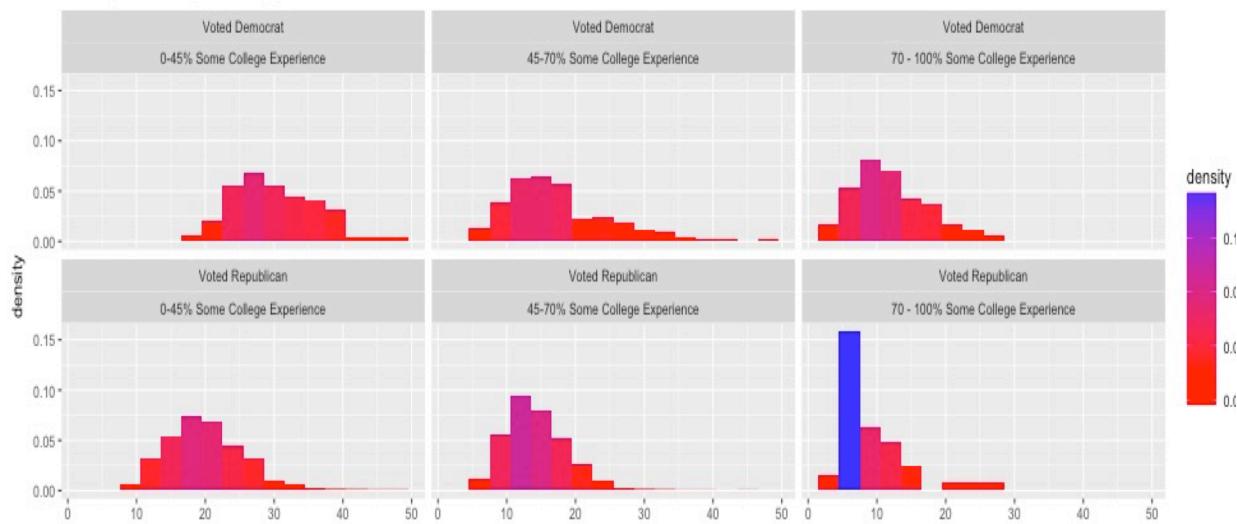
In the graph above, we observe the poverty levels in both blue and red counties. In counties that voted for Hillary Clinton, the average poverty level is 19.69%, while in counties that voted for Donald Trump the average poverty level was 16.25%. However, an even more interesting statistic is the difference in IQR's. For counties that voted Clinton, the IQR is 13.4 while the IQR for counties that voted Trump is only 7.6, meaning that counties that voted for Trump contains a much less deviated distribution in poverty levels. However, this could be due to the vast amount of counties Trump won which would lead to a lighter tailed distribution.

BoxPlot of Poverty Rates
Against Counties' Candidate



This boxplot exemplifies the disparity between the poverty rates of counties which supported Clinton and Trump. Clinton's median county poverty level was 17.7 percent compared to Donald Trump County's 15.5. We also see that the first quartile of Trump counties is slightly lower than that of Clinton's. Also the third quartile of Trump is 8% lower than the third quartile of Clinton; suggesting that a significant amount of impoverished counties voted for a democratic candidate. These two boxplots suggest that many of Trump's counties tend to have a similar poverty rates, the poverty rates of counties that voted for Hillary Clinton is much more dispersed.

Poverty Levels Against Party and Education Level



The graph above plots poverty levels against a county's general election vote and education. We binned education level by the percentage of Associate's Degrees in each county. Again, we can see a much heavier tailed distribution in counties that voted for Clinton instead of Trump. In lower educated areas (where only 0-45% of the population received an Associate's Degree), the average county that voted for Clinton has a significantly higher poverty level than those that have a lower poverty level. Some possible reasoning for this is that in areas with a higher poverty level and lower overall education, the population relies more on welfare and government programs (supported by Hillary Clinton) and farmers or blue collar works (counties with a lower poverty level, but never went to college) tended to vote for Donald Trump.

Looking at the graphs with a higher percentage of educated citizens, we can see there is a large spike in Trump support in areas with a lower poverty levels and high percentage of college educated voters. In these types of areas, people who tend to vote for the Republican party will value more conservative ideologies such as lowering income and corporate taxes. While Clinton also has support from counties with low poverty level and high education level, the distribution of poverty levels for those is definitely larger.

Conclusion:

In our attempt to see the effects county's education rate has on a county's choice for presidential candidate we found that there was a small correlation between education level and whether they voted for Trump or Clinton. Counties that tend to vote for Clinton were on average higher educated. However, there are definite outliers to this, many counties with uneducated immigrants also voted for Clinton. We also see from our histograms that poorer areas with low education rates tended to vote for Clinton; this may be due to the fact that these areas benefit greatly from larger welfare programs such as Social Security, Medicare, and the ACA, all programs sponsored by the Democrats. This also would help explain why we see in the histogram for education rate by voting preference, as bimodal for Clinton and unimodal for Trump. For each education level it appears that Trump's voting base tends to be more well off than Clinton's. This would also suggest that his voting base has less of a need for specific welfare programs like the Democrats would. To help further see how Education affected the results of each county, it would be beneficial to obtain ethnicity rates of each county, as well rates on gender, age, kids, immigrants, and religion. This would aid us in sub setting the data and eliminating the confounding variables which prevent us from seeing a significant relationship, if there is one, between education and voting results of the counties.