COVID-19 Vaccination Rates and Google Search Data

Arianna Schmid and Stacey Frank

2021-12-14

Contents

Introduction **1**

Data Collection **1**

1. Google Trends and Keywords 2
2. Vaccine Rates 7
3. State-level Demographics 8

Analysis **10**

1. Correlation Analysis 10
2. Regression Analysis 12

References **22**

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 Google socially stigmatized topics than they would be to admit such opinions in a survey or other form of data collection. As a result, 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 particular, are vaccine myths more commonly searched for in states that also have low vaccination rates? A secondary question we investigate is does the relationship between COVID vaccine rates and Google searches change between June and September 2021?

Data Collection

1. Google Trends and Keywords

The CDC provides lists of the most common questions about the COVID-19 vaccine (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, a list of 12 keyword search terms was constructed. We call this list "k" to signify "keywords." It consists of the two general searches "covid vaccine" and "covid vaccine near me," five meainstream searches such as "covid vaccine side effects," and 5 myth-related searches such as "covid vaccine microchip." Note that each element in "k" is going to be renamed based on its index (hits.1, hits.2, . . . hits.12) for code efficiency.

```r
# Gtrends keyword searches
# Info about keyword searches: https://github.com/PMassicotte/gtrendsR/issues/268
k <- c( "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")
```

Table 1: Google Search Terms

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 number of search hits 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, we look at results 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. The function get.hit.results() takes in the date range as an argument, calls gtrends() 12 times (once for each search term), and returns the hit results by region based on that date. 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.

location	hits.1	hits.2	hits.3	hits.4	hits.5	hits.6	hits.7	hits.8	hits.9	hits.10	hits.11	hits.12
Alabama	56	78	49	51	61	29	83	0	67	55	23	31

location	hits.1	hits.2	hits.3	hits.4	hits.5	hits.6	hits.7	hits.8	hits.9	hits.10	hits.11	hits.12
Alaska	56	58	98	42	79	0	86	0	33	0	38	0
Arizona	68	87	62	52	62	42	83	68	62	30	20	5
Arkansas	47	62	59	85	72	43	91	0	42	43	29	42
California	70	79	64	49	82	35	82	45	51	35	19	30
Colorado	68	70	52	43	72	25	78	51	76	64	36	28
Connecticut	93	73	76	49	87	29	91	78	56	0	34	59
Delaware	82	93	75	38	77	36	87	0	71	0	10	0
District of Columbia	73	53	53	25	56	0	63	0	0	0	0	0
Florida	72	96	55	60	50	37	70	46	44	28	18	46
Georgia	58	97	49	48	62	39	71	63	62	40	22	52
Hawaii	66	68	88	69	98	100	92	0	27	0	23	80
Idaho	50	66	57	61	46	32	83	0	36	0	34	47
Illinois	75	79	58	55	71	42	79	46	50	35	32	32
Indiana	55	67	56	63	54	33	78	0	52	53	23	63
Iowa	64	65	47	64	60	29	71	0	43	47	32	25
Kansas	52	61	62	42	78	47	87	0	56	97	37	70
Kentucky	60	66	53	51	58	34	78	42	83	70	40	100
Louisiana	46	67	37	35	52	24	70	0	90	35	23	54
Maine	79	63	69	100	80	28	99	0	55	0	24	55
Maryland	83	81	54	62	71	29	80	68	32	49	33	25
Massachusetts	93	76	62	58	70	33	80	40	64	34	34	17
Michigan	68	75	68	67	65	40	84	17	40	31	42	40
Minnesota	68	76	64	64	63	46	92	57	66	33	36	19
Mississippi	49	79	60	49	33	26	72	0	46	24	35	61
Missouri	58	70	57	59	60	29	81	85	59	62	20	37
Montana	44	45	53	74	37	0	72	0	77	0	44	38
Nebraska	51	54	35	55	67	31	84	0	37	0	59	18
Nevada	58	68	49	85	46	21	69	0	37	0	6	21
New Hampshire	67	43	70	91	84	34	92	0	100	0	29	74
New Jersey	100	87	73	54	77	38	83	47	55	26	30	34
New Mexico	52	65	50	39	66	24	80	0	46	0	37	0
New York	82	78	57	57	64	27	65	42	42	22	20	34
North Carolina	57	80	53	43	69	22	84	41	44	55	33	54
North Dakota	39	43	59	68	90	0	76	0	80	0	0	0
Ohio	68	75	49	75	54	36	90	26	51	49	32	43
Oklahoma	45	63	45	46	66	36	62	0	82	100	34	22
Oregon	74	76	54	57	85	47	95	0	58	0	33	8
Pennsylvania	86	98	68	64	65	26	93	12	60	26	35	36
Rhode Island	88	70	100	68	100	22	89	0	32	0	18	0
South Carolina	58	88	54	49	59	36	83	0	47	33	25	43
South Dakota	42	41	43	21	24	0	100	0	50	0	100	98
Tennessee	55	68	50	67	80	34	71	23	66	39	39	20
Texas	63	100	48	48	62	31	71	60	52	33	20	25
Utah	53	65	64	38	70	45	86	0	45	0	42	22
Vermont	79	53	54	26	59	0	67	0	0	0	18	0
Virginia	72	79	55	55	66	34	85	100	57	45	27	25
Washington	77	85	77	70	85	24	85	60	53	68	17	26
West Virginia	51	53	84	74	42	25	99	0	12	0	64	49
Wisconsin	61	70	61	87	69	51	88	85	62	48	34	43
Wyoming	41	40	75	32	0	0	51	0	0	0	0	0

location	hits.1	hits.2	hits.3	hits.4	hits.5	hits.6	hits.7	hits.8	hits.9	hits.10	hits.11	hits.12
Alabama	40	61	22	26	34	16	47	0	36	28	34	52
Alaska	54	56	55	0	41	0	36	0	100	0	0	87
Arizona	52	75	33	42	59	16	54	0	46	17	5	20
Arkansas	38	48	20	56	53	15	42	0	22	0	61	0
California	70	75	48	32	51	30	62	24	39	24	27	38
Colorado	72	74	38	63	56	36	63	100	61	39	55	23
Connecticut	94	78	41	57	53	53	76	50	52	0	100	45
Delaware	75	92	50	100	0	38	86	0	55	0	38	0
District of Columbia	69	62	21	84	0	0	59	0	0	0	0	0
Florida	60	83	35	36	39	30	49	42	39	25	38	49
Georgia	45	80	29	35	43	14	46	32	46	19	26	29
Hawaii	72	62	68	0	16	84	51	0	41	0	0	70
Idaho	43	48	34	13	59	50	62	0	0	0	26	0
Illinois	68	78	47	36	39	37	68	29	31	9	51	23
Indiana	53	63	27	32	29	12	60	34	44	100	25	84
Iowa	55	66	21	49	39	0	72	0	58	0	67	67
Kansas	43	48	20	23	35	0	44	0	43	0	45	37
Kentucky	44	49	29	72	23	39	43	0	56	32	59	61
Louisiana	29	53	33	15	17	9	37	0	55	63	19	47
Maine	98	80	48	63	52	0	87	0	43	0	30	37
Maryland	80	72	34	58	55	10	72	0	38	53	37	13
Massachusetts	95	68	46	32	74	23	77	77	47	0	28	47
Michigan	73	80	53	47	62	46	73	43	79	26	67	58
Minnesota	71	80	43	42	47	13	77	73	48	22	40	17
Mississippi	29	56	26	26	19	16	37	0	72	0	33	0
Missouri	47	61	34	28	42	19	56	0	28	22	26	49
Montana	46	48	56	22	0	0	83	0	60	0	0	52
Nebraska	49	53	36	21	46	39	54	0	28	0	79	25
Nevada	54	68	38	63	38	43	56	0	31	0	11	14
New Hampshire	71	55	57	71	47	27	100	0	0	0	27	0
New Jersey	84	80	53	36	57	21	64	40	42	36	18	32
New Mexico	47	65	15	60	100	37	57	0	0	0	57	24
New York	65	71	38	34	47	33	47	15	24	18	55	33
North Carolina	46	72	32	34	44	23	55	0	47	21	30	41
North Dakota	35	38	45	30	0	0	55	0	0	0	57	0
Ohio	55	67	37	49	32	15	66	0	53	10	25	38
Oklahoma	34	56	34	24	41	34	48	0	67	76	35	14
Oregon	87	86	28	32	62	17	83	0	51	0	35	22
Pennsylvania	72	89	62	51	40	35	77	30	70	36	46	47
Rhode Island	95	54	100	89	20	100	81	0	0	0	34	0
South Carolina	43	61	37	20	45	8	58	0	44	76	46	10
South Dakota	42	50	14	28	0	0	41	0	77	0	0	0
Tennessee	42	65	39	34	38	27	48	30	63	18	22	7
Texas	52	86	29	22	46	20	48	19	49	50	28	32
Utah	50	54	31	25	55	12	71	0	34	0	47	59
Vermont	100	61	18	0	39	0	78	0	0	0	0	0
Virginia	66	73	44	42	43	34	75	21	44	25	35	29
Washington	82	100	41	54	65	9	71	26	60	31	33	46
West Virginia	47	53	42	70	94	0	55	0	77	0	54	100
Wisconsin	58	66	46	28	43	13	68	0	29	22	27	17
Wyoming	38	36	0	0	0	0	31	0	0	0	0	0

location	hits.1	hits.2	hits.3	hits.4	hits.5	hits.6	hits.7	hits.8	hits.9	hits.10	hits.11	hits.12
Alabama	90	86	86	91	65	38	84	0	51	55	48	84
Alaska	93	80	81	35	41	0	68	0	0	0	0	0
Arizona	77	85	50	25	52	30	54	28	34	18	21	55
Arkansas	96	83	49	71	100	28	86	0	38	34	39	77
California	77	83	40	35	52	27	43	14	37	19	12	18
Colorado	66	74	33	49	42	16	44	0	51	39	11	0
Connecticut	74	75	36	28	59	8	35	46	22	0	23	0
Delaware	77	98	44	57	0	34	73	0	0	0	23	0
District of Columbia	57	84	31	0	24	0	16	0	0	0	0	0
Florida	82	91	52	38	45	34	61	0	34	22	18	39
Georgia	81	94	45	48	38	32	62	29	32	13	18	43
Hawaii	100	80	100	58	50	26	50	0	34	0	0	0
Idaho	84	82	63	27	78	48	100	0	32	0	17	0
Illinois	69	76	44	37	29	32	49	0	37	15	12	55
Indiana	76	81	42	38	89	6	55	0	53	34	16	31
Iowa	63	60	22	42	57	75	42	0	17	15	9	0
Kansas	75	69	24	47	27	42	50	0	19	100	29	0
Kentucky	79	79	58	81	70	27	79	100	36	21	44	98
Louisiana	89	81	83	62	65	64	78	0	0	0	25	100
Maine	74	67	43	99	49	50	63	0	0	0	0	0
Maryland	72	83	28	44	48	34	50	0	20	59	27	0
Massachusetts	67	76	24	53	36	17	28	0	45	15	12	23
Michigan	68	66	30	36	21	29	43	0	43	26	15	39
Minnesota	76	75	48	42	40	12	41	0	58	30	17	0
Mississippi	82	100	52	59	48	30	68	0	40	0	21	0
Missouri	83	86	40	49	48	18	66	68	25	44	21	67
Montana	63	56	48	84	72	0	68	0	49	0	26	0
Nebraska	64	68	47	31	35	55	55	0	73	0	0	0
Nevada	73	71	53	52	33	0	52	0	14	0	14	0
New Hampshire	66	62	85	55	32	0	53	0	0	0	34	0
New Jersey	75	86	47	34	43	23	41	0	22	20	21	0
New Mexico	74	77	46	61	58	18	35	0	96	0	0	0
New York	62	67	42	37	38	16	33	0	20	21	8	27
North Carolina	79	86	46	58	42	33	56	0	43	28	27	65
North Dakota	68	63	0	90	34	0	23	0	0	0	0	0
Ohio	71	72	49	37	30	26	41	32	23	21	12	32
Oklahoma	81	85	37	79	21	11	74	0	57	39	45	59
Oregon	86	77	64	74	69	16	49	0	33	0	17	89
Pennsylvania	69	75	47	41	43	26	42	0	31	22	16	0
Rhode Island	66	68	26	34	40	31	23	0	82	0	21	0
South Carolina	83	82	57	51	50	56	69	0	47	0	10	76
South Dakota	73	52	42	0	62	0	63	0	0	0	100	0
Tennessee	84	86	60	50	45	20	65	0	46	12	38	54
Texas	81	97	56	31	47	26	53	12	32	23	17	24
Utah	73	78	52	49	85	11	67	0	29	0	23	60
Vermont	59	51	26	100	76	0	51	0	0	0	0	0
Virginia	72	83	47	41	59	25	47	20	15	22	3	0
Washington	83	81	39	68	87	45	47	50	30	48	12	24
West Virginia	79	75	43	70	32	100	71	0	100	0	35	0
Wisconsin	75	68	40	38	48	25	50	0	41	59	4	0
Wyoming	65	76	0	40	0	0	77	0	0	0	0	0

Next, we are interested in getting the count of states that actually have a gtrends ranking for each search term. This is accomplished through the get.search.terms() function. It takes the hits results object from get.hit.results() as an argument, and returns the sum of states that are present 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 6 that for April-June, "covid vaccine," "covid vaccine near me," and "covid vaccine side effects" have hits in all 50 states and Washington DC. On the other hand, "covid vaccine microchip" was only searched for in 17 states.

Table 5: 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 6: 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 7: 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
hits.6	covid vaccine protect	42

Variable Name	Search Term	Number of States with Hit Rate
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. Vaccine 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. This data is acquired by using RSocrata to pull CDC COVID vaccine data through their API. After cleaning, the datasets vax.June21 and vax.Sept21 are created, where vax.June21 corresponds to the hit results from April-June, and vax.Sept21 corresponds to January-September and July-September.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 admitted for each 100,000th person.

```
x <- vax.June21 %>% select(location, series_complete_pop_pct, admin_per_100k) %>% arrange(location)
kable(head(x, n = 10), caption = "vax.June21")
```

Table 8: vax.June21

location	series_complete_pop_pct	admin_per_100k
Alabama	31.9	67006
Alaska	41.8	87653
Arizona	39.0	87211
Arkansas	33.4	72577
California	48.3	105246
Colorado	50.2	104991
Connecticut	59.0	121188
Delaware	48.0	103773
District of Columbia	50.7	119642
Florida	44.1	95343

```
y <- vax.Sept21 %>% select(location, series_complete_pop_pct, admin_per_100k) %>% arrange(location)
kable(head(x, n = 10), caption = "vax.Sept21")
```

Table 9: vax.Sept21

location	series_complete_pop_pct	admin_per_100k
Alabama	31.9	67006
Alaska	41.8	87653
Arizona	39.0	87211
Arkansas	33.4	72577
California	48.3	105246
Colorado	50.2	104991
Connecticut	59.0	121188
Delaware	48.0	103773

location	series_complete_pop_pct	admin_per_100k
District of Columbia	50.7	119642
Florida	44.1	95343

```
## Create two datasets for our vaccination dates of interest

vax.June21 <- cdc.df.50 %>%
  filter(date == "2021-06-21")
vax.Sept21 <- cdc.df.50 %>%
  filter(date == "2021-09-21")
```

3. State-level Demographics

Certain state-level demographic factors could potentially have an impact on our analysis. We turned to different sources for this information. Data of interest included state population counts, voter information, household income, age group, and race.

State population numbers were web scraped from Wikipedia using selector gadget (https://simple.wikipedia.org/wiki/List_of_U.S._states_by_population). These state counts were then joined with the vax.June21 and vax.Sept21 frames created in Part 2.

Table 10: State Level Percent of Population Fully Vaccinated by September 2021

State	Percent Pop. Fully vaccinated
West Virginia	40.2
Wyoming	40.8
Idaho	40.9
Alabama	41.6
Mississippi	42.5
North Dakota	43.4
Georgia	44.1
Louisiana	44.5
Tennessee	44.5
Arkansas	44.7
South Carolina	46.2
Oklahoma	46.6
Missouri	47.1
Indiana	47.8
Montana	47.9
North Carolina	48.8
Alaska	49.3
Ohio	49.7
Utah	49.8
Nevada	50.0
Texas	50.3
Kansas	50.4
Arizona	50.4
South Dakota	51.0
Kentucky	51.2
Michigan	51.8
Illinois	52.8

State	Percent Pop. Fully vaccinated
Iowa	53.4
Nebraska	54.0
Wisconsin	55.7
Florida	56.3
Delaware	56.8
Pennsylvania	57.0
Hawaii	57.1
Minnesota	57.6
California	58.1
Colorado	58.7
District of Columbia	59.3
Virginia	59.8
Oregon	60.0
New Hampshire	61.1
New Mexico	62.3
Washington	62.6
New York	62.7
Maryland	63.4
New Jersey	63.6
Rhode Island	67.1
Massachusetts	67.4
Maine	67.8
Connecticut	68.0
Vermont	69.0

In addition, voter information was downloaded from Cook Political https://cookpolitical.com/2020-national-popular-vote-tracker. Unnecessary columns were deleted and columns of interest were renamed. Here, we are primarily interested in the share of republican voters in the 2020 election. The final spreasheet was saved as vote.xlsx. Median household income was downloaded from Census https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households/h08.xls and saved as med.income.xlsx. 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 https://datacenter.kidscount.org/data/tables/6538-adult-population-by-age-group#detailed/2/2-53/false/574,1729,37,871,870,573,869,36,868,867/117,2801,2802,2803/13515,13516. 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 18 to 24, 25 to 64, and Ages 65 and over variables. The final spreadsheet was saved as percent.age.xlsx. Race data was pulled as a csv file from the Current Population Survey. https://www.kff.org/other/state-indicator/population-distribution-by-race-ethnicity-cps/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D 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' 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). Cov was then joined with the two CDC vaccine rate data collected 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 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.

```
## This function joins the gtrends dataset with vaccine info dataset
join.gtrends.vaccine <- function (hits.results.month,vax.month){

  month.analysis <- vax.month %>%
      select(location,date, admin_per_100k, series_complete_pop_pct,
             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)
```

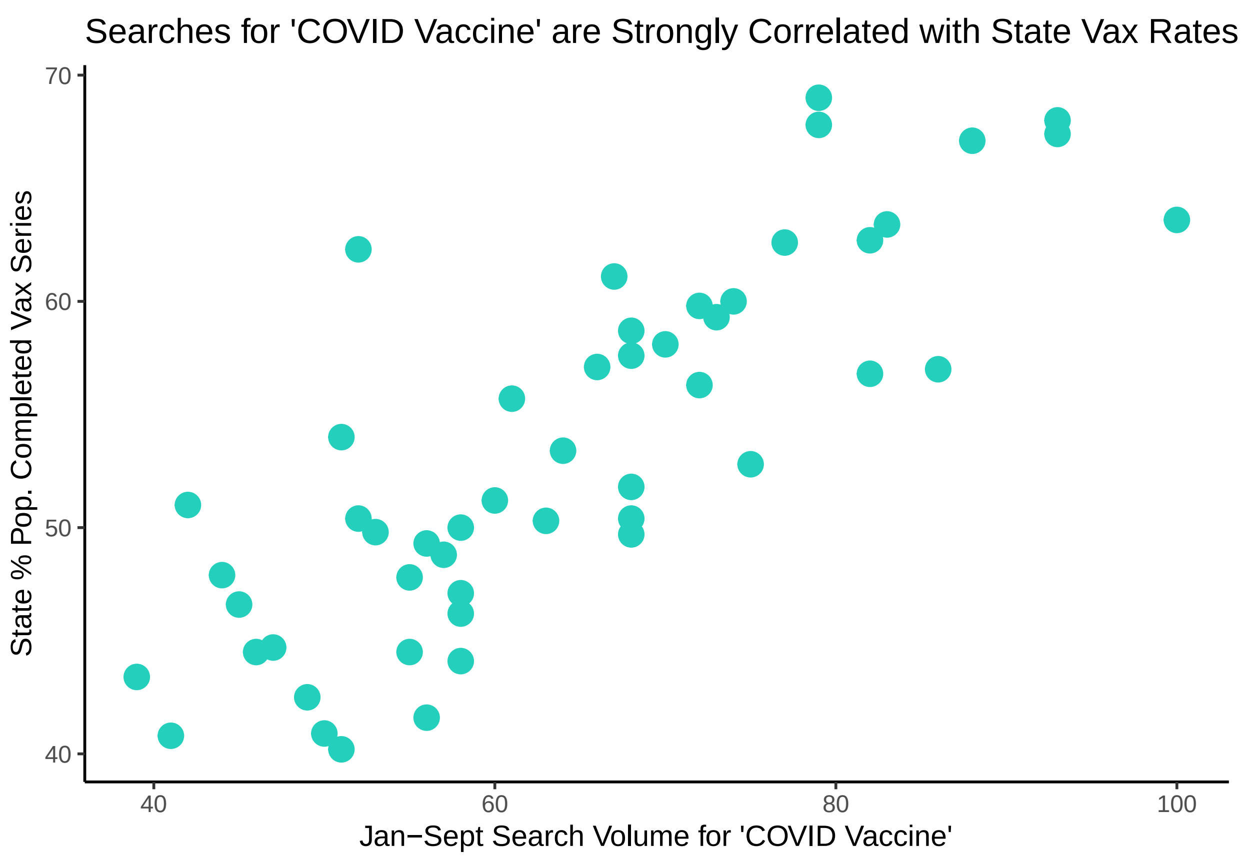
Analysis

1. Correlation Analysis

```
## 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)
```

```
##Plotting of correlations

#Jan-Sept Searches
# using series_complete_pop_pct as measure for state vaccination rate
ggplot(Jan01.analysis) + geom_point(aes(hits.1, series_complete_pop_pct), color = '#24d0bc', size = 4) +
      labs(y = "State % Pop. Completed Vax Series", x = "Jan-Sept Search Volume for 'COVID Vaccine'") +
      ggtitle("Searches for 'COVID Vaccine' are Strongly Correlated with State Vax Rates") +
      theme_classic()
```

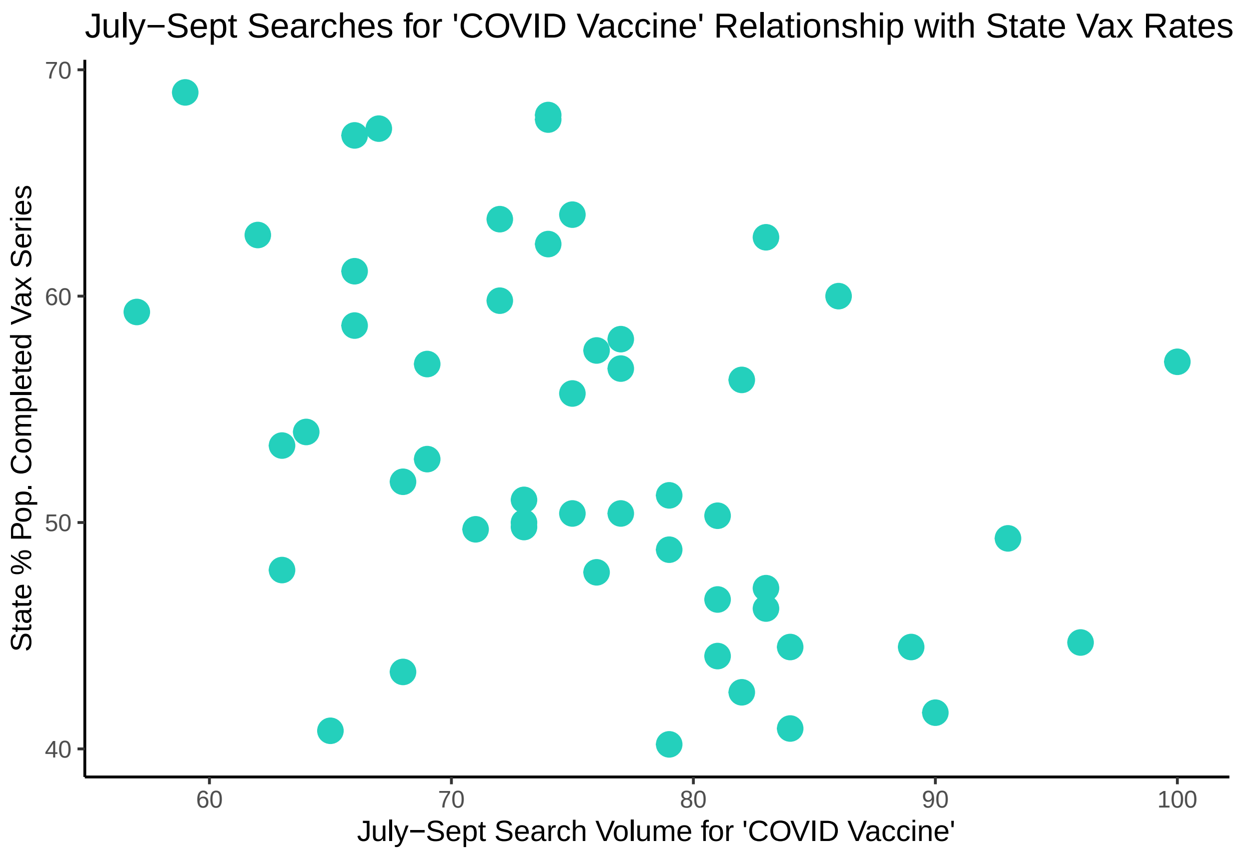



```
ggsave("covid.correlation.Jan.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
##Plotting of correlations continued

#July-September Searches
# using series_complete_pop_pct as measure for state vaccination rate
ggplot(Sept21.analysis) + geom_point(aes(hits.1, series_complete_pop_pct), color = '#24d0bc', size = 4)
```

```
    labs(y = "State % Pop. Completed Vax Series", x = "July-Sept Search Volume for 'COVID Vaccine'") +
    ggtitle("July-Sept Searches for 'COVID Vaccine' Relationship with State Vax Rates") +
    theme_classic()
```

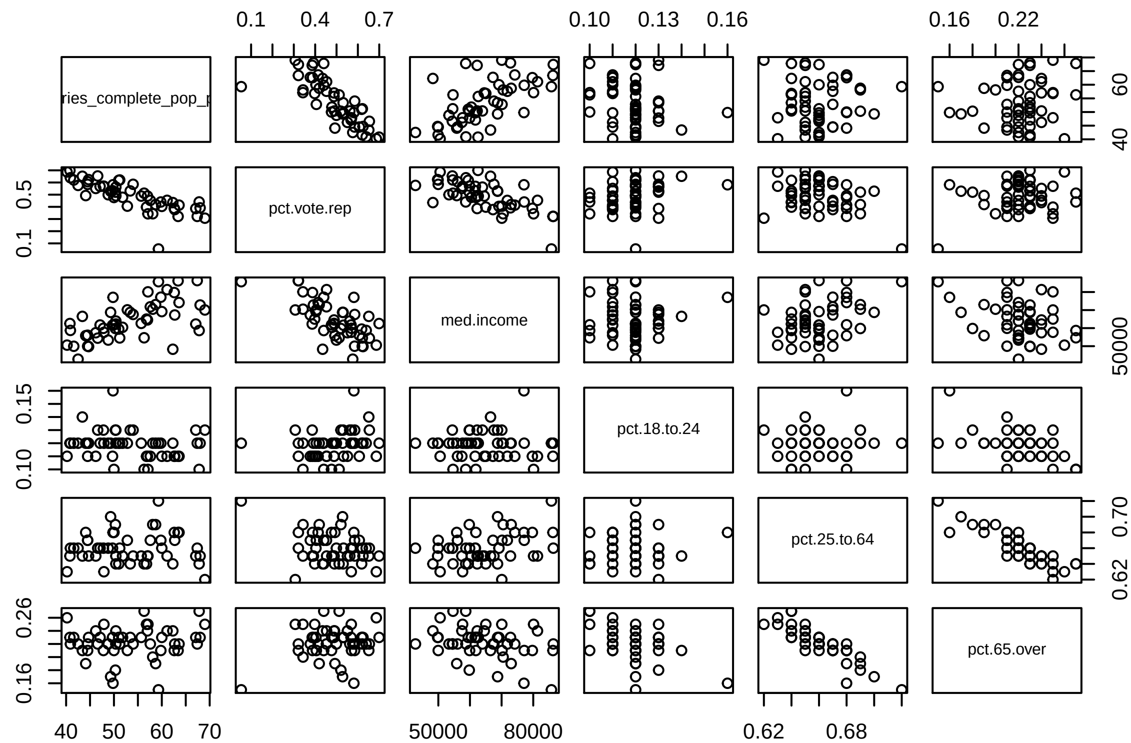



```
ggsave("covid.correlation.Sept.png")
```

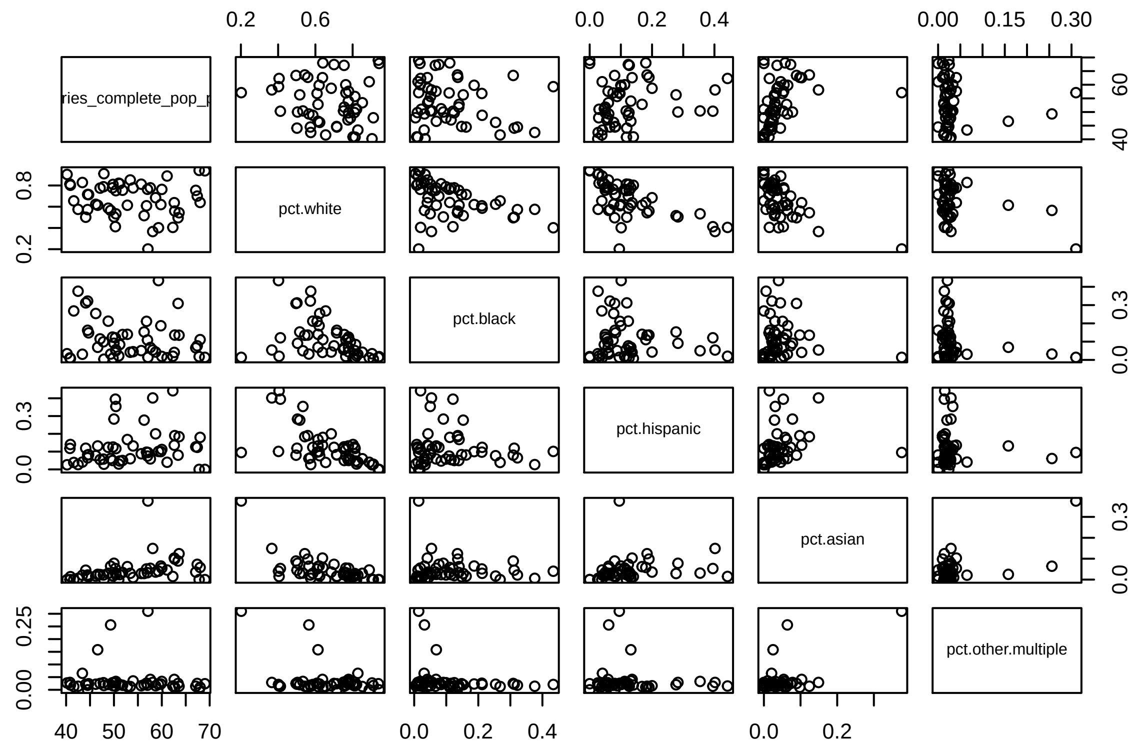
```
## Saving 6.5 x 4.5 in image
```

2. Regression Analysis

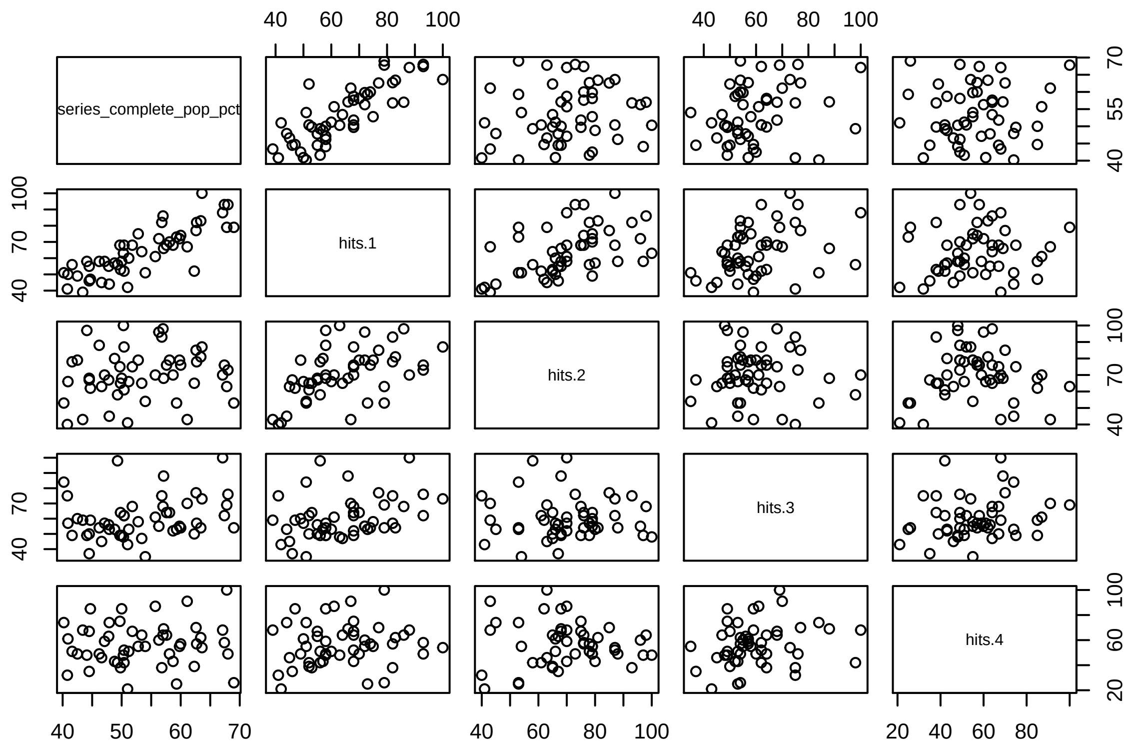
```
# Plotting and regression analysis

Jan01.analysis %>% select(series_complete_pop_pct, pct.vote.rep,med.income,
                          pct.18.to.24,pct.25.to.64,pct.65.over) %>%
                          plot()
```



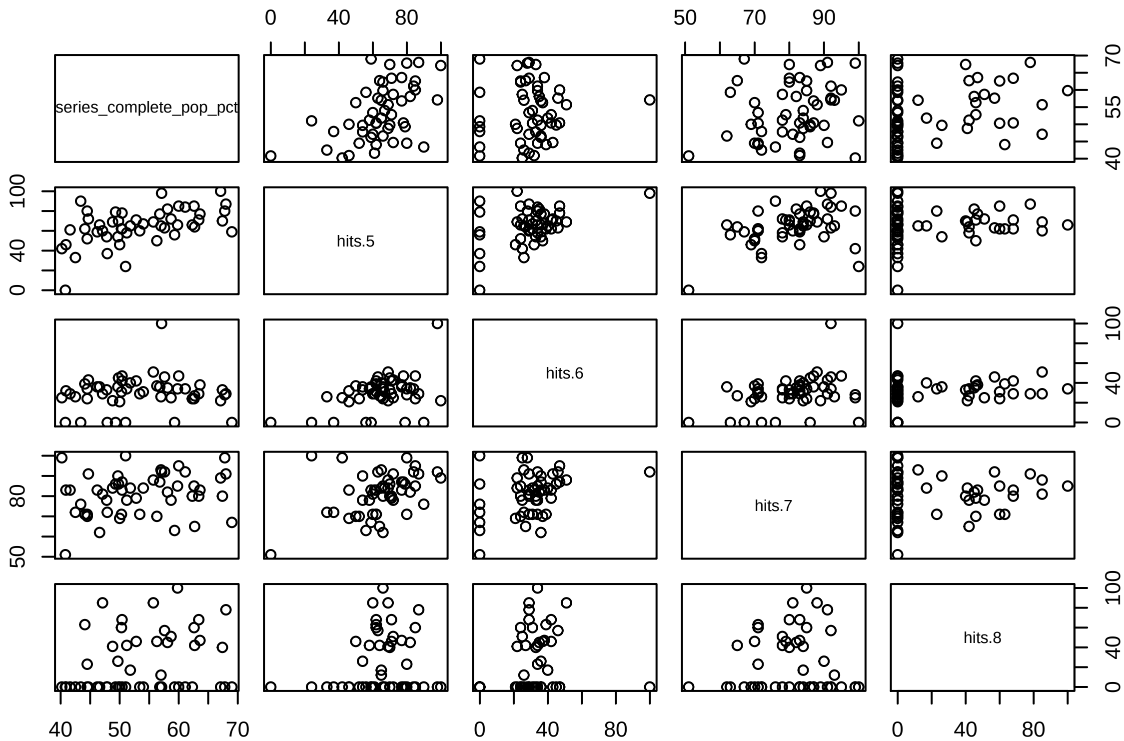

```
##percent republican and median income seem to have a linear relationship with series_complete_pop_pct;

Jan01.analysis %>% select(series_complete_pop_pct, pct.white,pct.black,pct.hispanic,pct.asian,pct.other
                         plot()
```



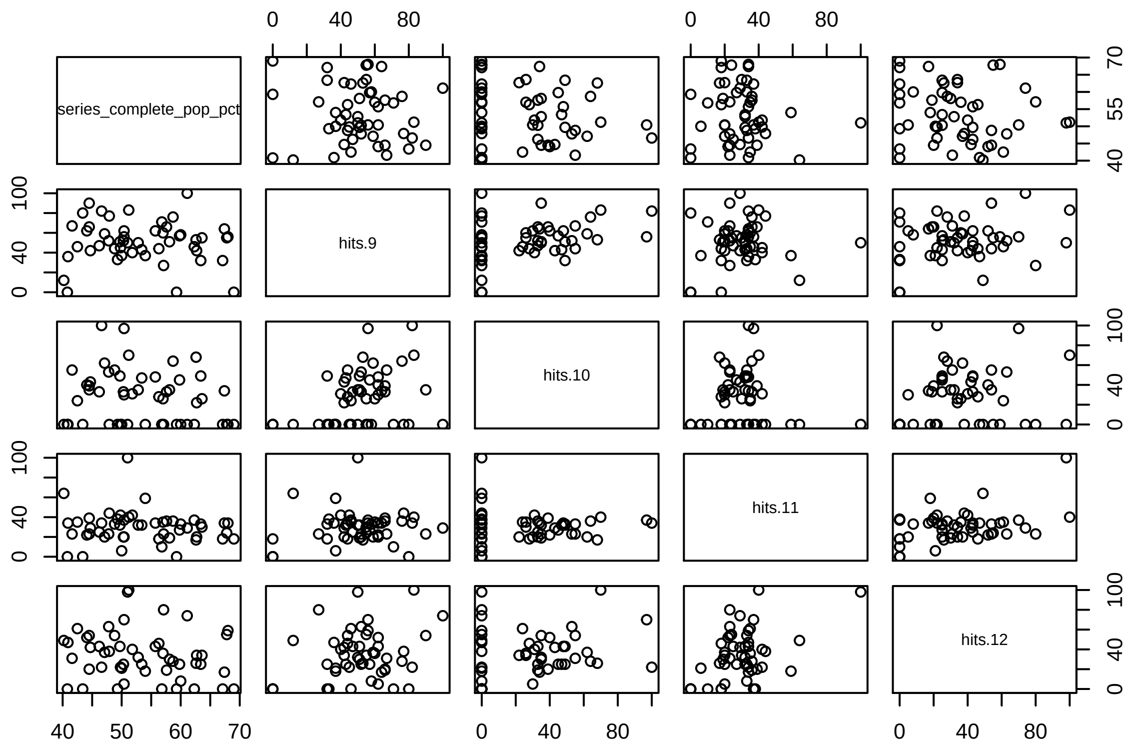

```
##None of the race variables seem to be related to vax rates


Jan01.analysis %>% select(series_complete_pop_pct, hits.1,hits.2,hits.3,hits.4) %>%
                   plot()
```



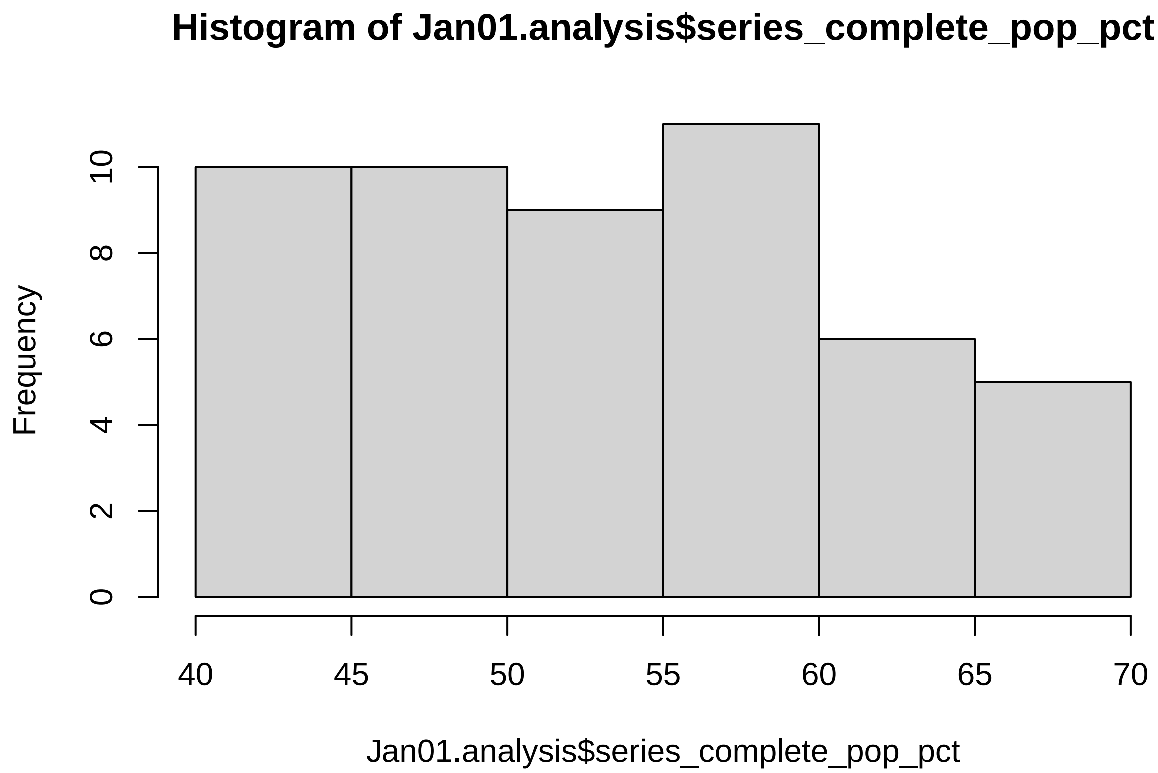

```
#Hits1 is related; other plots are widely scattered

Jan01.analysis %>% select(series_complete_pop_pct, hits.5,hits.6,hits.7,hits.8) %>%
                          plot()
```



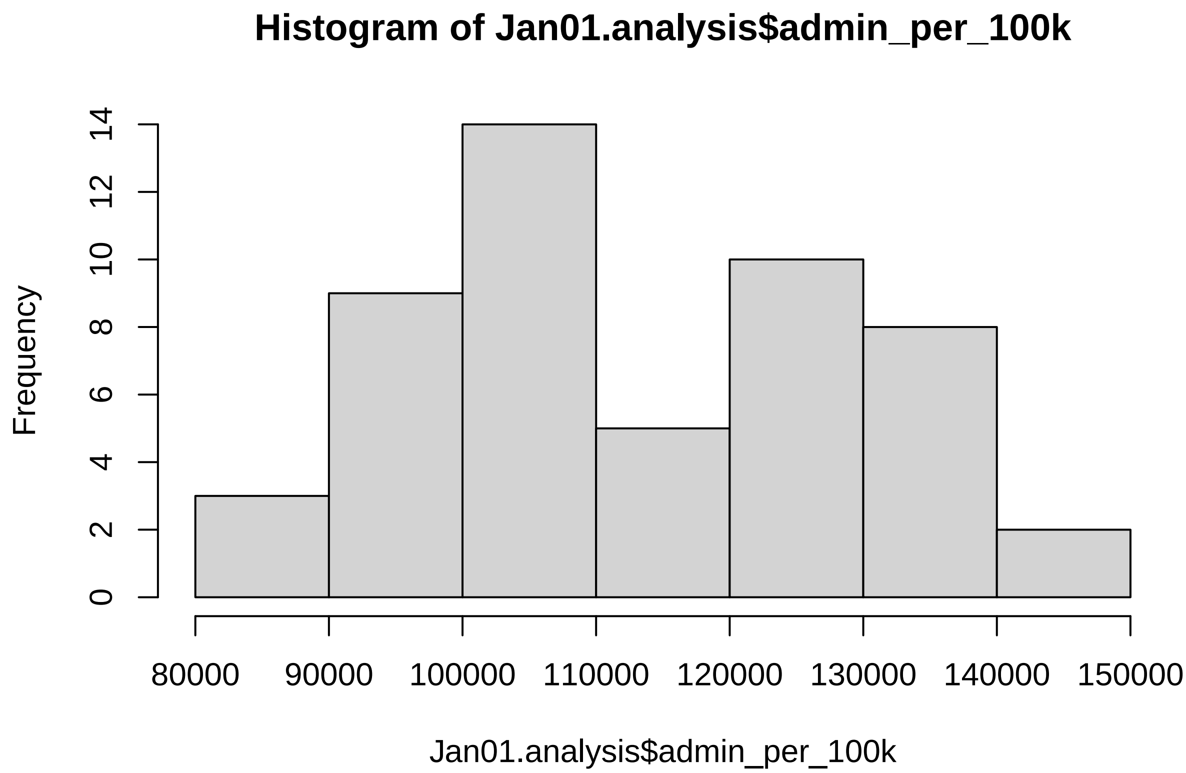

```
#Hits5 has some relationship; others not so much

Jan01.analysis %>% select(series_complete_pop_pct, hits.9,hits.10,hits.11,hits.12) %>%
                     plot()
```




```
#No strong relationships here


##histogram of outcome variable

hist(Jan01.analysis$series_complete_pop_pct)
```




```
hist(Jan01.analysis$admin_per_100k)
```

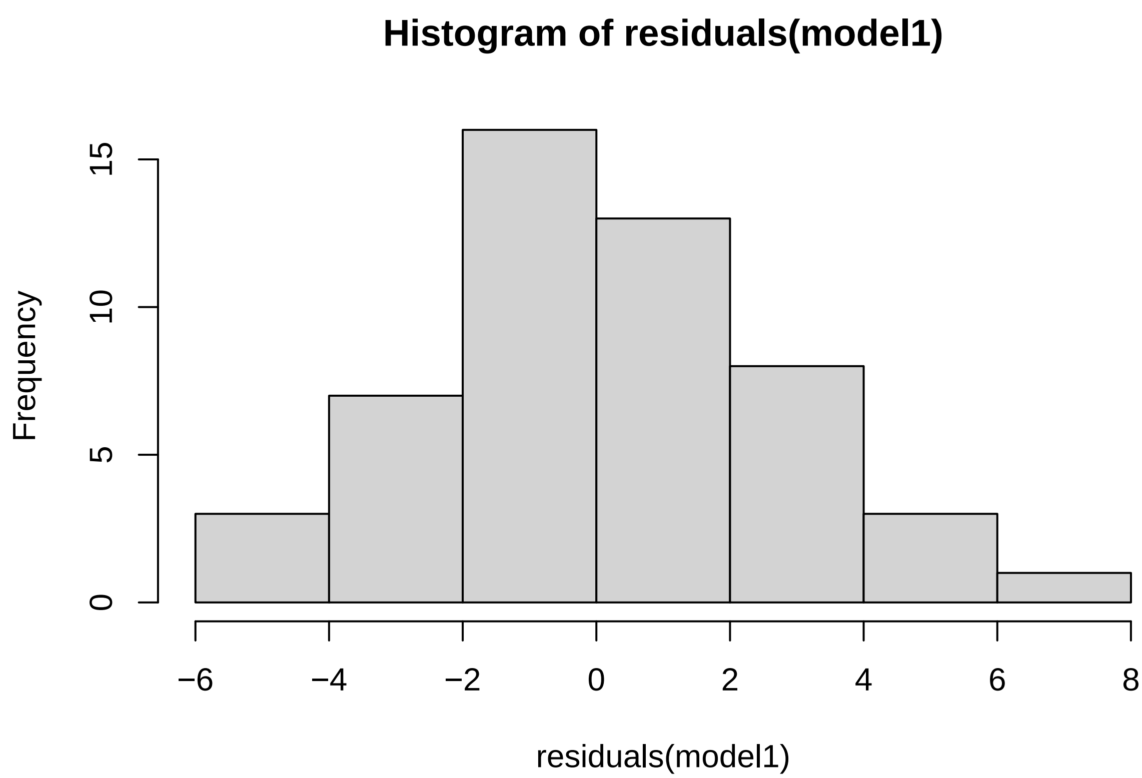



```r
##Linear model

model1 <- lm(series_complete_pop_pct ~ pct.vote.rep + pct.white + pct.black + hits.1 + hits.2 + hits.3 +

summary(model1)
```

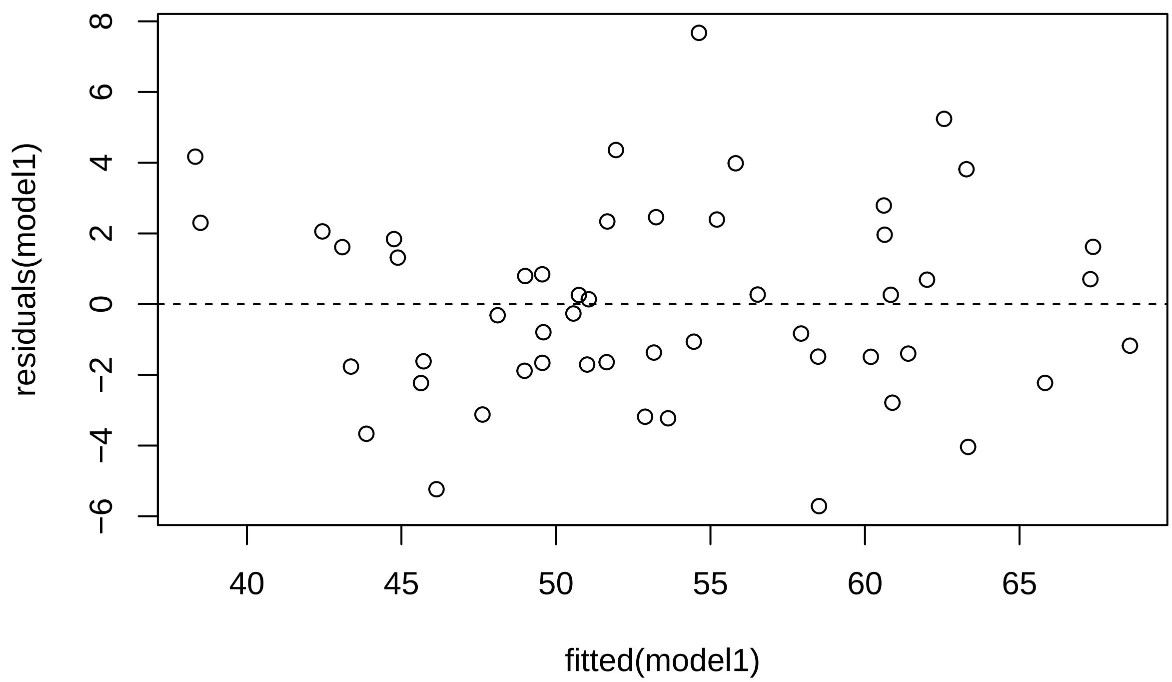
```
##
## Call:
## lm(formula = series_complete_pop_pct ~ pct.vote.rep + pct.white +
##     pct.black + hits.1 + hits.2 + hits.3 + hits.4 + hits.5 +
##     hits.6 + hits.7 + hits.8 + hits.9 + hits.10 + hits.11 + hits.12,
##     data = Jan01.analysis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7127 -1.6852 -0.2653  1.9016  7.6727
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.577e+01  6.716e+00   8.303 8.67e-10 ***
## pct.vote.rep -2.851e+01  9.199e+00  -3.099 0.003814 **
## pct.white    -1.523e+00  5.263e+00  -0.289 0.774004
## pct.black    -2.223e+01  6.187e+00  -3.593 0.000995 ***
## hits.1        3.474e-01  8.097e-02   4.290 0.000134 ***
## hits.2       -8.337e-02  6.730e-02  -1.239 0.223675
## hits.3       -8.722e-02  5.140e-02  -1.697 0.098609 .
```

```
## hits.4        -7.262e-03  3.543e-02  -0.205 0.838762
## hits.5         5.730e-02  4.619e-02   1.241 0.222954
## hits.6        -4.797e-02  4.233e-02  -1.133 0.264723
## hits.7         4.173e-04  8.533e-02   0.005 0.996126
## hits.8         5.368e-03  2.026e-02   0.265 0.792645
## hits.9         1.086e-02  3.036e-02   0.358 0.722698
## hits.10       -4.966e-03  2.417e-02  -0.205 0.838416
## hits.11        3.135e-02  4.718e-02   0.664 0.510760
## hits.12        2.003e-02  2.701e-02   0.742 0.463288
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.261 on 35 degrees of freedom
## Multiple R-squared:  0.8867, Adjusted R-squared:  0.8381
## F-statistic: 18.26 on 15 and 35 DF,  p-value: 2.981e-12
```

```r
#Check that residuals are normally distributed

hist(residuals(model1))
```




```r
#Check for homoskedasticity in residual variances (looks ok)

plot(fitted(model1), residuals(model1))
abline(h = 0, lty = 2)
```




```
#Linear model with interaction
#When adding interaction between hits.1 and % who voted republican, the main effects and the interactio

model2 <- lm(series_complete_pop_pct ~ pct.vote.rep + pct.black + hits.1  + hits.1*pct.vote.rep, data =

summary(model2)
```

```
##
## Call:
## lm(formula = series_complete_pop_pct ~ pct.vote.rep + pct.black +
##     hits.1 + hits.1 * pct.vote.rep, data = Jan01.analysis)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.5878 -2.1744 -0.2679  2.5307  7.9335
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            52.3502    12.8896   4.061 0.000188 ***
## pct.vote.rep          -25.7923    23.4115  -1.102 0.276326
## pct.black             -25.3249     4.7632  -5.317 3.01e-06 ***
## hits.1                  0.3237     0.1778   1.821 0.075133 .
## pct.vote.rep:hits.1   -0.1366     0.3496  -0.391 0.697849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 3.381 on 46 degrees of freedom
## Multiple R-squared:   0.84,  Adjusted R-squared:  0.826
## F-statistic: 60.36 on 4 and 46 DF,  p-value: < 2.2e-16
```

```
save.image(file = "shared_work_space.RData")
```

References

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.