# Paper

# Ankit Patel, Hallie Trial, and Prayag Gordy

# 7/26/2021

# Contents

| Introduction   | 1 |
|--|---|
| Data   | 1 |
| census_county and census_region  | 1 |
| Daily Covid counts   | 2 |
| Presidential Election Results by County  | 2 |
| Main Vaccination Dataset   | 2 |
| Texas Vaccination  | 3 |
| Small California Counties Vaccination  | 4 |
| Vaccine Hesitancy  | 4 |
| Methods  | 5 |
| Results  | 5 |
| County Vaccination Histogram by Region   | 5 |
| How are the prevailing political views of a county related to its rate of vaccination?         | 6 |
| What are the racial demographics of counties that do not conform to the expected relationship  |   |
| between politics and vaccination?  | 7 |
| Did the severity of the pandemic within a county affect residents' desire to get vaccinated? . | 8 |
| Discussion   | 9 |
| Appendix   | 9 |

# Introduction

# Data

# census\_county and census\_region

We used the tidycensus package in R to obtain American Community Survey 5-year data for the 2014-2019 span. The ACS is an annual survey that still collect large amounts of data from every single county in the US, so the data is more recent than the 2010 census and is still very comprehensive (2020 census data is not publicly available yet). We called two different data sets with different geographies: For one set, we got all the desired variables by county, while for the other set, we called the variables by region. The variables we called are as follows:

- 1. med\_income (B19013\_001 in tidycensus): the median income of the chosen geography's population (each county in census\_county, each region in census\_region).
- 2. male (B01001\_002): the total number of males in the selected geography's population.
- 3. female (B01001 026): the total number of females in the selected geography's population.
- 4. med\_age (B01002 001): the median age of the chosen geography's population.
- 5. white (B01001A 001): the total number of white residents in the selected geography's population.
- 6. black (B01001B\_001): the total number of black residents in the selected geography's population.

- 7. asian (B01001D 001): the total number of asian residents in the selected geography's population.
- 8. hispanic (B01001I\_001): the total number of hispanic residents in the selected geography's population.
- 9. total (B01001\_001): the total population of the selected geography.
- 10. other (B01001F\_001): the total number of residents of "other" races in the selected geography.

We then created several columns of our own data:

- 1. percent male: the percentage of the population of the selected geography that is male.
- 2. percent female: the percentage of the population of the selected geography that is female.
- 3. percent white: the percentage of the population of the selected geography that is white.
- 4. percent\_black: the percentage of the population of the selected geography that is black.
- 5. percent\_hispanic: the percentage of the population of the selected geography that is hispanic.
- 6. percent\_asian: the percentage of the population of the selected geography that is asian.
- 7. percent\_other: the percentage of the population of the selected geography that is of "other" race.

#### Daily Covid counts

We collected 537 days' worth of Covid data—cases and deaths—from the New York Times. These data are at the county level. Some counties, such as those in Washington State where Covid was first recorded in the US, report cases as early as January; others only start reporting in March and beyond.

#### Variables:

- 1. date the date of the reporting
- 2. county the county name
- 3. state the state name
- 4. fips the FIPS code for that county
- 5. cases the number of cumulative cases in that county as of that date
- 6. deaths the number of cumulative deaths in that county as of that date

#### Presidential Election Results by County

This dataset gives the number of votes cast for the major Democratic and Republican candidates for 2016 in 2020, along with the total number of votes cast, for each U.S. county. It was taken from the MIT Election Data and Science Lab's County Presidential Election Returns 2000-2020 dataset by filtering the year.

#### Variables:

- 1. year the year of the election results, numerical
- 2. state the state in which the county is located, character
- 3. state\_po the two-letter abbreviation for the state in which the county is located, character
- 4. county\_name the name of the county, character
- 5. county\_fips the county FIPS code, numerical
- 6. office the office the election results are for President for all rows, character
- 7. candidate the presidential candidate that the candidate votes in the row refer to, character
- 8. party the political party of the candidate, character
- 9. candidatevotes the number of votes cast for the candidate, numerical
- 10. totalvotes the total number of votes cast in the county for the election, numerical

#### Main Vaccination Dataset

We collected county-level vaccination data from the CDC's COVID-19 Data Tracker. Texas does not report county-level vaccination information to the CDC COVID-19 Data Tracker, and California does not report information for counties with fewer than 20,000 residents. All other US counties are included in this dataset. The dataset also contains county Social Vulnerability Index (SVI) information. SVIs are produced by the

CDC and Agency for Toxic Substances and Disease Registry based on US Census data and take into account factors like poverty, lack of vehicle access, and crowded housing.

#### Variables:

- 1. Date the date for which the data in the row is reported in the format mm/dd/yyyy, character
- 2. FIPS the county FIPS code, character
- 3. MMWR week the Morbidity and Mortality Weekly Report week, numerical
- 4. Recip County the county name, character
- 5. Recip State the two-letter state abbreviation for the state in which the county is located
- 6. Series\_Complete\_Pop\_Percent the percentage of the population that is fully vaccinated, numerical
- 7. Series\_Compete\_Yes the total number of people in the county who are fully vaccinated, numerical
- 8. Series\_Complete\_12Plus the total number of people in the county over 12 years old who are fully vaccinated
- 9. Series\_Complete\_12Plus\_Pop\_Percent the percentage of the population over 12 years old that is fully vaccinated, numerical
- 10. Series\_Complete\_18Plus the total number of people in the county over 18 years old who are fully vaccinated, numerical
- 11. Series\_Complete\_18Plus\_Pop\_Percent the percentage of the population over 18 years old that is fully vaccinated, numerical
- 12. Series\_Complete\_65Plus the total number of people in the county over 18 years old who are fully vaccinated, numerical
- 13. Series\_Complete\_65Plus\_Pop\_Percent the percentage of the population over 18 years old that is fully vaccinated, numerical
- 14. Completeness\_pct an estimate of how complete the vaccination data is for the county, numerical
- 15. Administered Dose1 Recip number of first vaccine doses administered, numerical
- 16. Administered\_Dose1\_Pop\_Pct percentage of the population that has received dose 1, numerical
- 17. Administered\_Dose1\_Recip\_12Plus number of first vaccine doses administered to people over 12 years old, numerical
- 18. Administered\_Dose1\_12PlusPop\_Pct percentage of the population over 12 years old that has received dose 1, numerical
- 19. Administered\_Dose1\_Recip\_18Plus number of first vaccine doses administered to people over 18 years old, numerical
- 20. Administered\_Dose1\_18PlusPop\_Pct percentage of the population over 18 years old that has received dose 1, numerical
- 21. Administered\_Dose1\_Recip\_65Plus number of first vaccine doses administered to people over 65 years old, numerical
- 22. Administered\_Dose1\_65PlusPop\_Pct percentage of the population over 65 years old that has received dose 1, numerical
- 23. SVI\_CTGY Social Vulnerability Index category (Low, Low-Mod, Mod, Mod-High, High), character
- 24. Series\_Complete\_Pop\_Pct\_SVI categorizes vaccination levels and SVI (Low, Low-Mod, Mod, Mod-High, High), character
- 25. Series\_Complete\_Pop\_Pct12Plus\_SVI categorizes vaccination levels in the population over 12 years old and SVI (Low, Low-Mod, Mod, Mod-High, High), character
- 26. Series\_Complete\_Pop\_Pct18Plus\_SVI categorizes vaccination levels in the population over 18 years old and SVI (Low, Low-Mod, Mod, Mod-High, High), character
- 27. Series\_Complete\_Pop\_Pct65Plus\_SVI categorizes vaccination levels in the population over 18 years old and SVI (Low, Low-Mod, Mod, Mod-High, High), character

#### Texas Vaccination

To complete our county-level vaccination data, we collected vaccination data from the Texas Department of Health and Human services COVID-19 Vaccination in Texas Summary.

Variables:

- 1. County\_Name character
- 2. PHR\_critical\_Pop Public Health Region in which the county is located, character
- 3. Estimated\_Coverage\_(12+\_1Dose) estimated percentage of the population over 12 that has received one dose (taking into account probable completeness of reporting), numerical
- 4. Estimated\_Coverage\_(12+\_Fully) estimated percentage of the population over 12 that has been fully vaccinated (taking into accouCOVIDnt probable completeness of reporting), numerical
- 5. Estimated\_Coverage\_(65+\_1Dose) estimated percentage of the population over 65 that has received one dose (taking into account probable completeness of reporting), numerical
- 6. Estimated\_Coverage\_(65+\_Fully) estimated percentage of the population over 65 that has been fully vaccinated (taking into account probable completeness of reporting), numerical
- 7. People\_Fully\_Vaccinated actual reported number of people who have been fully vaccinated, numerical
- 8. People\_with\_at\_least\_One\_Dose actual reported number of people with at least one vaccine dose, numerical
- 9. Percentage\_Vaccinated percentage of total population that has been reported as fully vaccinated, numerical
- 10. Population\_16Up total population of people at least 16 years old, numerical

#### **Small California Counties Vaccination**

To complete our county-level vaccination data, we obtained the total number of people vaccinated and the total number of people who have received at least one dose from all counties in California with fewer than 20,000 residents from democratandchronicle.com.

#### Variables:

- 1. County the name of the county, character
- 2. FIPS county FIPS code, numerical
- 3. One\_Dose number of people who have received at least one vaccine dose, numerical
- 4. Two\_Doses number of people who are fully vaccinated, numerical

## Vaccine Hesitancy

We obtained county-level estimates of the percentage of the population who is hesitant to receive the COVID-19 vaccine from the CDC. The CDC calculates and compiles these estimates based on Household Pulse Survey data. The dataset also includes racial demographics and SVI (Social Vulnerability Index) information (see Main Vaccination Dataset for details on SVI), as well as geographical information because it was designed to produce a map.

# Variables:

- 1. FIPS Code county FIPS code, numerical
- 2. County Name character
- 3. State state in which the county is located, character
- 4. Estimated Hesitant estimated percentage of the population with any degree of vaccine hesitancy, numerical
- 5. Estimated Hesitant or unsure estimated percentage of the population that is either hesitant to receive the vaccine or unsure about how they feel about the vaccine, numerical
- 6. Estimated Strongly Hesitant estimated percentage of the population that is strongly hesitant to receive the vaccine, numerical
- 7. Social Vulnerability Index (SVI) CDC/ATSDR Social Vulnerability Index, numerical
- 8. SVI Category Social Vulnerability Index category (Low, Low-Mod, Mod, Mod-High, High), character
- 9. level of concern for vaccination rollout CDC's estimated level of concern for getting the population of the county vaccinated on a scale of 0 to 1, numerical

- 10. CVAC Level of Concern categorical variable describing CDC's level of concern (Very High Concern, High Concern, Moderate Concern, Low Concern, Very Low Concern), character
- 11. Percent adults fully vaccinated against COVID-19 (as of 6/10/21) character
- 12. Percent Hispanic numerical
- 13. Percent non-Hispanic American Indian/Alaskan numerical
- 14. Percent non-Hispanic Asian numerical
- 15. Percent non-Hispanic Black numerical
- 16. Percent non-Hispanic Native Hawaiian/Pacific Islander numerical
- 17. Percent non-Hispanic White numerical
- 18. Geographical Point ordered pair (longitude, latitude) marking the starting point for the county boundary character
- 19. County Boundary a series of ordered pairs (longitude, latitude) marking the boundary of the county character

### Methods

Note: This is an outline of our methods section.

- Linear regression to find predictors of vaccine hesitancy, specifically race and partisanship
- Linear regression to find predictors of vaccination rate, specifically race, partisanship, vaccine hesitancy, and other demographics (e.g. income)
  - Together, this can suggest whether vaccine hesitancy is a mediator, moderator, or neither
- Linear regression to examine relationship between vaccination rate and cumulative deaths, controlling for race and partisanship
  - This can shed light on whether the severity of the pandemic in a county influences residents' desire to get vaccinated

### Results

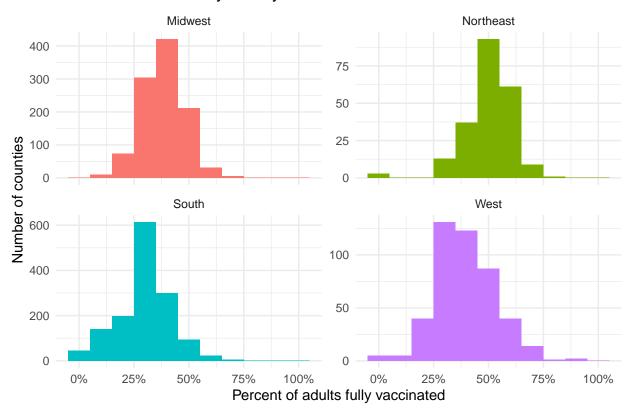
### County Vaccination Histogram by Region

The faceted histogram above demonstrates several important ideas. First, the Northeast region of the country displays a clear lead in the percentage of counties that have higher levels of full vaccination among residents. A large majority of counties in the Northeast are at least 65% fully vaccinated, and many are reaching the 70-75% full vaccination levels that can signal the beginnings of herd immunity. Furthermore, the data is more centralized in the Northeast compared to the other regions, suggesting that there is not a large level of hesitancy or belief in misinformation about the vaccine.

Meanwhile, the South is lagging behind the other three regions as a large portion of its counties are still in the 35-45% fully vaccinated range. Additionally, there is wider range in the data, and there are many counties that are still below a 25% full vaccination threshold. This could be a result of or correlated to several influencing factors which will be discussed in other results, including demographics, hesitancy, and politics.

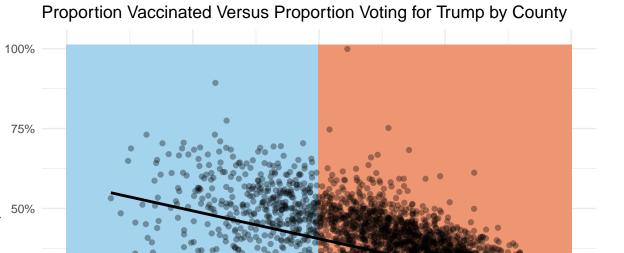
The Midwest and West are very similar in that they both have a peak around the 45-55% range. Both regions display a wide range of vaccination levels in both directions, suggesting that there is a wide array of willingness to get vaccinated within those regions.

# Percent vaccinated by county



# How are the prevailing political views of a county related to its rate of vaccination?

As anticipated, there appears to be a definite negative correlation between the percent of the county that voted for Donald Trump in 2020 and the percent of the population that is vaccinated. Donald Trump and other Republicans have consistently shown themselves to be less in favor of lockdown orders, mask mandates, and COVID-19 vaccinations than Democrats, so the counties in which a larger percentage of the votes were cast for Donald Trump in the 2020 presidential election are likely to be those for whom these anti-COVID precautions seemed reasonable. However, there are outlier counties in the lower left quadrant of the plot with far more or fewer people vaccinated than would have been predicted based solely on their political views. The next section explores race as a potential contributing factor predicting COVID-vaccination rates in these outlier counties.



Proportion of Population Vaccinated

25%

0%

0%

What are the racial demographics of counties that do not conform to the expected relationship between politics and vaccination?

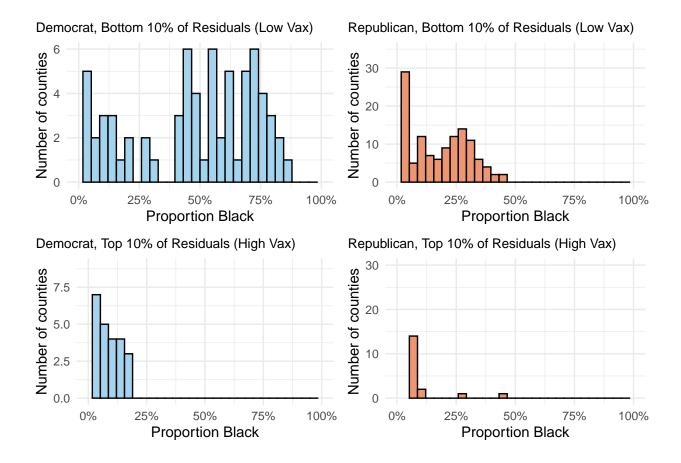
Proportion of Votes Cast for Trump

75%

100%

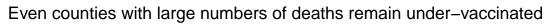
25%

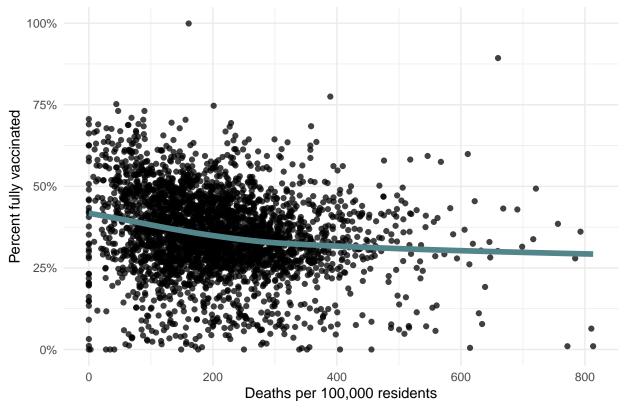
There appeared to be a distinct cluster of democratic counties with unexpectedly low vaccination rates in the lower left corner of the plot on the previous page. We hypothesized that these may be counties with high percentages of non-white residents. To investigate these and other outlier counties, we performed a linear regression on the relationship between proportion of votes cast for Trump and proportion of population vaccinated. We then examined both Democratic (<40% of all votes in the 2020 presidential race cast for Trump) and Republican (more than 60% of all votes cast for Trump) in the top 10% (unexpectedly high vaccination rates) and bottom 10% of residuals (unexpectedly low vaccination rates) from this linear regression. We then produced histograms of the proportion of Black residents in each of these four categories of counties. For the counties with unexpectedly high vaccination rates (both Democratic and Republican), the distribution of proportion Black residents peaked near 0. For the low-vaccination Republican counties, many counties also had a proportion of Black residents near 0, but there was another peak around 0.25. For the Democratic low-vaccination rate counties, the results were even more striking: a majority of counties had more than 35% Black residents. Clearly, counties with lower vaccination rates than predicted by their political leanings are much more likely to have more Black residents than counties with higher vaccination rates than predicted by their political leanings.



Did the severity of the pandemic within a county affect residents' desire to get vaccinated?

As far as we can tell from this chart, no. However, we haven't yet completed any regressions, so all we have is the plot. We want to regress controlling for demographics and vaccine hesitancy.





Discussion

Appendix