COVID-19 Vaccination Rates and Google Search Data

Arianna Schmid and Stacey Frank

2021-12-16

Contents

Introduction

Vaccines to control the Coronavirus Disease 2019 (COVID-19) became available to the public in the first half of 2021. Rejection and indecision towards being vaccinated is evident across the United States. The motivation for this study is to provide a better understanding of reasons for COVID-19 vaccine refusal in the United States. This can help public health messaging campaigns be more targeted and effective when promoting vaccination.

Google data is useful for exploring this topic because there is previous research that people feel freer to search for socially stigmatized topics on Google than they would be to admit such opinions in a survey or another form of data collection. In his study of racial animus's relationship to voting behavior, Stephens-Davidowitz notes that Google searchers are likely to be alone and thus more likely to search for private topics (Stephens-Davidowitz 2014). Since we are interested in understanding how myths and conspiracies are playing into people's choice to receive the vaccine or not, Google searches are a useful source of data about the prevalence and impact of these ideas.

Research Questions

Our primary research question is what is the relationship, if any, between state-level COVID-19 vaccine rates and the types of Google searches that are made about vaccines in each state? To assist in answering that question, we have two secondary research questions:

1. Are vaccine myths more commonly searched for in states that also have low vaccination rates?
2. Does the relationship between COVID vaccine rates and Google searches change between June and September 2021?

We will use Google search data, state level COVID vaccination rate information from the Centers for Disease Control, and state-level demographic information to address these questions.

Data Collection

1. Google Trends and Keywords

The Centers for Disease Control provides a list of the most common questions about the COVID-19 vaccine, including mainstream questions and myths (CDC 2021). Similarly, the Mayo Clinic provides information on the most common myths surrounding the vaccine ("COVID-19 Vaccine Myths Debunked" 2021). Using these two data sources, we constructed a list of 12 keyword search terms. It consists of the two general searches, "covid vaccine" and "covid vaccine near me," five mainstream questions, such as "covid vaccine side effects," and 5 myth-related searches such as "covid vaccine microchip."

We created a vector called "k" to signify "keywords" and referenced this vector in later code to loop over the search terms for which we wanted to collect data. Note that each element in "k" will be renamed based on its index (hits.1, hits.2, ... hits.12) for code efficiency.

Table 1: Google Search Terms in Vector 'k'

Search Term
covid vaccine
covid vaccine near me
covid vaccine safe
covid vaccine ingredients
covid vaccine pregnant
covid vaccine protect
covid vaccine side effects
covid vaccine microchip
covid vaccine dna
covid vaccine fetal
covid vaccine infertility
covid vaccine magnet

We are interested in studying the Google search hit rate for all 12 terms in "k" for all 50 states and Washington, D.C. Since vaccines were made available to states at different points in time, and the nature of people's concerns about the vaccine may have changed over time, we pulled search hit rates from 3 different time periods: 1/1/21-9/20/21, 4/1/21-6/20/21, and 7/1/21-9/20/21.The gtrendsR package was used to work with Google Trends queries.

We created the custom function get.hit.results() to iterate through all the Google Trends queries we wanted to run without repeating code. The function takes in the date range as an argument, calls gtrends() once for

each of the 12 search terms, and returns the hit results by state based on the given date range. The hit rates by state for each search term are then extracted from the list initially generated by the gtrendsr() query, and saved into a single tibble containing the hit rates for all 12 search terms for the specified time period. The function is called 3 times, once for each date range, and the result is saved in the 3 objects hits.results.jan, hits.results.june, and hits.results.sept.

```
#User-defined function to run all needed Google Trends queries

get.hits.results <-  function(date){
    for (i in 1:length(k)){
          new_frame <- paste("Keyword",i,sep = "")
          assign(new_frame, gtrends(k[i], geo = "US",
                   time = date, low_search_volume = T)
                 )
    }

    hits_results <- Keyword1$interest_by_region %>%
      left_join(Keyword2$interest_by_region, by = "location") %>%
      left_join(Keyword3$interest_by_region, by = "location") %>%
      left_join(Keyword4$interest_by_region, by = "location") %>%
      left_join(Keyword5$interest_by_region, by = "location") %>%
      left_join(Keyword6$interest_by_region, by = "location") %>%
      left_join(Keyword7$interest_by_region, by = "location") %>%
      left_join(Keyword8$interest_by_region, by = "location") %>%
      left_join(Keyword9$interest_by_region, by = "location") %>%
      left_join(Keyword10$interest_by_region, by = "location") %>%
      left_join(Keyword11$interest_by_region, by = "location") %>%
      left_join(Keyword12$interest_by_region, by = "location") %>%
      as_tibble() %>%
      dplyr::select(c(1,2,6,10,14,18,22,26,30,34,38,42,46))

    hits_results %<>% rename( hits.1 = hits.x,
                            hits.2 = hits.y,
                            hits.3 = hits.x.x,
                            hits.4 = hits.y.y,
                            hits.5 = hits.x.x.x,
                            hits.6 = hits.y.y.y,
                            hits.7 = hits.x.x.x.x,
                            hits.8 = hits.y.y.y.y,
                            hits.9 = hits.x.x.x.x.x,
                            hits.10 = hits.y.y.y.y.y,
                            hits.11 = hits.x.x.x.x.x.x,
                            hits.12 = hits.y.y.y.y.y.y)
    print(hits_results)

}

hits.results.jan <- get.hits.results("2021-01-1 2021-09-20")
hits.results.june <- get.hits.results("2021-04-1 2021-06-20")
hits.results.sept <- get.hits.results("2021-07-1 2021-09-20")
```

Next, we are interested in getting the count of states that actually have a Google Trends ranking for each search term. We do this to ensure we are using search terms that are popular enough to have generated a hit rate, which will allow us to actually conduct our analysis. We accomplish this through the user-created

get.search.terms() function. This function takes the hits results object from get.hit.results() as an argument, and returns the total number of states that have a reported hit rate for each search term in "k." Again, the function is called 3 times for each time period, and results are saved to the objects search.terms.jan, search.terms.june, and search.terms.sept. As an example, we see in Table 3 that for April-June, "covid vaccine," "covid vaccine near me," and "covid vaccine side effects" have reported hit rates in all 50 states and Washington DC. On the other hand, "covid vaccine microchip" was only searched often enough to generate a hit rate in 17 states.

Table 2: January-September: Number of States with Hit Rates by Search

Variable Name	Search Term	Number of States with Hit Rate
hits.1	covid vaccine	51
hits.2	covid vaccine near me	51
hits.3	covid vaccine safe	51
hits.4	covid vaccine ingredients	51
hits.5	covid vaccine pregnant	50
hits.6	covid vaccine protect	44
hits.7	covid vaccine side effects	51
hits.8	covid vaccine microchip	23
hits.9	covid vaccine dna	48
hits.10	covid vaccine fetal	31
hits.11	covid vaccine infertility	48
hits.12	covid vaccine magnet	43

Table 3: April-June: Number of States with Hit Rates by Search

Variable Name	Search Term	Number of States with Hit Rate
hits.1	covid vaccine	51
hits.2	covid vaccine near me	51
hits.3	covid vaccine safe	50
hits.4	covid vaccine ingredients	47
hits.5	covid vaccine pregnant	45
hits.6	covid vaccine protect	40
hits.7	covid vaccine side effects	51
hits.8	covid vaccine microchip	17
hits.9	covid vaccine dna	43
hits.10	covid vaccine fetal	26
hits.11	covid vaccine infertility	44
hits.12	covid vaccine magnet	40

Table 4: July-September: Number of States with Hit Rates by Search

Variable Name	Search Term	Number of States with Hit Rate
hits.1	covid vaccine	51
hits.2	covid vaccine near me	51
hits.3	covid vaccine safe	49
hits.4	covid vaccine ingredients	49
hits.5	covid vaccine pregnant	49

Variable Name	Search Term	Number of States with Hit Rate
hits.6	covid vaccine protect	42
hits.7	covid vaccine side effects	51
hits.8	covid vaccine microchip	10
hits.9	covid vaccine dna	41
hits.10	covid vaccine fetal	28
hits.11	covid vaccine infertility	42
hits.12	covid vaccine magnet	23

2. Vaccination Rates

We now have data for our Google search hits from January-September, April-June, and July-September. Next, we are interested in finding the corresponding information on vaccination rates for each state. This data is acquired by using RSocrata to pull COVID vaccine data from the Centers for Disease Control website via their API. After pulling the data into R, we did some basic data cleaning, including converting variables to a numeric type and dropping rows for U.S. territories and federal entities for which we did not have corresponding Google Trends data. We also recoded the state variable, which previously used the two-letter postal code for state (AK, AZ, etc.), to use the full state name. This was done so that the CDC data could be joined with the Google Trends data by state.

After cleaning, the datasets vax.June21 and vax.Sept21 are created, where vax.June21 will correspond to the Google Trends hit results from April-June, and vax.Sept21 will correspond to January-September and July-September Google Trends hit results. We are focusing on the vaccination rates in June and September 2021 because June is around the time when vaccines became widely available to anyone who wanted one, while September 21, 2021 is the day before boosters for some at-risk groups were approved by the CDC. By doing this, we are restricting our analysis to the time before boosters were available. Our variables of interest are series_complete_pop_pct, which is the percentage of the population in each state that has completed their vaccine series, and admin_per_100k, which is the number of vaccines administered in each state per 100,000 people.

The tables below list the top 10 states by percent of the population fully vaccinated in June and September 2021.

Table 5: Top 10 States for % Population Vaccinated: June 2021

State	Percent Pop. Fully vaccinated	Administered Per 100K
Vermont	64.3	134607
Massachusetts	60.0	125936
Maine	59.9	117770
Connecticut	59.0	121188
Rhode Island	57.3	118098
New Hampshire	54.9	110353
New Jersey	54.6	111391
Maryland	53.8	110157
Washington	52.9	109782
New Mexico	52.5	109500

Table 6: Top 10 States for % Population Vaccinated: September 2021

State	Percent Pop. Fully vaccinated	Administered Per 100K
Vermont	69.0	142674
Connecticut	68.0	139362
Maine	67.8	133752
Massachusetts	67.4	140598
Rhode Island	67.1	135818
New Jersey	63.6	128346
Maryland	63.4	129903
New York	62.7	130789
Washington	62.6	130222
New Mexico	62.3	130771

3. State-level Demographics

We now have our search hit results and vaccine rates. Before studying any correlations, we consider certain state-level demographic factors that could potentially have an impact on our analysis. Data of interest includes state population counts, voter information, household income, age group, and race.

State population numbers were web scraped from Wikipedia using Selector Gadget ("List of U.S. States by Population" 2021). These state counts were then joined with the vax.June21 and vax.Sept21 frames created in Part 2. In addition, information on the Trump vote share in the 2020 presidential election for each state was downloaded from the Cook Political Report ("2020 Popular Vote Tracker" n.d.). Unnecessary columns were deleted and columns of interest were renamed. The final spreadsheet was saved as vote.xlsx.

Median household income information from the 2018 Current Population Survey was downloaded from the U.S. Census Bureau website and saved as med.income.xlsx ("Current Population Survey 2019 Annual Social and Economic (ASEC) Supplement," n.d.). Unnecessary columns were deleted and columns of interest were selected: state (location) and median income (med.income).

Percent of state population by age group was pulled from the Kids Count Data Center ("Adult Population by Age Group" n.d.) The raw Excel data was downloaded. Columns from years 2011-2019 were deleted since we are focusing on the most recently available data, in this case 2020. The spread() function from the tidyr package was then used to reshape from long to wide so we can split up the categorical Age Group Variable into the 3 separate variables: ages 18 to 24, ages 25 to 64, and ages 65 and over. The final spreadsheet was saved as percent.age.xlsx.

Race data from the 2020 Current Population Survey was pulled as a csv file from the Kaiser Family Foundation website ("Population Distribution by Race/Ethnicity (CPS)" 2021). It was then converted from CSV to xlsx, unnecessary columns were deleted, columns of interest were renamed, and reported percentages of $<.01$ were replaced with 0 to make all cell values integers. Categories for American Indian/Alaska Native, Native Hawaiian/Other Pacific Islander, and Multiple Races were combined into a single 'Other' race category. This was saved as State Race Data.xlsx.

These four excel files were loaded into the FOCD Github public repository. They were then joined together in R and saved as the object "cov" (for covariates). This object was then joined with the two CDC vaccine rate data frames created in Part 2 (vax.June21 or vax.Sept21).

Finally, the three gtrends datasets from Part 1 (hits.results.jan, hits.results.june, hits.results.sept) are joined with the updated vax.June21 or vax.Sept21, depending on the dates the Trends are covering. A user-defined function called join.gtrends.vaccine() accomplishes this and saves the 3 final data sets as Jan01.analysis, Sept21.analysis, and June21.analysis. Now they are ready for analysis.

```r
## This function joins the gtrends datasets with vaccine info datasets

join.gtrends.vaccine <- function (hits.results.month,vax.month){

  month.analysis <- vax.month %>%
      dplyr::select(location,date, admin_per_100k, series_complete_pop_pct, population,
             pct.vote.rep, med.income, pct.18.to.24, pct.25.to.64, pct.65.over,
             pct.white, pct.black, pct.hispanic, pct.asian, pct.other.multiple) %>%
        full_join(hits.results.month, by = "location") %>%
        arrange(location)

  print(month.analysis)

}

Jan01.analysis <- join.gtrends.vaccine(hits.results.jan,vax.Sept21)
Sept21.analysis <- join.gtrends.vaccine(hits.results.sept,vax.Sept21)
June21.analysis <- join.gtrends.vaccine(hits.results.june,vax.June21)
```

Analysis

1. Correlation Analysis

The three data sets Jan01.analysis, Sept21.analysis, and June21.analysis are used as arguments for the user-defined function get.correlations() and results are saved to the three objects Jan01.correlations, Sept21.correlations, and June21.correlations respectively. This function calculates the correlation between the state level hit rate for each search term (as defined in vector "k") and series_complete_pop_pct (percentage of the population in each state that has completed their vaccine series), along with the p value.

```r
## This function pulls the correlations for all 3 data sets
get.correlations <- function(month.analysis){
      #Loop for correlations for each search term
      j <- c("hits.1", "hits.2",
      "hits.3", "hits.4",
      "hits.5", "hits.6",
      "hits.7", "hits.8",
      "hits.9", "hits.10",
      "hits.11","hits.12")

      correlations <- data.frame(estimate=numeric(26), p.value=numeric(26))

      for(i in 15:ncol(month.analysis)){
        test <- cor.test(month.analysis[, i], month.analysis$series_complete_pop_pct)
        correlations$estimate[i] = test$estimate
        correlations$p.value[i] = test$p.value
      }

      correlations %<>%
        slice_tail(n=12) %>%
        cbind(j,k) %>%
        relocate(estimate, p.value,.after = k)
```

```
      correlations %<>% rename(var_name = j, search = k)

      print(correlations)
}

Jan01.correlations <- get.correlations(Jan01.analysis)
Sept21.correlations <- get.correlations(Sept21.analysis)
June21.correlations <- get.correlations(June21.analysis)
```

Results from Jan01.correlations and June21.correlations show that the correlation patterns are fairly similar for January-September and for April-June. In both cases, the more mainstream searches are positively correlated, and most of the myth-related searches are negatively correlated (with the exception of April-June, where infertility has a small positive correlation). 'Covid vaccine microchip' has a positive correlation in both the January-September and the April-June time frame, so that search does not follow the pattern of the other myth or fringe-related searches.

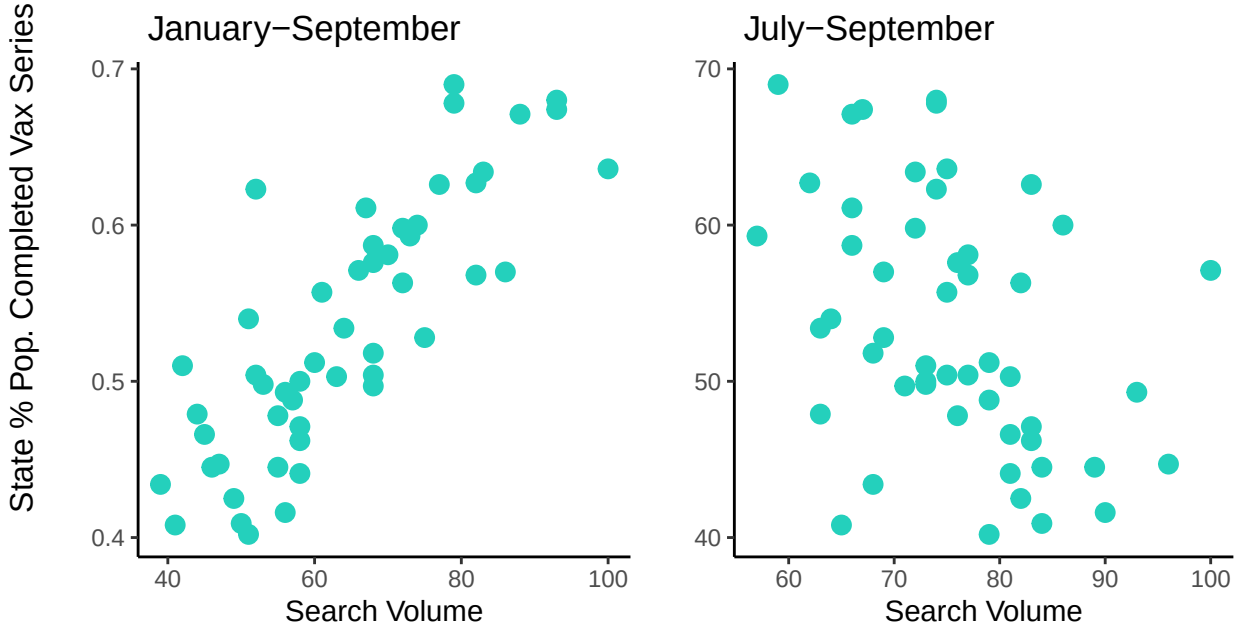
On the other hand, correlation patterns from Sept21.correlations are very different. Our results show for July-September searches, nearly all the search terms have a negative correlation. This is likely because most people in states with high vaccination rates had already been vaccinated by July. As a result, in these states, there are less people searching for vaccine information after July. Therefore in general, more of the vaccine searches (of all types) were happening in low vaccine rate states during July-September. This explains why 'covid vaccine' searches are positively correlated with vaccine rates when looking at the overall January-September searches– but when just looking at July-September, they are negatively correlated since vaccinated people were no longer searching for vaccine info (see Figure 1 below). Table 7 shows these correlation patterns for all three time periods.

Table 7: Correlations Between Search Hits and % Population Vaccinated

var_name	search	Jan-Sept	P-Value	Apr-June	P-Value	Jul-Sept	P-Value
hits.1	covid vaccine	0.819	0.000	0.910	0.000	-0.397	0.004
hits.2	covid vaccine near me	0.154	0.280	0.421	0.002	-0.271	0.054
hits.3	covid vaccine safe	0.253	0.074	0.373	0.007	-0.154	0.281
hits.4	covid vaccine ingredients	0.067	0.642	0.325	0.020	-0.061	0.668
hits.5	covid vaccine pregnant	0.504	0.000	0.224	0.114	0.014	0.923
hits.6	covid vaccine protect	0.070	0.626	0.243	0.085	-0.191	0.180
hits.7	covid vaccine side effects	0.221	0.119	0.680	0.000	-0.642	0.000
hits.8	covid vaccine microchip	0.270	0.055	0.299	0.033	0.017	0.904
hits.9	covid vaccine dna	-0.062	0.665	-0.186	0.191	-0.097	0.500
hits.10	covid vaccine fetal	-0.173	0.225	-0.194	0.173	-0.025	0.861
hits.11	covid vaccine infertility	-0.089	0.534	0.074	0.604	-0.285	0.043
hits.12	covid vaccine magnet	-0.168	0.239	-0.060	0.675	-0.362	0.009

Figure 1 shows the different correlation patterns between searches for 'COVID Vaccine' and state vaccination rates in two time periods. We see a statistically significant positive correlation for the entire span of January-September, and a negative correlation when we restrict ourselves to only searches done in the July-September time period.

Figure 1: Correlations Between Searches for
'COVID Vaccine' and State Vaccination Rates

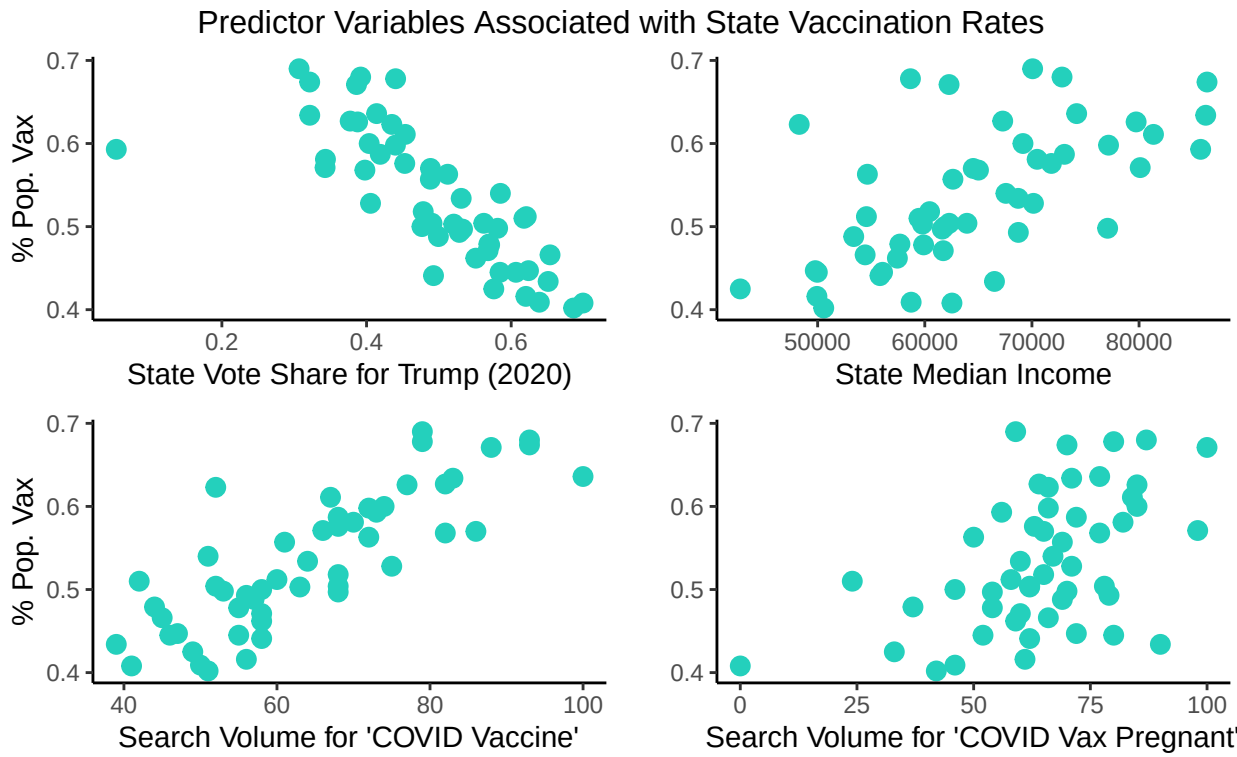


Still, there are very few cases where we have statistically significant correlations at level .05. For this reason, we will incorporate the social demographic variables from Data Collection Part 3 as covariates for a regression analysis to see if, after controlling for demographic variables, any of the search term hit rates are shown to be significant predictors of state vaccination rates.

2. Regression Analysis

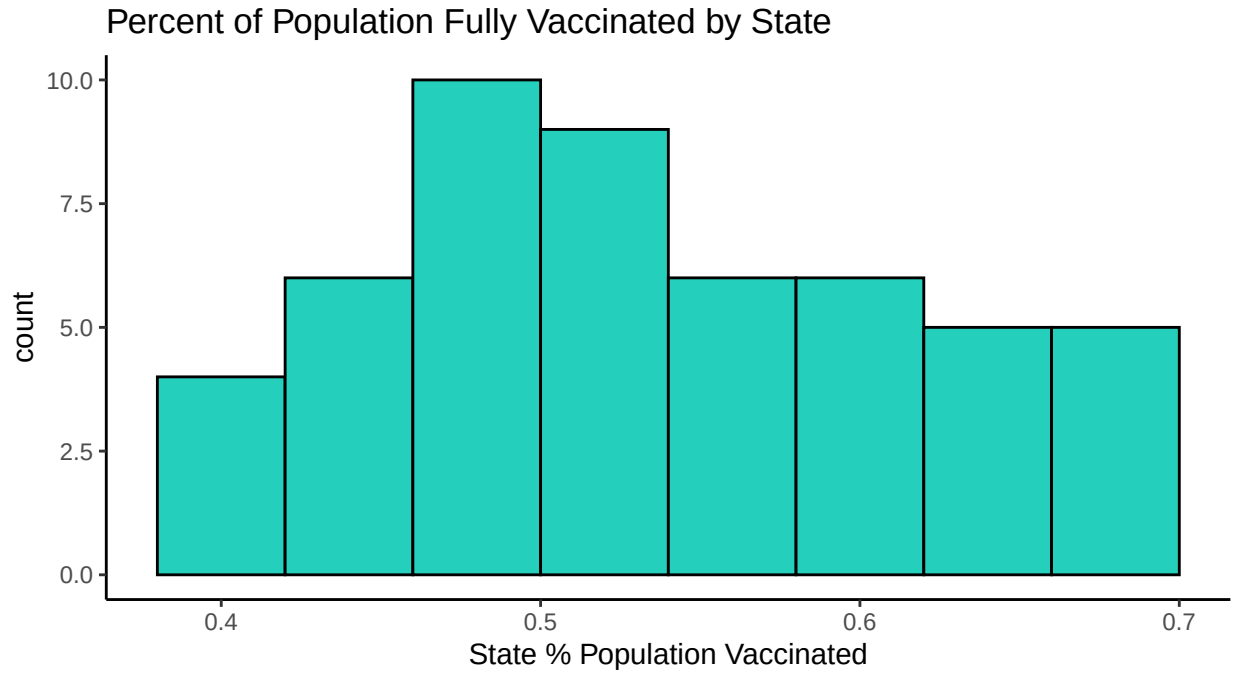
The goal of the regression analysis is to determine the primary predictors for state level vaccination rates, and to determine if, after controlling for state-level demographic factors, any of the Google search terms are significant predictors of vaccination rates. For the regression analysis, we will use the overall Google search hit rates from January-September, 2021.

We started with a descriptive analysis, by plotting all of our potential predictors against state level vaccination rates to determine if any of them have an association with our outcome of interest. This review showed that the 2020 vote share for Donald Trump and state median income had some association with state vaccination rates, but race and age did not seem to be related. As was indicated in our correlation analysis, only the search hit rates for 'COVID vaccine' and 'COVID vaccine pregnant' seemed to be related to state vaccination rates.



The plot of state vote share for Trump indicates that there is an extreme outlier in the data–this is Washington, DC, which had an extremely low vote share for Trump (about 5% of all votes cast in the presidential election). Therefore, we will run our regression analysis with and without that outlier to see how it impacts results.

Next, we created a histogram of our outcome variable to check its distribution and verify that it is not highly skewed. The histogram shows that there is no major skew in the distribution of the outcome variable.

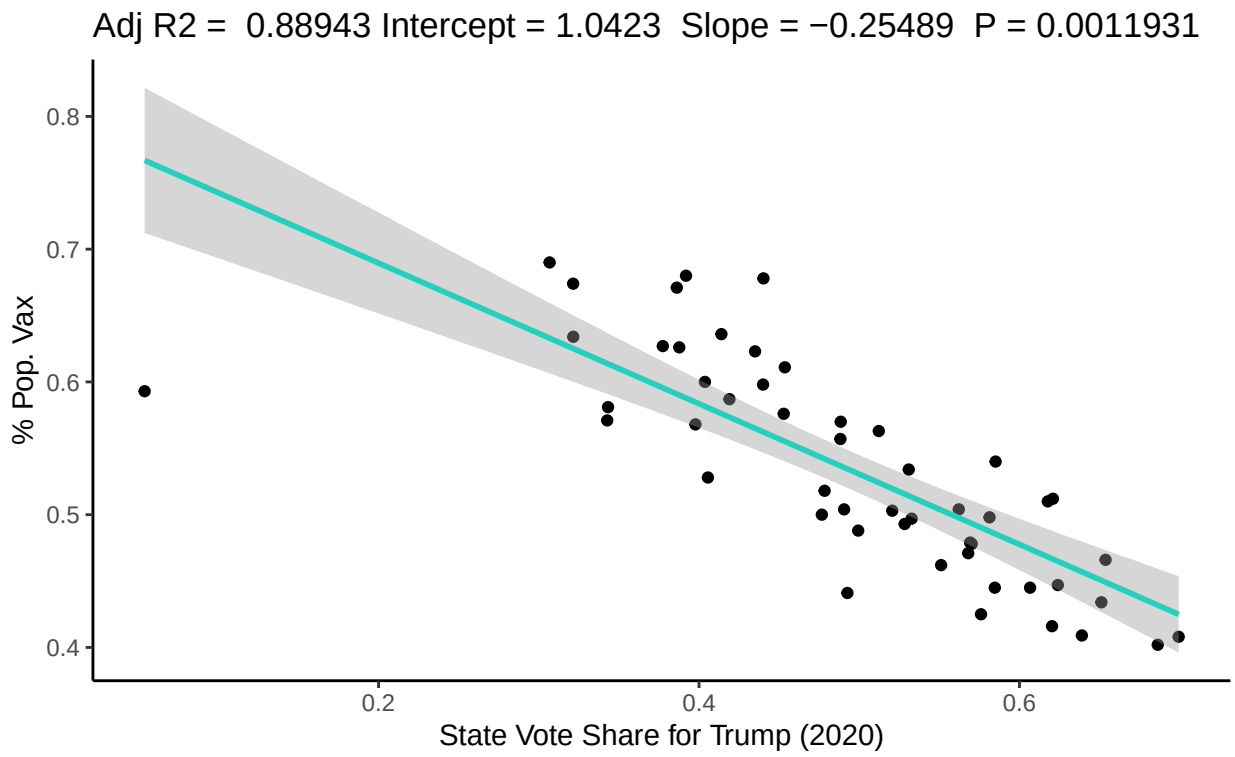


We fit our linear model using the backwards stepwise selection method, which starts with a full model, removes covariates one by one, and then determines which variables to retain based on pre-determined criteria. In this case, we employed the stepAIC() function, which utilizes the Akaike Information Criterion to determine which covariates to retain in the model.

The stepwise-selected model includes the following covariates: % Trump vote share, % White (race), % Black (race), % Hispanic (race), % Asian (race), % Other (race), % 18-24 (age), % 65+ (age), 'COVID Vaccine' search hits, 'COVID vaccine near me' search hits, 'COVID vaccine safe' search hits, 'COVID vaccine pregnant' search hits, and 'COVID vaccine microchip' search hits.

Trump vote share and race have negative coefficient estimates, while the age variables have positive coefficient estimates. 'COVID vaccine,' 'COVID vaccine pregnant,' and 'COVID vaccine microchip' have positive coefficient estimates, while the coefficient estimates for 'COVID vaccine near me' and 'COVID vaccine safe' are negative.

The adjusted R-squared value for this model is .8894, indicating that this model has a good fit. In addition, a histogram of the residuals shows that they are normally distributed, and a plot of the residual variances shows that there is no evidence of heteroskedasticity. We used a function created by Susan Johnston to make a plot of this model's regression line, as seen below [Susan E Johnston (2012)}.



Next, we removed Washington, DC, which is an outlier for the Trump vote share variable, from the data frame to evaluate the impact that would have on the regression model. When we applied the backwards stepwise selection to the updated dataset, which only included rows for the 50 states, a slightly different list of covariates were included in the final model, as seen in Table 8.

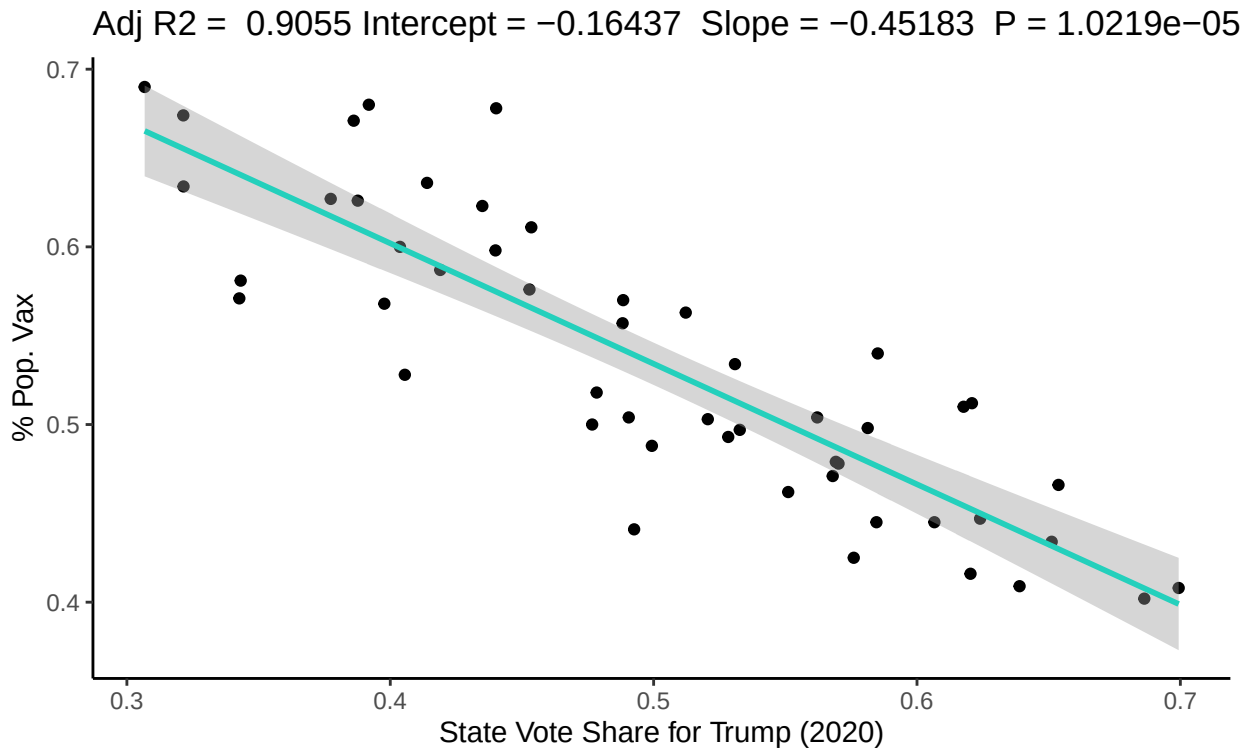
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 16, 2021 - 8:49:08 AM

The adjusted R-squared for this updated model is .905, indicating a good model fit, and the distributions of the residuals and the residual variances again indicate that the model is not violating any of the key linear regression assumptions. The coefficient estimate for Trump Vote Share in this model is -.45, while in the previous model (run on data that included Washington, DC), the coefficient estimate was -.25. This is a

Table 8: Regression Results Without Washington, DC

	Dependent variable:
	State Pct. Pop. Vaccinated
Pct. Trump Vote Share	-0.452^{***}
	(0.088)
Pct. White	-0.691^{***}
	(0.241)
Pct. Black	-0.733^{***}
	(0.229)
Pct. Hispanic	-0.611^{**}
	(0.251)
Pct. Asian	-0.991^{***}
	(0.247)
Pct. Other Race	-0.418
	(0.255)
Pct. 18-24	2.205^{**}
	(0.909)
Pct. 24-64	1.340
	(0.946)
Pct. 65+	1.441^{*}
	(0.850)
'COVID Vaccine' Search	0.004^{***}
	(0.001)
'COVID Vaccine Near Me' Search	-0.001^{***}
	(0.0005)
'COVID Vaccine Safe' Search	-0.001^{**}
	(0.0004)
'COVID Vaccine Ingredients' Search	0.0004
	(0.0003)
'COVID Vaccine Magnet' Search	0.0003^{*}
	(0.0002)
Constant	-0.164
	(0.882)
Observations	50
R^2	0.932
Adjusted R^2	0.905
Residual Std. Error	0.025 (df = 35)
F Statistic	34.536^{***} (df = 14; 35)
Note:	*p<0.1; **p<0.05; ***p<0.01

reflection of the impact Washington, DC was having on the regression line by dragging it downwards toward where that data point was located. Below, you can see that the regression line is steeper when Washington, DC is omitted from the analysis.



Ultimately, we decided to use the model that omitting Washington, DC as our final analysis model. Washington, DC is unique in that it is a small urban district, rather than a state that includes both urban and rural areas and a greater diversity of people. Therefore, if our goal is to understand how Google searches are impacting state-level vaccination rates, it makes sense to omit the entity that is not a state if it deviates from the trend of the states in meaningful ways.

The final model indicates that there are some vaccine search terms that are predictors of state vaccination rates, but the coefficient estimates are very small. A one unit increase in the search hit rate for 'COVID vaccine' in a state will result in a .4 percentage point increase in vaccination rates, when all other covariates are held constant. By contrast, a one percentage point increase in the 2020 vote share for Trump will result in a 45 percentage point decrease in vaccination rates in a given state, when other covariates are held constant. Overall, the regression analysis shows us that, when demographic factors are controlled for, some of the vaccine search terms are significant predictors for state vaccination rates. However, the coefficient estimates are so small that they are not practically significant, particularly in comparison to the much larger coefficient estimates estimated for Trump vote share, race, and age.

Conclusion

In this paper, we tried to determine what relationship there is between state-level COVID-19 vaccine rates and the types of Google searches that are made about vaccines in each state. Additionally, we tried to determine if vaccine myths more commonly searched for in states that also have low vaccination rates, and whether the relationship between COVID vaccine rates and Google searches change between June and September 2021.

We addressed these questions by collecting data from several web sources and conducting a correlation and regression analysis. We used gtrendsr() to run queries for 12 different Google search terms from Google

Trends, and joined that data with information about state level COVID-19 vaccination rates that we pulled from the CDC's data API. We also collected state demographic information from a variety of websites and loaded those data into R via Excel upload.

The correlation analysis showed that the for the January-September 2021 search time period, more mainstream searches are positively correlated with vaccination rates, and most of the myth-related searches are negatively correlated, which is the relationship you'd expect given the nature of the data. However, most of the correlations are not significant. When looking only at the searches that were done in July-September 2021, nearly all the search terms have a negative correlation. This is likely because most people in states with high vaccination rates had already been vaccinated by July. As a result, in these states, there are less people searching for vaccine information after July. Therefore in general, more of the vaccine searches (of all types) were happening in low vaccine rate states during July-September. But again, most of these correlations are not significant.

A linear regression that was built using backwards stepwise selection shows that there are some vaccine search terms that are predictors of state vaccination rates, but the coefficient estimates are very small. However, the coefficient estimates are so small that they are not practically significant, particularly in comparison to the much larger coefficient estimates estimated for Trump vote share, race, and age.

All of the code and data files for this analysis can be found on this public GitHub repository: https://github.com/ariannapschmid/FOCD

References

"2020 Popular Vote Tracker." n.d. *Cook Political Report*. Accessed December 15, 2021. https://www.cookpolitical.com/2020-national-popular-vote-tracker.

"Adult Population by Age Group." n.d. *KIDS COUNT Data Center*. Accessed December 15, 2021. https://datacenter.kidscount.org/data/tables/6538-adult-population-by-age-group.

CDC. 2021. "COVID-19 Vaccine Facts." *Centers for Disease Control and Prevention*. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/facts.html.

"COVID-19 Vaccine Myths Debunked." 2021. *Mayo Clinic Health System*. https://www.mayoclinichealthsystem.org/hometown-health/featured-topic/covid-19-vaccine-myths-debunked.

"Current Population Survey 2019 Annual Social and Economic (ASEC) Supplement." n.d. United States Census Bureau. https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households/h08.xls.

"List of U.S. States by Population." 2021. *Wikipedia, the Free Encyclopedia*. https://simple.wikipedia.org/w/index.php?title=List_of_U.S._states_by_population&oldid=7809509.

"Population Distribution by Race/Ethnicity (CPS)." 2021. *Kaiser Family Foundation*. https://www.kff.org/other/state-indicator/population-distribution-by-race-ethnicity-cps/.

Stephens-Davidowitz, Seth. 2014. "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data." *Journal of Public Economics* 118 (October): 26–40. https://doi.org/10.1016/j.jpubeco.2014.04.010.

Susan E Johnston. 2012. "A Quick and Easy Function to Plot Lm() Results with Ggplot2 in R." *Johnston Lab*. https://sejohnston.com/2012/08/09/a-quick-and-easy-function-to-plot-lm-results-in-r/.