Political Influence on Vaccine Acceptance

Caleb Powell, Kazi Fardin, and Owen Runge

January 5, 2022

Abstract

The current COVID-19 pandemic has created a high demand for the speedy production of potentially under-tested vaccinations. Concerns for potential risk of these vaccines have created strong political debates in the United States. As a result, an individual's political party affiliation in the 2020 general elections can be seen to correlate to their acceptance of the current vaccinations against COVID-19.

1 Introduction

Due to the United States 2020 general elections occurring simultaneously with the on-spread of a highly contagious and potentially fatal virus, vaccinations against COVID-19 were a main point of contention for the election. During the election, there was much debate over the differentiation of Republicans' and Democrats' opinions on the effectiveness and potential risks of the COVID-19 vaccines. Since then, vaccinations have become widely accepted and accessible to almost every citizen of the United States. However, the opinions presented in the debates of the 2020 election still effect the acceptance of vaccination for the citizens of the United States. By performing a simple linear regression fit between the vaccination data provided by the CDC and the election results provided by Harvard University for over 3000 counties, a correlation between citizen's political affiliation and their acceptance towards the COVID-19 Vaccine becomes evident.

2 Methods

Through the use of the Python programming language, we can use Harvard's election data to extract the total number of votes for each political party for each county. By dividing the the number of votes for each party by the total number of votes, we can determine a percentage of votes for each political party for every county in the United States. Similarly, we can extract the percentage of vaccinated citizens in each county as provided by the CDC's data set.

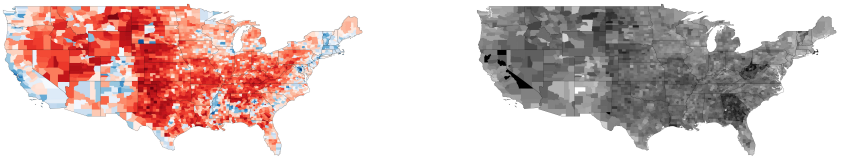

Figure 1: Map of Political Affiliation (Left) vs. Vaccination Percentage (Right).

Using map data provided by the United States Census Bureau, we can create a density map representing the Republican (Red) and Democratic (Blue) voting percentage and compare it to a map of vaccination percentage spread across the US. In these figures, the darker the color, the higher the percentage. By visual inspection we might be able to spot correlations in the darkness of the two images. However, these correlations require further analysis.

3 Analysis

To further analyze this data, we will implement a "simple linear regression". Simple linear regressions are a very useful tool in analyzing the relationship between two variables that produce a linear fit and two values defined as R and R^2. The R value is a correlation coefficient between two sets of data while the R^2 value is the variability of the data from the linear fit. For our analysis, we carried out a simple linear regression between vaccination and voting percentages for republican and democratic votes in every county of the US to analyze the R and R^2 values. We then produced scatter plots with the linear fit of the data sets.

3.1 Democratic Vaccine Acceptance

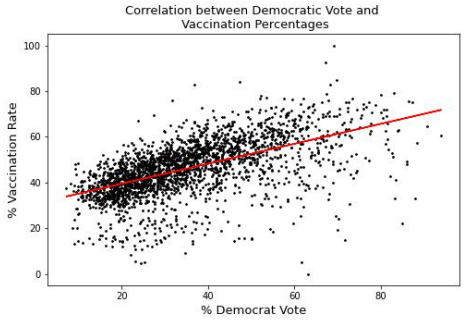

Figure 2: Scatter Plot of Democratic Vote and Vaccination Percentage with Linear Fit.

The linear fit in Figure 2 shows a positive correlation between Democratic voting percentages and vaccination percentages by county. From our regression, we have a R-value of 0.57. The R-value points towards a moderately positive correlation. This would mean the higher percentage vote a county has for a Democratic candidate, the more likely it is to have a high vaccination percentage. However, we found an R^2 value of 0.33 which would mean that our model's correlation value is only accurate 33% of the time. Even though we have a moderate correlation, it cannot be used to make a generalization for the entire data set.

3.2 Republican Vaccine Acceptance

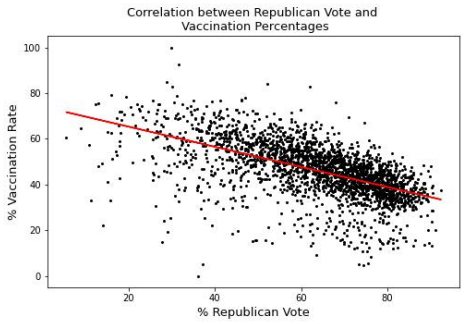

Figure 3: Scatter Plot of Republican Vote and Vaccination Percentage with Linear Fit.

In the linear fit between Republican voting percentages and vaccination percentages by county (Figure 3), we found an R value of the regression fit to be -0.58. This value shows a moderate negative correlation in the data. As for the R^2 value, it is 0.34 which is just 0.01 higher than the R^2 value in the linear regression fit for Democrats. Similarly, the correlation is only accurate 34% of the time.

4 Acknowledgement of Potential Counter-Arguments

Due to arguments presented by some individuals on the fabrication of the results of the 2020 election, it is important to analyze this possibility. In order to do so, we will implement Benford's law on the value of total votes presented in Harvard's data-set on the election. Benford's law is an observation in which distinct sets of naturally occurring data tend to have a greater frequency of lower leading digits. This observation is often used by the wider scientific community to test whether or not data has been manipulated. Although Harvard is a credible institution, from which we took this data, we still want to make sure the data has not been manipulated.

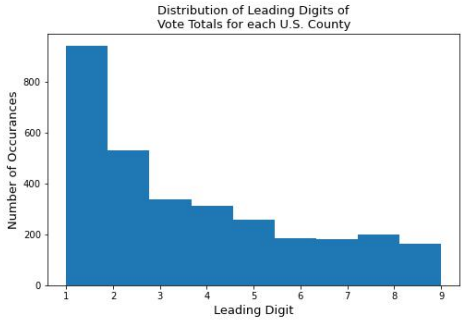

Figure 4: Death Total Leading Digits.

Figure 4 is a histogram plot of the Benford's law using the data of vote totals during the 2020 general election by each U.S. county.

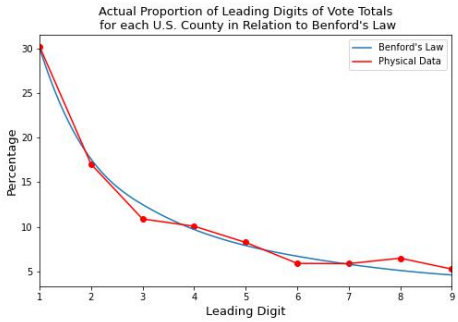

Figure 5: Expected Result vs Actual Result.

Figure 5 is a comparison between the expected Benford's Law curve and our data set. As the results show, our data set fits quite closely with the expected values. Thus, we can assume that the data was not doctored to suit certain results.

5 Conclusion

From our analysis, a correlation can be found such that the counties with majority Republican vote tend to have a lower vaccination rate while majority Democratic counties tend to have a higher vaccination rate. However, due to the low R^2 values found from our linear regression in both democratic and republican voting percentages, the correlation represented by our R value is only accurate around 33% of the time. This means that while there is a clear trend in the data, an exact correlation value cannot be accurately calculated.

$$fxx = dt * (df/res * *2) * (fpxy[:, 2:] + fpxy[:, 0:-2] - 2 * fpxy[:, 1:-1]) \tag{1}$$

$$\frac{\partial f}{\partial x} = dt(\frac{\partial f}{res^2}) \tag{2}$$