Paper

Ankit Patel, Hallie Trial, and Prayag Gordy

7/26/2021

Contents

Introduction

Background and Motivation

In the 18 months since COVID-19 first reached the shores of the United States, 35 million Americans have tested positive for the disease and over 600,000 Americans have died due to COVID-19. However, despite the pandemic's massive economic, social, and human capital toll, a large portion of the population has not taken advantage of the availability of free vaccinations, and large disparities in vaccination rates are prevalent.

This is especially alarming given that fully-vaccinated people accounted for less than 1% of all US COVID-19 deaths in May and close to the same percent of all hospitalizations, meaning that almost all current and future US COVID-19 deaths are entirely preventable. With new studies suggesting that the delta variant is as transmissible as chicken pox and deadlier than the original virus, speeding up the vaccination campaign is more crucial than ever.

Given the puzzling imbalance between the clear benefits of vaccination and the low rates of vaccination, we believe it to be in the interest of public health to examine and understand potential reasons and causes of said imbalance. Our investigation seeks to address this broad question by exploring the relationships between county-level vaccination rates and COVID-19 case and death rates in relation to variables including politics, race, and vaccine hesitancy.

Research Questions

In our exploration of the vaccination, cases, and demographic data, we sought to answer a number of questions that might elucidate possible relationships in the data. These questions include:

1. How do vaccination levels and hesitancy differ by geographic region in the US?
2. How is a county's Social Vulnerability Index (SVI) related to its rate of vaccination?
3. How are the prevailing political views of a county related to its rate of vaccination?
4. How does the percentage of Black residents in a county relate to its level of vaccination?
5. What are the racial demographics of counties that do not conform to the expected relationship between politics and vaccination?
6. What are the racial demographics of counties with high levels of vaccine hesitancy and low vaccination rates?
7. What are the racial demographics of counties with high SVI and low vaccination rates?
8. Did the severity of the pandemic within a county affect residents' desire to get vaccinated?

In this paper, we will present and discuss our findings and answers to these questions.

Data

See the appendix for details about each dataset's included variables.

`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 collect large amounts of data from every single county in the US, so the data are more recent than the 2010 census and are still very comprehensive (2020 census data are not publicly available yet). We created two data sets at different geographies: For one set, we got all the desired variables by county, while for the other set, we called the variables by Census region.

Daily Covid counts

We collected 556 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.

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.

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 SVI information (see appendix for details).

Texas Vaccination

To complete our county-level vaccination data, we collected vaccination data from the Texas Department of State Health Services.

Small California Counties Vaccination

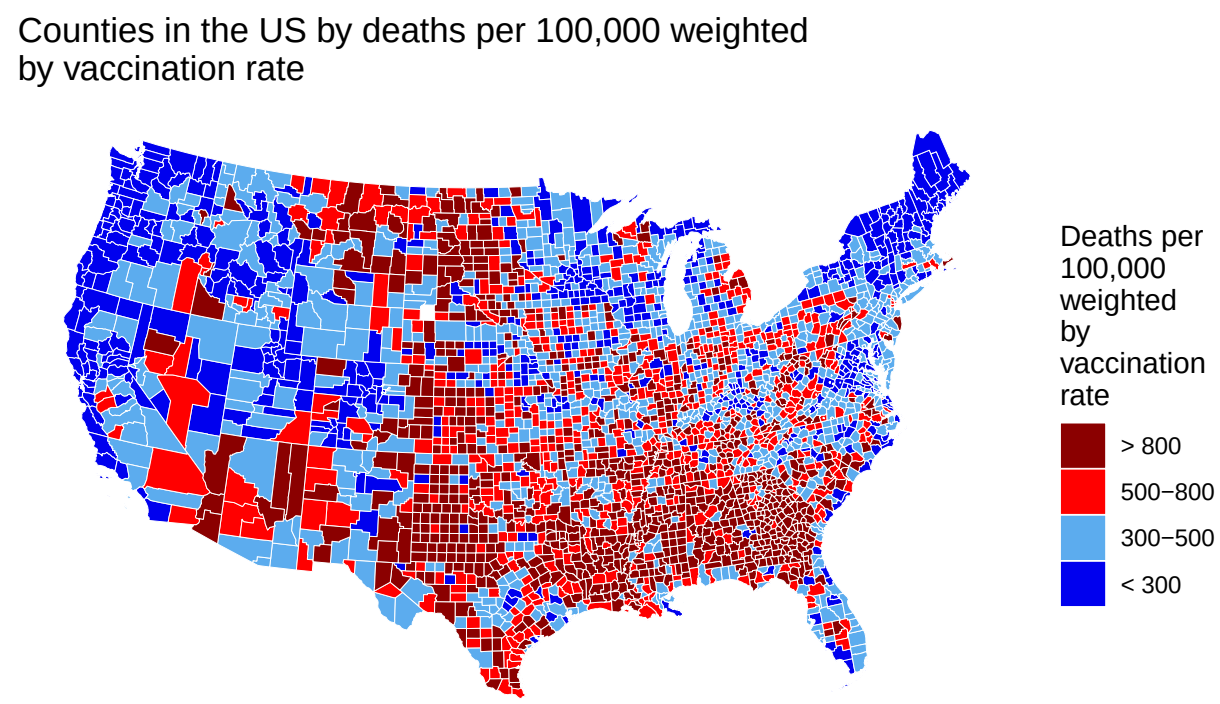
To complete our county-level vaccination data, we downloaded the total number of people vaccinated and the total number of people who have received at least one dose from all counties in California with fewer than 20,000 residents from democratandchronicle.com.

Vaccine Hesitancy

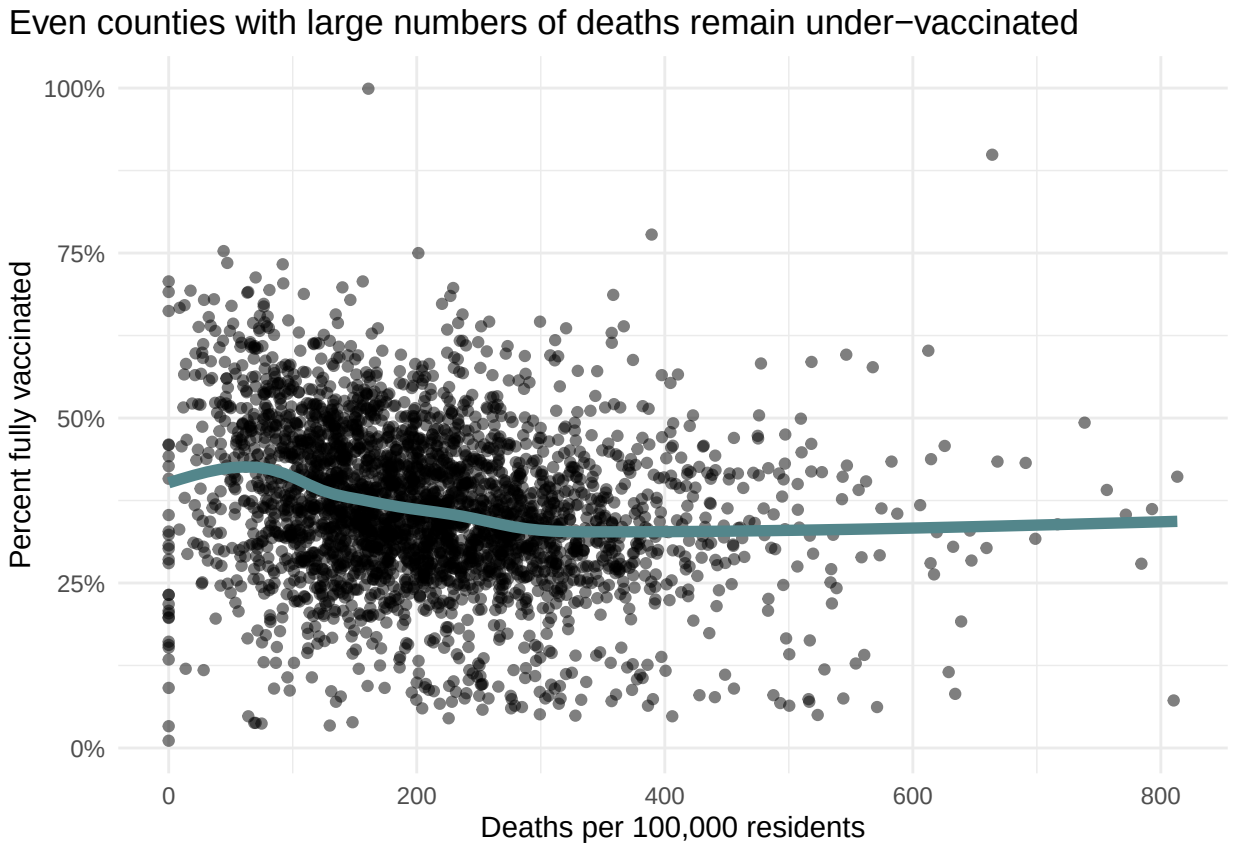
We obtained county-level estimates of the percentage of the population who is hesitant to receive the COVID-19 vaccine from the CDC (see appendix for details). The dataset also includes racial demographics and SVI information (see appendix for details), as well as geographical information because it was designed to produce a map.

Results

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

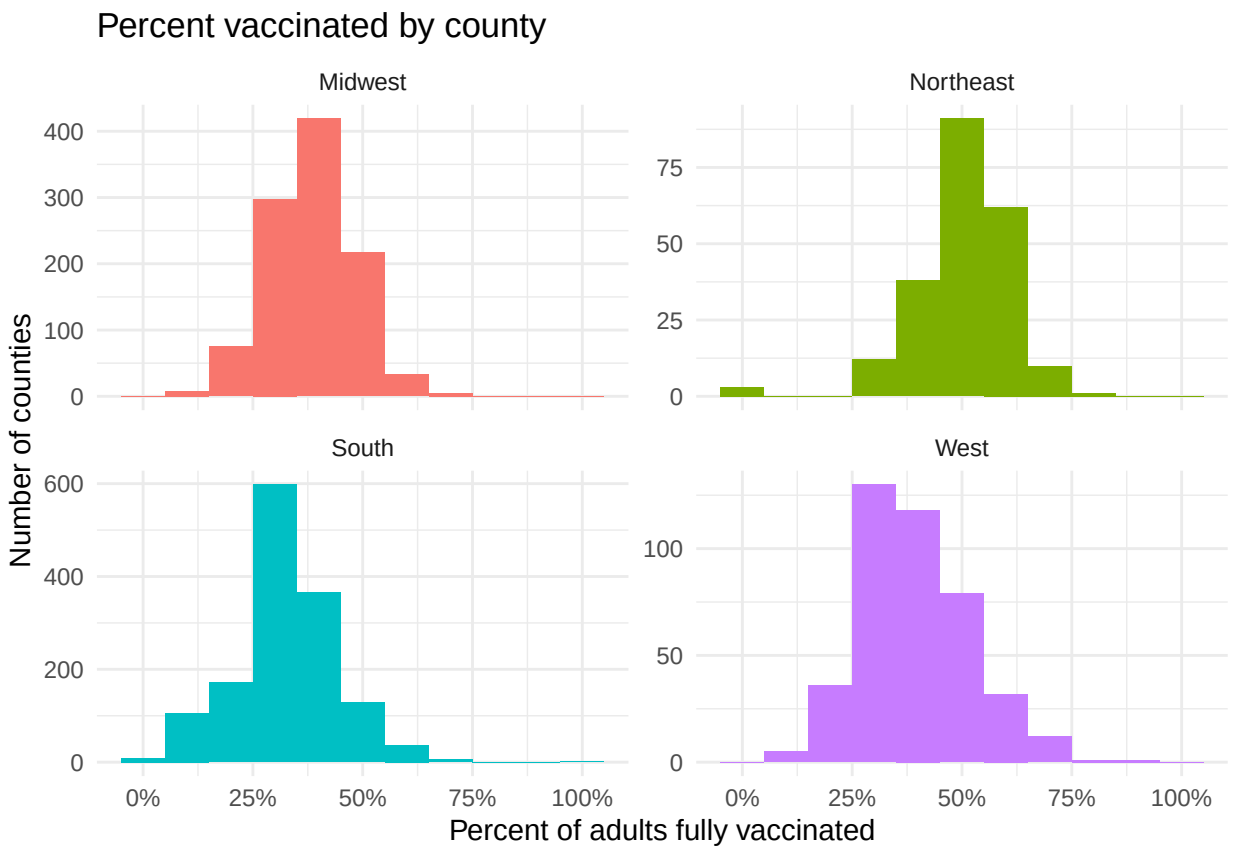


This map is colored based on the ratio of deaths per 100,000 to vaccination rate rather than the cases per 100,000 to vaccination rate. While this is a rudimentary form of analysis, there do seem to some differences in the two ratios, especially in the mountainous states of the West, such as Idaho, Wyoming, and Utah. Specifically, those states had a much lower ratio of deaths to vaccinations relative to their cases to vaccinations ratio, which might suggest that in those areas there might be some correlation, although there is nothing definitive. To further analyze this possibility, we plotted deaths per county versus vaccination rate:



This chart suggests that there is no correlation between the severity of the pandemic and residents' vaccination levels. Even some counties where close to 1-in-100 residents died from Covid are extremely under-vaccinated, while some counties that were not hard-hit are highly vaccinated.

How do vaccination levels and hesitancy differ by geographic region in the US?



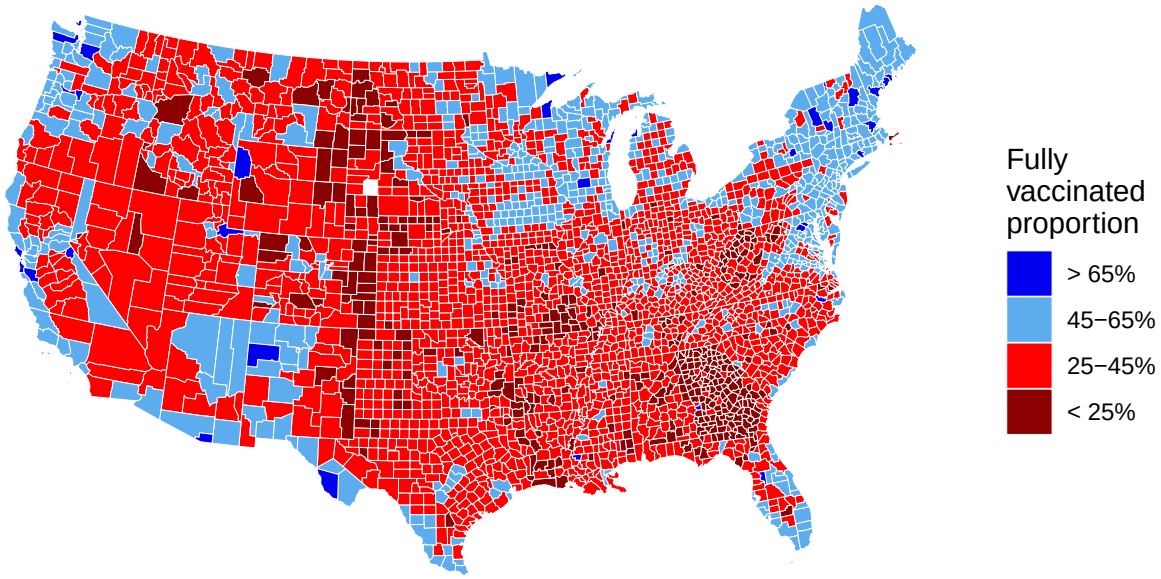
The faceted histogram above was obtained by combining census data at a county level with the county-level vaccination data from the CDC and state sources. We used the variable `fully_vax` as the data for the histogram and faceted by the census region (South, West, Midwest, Northeast) of the county.

The histogram 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.

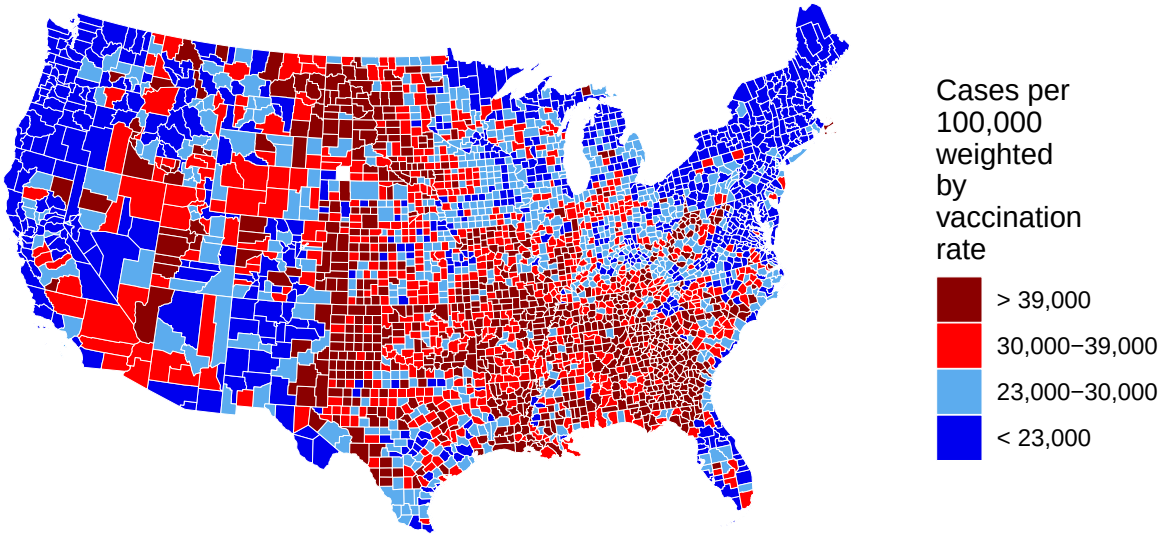
Counties in the US by fully vaccinated proportion



We created this map using `ggplot`'s `sf` geometry, the county-level total population census data, total case data, and total full vaccination data.

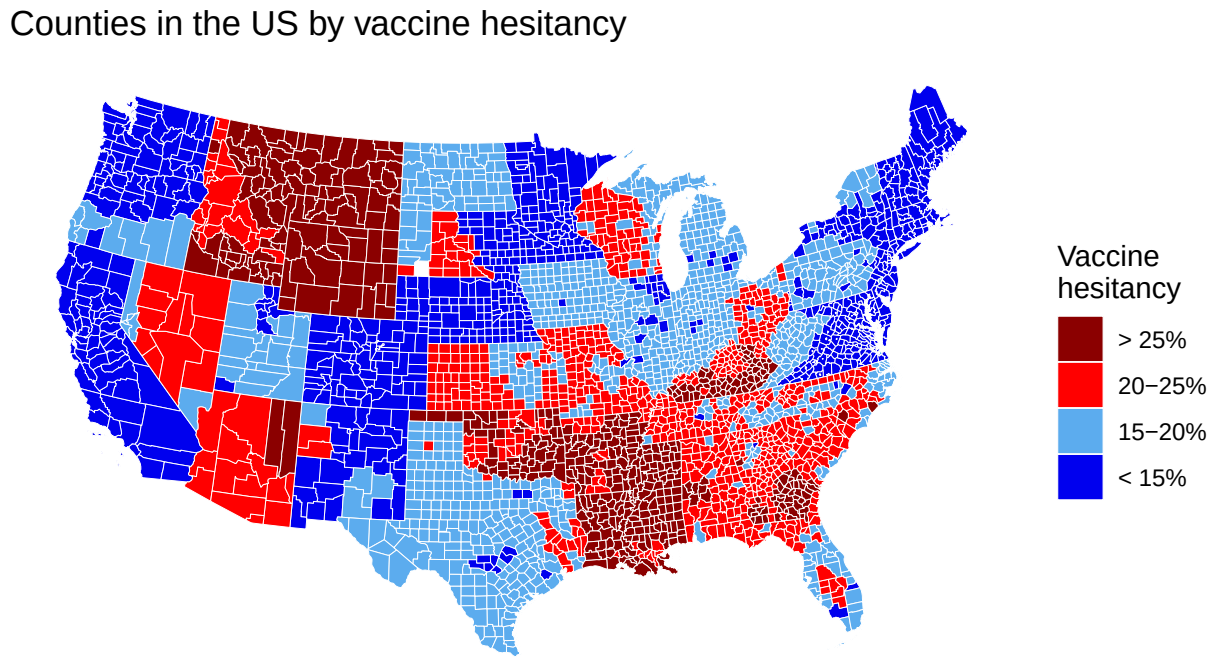
The map shows a county-level breakdown of the proportion of the population that is fully vaccinated. Once again, it is clear that there are regions of the country, especially the South (even more specifically, Georgia), that are severely under-vaccinated. In fact, only the Northeast region is reaching higher levels of vaccination, and while there are pockets of high vaccination in the West and Midwest, such as in the Seattle area, those regions do not have a uniformly high vaccination rate thus far. Perhaps most alarming is that despite many counties reaching a 45% to 65% full vaccination rate, only a few select counties have crossed the 65% full vaccination threshold thus far. This is important, as many scientists place the threshold for herd immunity in a community at around 65-70% full vaccination, suggesting that as a country, we are quite far from reaching a safe vaccination level.

Counties in the US by cases per 100,000 weighted by vaccination rate



This map is colored on a gradient based on the ratio of cases per 100,000 residents to the full vaccination percentage. In order to do this, we created a cases per 100,000 variable using a combination of case and census data and then a ratio variable using the cases per 100,000 and the fully vaccinated proportion.

We believed that this variable would be interesting in that it highlights the counties with either high case rates, low vaccination rate, or both. This map suggests that, once again, the South exhibits a much higher (worse) ratio of cases per 100,000 to vaccination proportion than the rest of the country, while the Northeast is the opposite, with a very low (better) proportion of cases to vaccinations. What is interesting in this map is that most of the West and Midwest also exhibits low ratios. In combination with the previous maps, this suggests that those areas are more average in both cases per 100,000 and vaccination rate, or have vaccination rates that reflect how bad the pandemic was in the county.



This map displays the county-level percentage of residents who are hesitant to receive the vaccine. There were certainly some surprises in this map. First, Arizona stands out as a state with very high levels of vaccine hesitancy despite being a particularly hard hit state throughout the pandemic and voting for Joe Biden in the 2020 presidential election. Another surprise came in the low hesitancy levels in Nebraska, Utah, North Dakota, and parts of South Dakota, despite those states being very Republican states who voted for Trump in 2020; it is worth noting that the Dakotas in particular had awful surges of Covid in the winter that perhaps motivated their residents to get vaccinated. This is also surprising since much of the South displays the opposite trend despite having similar politics. Thus, there might be other factors at play in these hesitancy levels, potentially race or vulnerability, which will be discussed in later section.

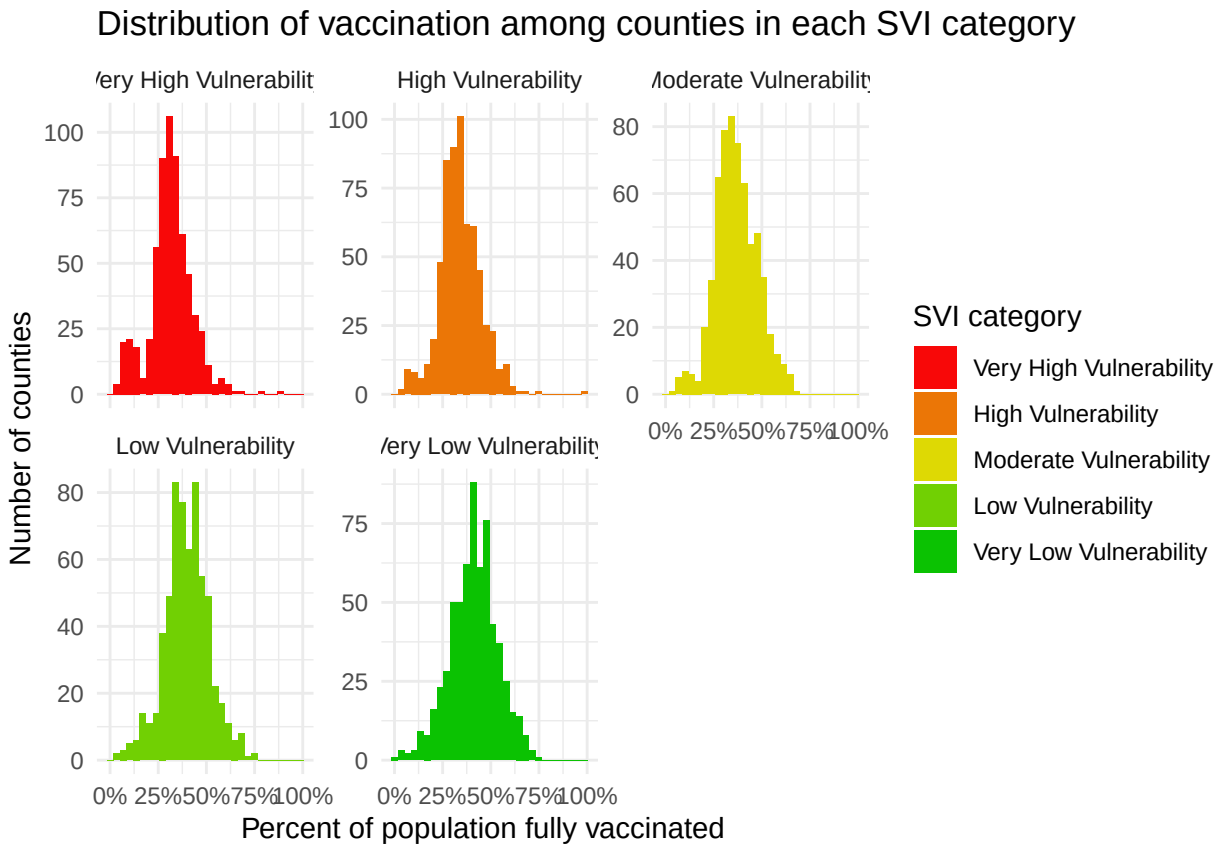
How is a county's SVI related to its rate of vaccination?

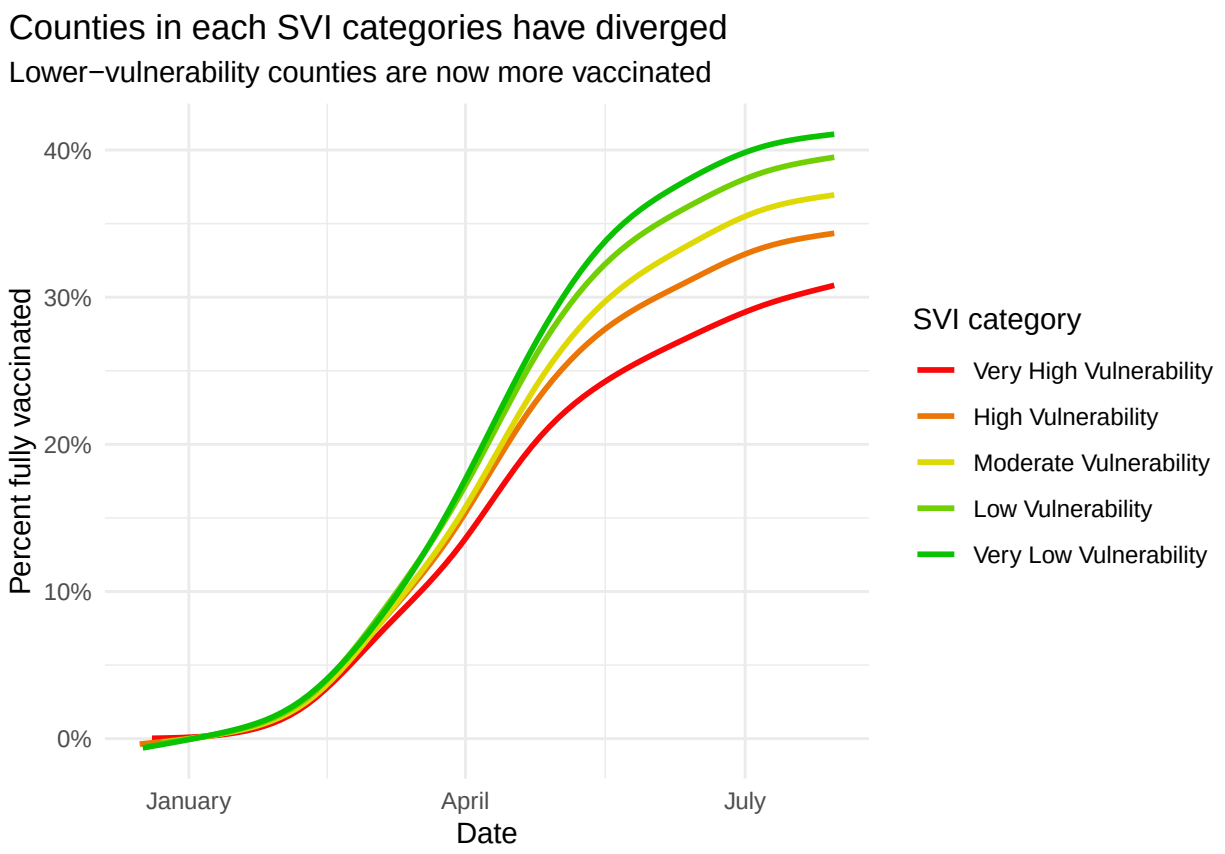
Here, we plotted county-level vaccination rate distributions by SVI category. The distributions appear approximately normal, with the exception of one intriguing feature: the very high, high, and moderate vulnerability counties all seem to have a bi-modal distribution, with a small peak of very low vaccination levels and a large peak of more moderate vaccination levels. This led us to consider whether there might be a subclass of vulnerable, very low vaccination rate counties in which vaccination rates are influenced by some other variable, a question we explore later.

The second plot, which shows average vaccination rate of counties over time grouped by SVI category, clearly demonstrates that average vaccination levels in higher vulnerability counties have tapered off at lower percentages than those of lower vulnerability counties, with divergence beginning around February 2020.

Present-day average vaccination levels for the different SVI categories smoothly increase with decreasing vulnerability.

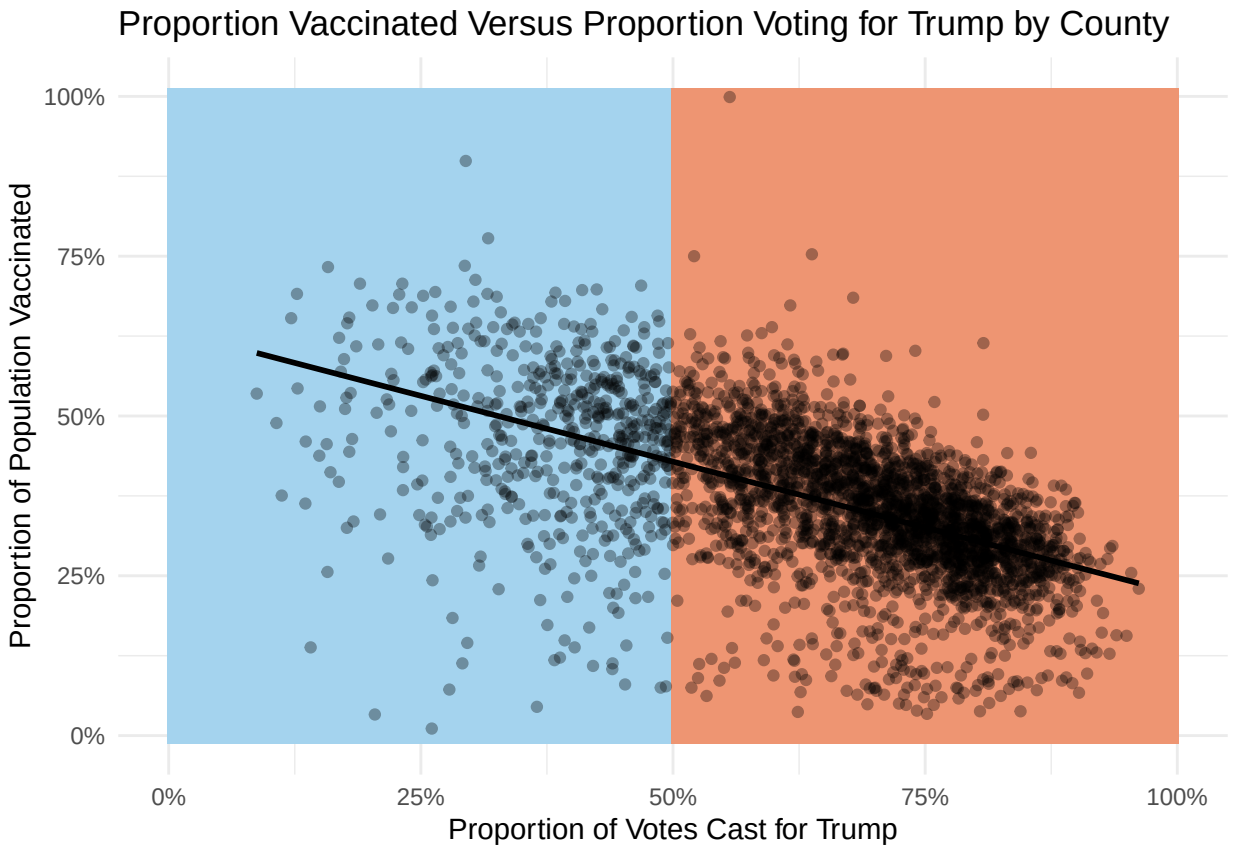
Given that the histograms for distribution of vaccination rates of broadly differing SVI categories overlap considerably, we wanted to determine whether the present difference in average vaccination rates between vulnerable (Very High, High, and Moderate Vulnerability SVI categories) and non-vulnerable (Low and Very Low Vulnerability) observed in the plot of vaccination rates over time was statistically significant, so we performed a one-sided t test with the alternative hypothesis that vulnerable counties have a lower average vaccination rate than non-vulnerable counties. The results were very significant, with a p-value near 0.





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

We plotted total percentage of county votes cast for Donald Trump in the 2020 presidential election versus total percentage of adult population vaccinated. As anticipated, there appears to be a definite negative correlation between the percent of the county that voted for Trump in 2020 and the percent of the population that is vaccinated. 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 Trump in the 2020 presidential election are likely to be those for whom these anti-COVID precaution sentiments seemed reasonable. However, there are outlier counties with far more or fewer people vaccinated than would have been predicted based solely on their political views. In particular, there was a cluster of Democratic counties with unexpectedly low vaccination rates. The next section explores race as a potential contributing factor predicting COVID-vaccination rates in outlier counties.

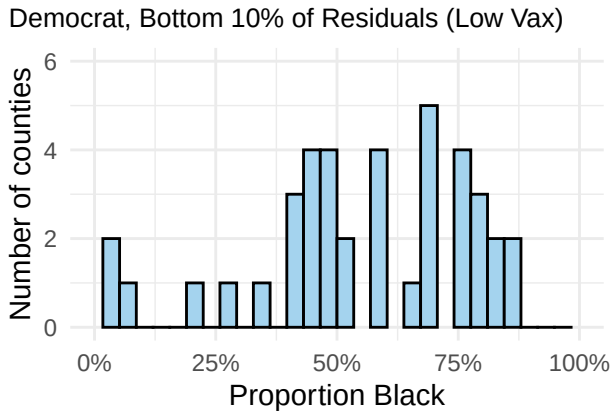


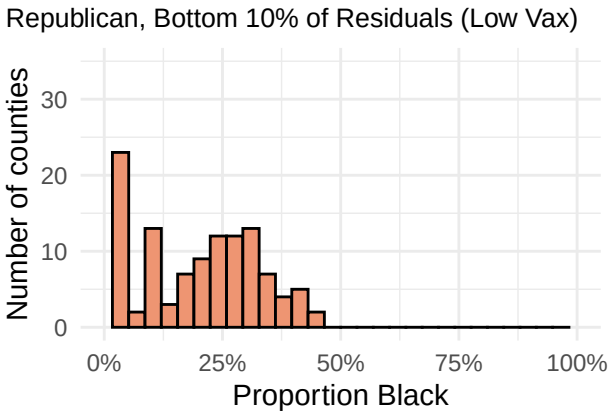
How does the percentage of Black residents in a county relate to its level of vaccination?

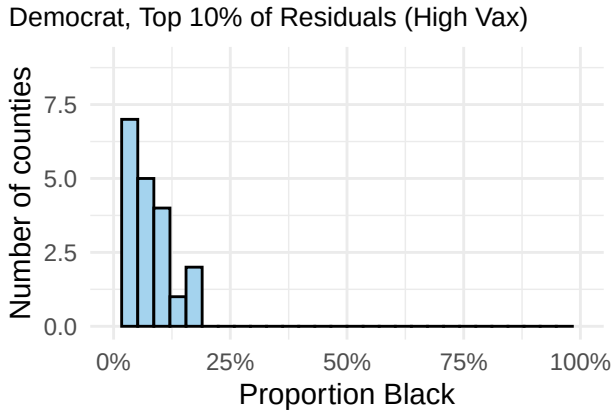
What are the racial demographics of counties that do not conform to the expected relationship between politics and vaccination?

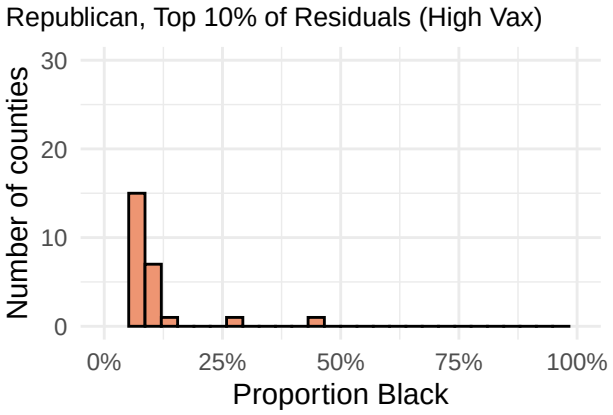
There appeared to be a distinct cluster of democratic counties with unexpectedly low vaccination rates in the lower left corner of the plot in the previous section. We hypothesized that these may be counties with high percentages of non-white residents. To investigate these and other outlier counties, we performed a robust linear regression (Huber 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 regression. We 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 25%. For the Democratic low-vaccination rate counties, the results were even more striking: a majority of counties had more than 35% Black residents, while Black people make up roughly 14% of the nationwide population. 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.

Next, we performed an ANOVA examining the mean percent Black for all four groups of outlier counties (low vaccination, Democrat; low vaccination, Republican; high vaccination, Democrat; high vaccination, Republican) followed by a Tukey Test pairwise comparison of means. All pairs showed significant difference in means at significance level 0.01 except high vaccination, Republican versus high vaccination, Democrat, demonstrating that the racial disparities visible in the histograms are significant.









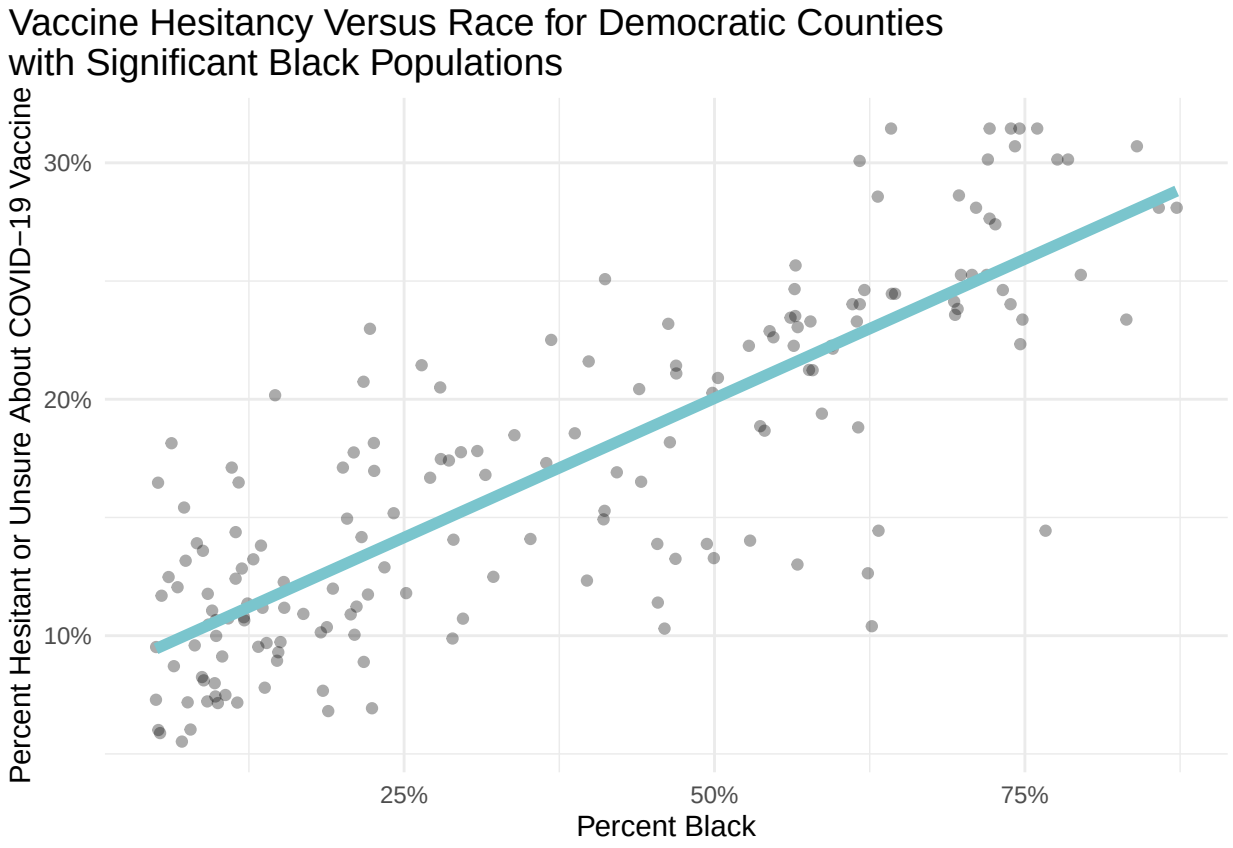
Comparison	Difference in Mean Proportion Black	95% Confidence Lower Bound	95% Confidence Upper Bound	p-value
Bottom 10% GOP vs bottom 10% Dem	-0.3634	-0.4144	-0.3124	0.0000
Top 10% Dem vs bottom 10% dem	-0.4734	-0.5444	-0.4023	0.0000
Top 10% GOP vs bottom 10% Dem	-0.4902	-0.5410	-0.4395	0.0000
Top 10% Dem vs bottom 10% GOP	-0.1100	-0.1743	-0.0457	0.0001
Top 10% GOP vs bottom 10% GOP	-0.1268	-0.1676	-0.0861	0.0000
Top 10% GOP vs top 10% Dem	-0.0168	-0.0809	0.0473	0.9057

How is race related to vaccine hesitancy?

Because political values have already been identified as a predictor of vaccination, we decided to filter by percent voting for Trump before examining race versus vaccine hesitancy. Additionally, we were particularly interested understanding the factors controlling vaccination in the previously identified cluster of low-vaccination Democratic counties, so we chose to examine only counties in which less than 40% of the votes in the 2020 presidential election were cast for Trump. Finally, because there are a large number of counties with close to no Black residents and these are not relevant to our examination of vaccine hesitancy in Black communities, we limited the counties to only those with a Black population of 5% or greater. Once we had selected these counties, we plotted the estimated percentage of the population that was hesitant or unsure about the Covid vaccine versus the percentage of the population consisting of Black residents and conducted a linear regression. Ultimately, we sought to investigate whether there was a correlation between the percent of the population

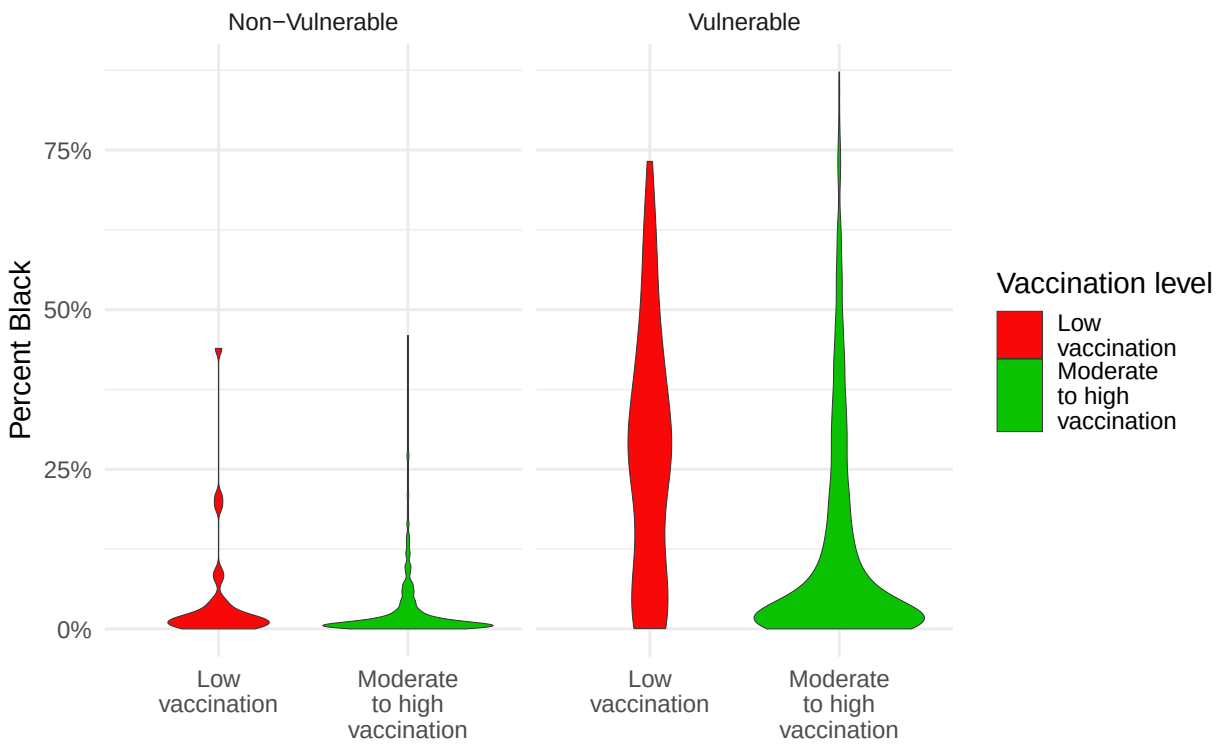
that was Black and levels of vaccine hesitancy in Democratic counties.

We found that among Democratic counties with more than 5% Black residents, vaccine hesitancy was indeed correlated with race. This suggests our previous observation that undervaccinated counties are more likely to have larger Black populations may result from greater reluctance of residents in these counties to accept the vaccine, rather than stemming from limited vaccine supply and distribution disparities, although the latter possibility cannot be ruled out.



What are the racial demographics of high SVI counties with low vaccination rates?

Because of our previous observations about the relationship between a county's proportion Black residents and its levels of vaccination and vaccine hesitancy, we hypothesized that the group of high SVI counties with especially low vaccination rates that we observed in the histograms of vaccination levels for counties in differing SVI categories might consist of counties with large Black populations. We assessed this theory by constructing two sets of violin plots showing distributions of percent Black residents based on SVI category and vaccination levels. The first set of violin plots shows counties in the "Very High" "High" or "Moderate" SVI categories and compares low vaccination counties (less than 15% vaccinated) versus all other counties, while the second set shows the same comparison but for "Low" or "Very Low" SVI categories. While the low vulnerability counties show similar distributions of percent Black residents for both the low and the high vaccination rate groups, for high vulnerability counties, the distribution is strikingly skewed towards more Black residents for the low vaccination rate counties in comparison with the high vaccination rate counties.



Conclusion

Summary of findings

1. Counties hit harder by the pandemic (counties with large numbers of cumulative deaths per 100,000 residents) do not seem to have significantly higher vaccination rates.
2. Among broad geographic regions of the nation, the Northeast has highest mean vaccination and tightest distribution, followed by the Midwest and West, while the South is lagging significantly behind the rest of the country in vaccination levels.
3. Counties in the South were more likely to have higher ratios of cases and deaths per 100,000 people to vaccination rates than were counties in the Northeast, Midwest, or West, suggesting that severity and vaccination levels might not be positively correlated or correlated at all.
4. Generally, much of the South exhibited high levels of vaccine hesitancy and much of the Northeast exhibited very low levels of vaccine hesitancy, inversely mirroring the trends in vaccination levels as expected. However, Texas had surprisingly low levels of vaccine hesitancy, as did Nebraska, Utah, and the Dakotas.
5. Counties in more vulnerable SVI categories had, on average, lower vaccination rates, and this difference was statistically significant.
6. The distribution of vaccination rates among counties in vulnerable SVI categories had a small peak at very low vaccination rates, indicating the possible existence of a subgroup of vulnerable counties in which vaccination rates are controlled by another factor.
7. The percentage of a county's votes in the 2020 presidential election cast for Trump was negatively correlated with the county's vaccination levels.
8. Among outlier counties in which vaccination levels were not well-predicted by political leanings, outliers with lower-than-predicted vaccination levels had higher average proportion Black populations than outliers with higher-than-predicted vaccination levels of the same political leaning and were more likely

to have larger Black populations.

9. For Democratic counties in which Black residents make up more than 5% of the population, vaccine hesitancy was positively correlated with race.
10. Very High, High, or Moderate Vulnerability counties (by SVI) with very low vaccination rates (<15%) were much more likely to have higher percentage Black populations than their higher vaccination counterparts with the same SVI values, while the difference between very low vaccination rate counties and their more vaccinated counterparts among Low and Very Low Vulnerability counties was far less pronounced.

Analysis and Recommendations

The concerning implication of our finding that vaccination levels are uncorrelated with cumulative county Covid death rates is that even in extreme cases, the severity of the pandemic did not influence residents' desire to receive the Covid-19 vaccine. This was a strong motivator to examine other factors that might be contributing to low vaccination levels and vaccine hesitancy.

Region

Ultimately, our analysis of ratios of case and death rates to vaccination levels, as well as of just vaccination levels and hesitancy, suggests that there is clearly some relationship between region and these data. Throughout the analysis, we see that the South has much worse ratio values than the rest of the country, especially as compared to the Northeast. Additionally, while the South and Northeast were at opposite ends of the spectrum, the Midwest and West exhibited far more middle-of-the-road ratio values and had a wide range of vaccination and ratio levels across counties.

While it may be tempting to then say geography is a strong determining factor in predicting these ratios and levels, it is in fact incorrect to do so. While there certainly seems to be a correlation, we know that geography is closely related to race and vulnerability levels, and especially to political views. As such, we cannot say just from looking at a regional analysis that geography by itself is an important factor rather than an indicator of the presence of another variable property that might in fact be the real reason or a more important reason for certain trends in the data.

Politics

There was a definitive negative correlation between the percentage of votes cast for Trump in the 2020 presidential election and county vaccination rate. This parallels a trend of conservative media coverage questioning the safety, efficacy, and necessity of Covid vaccination. As an extreme example, on July 27, conservative Fox show host Tucker Carlson compared the Covid-19 vaccine to "sterilization or frontal lobotomies" and pointed to rising cases in the US as evidence that "a rational person would assume there is something wrong with the vaccines," even though vaccines have shown to be effective at preventing infection and extremely effective at preventing serious illness and death.

There is hope that this trend could change, however. Recently, a growing number of conservative elected officials and media outlets have begun to speak out in favor of vaccination. Fox News has launched a public service announcement urging viewers to get vaccinated. On July 23, Republican governor of Alabama Kay Ivey urged her state to get vaccinated and called vaccines "the greatest weapon we have to fight Covid." In the same week, the Republican governor of Florida, Ron DeSantis, stated that "these vaccines are saving lives" and Senate minority leader Mitch McConnell said, "These shots need to get in everybody's arms as rapidly as possible, or we're going to be back in a situation in the fall that we don't yearn for." Our results suggest that vaccine outreach should be heavily targeted towards high-risk conservative communities, and Republican public officials could perhaps be the most effective mouthpiece; more Republican leaders, especially Trump, around the nation should follow the examples of Ivey, DeSantis, McConnell, and others in strongly endorsing the COVID-19 vaccine.

Race and Social Vulnerability

Counties with higher social vulnerability levels (Very High, High, or Moderate) had lower average vaccination levels than those with low social vulnerability levels (Low or Very Low), and the results were highly statistically significant. This may be attributable in part to larger Black populations in socially vulnerable counties, especially those with lower vaccination rates.

Throughout our report, the proportion of a county's population that identified as Black repeatedly emerged as a factor related to COVID-19 vaccination levels. It is important to disambiguate the influence of supply and demand on county vaccination levels to determine what strategies health officials can use to promote COVID-19 vaccination in lagging communities. We tentatively hypothesize that the relationship between race and vaccination results from higher levels of vaccine hesitancy among Black Americans because among predominately Democratic counties, we observed a correlation between percent Black residents and vaccine hesitancy.

Unfortunately, these results are unsurprising and echo other reports. In a survey by the CDC in December 2020, for instance, 46% of Black adults in the US stated that they probably would not choose to receive the COVID-19 vaccine, compared with 30% of white adults. Many have argued that vaccine hesitancy manifests a deeper distrust of modern medicine within Black American communities motivated by the historically exploitative relationship between medical science and Black patients. Notorious examples include the Tuskegee Syphilis Study, in which investigators withheld treatment from Black study subjects for decades; the harvesting of Henrietta Lacks' cancer cells to create a cell line for medical research without her (or or family's) consent or knowledge; and forced sterilization of Black women mandated by the North Carolina Eugenics Board.

Stronger outreach to vulnerable and historically mistreated communities is a key tool in vaccinating the American public.

Appendix

Variables in data

`census_county` and `census_region`

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.

Daily Covid counts

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

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

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

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

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

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

How does CDC estimate vaccine hesitancy?

The CDC calculates and compiles these estimates based on the U.S. Census Bureau's Household Pulse Survey data. Note that they provide the following note about the most updated version of the vaccine hesitancy data, which was used in our report: "Due to the change in the survey instrument regarding intention to vaccinate, our estimates for "hesitant or unsure" or "hesitant" are not directly comparable with prior Household Pulse Survey data and should not be used to examine trends in hesitancy prior to April 14, 2021." Estimates are based on the answers to the following question: "Once a vaccine to prevent COVID-19 is available to you, would you... get a vaccine?", the answer to which is chosen from the following opens: 1) "definitely get a vaccine"; 2) "probably get a vaccine"; 3) "unsure"; 4) "probably not get a vaccine"; 5) "definitely not get a vaccine." Respondents were classified as "strongly hesitant" if they indicated that they would "definitely not" receive a vaccine. The category "hesitant" includes responses "probably not" or "definitely not". The category "hesitant or unsure" is the broadest and includes "probably not", "unsure", or "definitely not".

How are Social Vulnerability Indices calculated?

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, education levels, ethnicity makeup, family characteristics, lack of vehicle access, and crowded housing. SVIs are assigned a value between 0 and 1 according to this information, where 0 is least vulnerable and 1 is most vulnerable.

Counties are classified as having "Very Low" social vulnerability if their SVI is from 0-0.19, "Low" if it is between 0.20 and 0.39, "Moderate" if it is between 0.4 and 0.59, "High" if it is between 0.60 and 0.79, and "Very High" if it is 0.8 or above.

References

Butanis, Benjamin. n.d. "The Legacy of Henrietta Lacks." Accessed July 30, 2021. https://www.hopkinsmedicine.org/henriettalacks/.

"California COVID-19 Vaccine Tracker." n.d. *Democrat and Chronicle*. Accessed July 30, 2021. https://data.democratandchronicle.com/covid-19-vaccine-tracker/california/06/.

CDC. 2020. "Covid Data Tracker." *Centers for Disease Control and Prevention*. https://covid.cdc.gov/covid-data-tracker.

"Cho-Alb_vaccinations Virginia Open Data Portal." n.d. Accessed July 30, 2021. https://data.virginia.gov/dataset/CHO-ALB_Vaccinations/qv6b-8qf3.

CNN, Paul LeBlanc, Maggie Fox and Elizabeth Cohen. 2021. "CDC Document Warns Delta Variant Appears to Spread as Easily as Chickenpox and Cause More Severe Infection." *CNN*. https://www.cnn.com/2021/07/29/politics/cdc-masks-covid-19-infections/index.html.

"Covid-19 County Hesitancy Data Centers for Disease Control and Prevention." n.d. Accessed July 30, 2021. https://data.cdc.gov/Vaccinations/COVID-19-County-Hesitancy/c4bi-8ytd.

Ellis, Ralph. 2021. "McConnell, Other Conservatives Now Backing Vaccinations." *WebMD*. https://www.webmd.com/vaccines/covid-19-vaccine/news/20210723/mcconnell-other-conservatives-now-backing-vaccinations.

"For Eugenic Sterilization Victims, Belated Justice." 2014. *MSNBC.com*. https://www.msnbc.com/all/eugenic-sterilization-victims-belated-justice-msna358381.

"Nearly All COVID Deaths in US Are Now Among Unvaccinated." 2021. *AP NEWS*. https://apnews.com/article/coronavirus-pandemic-health-941fcf43d9731c76c16e7354f5d5e187.

"Republican Governor Says 'Time to Start Blaming Unvaccinated' for Rise in Cases." 2021. *The Guardian*. http://www.theguardian.com/us-news/2021/jul/23/kay-ivey-republican-blame-unvaccinated-covid-cases.

Subbaraman, Nidhi. 2021. "This COVID-Vaccine Designer Is Tackling Vaccine Hesitancy — in Churches and on Twitter." *Nature* 590 (7846): 377–77. https://doi.org/10.1038/d41586-021-00338-y.

"Texas Covid-19 Data." n.d. Accessed July 30, 2021. https://dshs.texas.gov/coronavirus/AdditionalData.aspx.