

COVID responses and effects

A study of socioeconomic influencing factors in the United States

Authors: Simi Talkar, Ali Tobah, Matthieu Lienart

Motivations

The Nigerian Igbo word “igwebuike”, which means “there is strength in community.” paraphrases the ideal response to the COVID pandemic. To build herd immunity and stop its spread, within and outside borders, a cohesive and co-operative front is called for. However, amidst this pandemic, there is a suggestion that the U.S. has never been as divided as in recent years.

In a year of presidential elections, the leadership rhetoric and messaging regarding COVID response and behavior, as projected by the media, appears to starkly differ along political ideological lines, specifically, Democrat versus Republican. Usage of masks, openness to vaccination and a general appreciation of the severity of the pandemic, as portrayed to the public, arouses curiosity as to whether the response is truly divided along party lines.

We will also explore if other factors, commonly associated with political affiliation, differentiate the response by communities towards US Centers for Disease Control guidelines. Amongst those, employment rates and the urban/rural demographic of the voting population, will be considered.

Question to answer: Is there a significant correlation between political affiliation and population response to the COVID pandemic, and the confirmed case and death rates?

Additional analyses:

- Did this trend continue during the more recent rise of the Delta variant?
- Are there other factors that could have affected this correlation?
- Two prominent ones often mentioned with the political divide are unemployment and the urban/rural demographic.

Notes:

- In all visualizations, county and state political affiliation outcome in a presidential election is represented by these colors:
 Democrat Republican
- This report is a static version and an analysis summary. The analysis was conducted with interactive charts and the entire analysis can be found [here](#).

Is There an Observable Correlation Between Political Affiliation and Population Response to the COVID Pandemic?

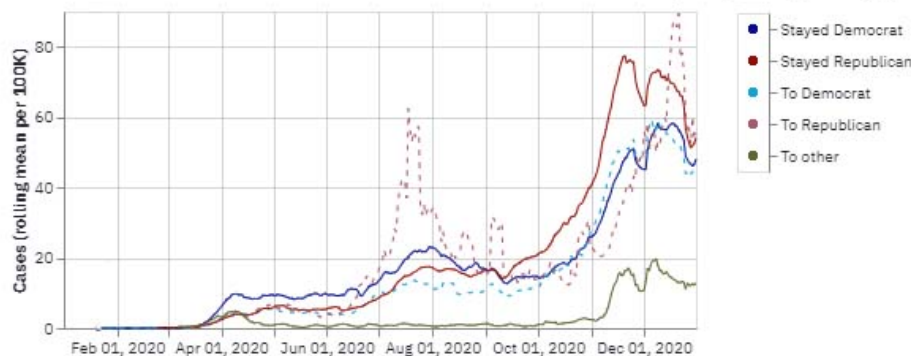
Party affiliation and COVID case trend in Presidential election year of 2020

The basis of this project began with the observations derived from the visualization below. The task is to determine if there is a significant difference in the rate of rise of COVID cases between normalized (per 100K) populations that profess affiliation to one party or another (Republican, Democrat and Other).

We differentiate between populations that voted Republican, Democrat or Other. The affiliation was further sectioned into those that stayed loyal to a party from 2016 to 2020, in contrast to those that were perhaps more influenced by the political messaging and switched loyalties. The number of cases plotted along the y-axis is the seven-day rolling average rather than raw numbers which can reflect data collection and delay in recording errors.

From the visualization we observed that the rate of rise of the infection was markedly higher, towards the end of the first year of the pandemic, in the populations that voted Republican.

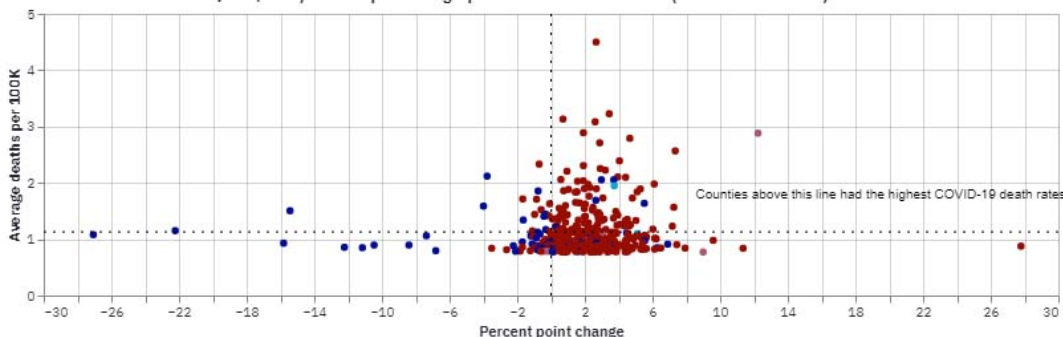
Year 2020 Timeseries of Confirmed Covid Cases Per 100K Residents (7 day rolling average)



Strength of affiliation and COVID deaths at County level

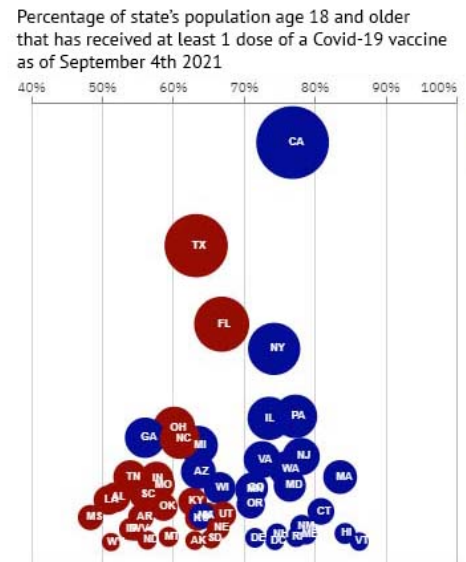
Does strength of affiliation, as determined by the percentile point change in votes received by a party in 2020 over 2016, show any correlation to the number of COVID related deaths in that county? We compared the top 400 counties voting for each party that suffered the most deaths per 100K population. It can be observed that the ratio of Republican counties to Democratic counties that switched with greater percentile points had a slightly higher average death rate but very marginally so.

Covid deaths in election year (2020) versus percentage point difference in votes (from 2016 to 2020)



Affiliation and Vaccine Adoption Rates By State

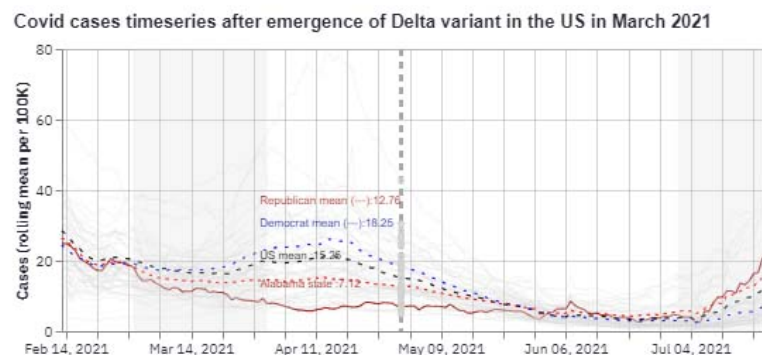
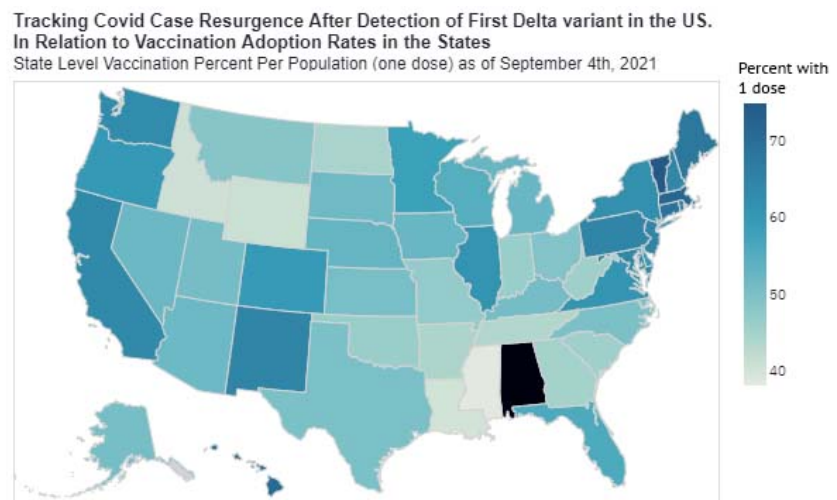
In this visualization, we notice clearly that "blue" states surged forward in vaccinations earlier than the red states after about three months after its release. The size of the bubbles in the chart below is scaled by population of each state. The X-axis positions each state by the percent of its population with at least one shot (since some vaccines require only one shot).



Vaccinations and the Delta Variant Case Resurgence

The chart below allows for selection of a state in the map to learn about case trends for the period after the first Delta variant was detected in the US. The US average and per political affiliation averages are also plotted for baseline comparison.

Efficacy of the vaccine over time, for this pandemic, can only be measured as time progresses, since we do not have a precedent. Mutation of the virus is inevitable and immunity response is constantly being monitored. This visualization presents the resurgence in COVID cases that are observed to be rising especially in states with low adoption. We compare an individual state's trend with mean trends pertaining to the US, combined mean of states that voted Republican and states that voted Democrat. It was noted that the resurgent rising trends were notably seen in states such as Louisiana, Missouri, and Florida where the numbers rose well above all the means. These are "red" states. Vermont has a high vaccination adoption and a



notably lower trend. Most of the states that voted Democrat also have trends that initially spiked and later settled closer to their mean.

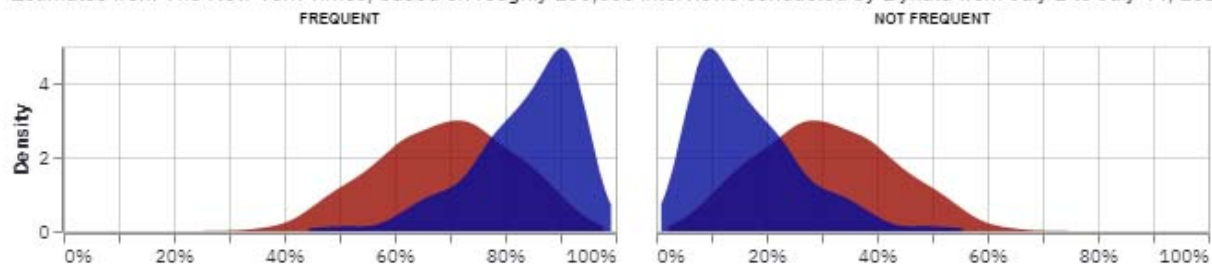
Frequent and Infrequent Mask Usage by Political affiliation

Mask usage data was collected in a survey by New York Times (through a professional survey firm). The data was gathered from 250,000 people surveyed in a two weeks period in July 2020 (please see details in [Appendix in the Streamlit app](#)). The five choices offered: Never, Rarely, Sometimes, Frequently, Always. The estimations of all five for every county, provided as a float adds up to 1. This was binned into Not Frequent (Never, Rarely, Sometimes) and Frequent (Frequently, Always) usage which for every county now adds up to 1. Binning is a common technique used to convert a quantitative data into a categorical value and this technique was once again applied to bin the estimates into ranges of Low, Medium and High chance of frequent and infrequent usage of masks.

The density plot of mask usage shows the distribution of infrequent and frequent mask usage among counties (that the surveyed participant resides in) that voted Democrat or Republican. We see that relatively, there is a higher probability of mask usage among the counties voting for Democrats over the Republicans.

Mask Usage Survey Response Distribution by Political Affiliation

Estimates from The New York Times, based on roughly 250,000 interviews conducted by Dynata from July 2 to July 14, 2020.



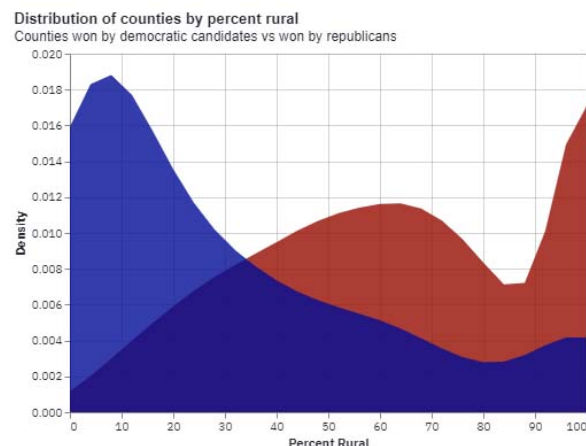
Does the Urban/Rural Demographic Influence the COVID Response?

Among the socioeconomic factors frequently and strongly associated with political affiliation is the urban/rural demographic of the voting population. Could this factor influence the pattern we have seen above, of the COVID-19 effects being split along party lines?

The Census Bureau defines urban areas by population size, specifically 2,500 or greater. Also from the Census Bureau, we obtained a list of US counties with a 'percent rural' designation, meaning the percentage of county population living in rural areas.

Political affiliation

Merging on FIPS codes with the presidential election county results, we see there seems to be a clear divide in political affiliation between rural



areas and urban centers. Counties less than about 32% rural were more likely to vote Democrat, while those above were more likely to vote Republican.

Since we want to see the effect of political affiliation on the COVID response, the next question is: How strongly democrat or republican were those counties? For that, we used the ratio of winning party votes to total votes, which we call the "vote fraction". Plotting against the Percent Rural designation of each county ([see the Streamlit app](#)), we find no correlation (-0.07) for counties won by the Democrat candidate, and a weak positive correlation (0.25) for counties won by the Republican candidate.

This is a positive initial result, since it implies that the urban/rural nature of a county has little to no effect on the strength of the affiliation.

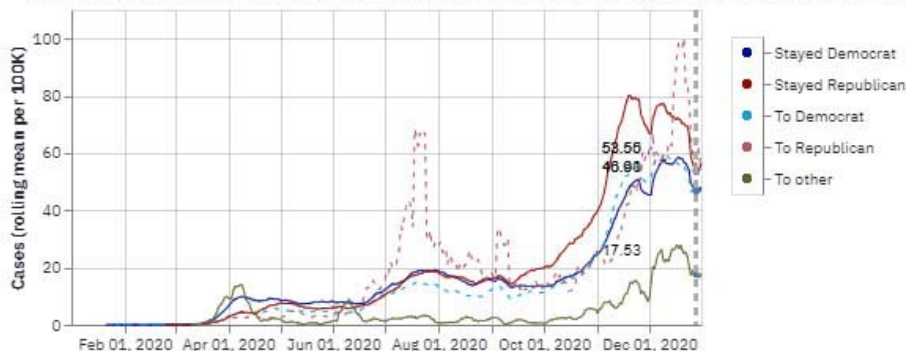
Finally, we come to the main question, which is whether the urban/rural nature has an influence on the effects of - and responses to - the COVID pandemic, which were investigated above.

COVID effects

The Census Bureau classifies counties with 50% or more of their population living in rural areas as 'mostly rural', while the remainder are classified as 'mostly urban'. We split the counties by that designation, and for each, we repeat the analyses performed above.

Urban Counties

Year 2020 Timeseries of Confirmed Covid Cases Per 100K Residents (7 day rolling average)



Rural Counties

Year 2020 Timeseries of Confirmed Covid Cases Per 100K Residents (7 day rolling average)



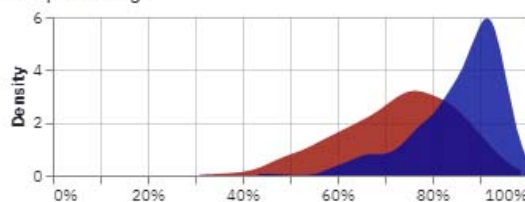
For COVID rolling case average, we find that the same trends hold as those found earlier for all counties: Initially, the case numbers are higher for counties that voted Democrat, but in September 2020, the rise in case numbers for counties that voted Republican is steeper.

As for COVID deaths, we also find the same trends hold for both urban and rural counties as those found earlier for the complete set of counties: The number of deaths per 100K population is higher for more counties that are strongly republican.

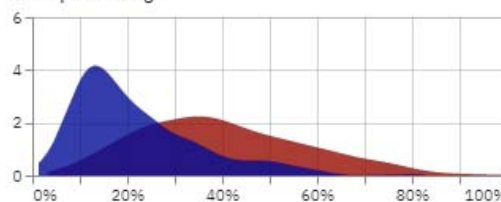
COVID response

For influence of the urban/rural nature of the counties on the COVID response, we look at frequency of mask usage, as defined earlier.

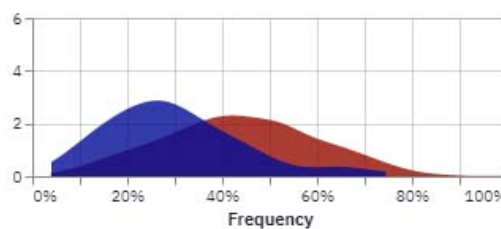
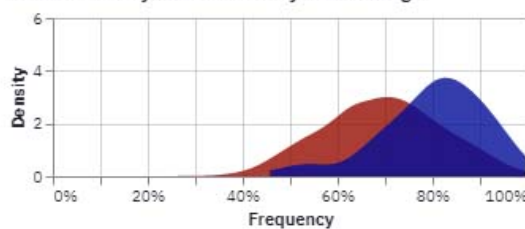
URBAN county distribution by mask usage
Frequent usage



Infrequent usage



RURAL county distribution by mask usage
Frequent usage



We find the same trends, unchanged, as in the analysis of total counties: Counties won by the Democratic candidate, compared to those won by the Republican candidate, were more likely to use masks frequently, and less likely to use them infrequently. This trend held for both counties classified as urban and those classified as rural.

We can conclude that the urban or rural nature of a county does not have an influence either

- on effects of the COVID pandemic as represented by the rolling case average and the death rate
- or on response to the COVID pandemic as represented by the frequency of mask usage.

Does Unemployment Influence the COVID Response?

COVID effects

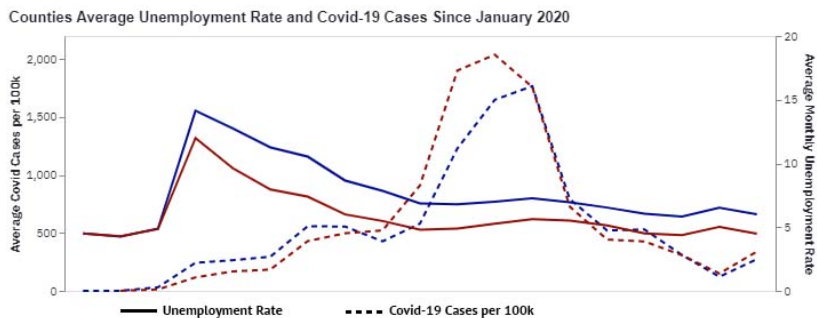
Stark unemployment increase was a major side effect of the COVID pandemic. In December 2019 the mean of the unemployment rate was 3.79% with an interquartile range of 2.7% to 4.4% (1.7 points difference). By April 2020, unemployment rate had jumped to a mean of 12.38% with an interquartile range of 8.7% to 15.5% (6.8 points difference) showing higher disparities between counties.

Could the unemployment rate have a bigger impact than political affiliation on the response to the COVID, by pushing more people to wear masks or get vaccinated?

COVID response

The Republican and Democrat counties' monthly average unemployment rate and COVID cases since the beginning of the pandemic clearly show different trends. This suggests that there is no correlation between unemployment rate and COVID cases.

What we see however is that the average unemployment rate is constantly higher in Democrat counties, while at the peak of the pandemic the average COVID cases number was higher in Republican counties.



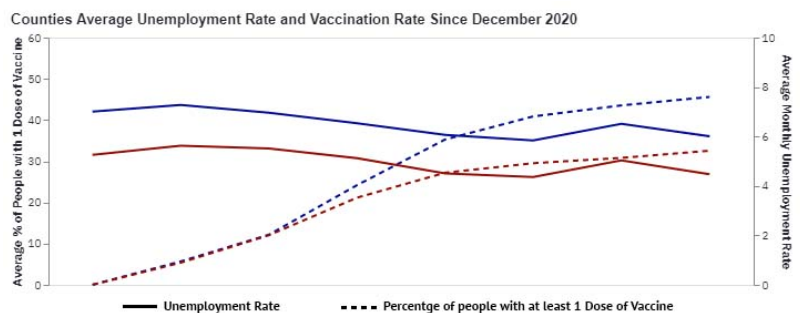
If there is no correlation between unemployment rate and COVID cases, could there still be one with the COVID response like mask usage and vaccination?

Looking at July 2020 data, a slightly positive correlation (0.348) between unemployment rate and frequent mask-wearing habits doesn't really help to validate or reject the hypothesis of a correlation. From the plot (see the [Streamlit app](#)), it seems however, that political affiliation is a stronger differentiator in following CDC mask-wearing guidelines than unemployment rate.

When we look at vaccination, the above pattern seems even more clear. The correlation between the percentage of the population with at least 1 dose of vaccination and the unemployment rate is even less clear. However, as for mask-wearing behaviors, we also see that

- the counties with high vaccination and high unemployment rates are mainly Democrat
- the counties with low vaccination and low unemployment rates are mainly Republican

As for mask-wearing, the political affiliation seems to be a stronger differentiator in following CDC vaccination guidelines than unemployment rate.



As we wanted to verify if unemployment rate could have a stronger impact on the COVID response than political affiliation, we see that

- there seem to be no correlation between COVID case and unemployment rate, and between unemployment rate and COVID response behaviors like wearing a mask or getting vaccinated
- On average, Democrat counties have higher unemployment rate, less COVID cases but better follow COVID response CDC guidelines while Republican counties show opposite trends and behaviors

Conclusion

Preliminary exploratory data analysis clearly indicated a distinct difference in rates of infection along party lines at the county levels. Towards the beginning of July of 2021, we show evidence of cases rising steeply in states with low vaccination rates. These states that predominantly voted Republican (such as Louisiana and Mississippi), are clustered in the Southern parts of the United States and can visually be spotted as states with low vaccination rates. The short-term New York Times mask usage survey data also showed marginally higher adherence rates among counties that voted Democrat. We also noted that socio-economic factors such as unemployment and urban or rural demographics only served to further highlight the split in response, and ultimately the effects of COVID, along political lines.

Statement of Work

Simi Talkar:	Initiated project with data collection, manipulation, analysis, and visualizations on political affiliation and COVID response. Enabled deployment to Streamlit
Ali Tobah:	data collection, manipulation, and visualizations for the second section on the response to COVID in urban/rural counties. Proofing and re-writing of all documents.
Matthieu Lienart:	data collection, manipulation, and visualizations for the third section on unemployment. Setup and optimization of Streamlit. Report formatting.
All members:	writing content for the project proposal, the final report and the Streamlit app.

Noteworthy Visualization Functions

Why Interactivity? The reason is best expressed by Cassie Kozyrkov, here: "As an analyst, I'm not here to funnel you towards my opinion. I'm here to help you form your own."

Interactivity we offer:

1. Ability to pause in time by setting duration from date of origin of an event. (VizVaccine.py/[createStateVaccinationChart](#))
2. Select a category such as state or segment of political affiliation to focus on. (VizCOVID/[createPercentPointChangeAvgDeathsChart](#))
3. Find an insight and drill into the selection through cross-chart interactivity. (VizVaccine.py [createCombinedVaccinationAndDeltaVariantTrend](#))
4. Exploration of a category in spatial map. (VizMask.py/[createFreqCountyMaskUsageWithRanges](#))
5. Offering hover and rule tooltips in all charts. (VizUnemployment.py/[createUnemploymentCorrelationLineCombinedChart](#))

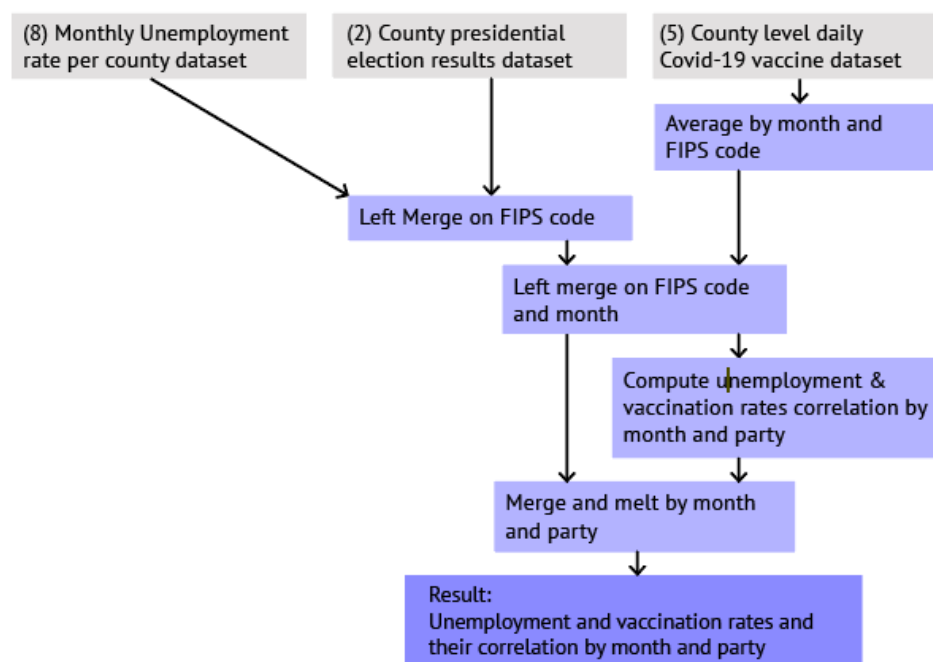
Data Sources

Name	Description	Details	Access
(1) State presidential election results	Constituency (state-level) returns for elections to the U.S. presidency from 1976 to 2020	size: 500KB shape: 4287 x 15 format: CSV file	Harvard Dataverse website
(2) County presidential election results	County-level returns for presidential elections from 2000 to 2020	size: 7.4MB shape: 72603 x 12 format: CSV file	Harvard Dataverse website
(3) COVID cases and death rolling averages	Daily number of new cases and deaths, the seven-day rolling average and the seven-day rolling average per 100,000 residents for all counties in the U.S.	size: >85MB shape: >146M x 10 format: CSV file	The New York Times GitHub page
(4) State level total COVID vaccine	Total COVID Vaccine deliveries and administration data at the state level.	size: 28KB shape: 63 x 62 format: CSV file	The U.S. Centers for Disease Control website
(5) County level daily COVID vaccine	Daily COVID Vaccine deliveries and administration data at the county level.	size: 139MB shape: >840,000 x 27 format: CSV file	The U.S. Centers for Disease Control website
(6) Mask-wearing survey	Estimate of mask usage by county in the United States released by The New York Times. It “comes from a large number of interviews conducted online” in 2020 between July 2nd and July 14th.	size: 109KB shape: 3143 x 6 format: CSV file	The New York Times GitHub page
(7) Population census and estimates	2010 population census data per county and the 2011~2020 population estimates. We are mainly interested in the 2020 estimates	size: 3.7MB shape: 3195 x 180 format: CSV file	U.S. Census Bureau website
(8) Unemployment rate dataset	Collection of labor force county data tables for 2020 issued by the U.S. Bureau of Labor Statistics	size: 7.69MB shape: >96,000 x 7 format: CSV file	To capture the unemployment rate for the desired period, we used the BLS v2 public APIs. Registration is required here . Use the
(9) Census Urban and Rural	Classification of all the counties in the U.S. as rural or urban areas	size: 302KB shape: 4287x15 format: 3142 x 8	U.S. Census Bureau website

Data Manipulation: Merging everything using FIPS codes

The datasets were joined using FIPS codes. FIPS stands for Federal Information Processing Standards, which are published by the National Institute of Standards and Technology (NIST) that every tract, county and state in the US is assigned. The code had to be extracted from a combined Geo code or had to be formed by concatenation if state and counties were separately coded and served as the primary key to merge the datasets.

For each visualization, the county/state level election data was combined with a feature of interest, such as vaccination rate, COVID cases, unemployment, urban/rural demographics. Below is an example on the data transformation flow used to generate the data for the first figure on the unemployment rate effect on the COVID response "Counties Average Unemployment Rate and Vaccination Rate Since January 2021" (page 7).



Thanks to the clean nature of these already curated datasets coming from reliable sources and the FIPS codes, there was no major challenge of missing or incorrect data. The main challenges we faced were:

- **FIPS codes:** Alaska has 29 counties + 1 unorganized borough. However, for the elections, unlike in other states, the results are reported by election districts and not by counties. And election districts have their own geographical boundaries and FIPS code which do not match counties (except for 3 counties). It is as such very hard to define the political affiliation of a county in Alaska and such counties were dropped. Out of a total of more than 3,000 counties, we considered this acceptable.
- **Vaccination:** States such as Texas and Hawaii only record vaccinations administered at the state level, in contrast to other states which report it at the county level. For visualizations at the county level, these states have been ignored.