Economic Indicators and Voter Turnout

Group 12

Business Question

With the recent midterm elections and all the campaigning that proceeded it, voter power has been highly publicized. Although a single vote seems insignificant, enough people voting based on the same interests can create real change. If politicians and interest groups can identify the popular voter concerns, they can focus their campaigns around these topics and ideally inspire a large voter turnout in their favor. One such theme in political debates for democrats and republicans alike is the economy. The goal of this project is to provide insights on the validity of appealing to voters' economic concerns. More specifically, is there a relationship between the economic performance of a state and voter turnout? If the economy is down, are voters motivated to visit the poles and attempt to change our political situation, or do they feel a sense of hopelessness and decide to stay home?

We focused on data from two years – 2012 and 2016 – to do our analysis. The economy during the 2012 presidential election was still recovering from the crash 2008. It was no doubt improving, but the American public was still suffering from unemployment as a result of crash-related layoffs; the national unemployment rate hovered at around 8%. Additionally, growth slowed significantly in the second half of the year, primarily due to a 10% decrease in government spending. This governmental effect on the economy was a hot topic for the November election (Brown, 2013). In 2016, the economy had been consistently strong for a few years, but growth was starting to slow, and performance indicator data was falling behind economist predictions. According to report by the Washington Post, bond yield and interest rates were suggesting the beginning of a recession, and the US economy experienced the worst growth since 2011 at just 1.6% (Swanson, 2017). For both elections, the economy left something to be desired.

Based on the results from our analysis, politicians and interest groups can decide how to prioritize their campaign, potentially even adjusting how they publicize their economic agendas from location to location depending on the current economic health in the area. They can target certain states with commercials and speeches highlighting economic policies and emphasize their social politics in others. While the reasons Americans vote are complex with multiple different factors, for the purpose of this project we will focus on economic indicators – unemployment rate and gross domestic product per capita – to attempt to explain one motivation behind why people take time out of their day to visit the polls.

Raw Data

We used three separate data sources to pull voting data, unemployment data, and GDP data for each state in the years 2012 and 2016.

The election data comes from the United States Election Project, a subset of the united states electoral system focused on provided accurate voter statistics and electoral laws. The dataset includes year, state, ballots counted, voting eligible population, voting age population, and the percent of the population that are non-citizens and felons. You can find the original data here:

http://www.electproject.org/home/voter-turnout/voter-turnout-data (Links to an external site.)

The second dataset comes from the National Conference of State Legislatures and consists of the state name and the unemployment rates for each month of the year. The original data can be found here:

www.ncsl.org/research/labor-and-employment/2012-state-unemployment-rates.aspx

The last dataset on GDP comes from the Bureau of Economic Analysis which is a part of the U.S. Department of Commerce. This data was loaded directly from the table on the webpage and includes only the state and the GDP for the year and an Identifier ID for each state. To find the original data, visit this link:

2012:
https://apps.bea.gov/itable/drilldown.cfm?reqid=70&stepnum=40&Major_Area=3&State=00000&Area=XX&TableId=505&Statistic=1&Year=2012&YearBegin=-1&Year_End=-1&Unit_Of_Measure=Levels&Rank=0&Drill=1

2016:
https://apps.bea.gov/itable/drilldown.cfm?reqid=70&stepnum=40&Major_Area=3&State=00000&Area=XX&TableId=505&Statistic=1&Year=2016&YearBegin=-1&Year_End=-1&Unit_Of_Measure=Levels&Rank=0&Drill=1

Data Collection and Cleaning

There was a lengthy data cleaning process that needed to take place before any work could be done with the data. Because our question is very specific there are many parts of the data we did not need.

For the election 2016 dataset:

- Before reading in the file the data was examined and we assigned columns names in the dataset
    - These names include: “Date”, “State”, “Election Type”, “Party”, “Turnout Rate”, “Population VEP”, “Population VAP”, “Democrat”, “Republican”, “Minor”, “Total Ballots Counted”, “Notes”, and “Additional Notes”.
- The next step was to read in the csv file with the column names above
- The third step was to delete any unnecessary columns
    - This step includes deleting the following columns: “Notes”, “Additional Notes”, “Party”, “Democrat”, “Republican”, “Minor”, and “Total Ballots Counted”.
- After deleting any unnecessary columns, the first row was deleted since they have some headers

- The final two steps to clean the data before we can use it was to get rid of all the NAs in the dataset and to change the date to include on the year.

For the election 2012 dataset:

- Before reading in the file the data was examined and we assigned columns names in the dataset
    - These names include: “Date”, “State”, “Election Type”, “Turnout Rate”, “Population VEP”, “Population VAP”, “Democrat”, “Republican”, “Minor”, and “Total Ballots Counted”.
- The next step was to read in the csv file with the column names above
- The third step was to delete any unnecessary columns
    - This step includes deleting the following columns: “Democrat”, “Republican”, “Minor”, and “Total Ballots Counted”.
- After deleting any unnecessary columns, the first row was deleted since they have some headers
- The final two steps to clean the data before we can use it was to get rid of all the NAs in the dataset and to change the date to include on the year.

For the unemployment dataset:

- The unemployment data for 2012 and 2016 was initially only available as a PDF, so it was first copied into Microsoft Word. From there, it was copied into Microsoft Excel and saved as two .csv files.
- The columns created for this dataset include “State” and then each of the 12 months of the year. These column names were used for both the 2012 and 2016 datasets. The data was read in to two data frames, dfu12 and dfu16, using these names.
- A rows variable was created to delete unnecessary rows.
- A new columns variable, cols.num, was created so that when unemployment rate was calculated for each month, the columns were numeric. The numeric columns were applied to the 2012 and 2016 unemployment data frames.
- Next, the Average Unemployment column was created to calculate the average unemployment for each state.
- Lastly, “D.C.” was renamed to “District of Columbia” for uniformity.

For the GDP dataset:

- The GDP data (GDP by state) was pulled from the web. The 2016 GDP webpage was assigned to the variable “url” and then read into an HTML table. The same was done for the 2012 GDP webpage.
- These HTML tables were then converted into data frames and unnecessary columns were taken out. The columns containing the numerical GDP data were renamed “GDP” for clarity.

<u>Data Merging</u>

- The 2012 election and GDP data were horizontally merged using the merge function. The same was done for the 2016 election and GDP data. The resulting data frames were named “dfGDPelection12” and “dfGDPelection16.”
- Next, dfGDPelection12 and 16 were merged horizontally with the 2012 and 2016 unemployment data (dfu12, dfu16). Before this merging occurred, the trim function was used to trim white space in each of the column headers that were merged together. The resulting merged data frames were named dfAll12 and dfAll16.
- DfAll12 and dfAll16 were then vertically merged together using rbind to get dfAll.
- With all the data combined, it was ready to be exported to a .csv using the “write.csv” function. With all the metrics in one place, data visualization can now be performed.

<u>Data Visualizations and Results</u>

In order to create effective visualizations, multiple columns needed to be recoded as a numeric data type. To achieve this, some extra data cleaning was necessary. Please see the R script for exact code.

- First, gsub was used to remove punctuation like commas and percent signs.
- GDP, voter turnout rate, and both voter eligible and voter age population were changed the numeric type
- Clean the Election type data by using the grep function to determine the election type (Primary or Caucus)
- Additionally, problems with the data source caused a need to delete an extra row for Minnesota, as well as forcing NAs for Wyoming and Colorado since the original data was missing these two states.

After the data was adequately prepared, we gained some background on the economic and political climate at the time through a handful of visualizations. The first compared average voter turnout for both the primary election and caucus in years 2012 and 2016.

- We first grouped dfAll by election type and date using groupby
- With the grouped data frame, we made a summary table using the summarize function to calculate average voter turnout rate while ignoring NAs.
- Using as.data.frame, we turned the summary table into its own data frame which was then leveraged in a qplot function to create a simple scatterplot
    - The x-axis was election type, the y-axis was average turnout rate
    - Facets separated the graph into separate plots for each year
    - The graph was colored by election type
    - The axis names were added using + xlab and ylab, the font sizes were manipulated using + theme, and the plot was saved using ggsave.

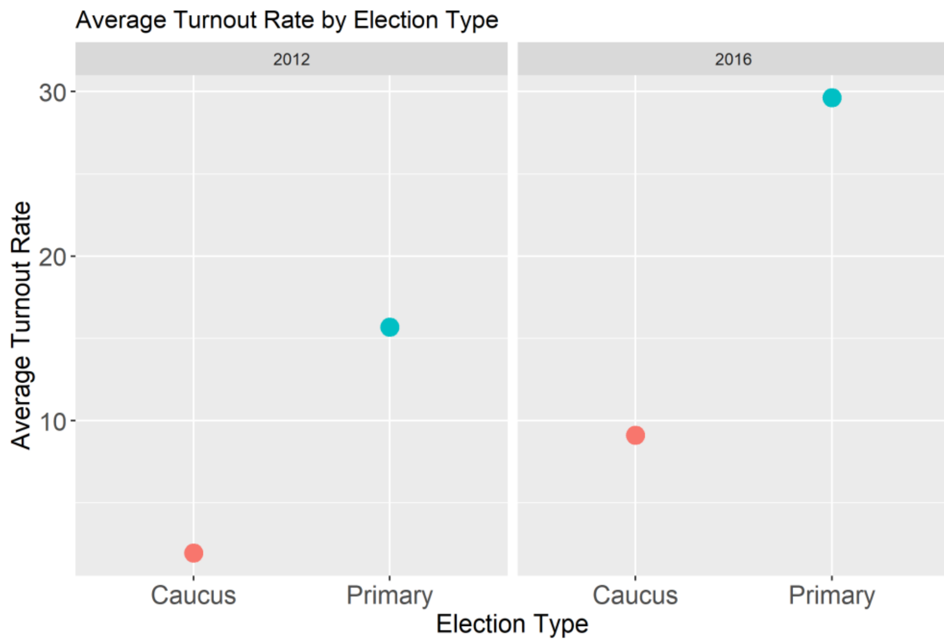


Figure 1: The Average Turnout Rate by Election Type for 2012 and 2016

This simple scatterplot shows that in both years, the turnout to the primary election was significantly better than the caucus. Additionally, **voter turnout was greater in 2016 than in 2012**, perhaps as a result of the divisive and intense presidential campaigns.

The second plot, a choropleth map, focuses on the average unemployment for each state. The same steps were repeated to create both a 2012 map and a 2016 map

- We again conducted minor data cleaning to remove all missing values and duplicate information
- To focus on one year at a time, we created the data frames 'mapAvgUnemp2012' and 'mapAvgUnemp2016' by selecting only rows from dfAll with the corresponding date.
- The average unemployment rate column's name was changed to 'value' and the state column to 'region' for compatibility with choropleth.
- Next, we loaded the state information from the choropleth package in RStudio and, for both the state and the 'mapAvgUnemp2012/2016' data frame, reordered state name to ascending order and forced all state names to be lowercase for consistency.
- Finally, we used the state_choropleth function to create the map, added a title using ggtitle, and saved it with ggsave.

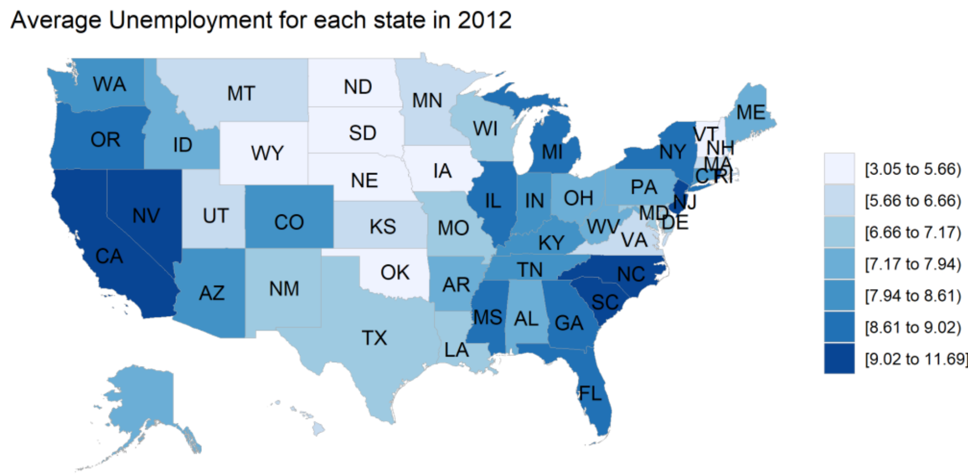


Figure 2: Average Unemployment Rate in 2012

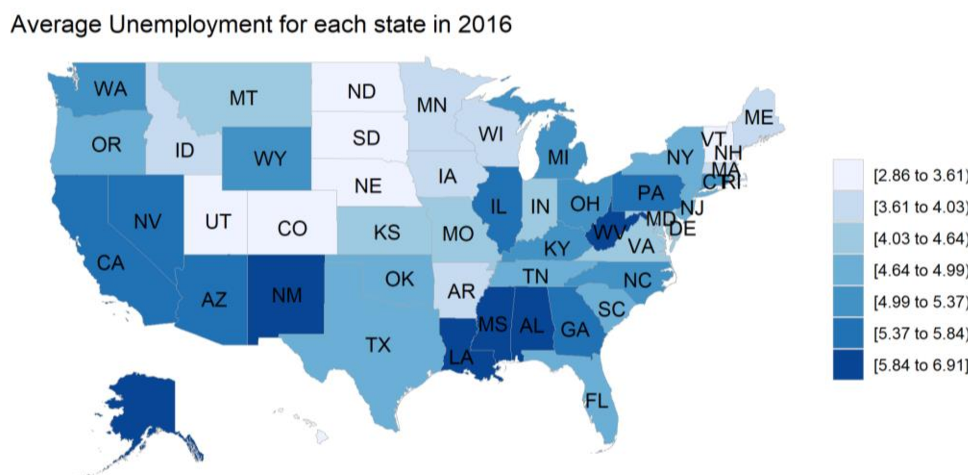


Figure 3: Average Unemployment Rate in 2016

Looking at the ranges for each color, **2012 had higher unemployment than 2016**. Unemployment was most heavily felt in the deep south and New Mexico in 2012, and in the Carolinas and California in 2016.

To begin our analysis, we focused on changes between 2012 and 2016 to gain insight. We completed a similar process for GDP, unemployment, and voter turnout to create more choropleth maps.

- For all three maps (GDP, unemployment, and turnout), necessary data was pulled for each year (2012 and 2016) from dfAll. Certain columns were renamed to specify the year for clarity after merging. For example, after pulling data for average unemployment from rows with a 2012 date, the column name ‘AverageUnemployment’ was changed to ‘AverageUnemployment12’ to avoid confusion later.

- Next, the data for 2012 and 2016 in all three cases were merged horizontally with the merge function.

After this step, the remaining processes diverged depending on the topic of the map. For unemployment data:

- We added a new column called ‘IncOrDec’ and assigned a 1 if average unemployment was greater in 2016 than in 2012 and a –1 if 2016 was less than or equal to 2012.
- Then, after ensuring compatibility with the state information downloaded from RStudio, we used state_choropleth to create the map and ggsave to save it.
    - Changed legend title to ‘Increase or Decrease’
    - Assigned two for the number of colors
    - Used scale_fill_brewer to set palette to ‘Set1’ and labels to ‘Decrease’ or ‘Increase’

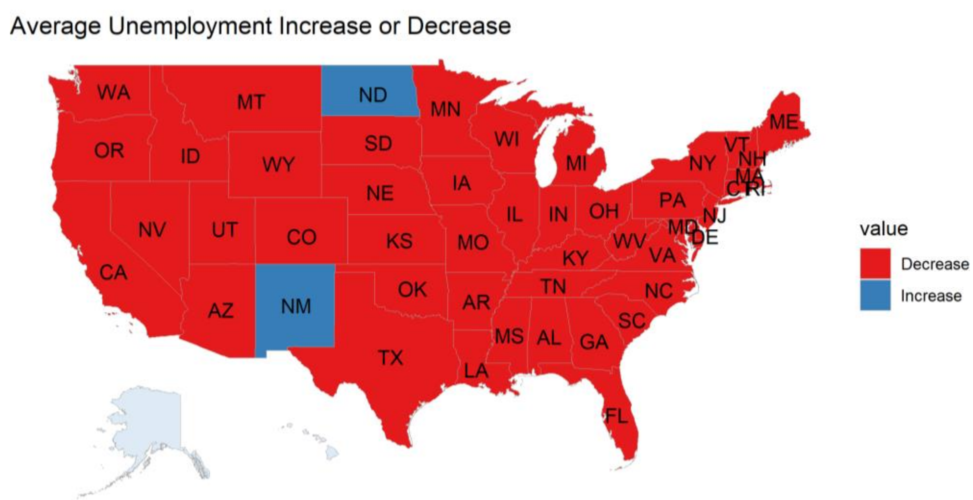


Figure 4: The changes in Average Unemployment rate for each state

For GDP data:

- Since this was the first-time using GDP data, we needed to create a GDP per capita column to account for the differences in population density from state to state. This was achieved by dividing the GDP column by the population data for each year.
- We then subtracted 2012 GDP from 2016 GDP to find the change.
- Finally, we repeated the steps to create a choropleth map based on state and saved with ggsave.

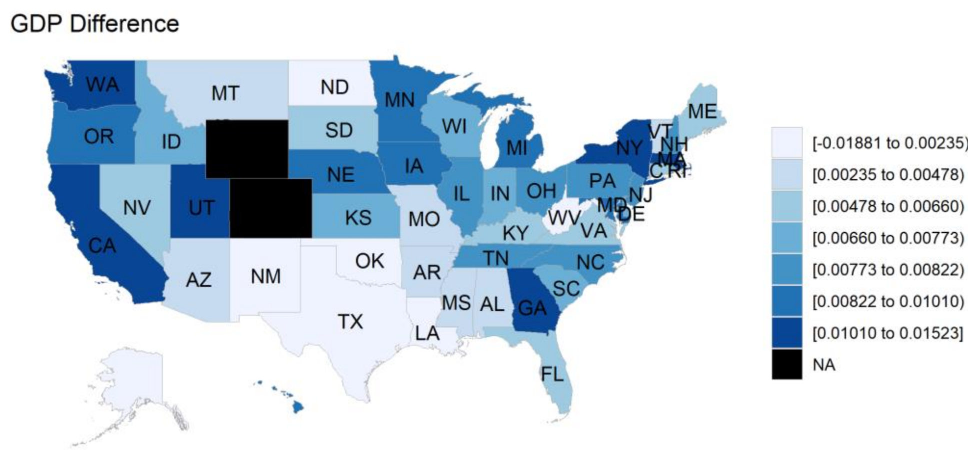


Figure 5: The GDP Difference of 2016 from 2012 in each state

For voter turnout data:

- First, we had to group the merged 2012 and 2016 turnout data by GeoName (state name).
- Next, we created a summary table using the sum function on both the 2012 and 2016 turnout rate columns to calculate the total turnout rate between both caucus and primary for each state.
- Once again, we added a ‘IncOrDec’ column and assigned either ‘Increase’ or ‘Decrease’ depending on if 2016 or 2012 had a larger sum.
- Lastly, we created a state choropleth map with ‘Set1’ color palette

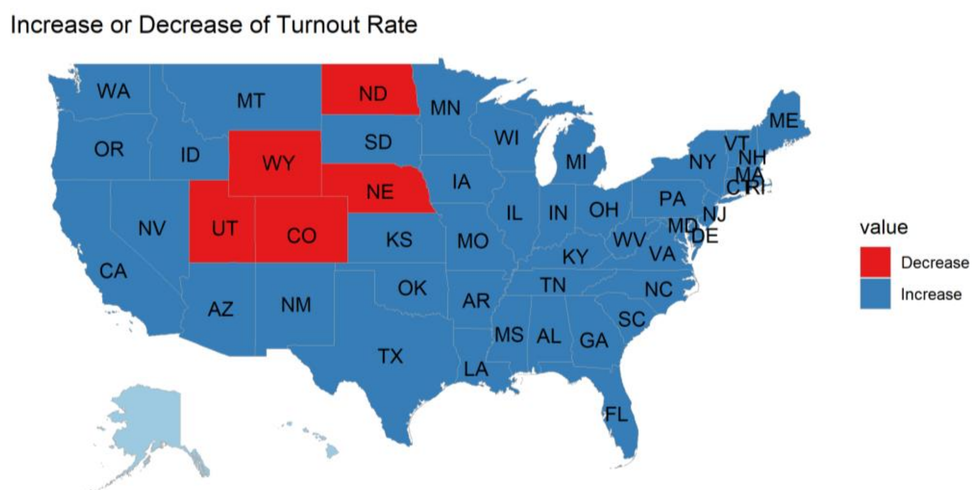


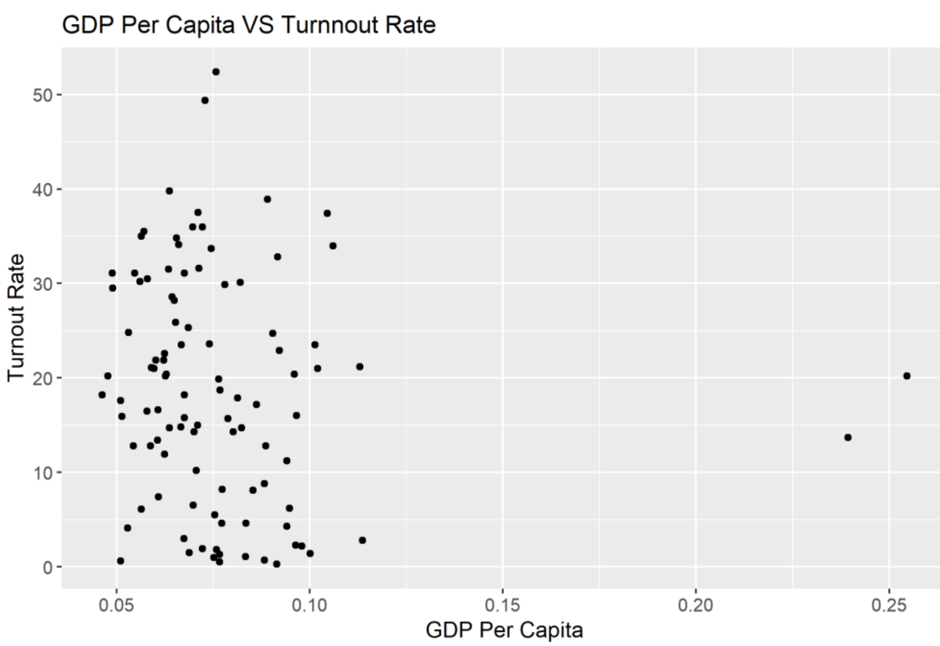
Figure 6: The changes in Turnout Rate for each state

Based on these maps, there are **no obvious conclusions** to draw regarding correlation, especially because the turnout ratio and unemployment maps show a strict increase or decrease. **Almost all states had a decrease in unemployment and an increase in voter turnout**, so it's difficult to determine if states with a large decrease in unemployment had a large increase in turnout or vice versa. It's also difficult to compare with GDP given that the GDP map indicates to what degree GDP changed. Some notable observations include North Dakota, as the GDP fell and Unemployment rose, but voter turnout decreased. This alone suggests that poor economic performance decreased voters' motivation to vote, but a single state cannot indicate an actual relationship by itself.

To gain more insight on the data, we also plotted the data as a scatter plot to determine whether there was any correlation between the three main features.

For GDP Per Capita VS Turnout Rate:

- First, we plotted a scatter plot using the qplot function with the GDP divided by Population data for the x axis and the Turnout Rate data for the y axis
- Then, we added the x axis label, y axis label, and the title for plot
- Next, the size of the axes label and the breaks were set to 12 and 10 respectively to ensure it is readable.
- Last, the plot was saved as a PNG file using the ggsave function



c

Figure 7: The GDP Per Capita against Turnout Rate for both 2012 and 2016

For Average Unemployment VS Turnout Rate:

- First, we plotted a scatter plot using the qplot function with the Average Unemployment for the x axis and the Turnout Rate data for the y axis
- Then, we added the x axis label, y axis label, and the title for plot
- Next, the size of the axes label and the breaks were set to 12 and 10 respectively to ensure it is readable.

- Last, the plot was saved as a PNG file using the ggsave function

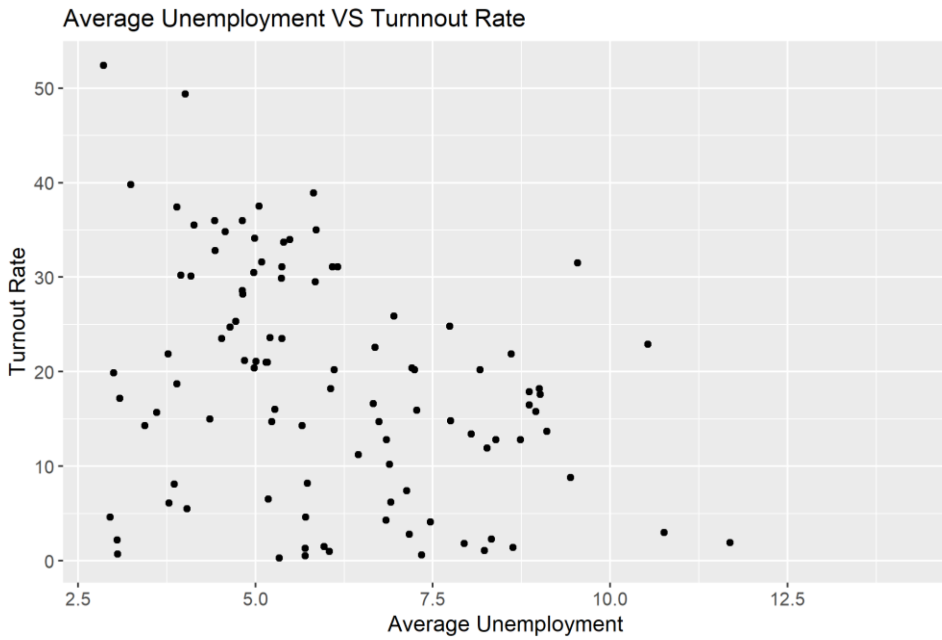


Figure 8: Average Unemployment against Turnout Rate for both 2012 and 2016

For GDP Per Capita VS Average Unemployment:

- First, we plotted a scatter plot using the qplot function with the GDP divided by Population data for the x axis and the Average Unemployment data for the y axis
- Then, we added the x axis label, y axis label, and the title for plot
- Next, the size of the axes label and the breaks were set to 12 and 10 respectively to ensure it is readable.
- Last, the plot was saved as a PNG file using the ggsave function

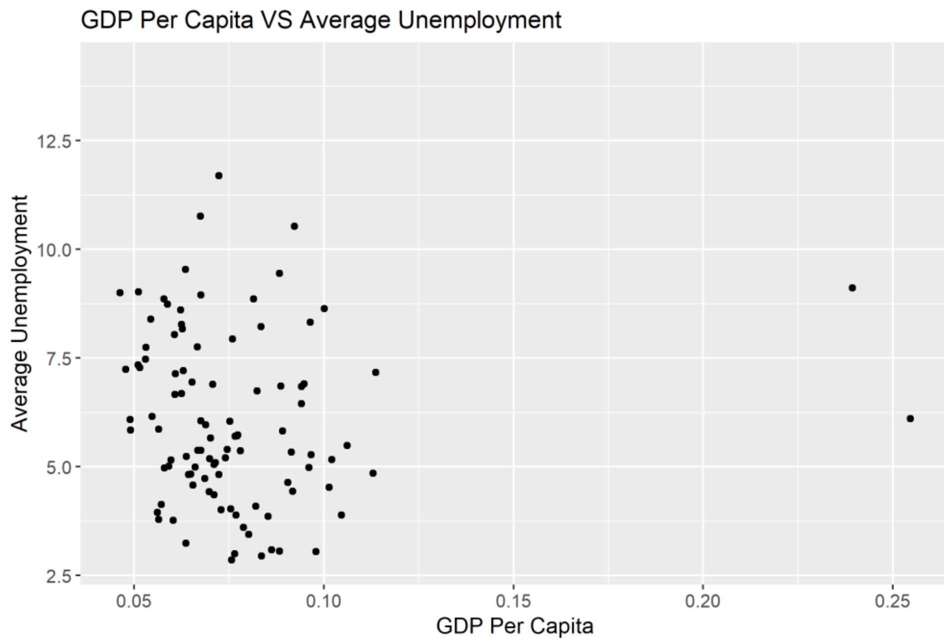


Figure 9: The GDP Per Capita against Average Unemployment for both 2012 and 2016

Based on the above scatterplots, our data suggests that **GDP and voter turnout are not related**. However, it seems that as average unemployment increases, voter turnout data actually decreases, supporting a **negative correlation between the two**.

Given the small scope of this project, there are a multitude of improvements we could make if given more time. The first problem stems from the original datasets, which had missing values for Colorado and Wyoming. The first improvement we would make is to go back and find the necessary data for these states to help the completeness of our data. Most importantly, we would expand on this analysis by including more elections. Trends require time, so in order to really prove the negative relationship between average unemployment and voter turnout, we would have to observe voting behaviors over many years and perhaps many different types of elections (e.g. midterm, state).

Should we correct these flaws, however, we could give politicians and interest groups interesting insights on their constituents. Knowing that citizens in a struggling economy are less motivated to vote, politicians can adjust their campaigns in these areas. Perhaps politicians should focus adds on simply inspiring locals to vote in general rather than to vote for them and not their opponent.

References

Brown, A. (2013, Jan 30). *U.S. economy stumbled to end 2012: Q4 GDP down 0.1%*. Retrieved from https://www.forbes.com/sites/abrambrown/2013/01/30/u-s-economy-stumbled-at-the-end-of-2012-q4-gdp-down-0-1/#35a4c4f7656a

Swanson, A. (2017, Jan 27). *The U.S. just had its worse year of economic growth since 2011*. Retrieved from https://www.washingtonpost.com/news/wonk/wp/2017/01/27/u-s-economy-expanded-1-9-percent-in-fourth-quarter-2016-gdp-up-1-6-percent/?noredirect=on&utm_term=.e6d9af10d3e2