An Election Year During A Politicized Global Pandemic

Group 315C

Kevin Joseph, Jim Lee, Peter Kennedy, Evan O'Keefe

Introduction

2020 has been far from a normal year and two main focuses within the United States were the Coronavirus Pandemic and the Presidential Election between incumbent Republican Donald Trump and Democratic Candidate Joe Biden. The goal of our group was to analyze how these two major events impacted each other specifically in the state of Wisconsin by observing a few different factors. A key aspect that we plan to observe is the impact of increased mail-in/absentee voting. To achieve this analysis, data was collected describing Wisconsin election data (including vote totals and absentee ballot data), COVID-19 cases data (including number of cases and deaths), and general census information. Using this data we hoped to answer the following questions:

1. How has the COVID-19 pandemic impacted how people vote in the 2020 presidential election?
2. Is there a correlation between the majority political party and the pandemic impact?
3. Are there any demographics that are impacted more signifcanly than others?

This project will test the hypothesis that conservative areas of Wisconsin will have higher COVID-19 infection rates and lower absentee voting due to the general rhetoric and actions of the American Republican Party.

Background

About the Data

Our main dataset was connected from the internet using Kaggle[1]. It is a compilation of data on both the 2016 and 2020 Presidential Elections, the ongoing COVID-19 Pandemic, and key population demographics. The data is only from the United States and is sorted based on county and state. There are many variables in the data set, including many that we do not need. Due to the scope of the project only variables relating to the 2020 election and the pandemic and those used for other references were selected. We did not utilize the majority of the demographic variables; however, those presenting interesting trends such as Unemployment were included.

Besides direct pandemic impact (through number of cases and deaths) and the election results, our group sought to analyze the change in how people vote as a result of the public health crisis. To analyze this change, we chose to focus on the differences in absentee/main-in voting. Since our main dataset did not include this information, auxiliary datasets were obtained from the Wisconsin Election Comission The comission releases information after each election detailing the number of absentee ballots sent and recieved in each county (2016[2] and 2020[3]).

To make the data more comparable between counties, we had to manipulate variables such as number of COVID cases to standardize to the population. This allowed us to reasonably compare counties which have population ranging from 4,000 to just under 1,000,000. For our analysis, we filtered the data to only include counties in Wisconsin. The main variables we used are described below based on their dataset of origin:

County Statistics Dataset

1. County: (str) Name of the county for which the data was collected.
2. State: (str) Two letter intials for each US state. Only data from Wisconsin (WI) was used in this project.
3. percentage16_Donald_Trump: (float) Percent of the county vote for Donald Trump in the given election cycle. Given in the decimal form, maximum values of 1.

[1] https://www.kaggle.com/etsc9287/2020-general-election-polls
[2] https://elections.wi.gov/node/4414
[3] https://elections.wi.gov/node/7236

4. percentage16_Hillary_Clinton: (float) Percent of the county vote for Hillary Clinton in the given year. Given in the decimal form, maximum values of 1.
5. total_votes16: (int) Total number of county votes for the given election cycle. Includes democratic, republican, and any third party votes.
6. votes16_Donald_Trump: (int) Votes cast for Donald Trump in the given election cycle.
7. votes16_Hillary_Clinton: (int) Votes cast for Hillary Clinton in the given election cycle.
8. percentage20_Donald_Trump: (float) Percent of the county vote for Donald Trump in the given year. Given in the decimal form, maximum values of 1.
9. percentage20_Joe_Biden: (float) Percent of the county vote for Joe Biden in the given year. Given in the decimal form, maximum values of 1.
10. total_votes20: (int) Total number of county votes for the given election cycle. Includes democratic, republican, and any third party votes.
11. votes20_Donald_Trump: (int) Votes cast for Donald Trump in the given election cycle.
12. votes20_Joe_Biden: (int) Votes cast for Joe Biden in the given election cycle.
13. cases: (int) Number of COVID-19 cases in the county as of the dataset being published.
14. deaths: (int) Number of COVID-19 deaths in the county as of the dataset being published.
15. TotalPop: (int) Total population of the county including those not eligible to vote,
16. VotingAgeCitizen: (int) Number of citizens in the county that are eligible to vote in the election.
17. Unemployment: (float) Percent of the county that is not currently employed. Given in the percent form, maximum values of 100.

Absentee Ballot Report 2016

18. County: Same as previous *County*
19. BallotReturnedCount: (int) The number of absentee/mail-in ballots returned and counted during the election cycle.

Absentee Ballot Report 2020

20. Jurisdiction: Same as previous *County*
21. BallotsReturned: Same as previous *BallotReturnedCount*

Calculated Values

22. maj_party_16: (str) The party that had the majority of the vote in the given election cycle.
23. maj_party_20: Same as *maj_party_16*
24. cases_thou: (int) Number of COVID-19 cases in the given county, standardized to the county population.
25. deaths_thou: (int) Number of COVID-19 deaths in the given county, standardized to the county population.
26. pct_absentee_16: (float) Percent of the county votes that come from absentee/mail-in ballots for the given election cycle. Given in the percent form, maximum values of 100.
27. pct_absentee_20 = Same as *pct_absentee_16*
28. pct_absentee_diff: (float) Percentage change of absentee votes from 2016 to 2020. Given in the percent form, no maximum value.
29. third_party_16: (int) Number of votes cast for the third part canidate in the given election cycle.
30. third_party_20: Same as *third_party_16*
31. win_margin_16: (int) Difference in number of votes between the winning canidate and the losing.
32. win_margin_20: Same as *win_margin_16*

Unusual Factors Present

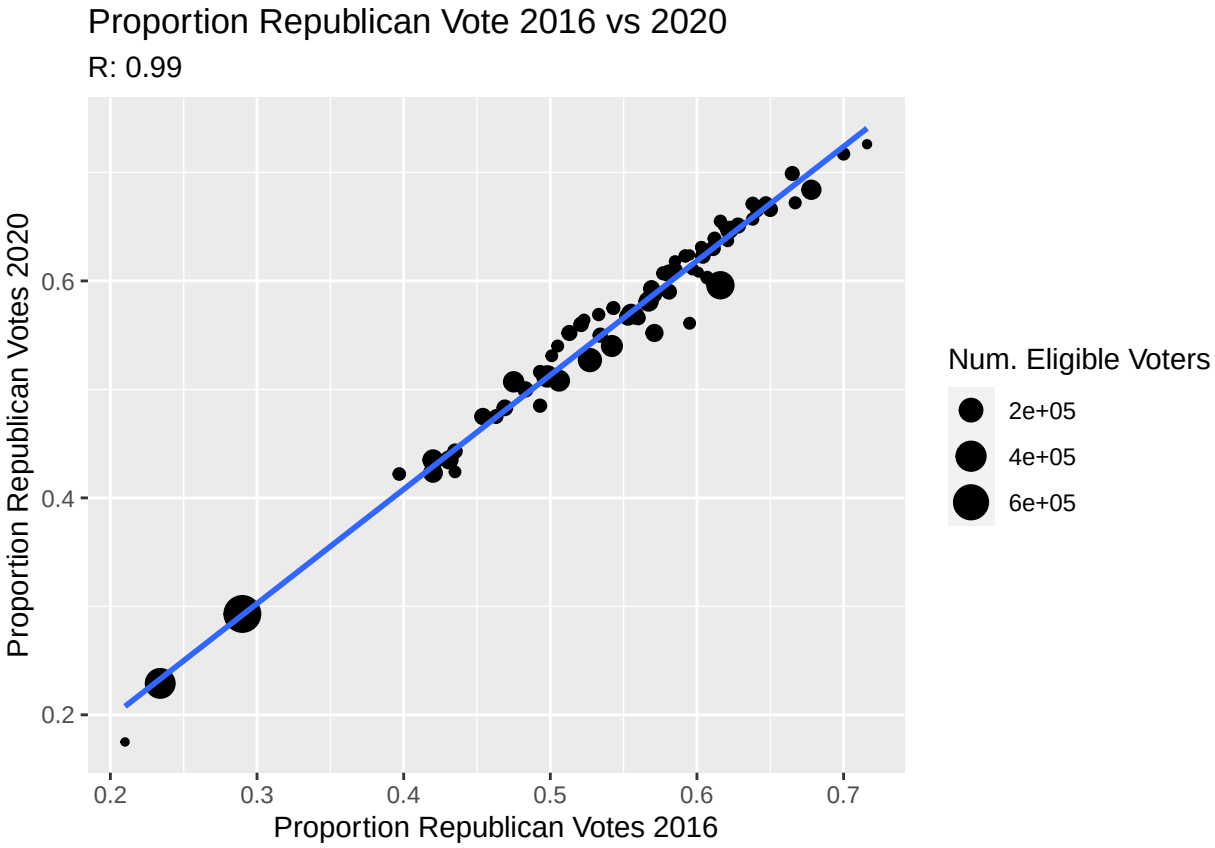
There discrepancies within the data that had to be considered before analyzing the data. Initially, between the datasets the Counties had different naming formats. This issue was resolved through some data processing before joining the datasets. It is also necessary to note that the COVID data presented is from November 1st, so it does no represent cases that have occurred in the last month. Additionally, a majority of the demographic data was collected in 2017, so it is not completely up to date. Most of this was data we did not plan to use anyways. Finally, when looking at the COVID cases and deaths in each county it is not made explicitly clear whether cases are assigned to the county they live in or the county they are tested in.

Project Goal

To observe what correlations, if any, exist between election data, pandemic impact, and general demographics at the county level in Wisconsin. Additionally, some general election data will be analyzed to compare the results of the 2020 Presidential Election to the 2016 election.

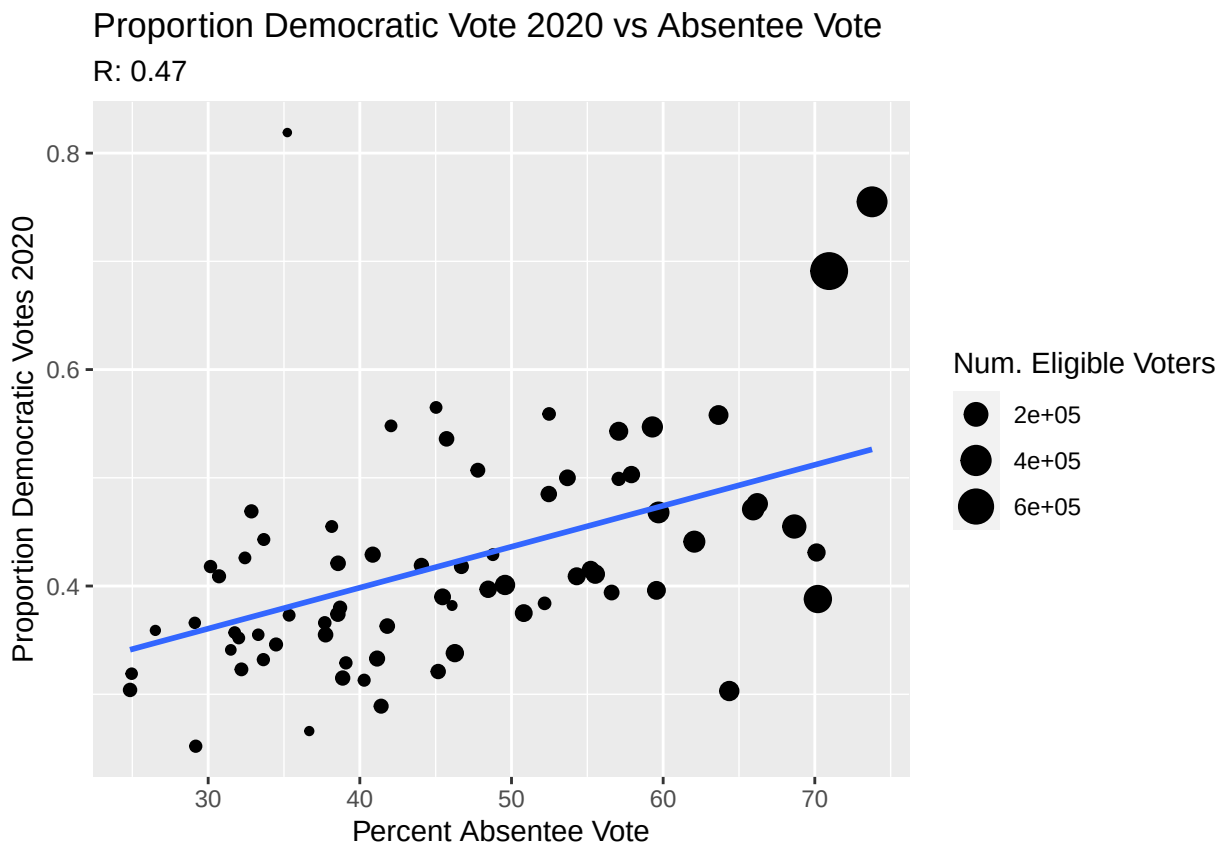
Analysis

Initially, we wanted to look if there was any significant change in how Wisconsin counties voted in the 2020 Presidential Election from the 2016 Election. For this analysis we plotted the proportion of Republican votes in 2020 vs those in 2016 for each county and calculated the line of best fit.



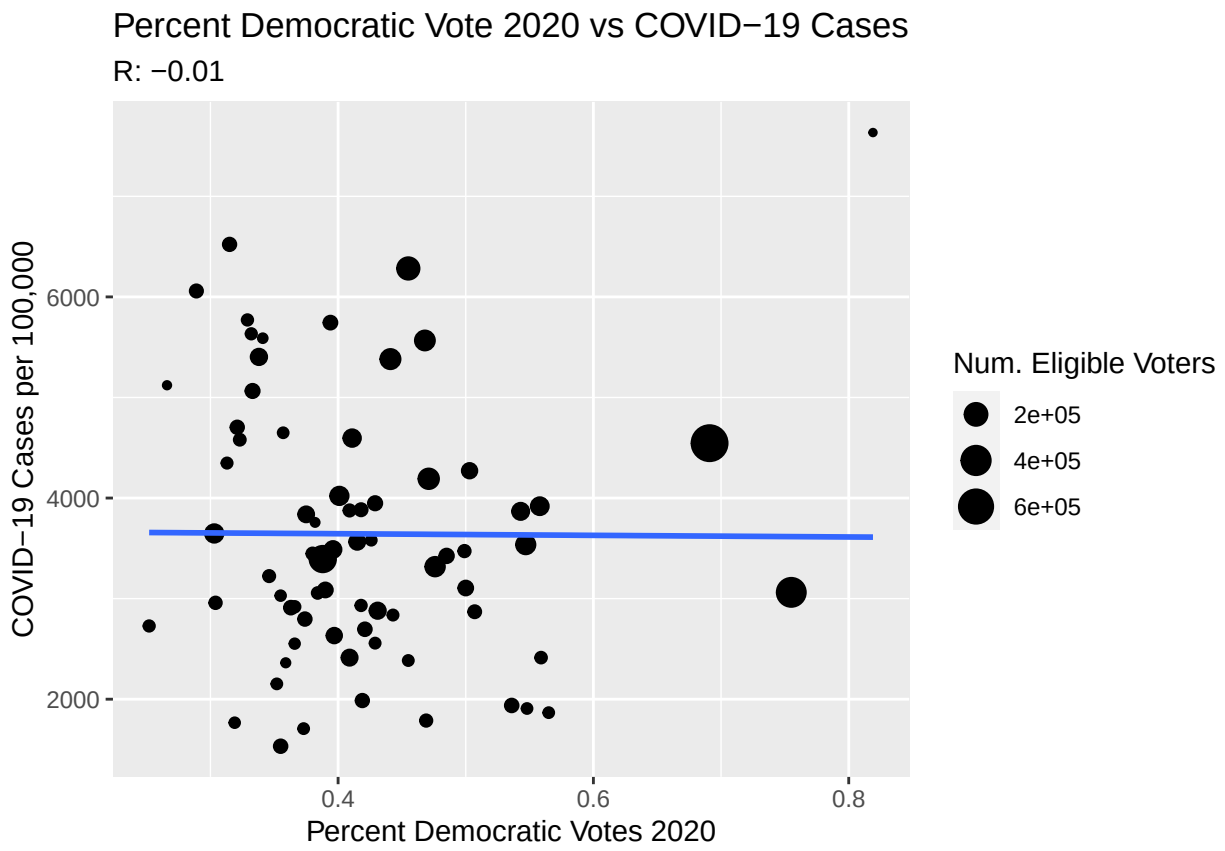
This data was highly correlated (R = 0.99), showing that counties did not have any significant change in the portion of voters who vote Republican.

To observe if there was a signifcant difference in how people voted (in-person vs mail-in) bases on county majority political party, we plotted the proportion of Democratic vote against the mail-in percent of the total vote.



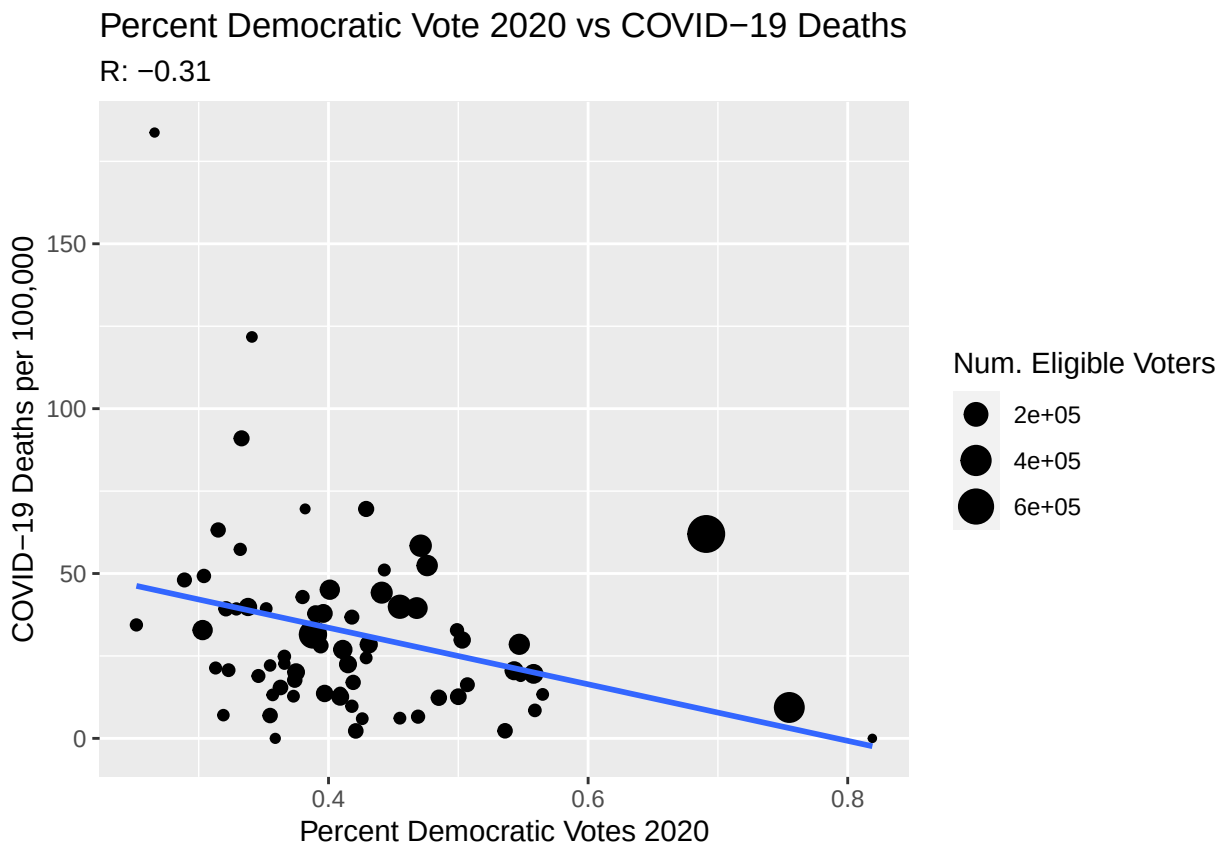
This data was somewhat correlated (R = 0.47); however, it is interesting to note that Wisconsin's two most populous counties, Milwaukee and Dane, are in the top values for both proportion democratic vote and percent absentee vote.

To observe if there was a signifcant difference in the number of COVID-19 cases in a county based on the majority political party, we plotted the proportion of Democratic vote against the number of cases per hundred thousand.



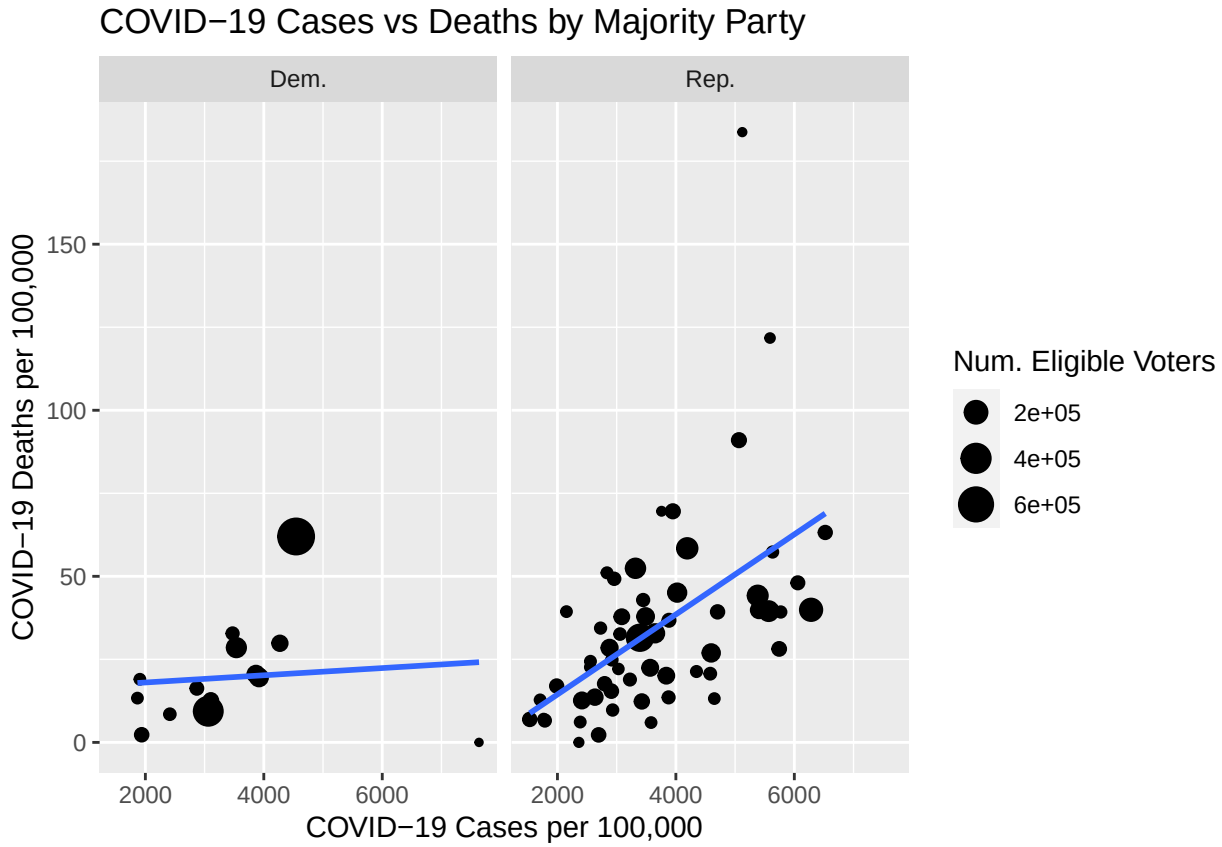
We found essentially no correlation in this data (R = -0.01) showing the party a county votes for does not impact how many people are infected with COVID-19.

To observe this question from another perspective, we plotted the proportion of Democratic vote against the number of COVID-19 deaths per hundred thousand.



Contrary to the previous result there was some correlation between the number of deaths in a county and the majority party (R = -0.31)

Finally to analyze further the data was split by majority political party and separate COVID-19 cases vs death plots were constructed.

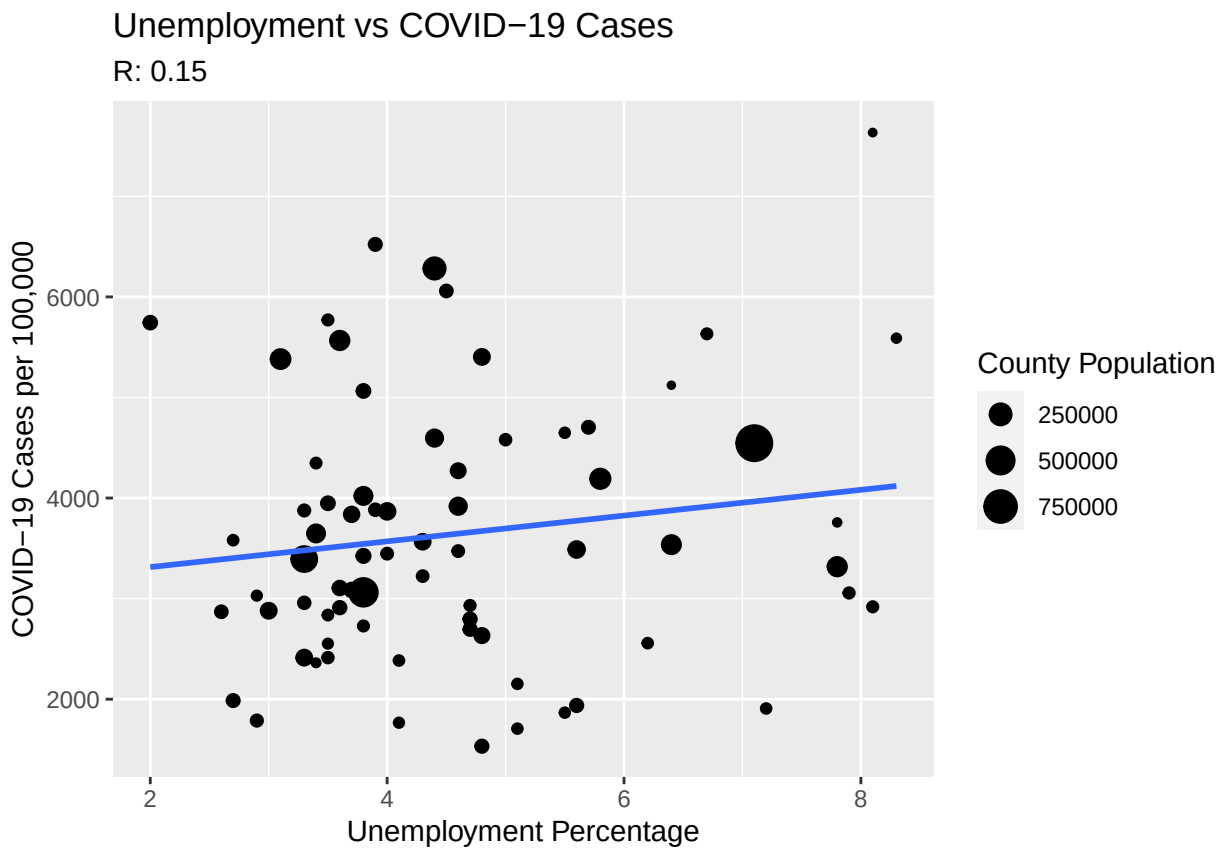



```
## [1] "Dememoratic R: 0.1"
```

```
## [1] "Republican R: 0.51"
```

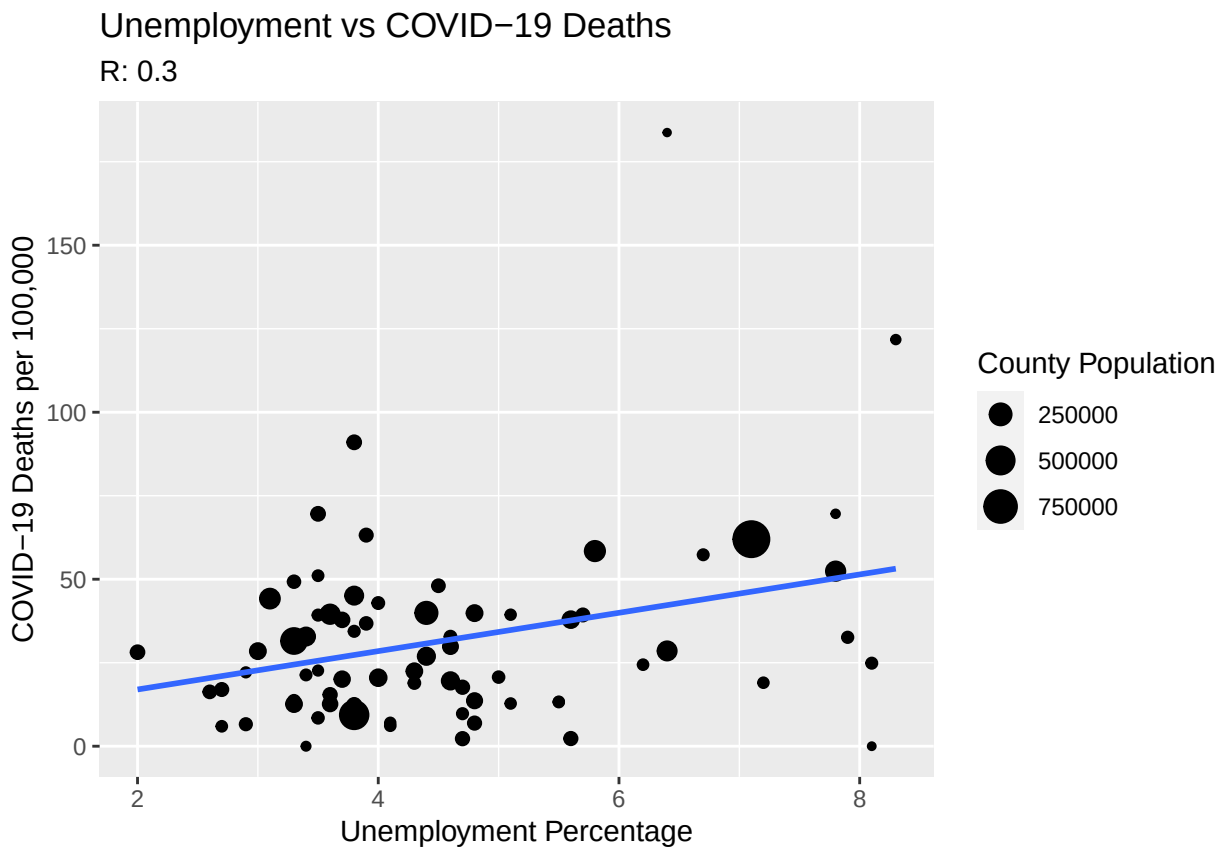
The appears to be a major difference in correlation between the split data (R = 0.1 vs R = 0.51). However, the outlier of Menominee County with over 7,500 cases per 100,000 and 0 deaths may be a significant contributing factor in this.

> To observe if there was a unemployment had an impact on the number of COVID-19 cases in a county, we plotted the unemployment percentage against the number of cases per hundred thousand.



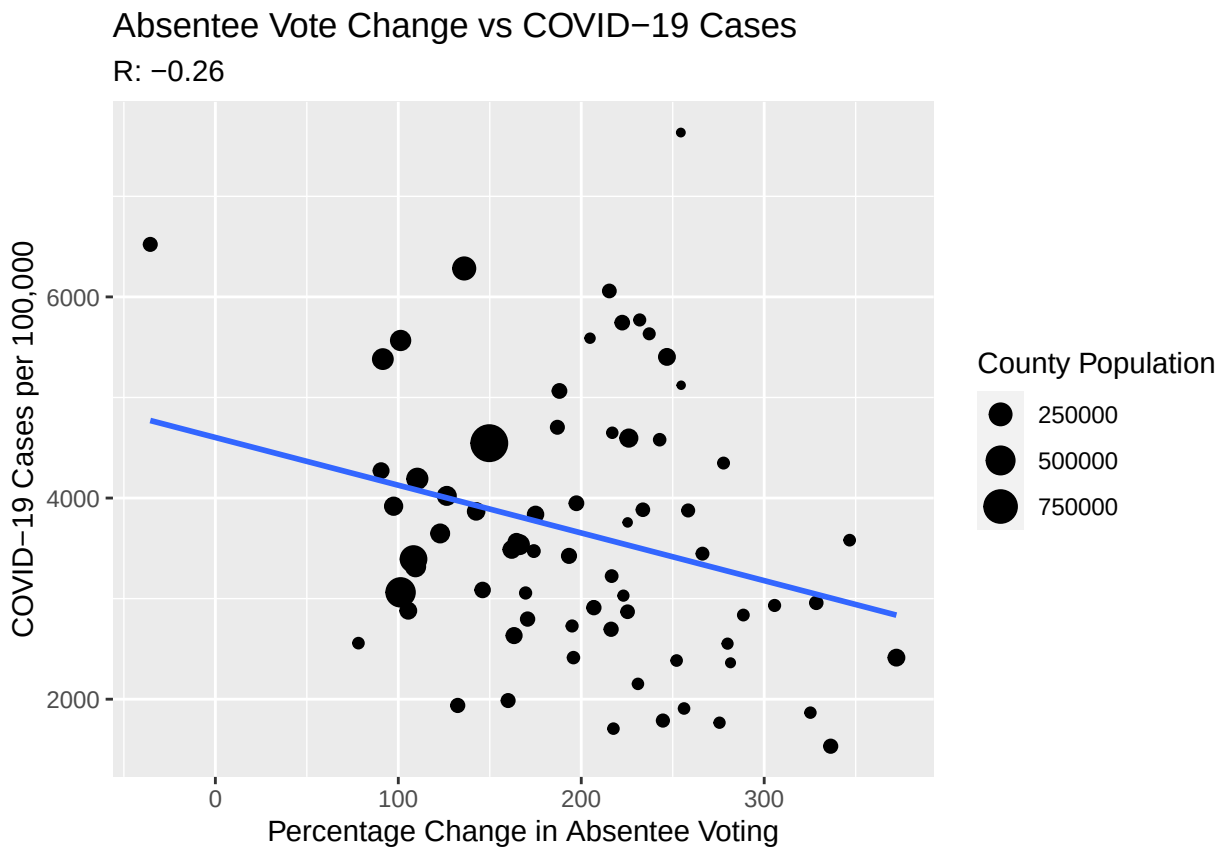
There was not a very large correlation between these factors (R = 0.15).

Again for further analysis, we switchined cases for deaths and plotted the unemployment percentage against the number of deaths per hundred thousand.



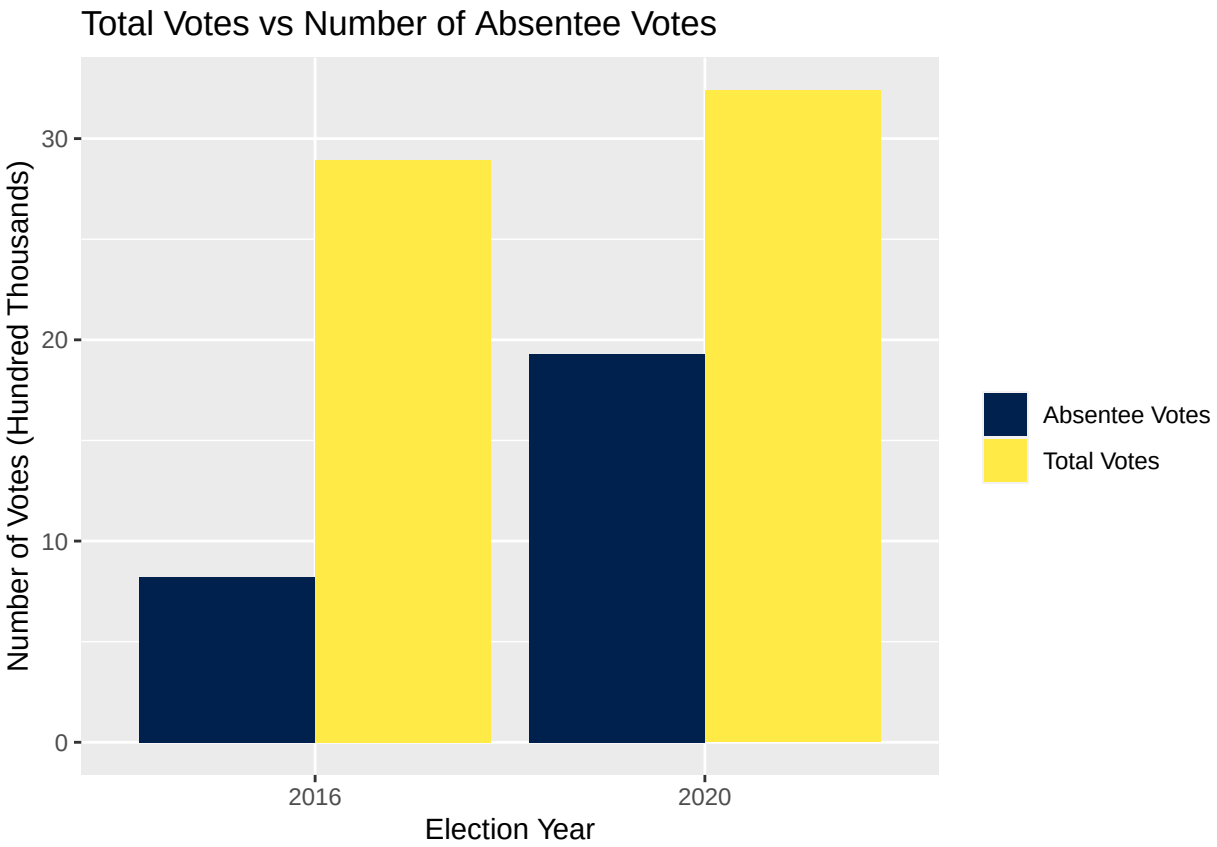
This yielded a slightly higher correlation (R = 0.3) but is still not very significant.

To observe if there was a correlation between increase in absentee voting this year vs 2016 and pandemic impact, we plotted the percent change in absentee voting against the number of COVID-19 cases per 100,000.



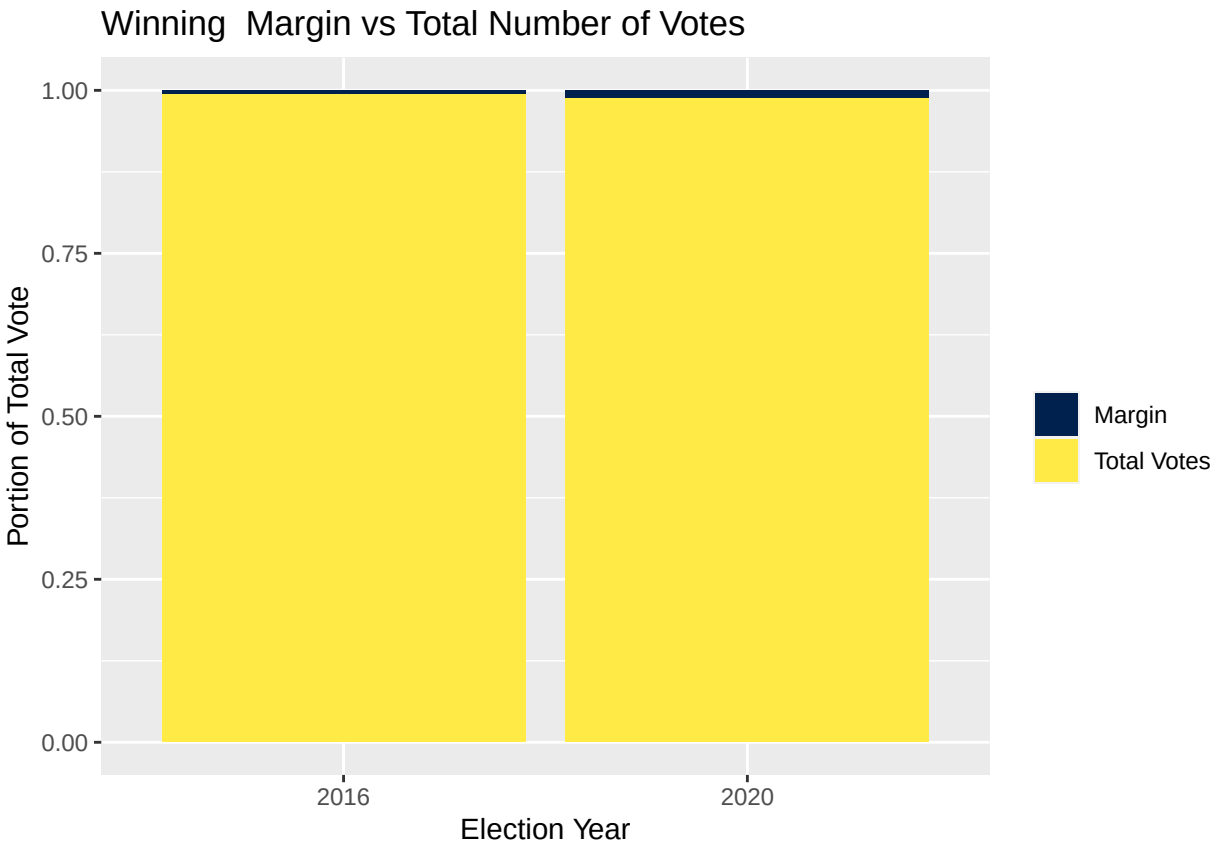
There was a very low correlation between for this comparison (R = -0.26). However, the only county to decrease in absentee votes, Shawano County, had one of the highest values for cases per hundred thousand. Additionally, Sauk County was not included as it has an increase in absentee voting of over 1000% but only a marginal impact on correlation.

For a general overview, we wanted to look at trends in total vote count and number of absentee votes.



It can be seen that there was a significant increase in the absentee vote count and a marginal increase in total vote counts.

Given that Wisconsin is always considered a swing state (is only won by either party each year by a slim margin), we wanted to analyze how close the last two elections were. In order to do this analysis, we plotted a stacked bar of the

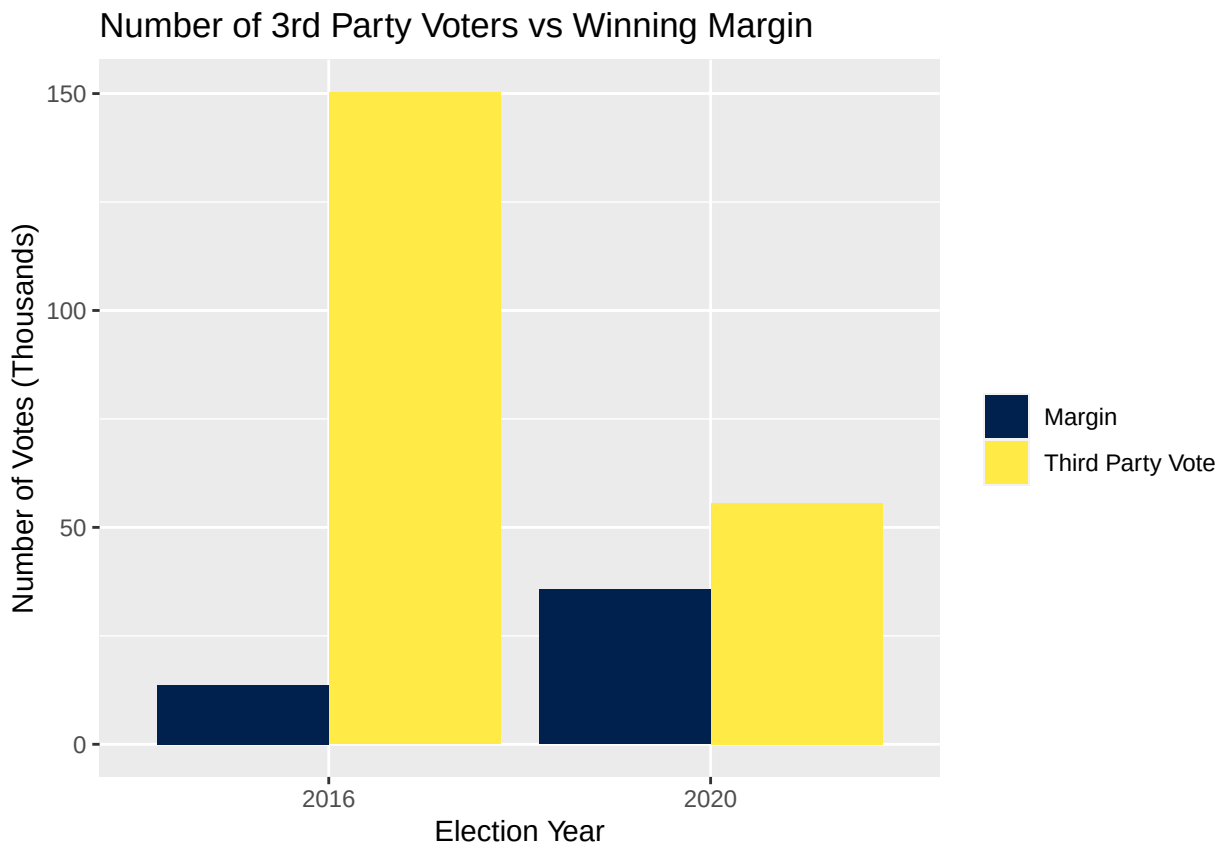



```
## [1] "Margin of Victory 2016 0.47%"
```

```
## [1] "Margin of Victory 2020 1.1%"
```

This graph shows just how close elections are in Wisconsin, showing it is a swing state.

> With the current voting system in Wisconsin, third party canidates can be spoilers (take a winning majority of voters away from a canaditate with a similar platform). To see this effect, we plotted the total number of third party votes and winning margin for 2016 and 2020.



This graph shows how the third party vote played a much larger role in the 2016 election than the 2020, with a much higher number of votes and much slimmer margin. This spoiler effect could be further analyzed by breaking the third party vote into their respective candidates to see which way third party voters may have leaned in this election. (Traditional sides of the third parties:Green party -> Democratic leaning, Libertarian party -> Republican leaning)