An analysis of the changing sentiments concerning COVID vaccines

Group Members:
Elizabeth Choi, Yena Kim,
Arjun Ananda Padmanabhan, Kashmira Sawant, Sunny Singh

LHS 712: Natural Language Processing in Health Data
April 28, 2022

Abstract

Background

The COVID-19 pandemic has had devastating effects on people's wellbeing. As there is no empirical research done to fully explore the public's vaccine awareness, there is limited information about the quantitative evolution of public attitude since the rollout of COVID-19 vaccines. The rollout of the vaccine has produced complicated feelings and emotions due to misinformation, concern about its fast rollout, religious observances, etc - many of which have been expressed through social media. The goal of this project is to better understand sentiments regarding the COVID-19 vaccine throughout the vaccine rollout process, and the factors that potentially impacted them.

Objective

The focus of this study is to analyze tweets about COVID-19 vaccines available in the United States (Pfizer, Moderna, and Johnson & Johnson) before and after vaccines became publicly available. We aimed to explore the overall change in sentiments and topics of tweets about COVID-19 vaccines, as well as how such sentiments and main concerns evolved. To find the sentiment toward COVID-19 vaccines, we will sample a set of tweets that contain keywords from a set of words determined to be representative of COVID-19 vaccinations and develop a fine-tuned BERT transformer model to predict sentiment on those tweets.

Our three main hypotheses are as follows:

1. The status of states as either Blue or Red will influence the average sentiment towards vaccines
2. Geographical locations with lower socioeconomic status will have lower rates of vaccination and more negative sentiments towards it.
3. As vaccinations become authorized by health organizations, and widely used, sentiments regarding the vaccine will become more positive.

Results

1. Sentiment analysis in Red versus Blue states revealed that political affiliations may have little to no effect on the public's sentiment of COVID-19 vaccines.
2. Sentiments towards COVID-19 vaccines may not be affected by the socioeconomic status of the geographic location and a lower rate of COVID-19 vaccination.
3. Analysis of state-wide vaccination progress revealed that although there was generally a decline in positive sentiment regarding COVID-19 vaccines as the FDA issued EUA approval of COVID-19 vaccines, there was an increase in vaccination rates. It was also found that the pause in the use of a particular vaccine brand (J & J vaccine) did not negatively impact sentiment regarding COVID-19 vaccines, nor did it slow down vaccination rates across the United States.

Table of Contents

Introduction

COVID-19, caused by the virus SARS-CoV-2, was discovered in December of 2019 in Wuhan, China, and has quickly spread throughout the world, having been declared a pandemic in March 2020 by the World Health Organization ("Covid-19 basics", 2022). It is very easily spread through droplets projected out of a person's mouth or nose when they breathe, cough, sneeze or speak ("Covid-19 basics", 2022). COVID-19 symptoms range from no symptoms to severe symptoms such as high fever, shortness of breath, neurological symptoms, and gastrointestinal symptoms ("Covid-19 basics", 2022). The World Health Organization, as well as the Center for Disease Control, has recommended that anyone eligible for a vaccine, get one to protect themselves as well as their community stating protection against moderate to severe disease, hospitalization, and death ("Covid-19 basics", 2022).

In December of 2020, the Center for Disease Control issued the Emergency Use Authorization (EUA) for both the Pfizer-BioNTech and Moderna COVID-19 vaccines with the Johnson and Johnson vaccine also being issued a EUA on February 27, 2021 (*Historic Dates and Events Related to Vaccines and Immunization*, n.d.). Out of the 22 COVID-19 vaccines, the top 5 manufacturers include Oxford-AstraZeneca (181 countries), Pfizer-BioNTech (154 countries), Sinopharm-Beijing (86 countries), Moderna (86 countries), and Johnson & Johnson (81 countries) (Holder, 2021). Since the initial rollout of the vaccine, more than 10.7 billion doses have been administered across 184 countries, making this the biggest vaccination campaign in history ("More Than 10.7 Billion Shots Given," 2022). Although millions of vaccinated individuals have gotten sick from Covid-19, the vaccines were successful in preventing severe illness, reducing the chances of hospitalization and death by more than 90% ("More Than 10.7 Billion Shots Given," 2022).

An ongoing research project by the Kaiser Family Foundation tracks the public's attitudes and experiences with COVID-19 vaccinations. The report states that unvaccinated adults tend to be younger, with six in ten adults under the age of fifty not being vaccinated. The report also states that six in ten adults identify as Republican or Republican-leaning. In terms of education, almost half of unvaccinated adults have a high school education or less (49%) compared to the 33% of adults who completed some college and the 17% with a college degree or more. Seven in ten individuals who are unvaccinated report concerns regarding the safety of COVID-19 vaccines, stating they are either "not too confident" or "not at all confident" that they are safe for adults ("KFF COVID-19 Vaccine Monitor Dashboard," 2022).

Misinformation regarding the COVID-19 vaccine has led to the undermining of public trust, posing a great threat to global public health. To ameliorate the spread of misinformation from its platform, in March of 2020, Twitter, in coordination with other social media platforms, stated that false or misleading information about COVID-19 would no longer be able to be shared. Twitter stated concerns that misinformation could prevent individuals from making informed decisions, putting themselves and their communities at risk. This includes information about the nature of COVID-19, preventative measures including the vaccine, treatments, precautions, official regulations, prevalence, etc. In terms of the vaccine, tweets that insinuate the following information will be required to be deleted (COVID-19 Misleading Information Policy, n.d.):

- Vaccines are part of global surveillance, population control, depopulation efforts, etc.
- Vaccines are dangerous and its side effects are being covered up by the government and the medical industry
- The vaccine is being experimented on vulnerable groups

- Vaccines cause magnetic reactions
- Vaccines did not receive full approval and authorization and are thus unsafe and experimental

Methods

Data Source

Data was collected in the form of "tweets" from the social media platform Twitter. This was done through the use of snscrape, a social networking services scraper. The keywords used in our query include:

CovidVaccine	modernavaccine	vaccinepassports
COVIDVaccine	FullyVaccinated	vaccinatedconnections
COVIDVaccines	COVIDvaccines	coronavirusvaccine
vaccinedreams	EndVaccineApartheid	ModernaVaccine
GetVaccinated	VaccinesWork	TakeTheVaccine
vaccinate	VaccineConfidence	vaccination
vaccines	COVID19Vaccines	VaccineEquityNow
COVID19Vaccine	COVIDVaccinesSaveLives	VaccinEquity
vaccinated	vaccinedeliberation	VaccinatetheWorld
VaccineApartheid	VaccinateALL58	vaccinesforall
vaccine	freethevaccine	thepeoplesvaccine
Vaccine	COVID19Vaccination	vaccinatetheworld
VaccinateAll58	Vaccin8	vaccineequity
CovidVaccines	COVIDvaccine	mRNAvaccines
COVID19vaccine	VaccinesSaveLives	ivaccinatebecause
VaccineDreams	getvaccinated	Covidvaccine
vaccinehesitancy	PleaseGetVaccinated	VaccineEquity
GetVaccinatedNow	Vaccines	COVIDVaccination

Tweets were collected from March 1, 2020 until April 6, 2022, when the data was collected. We collected 10 tweets per day per phrase. To ensure that we weren't collecting duplicate tweets, we made sure that every tweet we appended was not already in our collection by using the tweet's ID. We also made sure we were only collecting tweets written in English. We included the following fields:

Search phrase	Tweet Place
Tweet date	Tweet Coordinates
Tweet ID	Tweet User Verified
Tweet content	Tweet User Follower Count
Tweet User Username	Tweet Retweet Count
Tweet User Location	Tweet Language

Using multithreading, we collected 222,232 tweets. For model training purposes, we collected 1 tweet per phrase per day, and performed .sample() to obtain a list of 3,500 randomly selected tweets that we later split into a training dataset of 2,000 tweets and a testing dataset of 1,500 tweets.

Coding Strategy

Codes were adapted from themes previously identified in the literature. A final codebook was created, which presented the definition of the sentiment categories (positive, negative, neutral sentiment regarding COVID vaccines). Irrelevant tweets, or tweets unrelated to the sentiment on COVID vaccines, were flagged for removal. Themes, sub-themes, and examples for positive and negative tweets were also made available. <u>Figures 1.1 – 1.3</u> display the codebook.

Everyone in the group was responsible for manually labeling sentiments to identify if the tweet was positive, negative, neutral, or irrelevant and whether the location of the tweet was in the United States or not. If it was, the name of the state was specified. Each person was assigned 400 tweets from the training dataset and 300 tweets from the test dataset.

To establish inter-rater reliability, a qualitative assessment was performed where two people in the group examined and labeled tweets that were labeled by other people in the group. Combined, a 90% accuracy was achieved.

Geolocation

The geolocation for many of the ingested tweets was unknown or missing. The initial cleaning was done in Microsoft Excel by filtering the tweets with known geolocation. The final tweet count after removing all tweets with missing geolocation was 123,021 tweets. All tweets with irrelevant geolocation such as "Earth", "global", and "%625" were also removed using the filter function in Microsoft Excel. The count of tweets with irregular locations was 6,199. The irrelevant geolocations were replaced with None in the dataset.

After initial processing, we had 116,822 tweets with relevant geolocations. Further validation and cleaning of the geolocation were done using geocoders from the GeoPy module. The "exactly_one = True" parameter was used to only return one address. The geocoder was used to geocode the remaining tweets along with the latitude and longitude.

After validation and standardization of the geolocation, we had 60,605 tweets with geolocation as "{States}, United States". These tweets were then used in our models and analysis.

Neural Network Model

1. Methodology

When going through the model-architecture decision process, the 2 architectures that were being considered were transformers and LSTMs. These 2 architectures were chosen over the standard RNN since literature has shown they're better suited for deriving context from the given tokens in an input (Colón-Ruiz, 2020). Ultimately, we decided to utilize a transformer architecture for our model, as we felt that it would be able to properly capture the peculiarities of our data source, especially given that the derived context of specific search phrases is extremely important. Based on our analysis of the data, we have classified the tweets as either **Positive**(1), **Neutral**(2), **Negative**(3), or **Irrelevant**(4). The irrelevant class was created to accurately identify the irrelevant tweets in our actual prediction dataset and remove them as appropriate. An example of this is how we were selecting tweets that spoke *specifically* about COVID-19 vaccines; however, we had many tweets speaking to vaccines *in general.*

These tweets were labeled as irrelevant, as they could not be used in our analysis against sentiment towards COVID-19 vaccines; therefore, it was of the utmost importance to have a high degree of sensitivity towards this class. The same could be true for tweets labeled as negative, as there was a fine line between a tweet speaking negatively about COVID-19 vaccines, and a tweet speaking negatively about those that chose *not* to get a vaccination against COVID-19.

When trying to decide between LSTMs and transformers, the ultimate deciding factor was the robustness of our training data. Due to the limitations of our data ingestion and labeling process, we were only able to create a training and testing set, with fully labeled data, to the tune of ~3500 tweets. This is an *extremely* small source dataset for training a fully validated and customized LSTM model. As such, we decided to move forward with a pre-trained BERT model; however, we decided to test several different established models from the Hugging Face model repository and compare their results. The models chosen included:

- **model_ctb_v1** = 'digitalepidemiologylab/covid-twitter-bert'
- **model_ctb_v2** = 'digitalepidemiologylab/covid-twitter-bert-v2'
- **model_bcb_c** = 'vinai/bertweet-covid19-base-cased'
- **model_bcb_u** = 'vinai/bertweet-covid19-base-uncased'
- **model_bb** = 'vinai/bertweet-base'
- **model_vct** = 'ans/vaccinating-covid-tweets'
- **model_msc** = 'clampert/multilingual-sentiment-covid19'

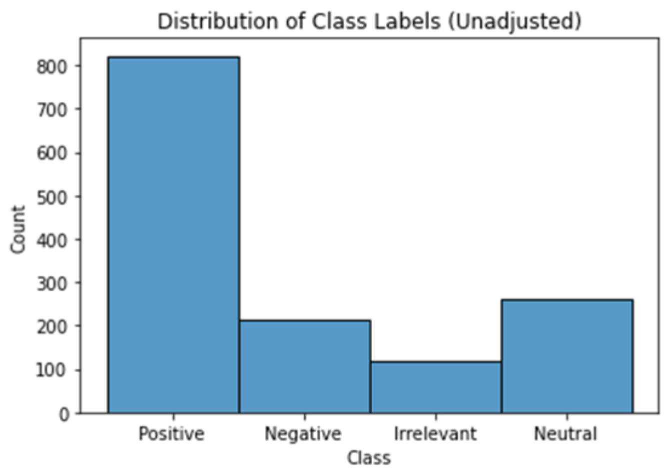
We then tested the weighted F1 scores of each of these models to determine the appropriate one for use in our project. A full description of each of the models can be found in the appendix under Table 2.1. Since we are using pre-trained models, we can leverage our smaller training dataset to fine-tune these established models. This reinforcement learning, while not as robust as custom training on a personalized tweet corpus, will still result in a high degree of model specification to our stated project area.

The tokenizers for each model are defined below.

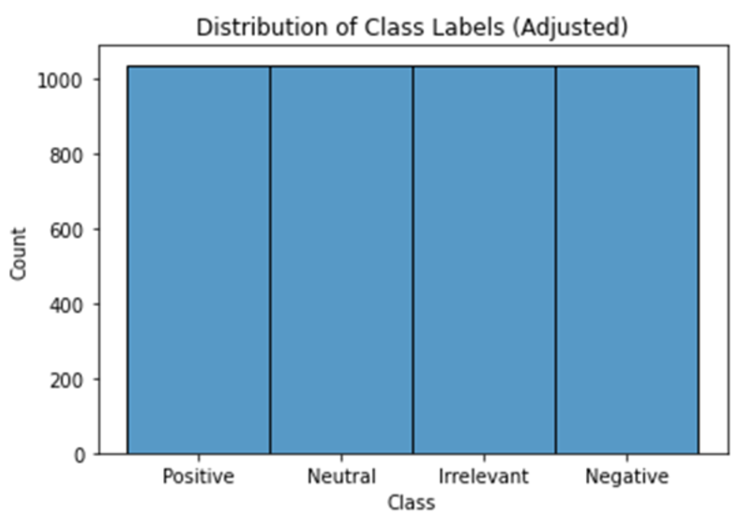
- **tokenizer_ctb_v1** = 'digitalepidemiologylab/covid-twitter-bert'
- **tokenizer_ctb_v2** = 'digitalepidemiologylab/covid-twitter-bert-v2'
- **tokenizer_bcb_c** = 'vinai/bertweet-covid19-base-cased'
- **tokenizer_bcb_u** = 'vinai/bertweet-covid19-base-uncased'
- **tokenizer_bb** = 'vinai/bertweet-base'
- **tokenizer_vct** = 'ans/vaccinating-covid-tweets'
- **tokenizer_msc** = 'clampert/multilingual-sentiment-covid19'

We sought to reduce the number of tokenizers, and by proxy the number of models, that we had in our analysis. We attempted to do this by getting rid of models that had similar architecture and encodings, as we felt there would not be a significant amount of information gained by including these redundant models. This was done by running an initial training and testing iteration for the models after a single round of fine-tuning, and unifying models that had the same embedded tokenized tensors and 99% similarity for predicted labels. The results of this reduction in model dimensionality are discussed further in the next section.

Prior to splitting the data, we looked at the level of class imbalance for our target labels. We saw that there was a high degree of class imbalance in our labeled testing dataset, as seen below.



To fix this issue, we decided to use oversampling of our minority classes to build a more robust training set. We chose oversampling vs. resampling to create the same distribution, as we felt that there was an appropriate signal within the tweets of each minority category (with appropriate signal defined as cosine similarities between tokenization embeddings for minority classes) where appropriate trends would be amplified. The distribution of class labels after oversampling is seen below.



Before running through all of the model embeddings, we created a 70-30 train-validation split for each tokenized dataset for our model. The feature space will be the content of the tweet (cleaned) and the target will be one of the multi-class labels for the sentiment of the tweet.

2. Analysis

Initially, we saw that after initial encoding, we had 4 independent sets of tensors. This meant that we had models that encoded the tweets into the same word embeddings, and as such, we sought to pick a single tokenizer/model for each of the embeddings and architectures. We see that the embeddings created by tokenizer_ctb_v1 is equal to that of the embeddings created by tokenizer_ctb_v2, and the embeddings created by tokenizer_bcb_c, tokenizer_bcb_u, and tokenizer_bb are all equal. We can therefore eliminate the redundant word embeddings and move forward with a group of 4 different embeddings. Ultimately, we decided on using the following models for our final analysis.

- **model_ctb_v2** = 'digitalepidemiologylab/covid-twitter-bert-v2'
- **model_bcb_u** = 'vinai/bertweet-covid19-base-uncased'
- **model_vct** = 'ans/vaccinating-covid-tweets'
- **model_msc** = 'clampert/multilingual-sentiment-covid19'

The model summaries can be found in the appendix under Figure 2.3. We can see that each of the models had the same baseline architecture, with only the number of tunable parameters per-layer changing between each of the models. In the case of **model_bcb_u** and **model_vct**, they had the same number of tunable parameters in the architecture. Nevertheless, they both performed differently, albeit similarly.

Our hyperparameters when running the models included the following specifications:

- Using an AdamW optimizer with a learning rate of 1*10^-5 and a EPS of 1*10^-8
- Epochs = 4
- Scheduler with number of warm-up steps = 0 and the number of training steps equal to the length of our data loader multiplied by the number of epochs (4).

Additionally, since there was a wide-ranging distribution of vector lengths for the tokenized content, we tried to find an optimal max_length such that we had a strong representation of our data, without necessarily increasing the memory space of our algorithm extensively. To accomplish this we initially tried to find a max_length (k) such that 90% of all tweets in our training dataset had a length below k. We found this value to be 47 ; however, the model performance using this max_length was *extremely* poor. This was tied back to the basis of how BERT neural networks tokenize content– BERT models employ pseudo-character-level tokenization. This means that a given word can be broken up into several tokens, such as a prefix and gerunds. This better enables the transformer model to understand the given context of a set of tokens, adding to performance and generalizability. By truncating the model to only 47 tokens, we were significantly limiting the performance of our model by training on only a small subset of tweet content. The remedy to this problem was to extend the length of tokenized vectors to 512, as this was the max length for input to the BERT model. This resulted in a roughly 20-point gain in weighted F1 evaluation across each model.

Our evaluation metrics for the models were the weighted F1 score and the accuracy per class. To conform to system settings and specifications, we initially had to set the batch size to 2 and run the analysis on 2 epochs. However, we optimized the memory storage to the GPU to allow us to run 4 epochs on each of the identified models. The results of each model are outlined in Table 2.2**.**

In general, there was equal classification performance across each class in our model. We can see by the confusion matrices, figure 2.5 and figure 2.6, that no one class performed *significantly better or significantly worse*, although there was noticeable specificity for the 'Positive' encoding class. The final output dataset was filtered for irrelevant tweets; therefore, only tweets labeled 'Positive', 'Negative', and 'Neutral' were included.

Red vs Blue States

To determine whether there is a difference in sentiment between Blue or Red states, we used BeautifulSoup to scrape a table on a Wikipedia page on presidential election results. This will be used to compare whether there is a difference in overall sentiment by state status and whether there are states that have changed with the most recent election. Our table contains data on the presidential election results by state and was filtered to only include the years 2016 and 2020 to represent the two most recent elections. We also wanted to look at the difference in vaccine sentiment depending on whether the governor was Red or Blue. To do this, we also used BeautifulSoup to scrape a Wikipedia article on the list of current United State governors. Data regarding political majorities in the State Senate and State House was also ingested using BeautifulSoup and was filtered to only include data from the most recent elections.

For data processing purposes, tweets that were classified as irrelevant were removed from the dataset. Sentiments were later grouped according to their political affiliation, and visualizations were used to compare the Twitter sentiment analysis to the COVID-19 vaccination rate across different political affiliations.

Two sample t-tests, also known as Independent t-tests, were performed to compare the means between Red and Blue states to determine whether there is statistical significance between the sentiments of the two variables.

Further analysis was done for misaligned states. A state was considered misaligned if it did not have the same party majority in the State Senate, State House and the governor did not belong to the same party (e.g., The political majority in the State Senate is Republican, State House is Democratic and the governor belongs to the Republican party.). There were 12 misaligned states and they are Alaska, Kansas, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, North Carolina, Pennsylvania, Vermont, Virginia, and Wisconsin.

Additional two sample t-tests were performed to compare the means between Red and Blue misaligned states to determine whether there is statistical significance between the sentiments of the two variables.

Socioeconomic Status Data Ingestion

To determine whether there is a difference in sentiment between geographical locations of significantly different economic statuses, we have selected two cities; San Francisco and Detroit, which have significantly different socioeconomic statuses in their populations. We used the American Community Survey data (2020) from the Census Bureau to obtain all the socio-economic data for San Francisco and Detroit including poverty levels, educational attainment, and median household income between different race/ethnic groups. All data was obtained from Census Bureau API. We obtained the vaccination data for San Francisco and Detroit from the CDC webpage.

We have removed tweets that were classified as irrelevant from the dataset. The sentiments were grouped by the city for further visualizations and analysis including word clouds. To normalize the median income for both cities, income below the 25th percentile was coded as "low", income between the 25th and 75th percentile was coded as "medium", and income higher than the 75th percentile was coded as "high".

Two sample t-tests were performed to compare the means between Detroit and San Francisco city and to determine whether there is any statistical significance to the sentiments of the two cities.

Statewide Vaccine Progress

To view vaccine progress by state, we first located a dataset on Kaggle (which retrieves data from covid.ourworldindata.org) that provides state-by-state data on COVID-19 vaccinations in the United States from December 20, 2020, to March 9, 2022. This data provides information on total vaccinations, people fully vaccinated per hundred, people vaccinated per hundred, and total boosters per hundred. Two other Kaggle datasets were used to pull and link the state code and region/division information to the vaccination rate data. One minor change was made to clean up the data such as changing "New York State " to "New York". Further, the data included all 50 states including the District of Columbia and Puerto Rico.

Data processing for sentiment analysis involved removing irrelevant tweets and coding positive tweets as '1', neutral tweets as '0', and negative tweets as '-1'. Sentiments were later grouped by state and periods before and after specific dates identified:

- **January 25, 2021:** First case of Gamma variant in the U.S.
- **January 28, 2021:** First case of Beta variant in the U.S.
- **April 13, 2021**: The CDC and the FDA recommended a pause in the use of the Janssen (Johnson & Johnson) COVID-19 vaccine in the U.S. out of an abundance of caution. A CDC Health Alert Network (HAN) was issued with recommendations.
- **May 10, 2021**: The FDA expands the emergency use authorization of the Pfizer-BioNTech COVID-19 vaccine to include adolescents 12-15 years of age.
- **July 7, 2021:** The Delta variant becomes the dominant COVID-19 strain in the U.S.
- **August 23, 2021**: The FDA approves the first COVID-19 vaccine Comirnaty (Pfizer-BioNTech) for individuals 16 and older. (The EUA remains in effect for individuals 12 years of age and older and a third dose for immunocompromised individuals 12 years of age and older).
- **September 22, 2021**: The FDA authorizes a booster dose of Pfizer-BioNTech COVID-19 Vaccine for aged 65 years and older, aged 18 through 64 at high risk of severe COVID-19, aged 18 through 64 who have institutional or occupational exposure to SARS-CoV-2.
- **October 29, 2021**: The FDA authorizes EUA for Pfizer-BioNTech COVID-19 vaccine for children 5-11 years.
- **December 1, 2021:** The U.S. announces the first cases of Omicron variant.

Visualizations were generated which displayed percent fully vaccinated and percent vaccinated (those with at least 1 dose) by state and by a date specified. These visualizations were resourceful for comparing average sentiment regarding COVID vaccines across different states or regions in the United States.

Two sample t-tests were performed to compare the mean sentiment before and after the dates specified and determine whether there was statistical significance.

1. *Smoothing of Observation*

A time analysis was done for the sentiments. It was done to analyze the change in sentiments over a time period. The period which was assessed was from March 1, 2020 until April 4, 2022. The y-axis consisted of the sentiments predicted by the model and could take any continuous value between -1 (fully negative) and 1 (fully positive).

The benefits of using time series are:

- Helps analyze historical sentiment change.
- Understanding and matching the current situation with patterns derived from the previous sentiment analysis
- Understanding the factor or factors influencing the sentiments and causing fluctuations in the graph

To appropriately analyze the time-series data of sentiment, we applied a time-series decomposition to our dataset. This resulted in the following components of our data being analyzed individually:

- <u>Trend:</u> In trend, there is no fixed interval and any divergence within the given dataset is a continuous timeline. The trend would be Negative, Positive, or Null Trend.
- <u>Seasonality:</u> The Seasonality of the data is defined as a regular or fixed interval shift within the dataset in a continuous timeline.
- <u>Irregularity/Noise:</u> It can be defined as unexpected situations/events/scenarios and spikes in a short period within the data.

2. *Smoothing of Observation*

Smoothing is a technique applied to time series to remove the fine-grained variation between time steps. Moving averages are a simple and common type of smoothing used in time series analysis and time series forecasting. A commonly used time series method is Moving Average. The Moving Average (MA) or Rolling Mean is calculated by taking averaging data of the time series, within k periods.

Calculating a moving average involves creating a new series where the values are the average of raw observations in the original time series. A moving average requires that you specify a window size called the window width. This defines the number of raw observations used to calculate the moving average value. The "moving" part in the moving average refers to the fact that the window defined by the window width is slid along the time series to calculate the average values in the new series. The moving average period or rolling window period of 14 days was used in the time series analysis of the sentiments.

Results

Red vs Blue States

Political affiliation is known to greatly influence views of COVID-19 vaccines (Albrecht, 2022). Since May 2021, people living in counties that heavily voted for President Trump during the last presidential election are three times as likely to die from COVID-19 as those who live in areas that heavily voted for President Biden, showing that states with higher vote shares for Trump have lower vaccination rates (Wood et al., 2021). Vaccinated Republicans are also less likely to acknowledge the seriousness of the pandemic and are more likely to say the pandemic has been exaggerated in the news (Kirzinger et al., 2021). An overwhelming majority of unvaccinated Republicans (96%) and vaccinated Republicans (73%) state that vaccinations are a personal choice, while the majority of vaccinated Democrats (81%) see it as a responsibility to protect the health of the community (Kitzinger et al., 2021). In general, Biden counties have had higher COVID-19 vaccination rates than Trump counties; as of January 11, 2022, 65% of those in Biden counties have been fully vaccinated compared to 52% of those in Trump counties and the vaccine gap has widened over time (Kates et al., 2022).

Misinformation seems to be a major factor in the slow uptake of vaccinations (Wood et al., 2021). Polling conducted by the Kaiser Family Foundation shows that Republicans are more likely to believe false statements about COVID-19 and vaccines, with 94% of Republicans believing in one or more erroneous statements about COVID-19, and 46% believe more than four false statements compared to 14% of Democrats (Wood et al., 2021). This uptake of misinformation seems to be tied to exposure to right-wing media (Albrecht, 2022). Many Republican communities see the act of wearing masks or getting a vaccine as a political statement, unnecessary, and a violation of individual freedoms (Albrecht, 2022).

1. ***Presidential Election***

In 2016, there were 30 Republican states and 21 Democratic states, while in 2020 there were 26 Republican states and 25 Democratic states. Five states changed status - Arizona, Georgia, Michigan, Pennsylvania, and Wisconsin. All five states transitioned from Red states to Blue states from 2016 to 2020.

The mean sentiment score for Red states according to the 2016 elections was 0.17 while the sentiment score for Red states according to the 2020 elections was 0.44, which indicates a 158.824% change in sentiment. In terms of Blue states, the sentiment score for Blue states according to the 2016 elections to those in the 2020 elections changed from -0.1 to 0.50 - a 600% change. Please refer to Figure 3.1.

2. ***Governors***

In terms of governors, throughout the pandemic, 28 states had Republican governors while 22 had Democratic governors. Sentiment scores for states that had a Republican governor were 0.38 while those for states with a Democratic governor were 0.34. Please refer to Figure 3.2.

3. ***State Legislature***

32 states have a majority Republican State Senate, with only 18 states having a majority Democratic State Senate. 18 states had a majority Democratic State House, 31 states had a majority Republican State House and 1 state (Alaska) had a split State House. State senates with a

majority Republican legislature had a mean sentiment of 0.36, while those with a majority Democratic legislature had a mean sentiment of 0.37. In terms of State House, states with a majority Republican legislature had a mean sentiment of 0.35, compared with the 0.37 for a majority Democratic legislature and 0.48 for Alaska which has an equally split legislature. Please refer to Figure 3.3.

Contrary to what we had expected, we did not see a significant difference in sentiment between political affiliations. The visualizations indicate that in terms of the presidential election and governor party, Republicans seem to have a more positive sentiment towards vaccines than Democrats. Using an alpha of 0.05, we fail to reject the null hypothesis that Red and Blue states have significantly different sentiments. The p-value for the t-test comparing Red and Blue states in the 2016 election was 0.078, while that of the 2020 election was 0.305. In terms of Red and Blue states according to the current governor, we obtained a p-value of 0.49. The p-value for Red and Blue states according to the State Senate was 0.88, while that of the State House was 0.74. Some of this discrepancy could be attributed to Twitter's decision to no longer allow the sharing of false and misleading information about COVID-19, different counts of tweets per state, and inherent limitations of the use of Twitter data.

The sentiment analysis for misaligned states showed diverse results. Using an alpha of 0.05, we reject the null hypothesis that Red and Blue misaligned states according to the current governor and the State House have significantly different sentiments (p-value = 0.037, p-value = 0.0074). On the other hand, the p-value for Red and Blue misaligned states according to the State Senate was 0.20, therefore we failed to reject the null hypothesis and concluded that Red and Blue misaligned states do not have significantly different sentiments according to the State Senate.

Socioeconomic Status

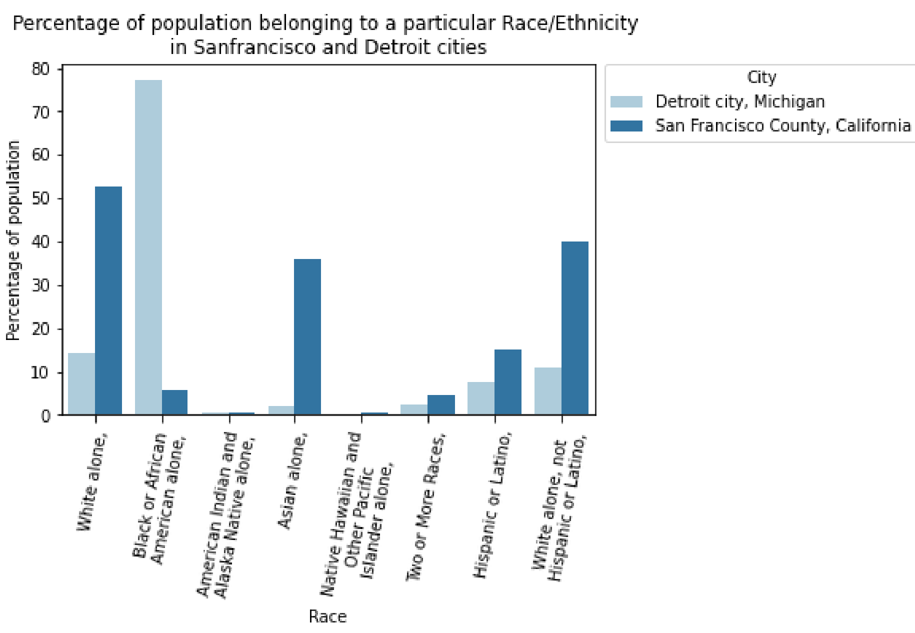
Socioeconomic status has an overall impact on well-being including physical and mental health. Low socioeconomic status along with low educational attainment, poverty, and poor health can ultimately affect society. Research has shown that race and ethnicity in terms of stratification often determine a person's socioeconomic status (*Ethnic and Racial Minorities & Socioeconomic Status*, 2017). Studies have shown that counties with more Black residents have significantly lower uptake of COVID-19 vaccination (Beusekom, 2021). COVID-19 vaccination disparity is associated with income, poverty, and education. People in counties with higher median income and educational attainment have less COVID-19 vaccine disparity (Agarwal et al, 2021). We decided to analyze the COVID-19 vaccination trend and COVID-19 vaccine sentiment analysis between two geographical locations that have significant differences in socioeconomic status - Detroit, Michigan, and San Francisco, California.

1. Social Vulnerability Index:

Social Vulnerability Index was developed by CDC and ATSDR's Geospatial Research, Analysis & Services Program (GRASP) to assist public health officials and emergency response teams in identifying and mapping the communities that most likely require assistance during a public health event. SVI ranks the census tracts on 15 social factors, including unemployment, minority status, and disability, and is further divided into four themes (Figure 4.1). Poverty, overcrowding, and other neighborhood characteristics linked to social vulnerability raise the likelihood of negative health outcomes during and after a public health event. Counties with high social vulnerabilities, particularly with a higher percentage of racial and ethnic minority residents were at higher risk of becoming a COVID-19 hotspot (Dasgupta et al., 2020). The overall SVI for San Francisco is 0.263 (Centers for Disease Control and Prevention/ Agency for Toxic Substances and Disease Registry/ Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index [2018] Database [California]., n.d.) whereas the overall SVI for Wayne County (of which Detroit is part) is 0.9878 (Centers for Disease Control and Prevention/ Agency for Toxic Substances and Disease Registry/ Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index [2018] Database [Michigan]. , n.d.).

2. *Race/Ethnicity:*

San Francisco is home to a population of 882,000 people, of which 88.1% are citizens. In 2019, there were 1.15 times more White (Non-Hispanic) residents than any other race or ethnicity. Detroit is home to a population of 670,000 people, of which 96.5% are citizens. In 2019, there were 7.31 times more Black or African American residents in Detroit than any other race or ethnicity. As per the American community survey (2020), 52.8% of the population in San Francisco are White (Non-Hispanic), 5.6% are Black or African American, 0.7% American Indian/Alaska Native, 36.0% Asian, and 15.2% Hispanic. In Detroit, 14.4% are White (Non-Hispanic), 77.1% Black or African American, 0.4% American Indian/Alaska Native, 1.9% Asian, and 7.7% Hispanic or Latino (*U.S. Census Bureau QuickFacts: Detroit City, Michigan; San Francisco County, California*, n.d.).

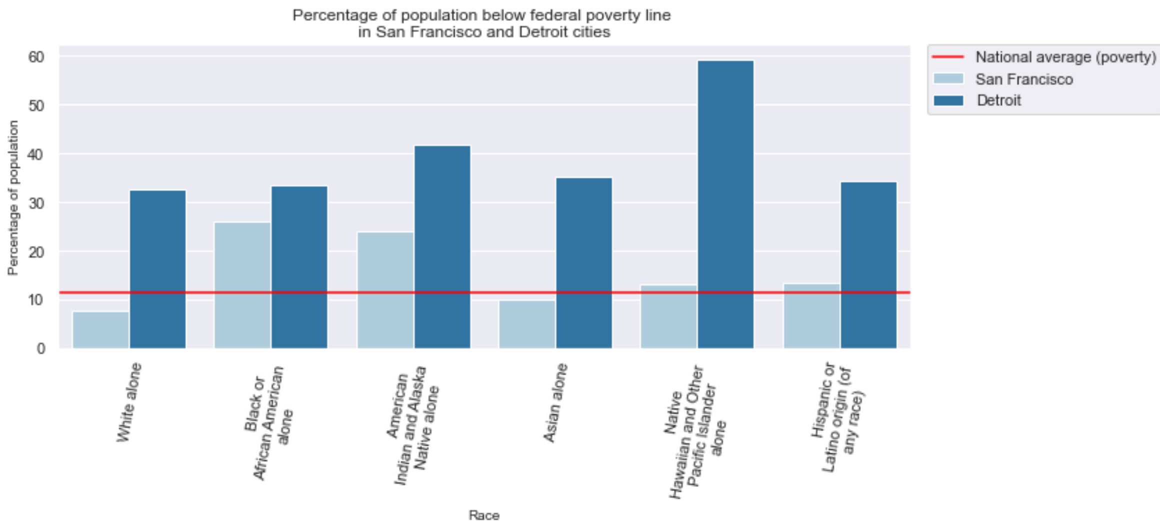


3. *Poverty*

The majority of the population who are below the federal poverty line in San Francisco are from Black or African American communities (26.1%) followed by American Indian/Alaska Native (23.9%). Only 7.7% of the people of White ethnicity are below the poverty line in San Francisco. In Detroit, the majority of the population who are below the federal poverty line are Native Hawaiian (59.4%) followed by American Indian/Alaska Native (41.7%), and Black or African

American (33.5%) (U.S. Census Bureau (2020)). The percentage of Black or African Americans in Detroit is around 77.1% and around 33.5% of them live below the poverty level.

The red horizontal line represents the national average of the percentage of the population who are below the federal poverty line. The percentage of the population who are below the federal poverty line is significantly higher than the national average for all races in Detroit.

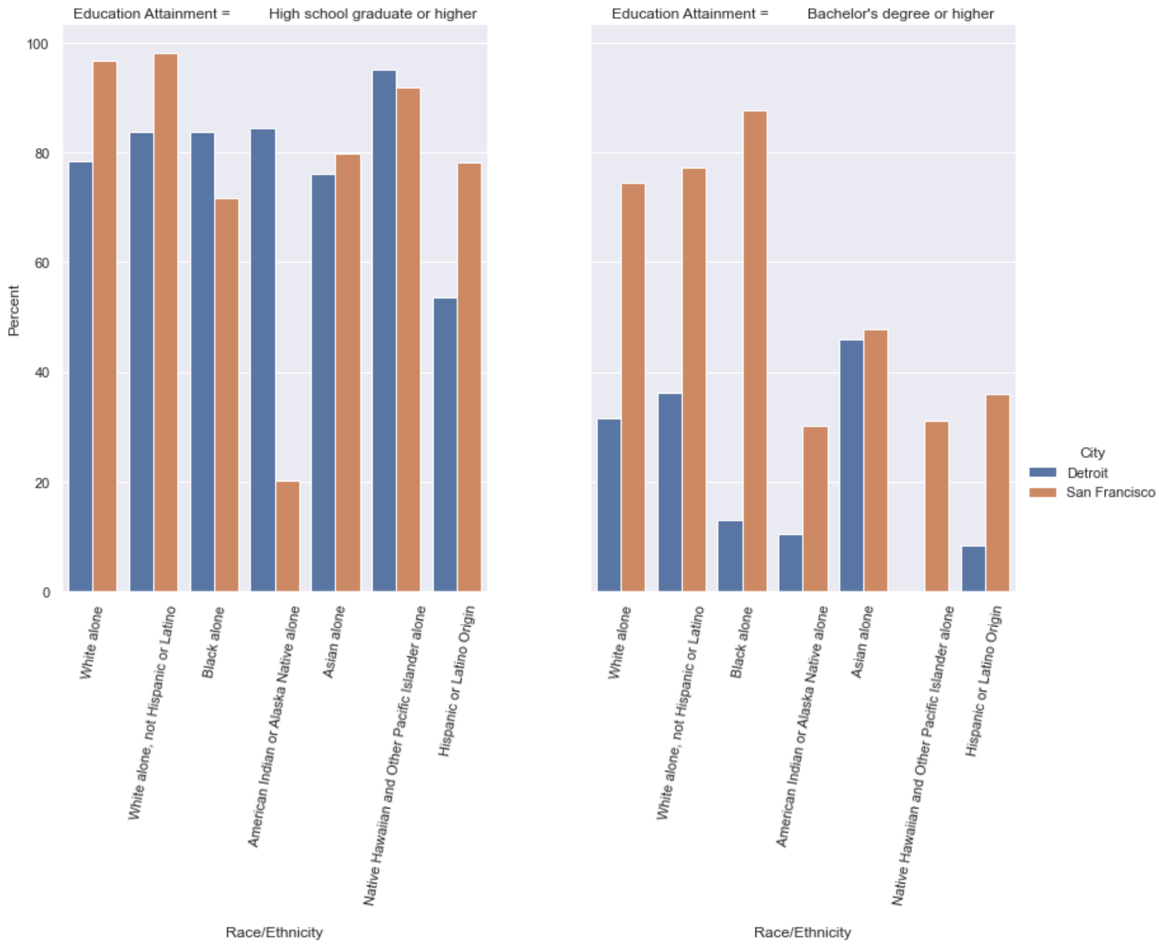


4. *Education attainment*

Around 96.8% of White (Non-Hispanic), 71.8% of Black or African American, 20.2% of American Indian/Alaska Native, 79.9% of Asian, 92.0% Native Hawaiian/Other Pacific Islander, 78.2% of Hispanic/Latino has completed high school or higher in San Francisco. Whereas around 78.5% of White (Non-Hispanic), 83.9% of Black or African American, 84.6% of American Indian/Alaska Native, 76.2% of Asian, 95.2% of Native Hawaiian/Other Pacific Islander, 53.5% of Hispanic/Latino have completed high school or higher in Detroit (U.S. Census Bureau (2020)).

The high percentage of high school graduates belonging to Black or African American in Detroit is related to the higher percentage of Black or African Americans in the city.

The percentage of the population who have a bachelor's degree or above is only 58.5% in San Francisco and 20.2% in Detroit. Only 13% of the Black or African American population in Detroit have a bachelor's degree or above as compared to 87.7% in San Francisco.



5. Median Income

The average median income for San Francisco county for all races is $50,682 and the average median income for Detroit for all races is $20,461 (U.S. Census Bureau (2020)). The median income for White and Asian ethnicity is High in both Detroit and San Francisco. The median income for Black or African American communities is "Medium" in San Francisco county and "Low" in Detroit city.

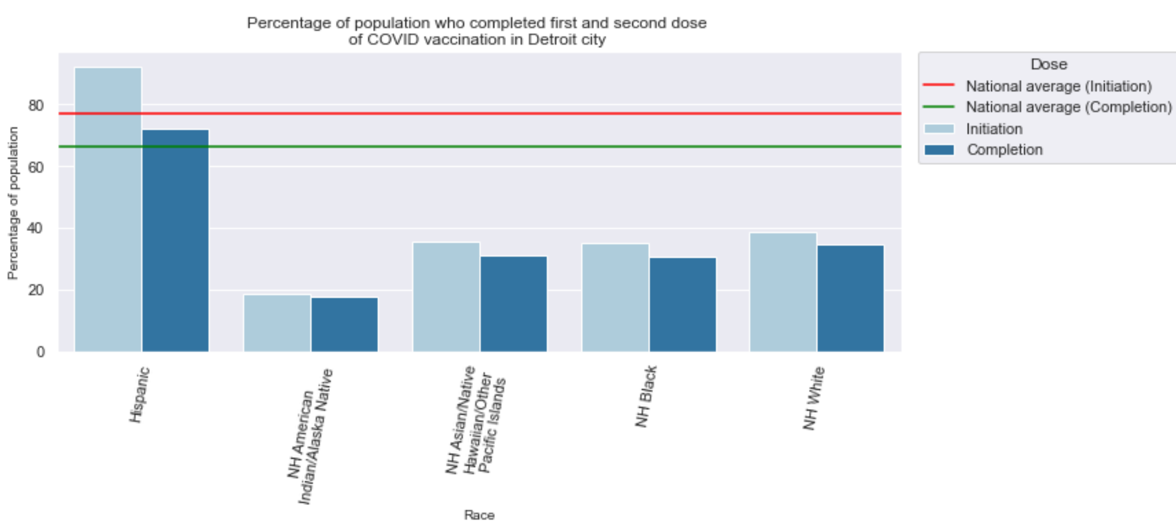
Race/Ethnicity	San Francisco, California	Detroit, MI
White	High	High
Black or African American	Medium	Low
American Indian and Alaska Native	Low	Medium
Asian	High	High
Native Hawaiian and Other Pacific Islander	Low	Low
Hispanic or Latino origin (of any race)	Low	Low

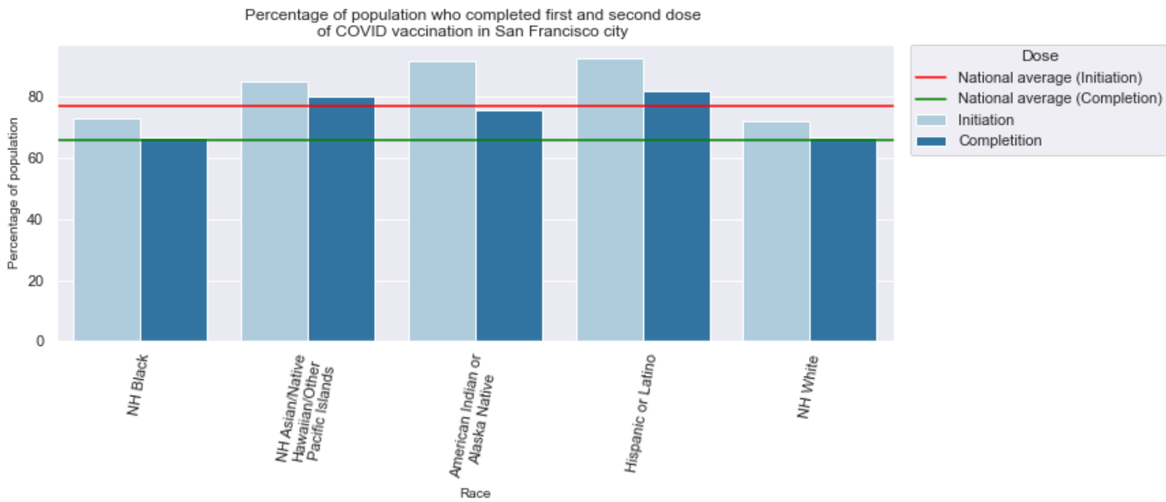
Comparing the socioeconomic status of all the races between San Francisco and Detroit, the population belonging to the Black or African American ethnicity in Detroit has a high poverty rate and low average median income as compared to San Francisco. Additionally, the population belonging to the White ethnicity has a low poverty rate and high median income in both the cities. The education attainment is almost comparable between the two cities for both these races in terms of high school graduation. However, the population belonging to Black or African Americans who have completed bachelor's degrees is very low in Detroit as compared to San Francisco.

6. COVID Vaccination Progress

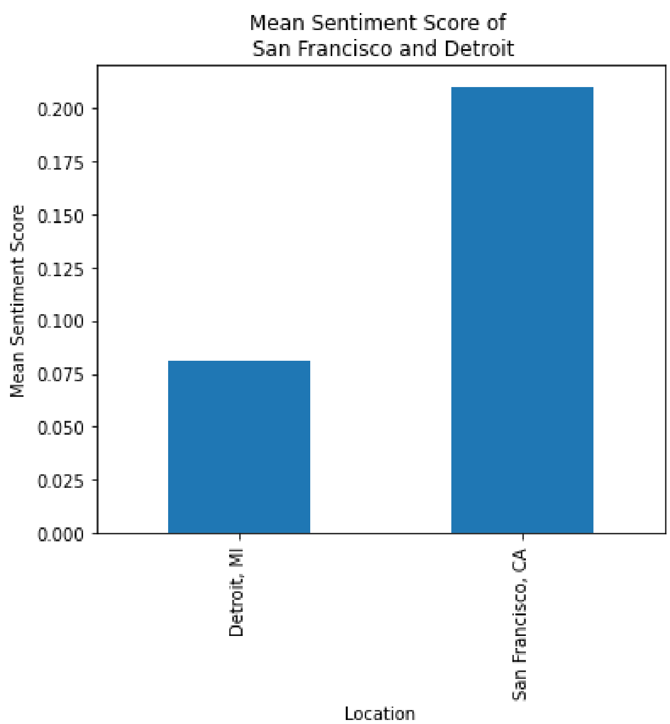
As of April 2022, the overall vaccination rate for San Francisco city is 97.5% and for Detroit is 49.6%. The vaccination rate is highest for Hispanic/Latino communities in both Detroit and San Francisco. The vaccination rate for Black or African communities is only 33.6% in Detroit City. Research published by the Ford School of Public Health, University of Michigan has reported that safety is the main concern among the Black or African American community to get vaccinated (*Concern about Safety Is Main Reason Many Detroiters Are Not Getting Vaccinated, U-M Survey Finds*, 2021).

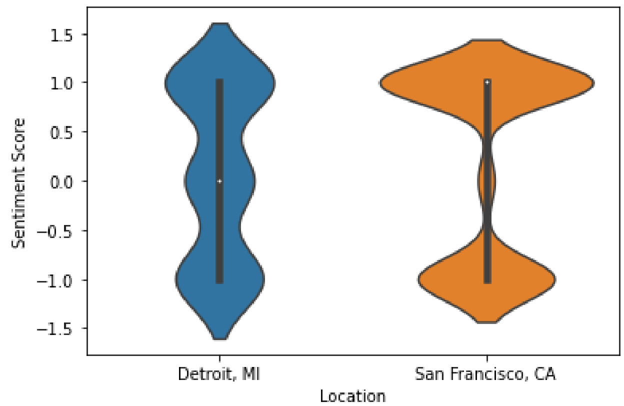
Additionally, San Francisco in March 2021 has taken significant steps toward an equitable vaccination strategy. The city of San Francisco has focused on reaching populations that have been disproportionately affected by COVID-19 (especially Hispanic and Black or African American) (*San Francisco to Expand COVID-19 Vaccinations to People with Disabilities and Severe Underlying Conditions and Those in High-Risk Congregate Settings on March 15 | Office of the Mayor*, n.d.). This has led to an increase in vaccination rate among the Hispanic (92.4% (initial dose) and 81.9% (second dose)) and Black or African American (73.6% (initial dose) and 66.8% (second dose)) communities.





The red horizontal line represents the percentage of the population who has completed at least one dose of COVID vaccination all over the US and the green horizontal line represents the percentage of the population who has completed both doses of COVID vaccination all over the US.





The mean sentiment score for San Francisco is 0.20 and for Detroit is 0.08. The median score for San Francisco is around 1.0 (positive) and the median sentiment score for Detroit is around 0 (neutral). From the violin plot, we can observe that the probability of positive sentiment in San Francisco is more than the negative sentiment whereas it is equally distributed in Detroit (the median is around Neutral).

We have hypothesized that there will be more negative sentiment and a lower vaccination rate around vaccination in counties with lower socioeconomic status. Even though the vaccination rate is lower in Detroit, there is equal distribution of positive and negative sentiments around vaccination. On the other hand, as hypothesized, San Francisco has more positive sentiment towards vaccination than negative sentiment.

As discussed in the previous section, "safety" is the main concern among the Black or African American community getting vaccinated. When we analyze the word cloud for Negative tweets from Detroit, we can see that the words like "fear", "explanations", "scientists", "death", "covidvaccineinjuries", "realdonaldtrump" and "deadly" are visible in the word cloud (Figure 4.2).

The interquartile range for the mean sentiment score for both San Francisco and Detroit (from the Violin Plot) is the same. The standard deviation of the sentiment score for Detroit is 0.85 and the standard deviation for the sentiment score in San Francisco is 0.95. The t-test performed on the sentiment scores from Detroit and San Francisco has given a p-value of 0.058. The alpha that we chose for hypothesis testing was 0.05. Hence, we fail to reject the null hypothesis that there is no difference in mean sentiment towards COVID vaccination between geographical areas with different socioeconomic factors and COVID vaccination rates.

Even though Detroit city has a higher percentage of racial minorities, a high poverty rate, low median income, and a lower rate of COVID-19 vaccination progress, there were no overall negative sentiments toward the COVID-19 vaccination.

Statewide Vaccine Progress

To analyze statewide vaccine progress, we referred to Immunize.org's vaccine timeline, which provided important events regarding COVID vaccines, and also researched the dates for initial cases of COVID variants in the United States. To examine the vaccine progress for particular dates (Figure 5.1), bar charts, choropleth maps, and tables were created to illustrate the overall vaccine progress before and after these events. Please note that some states may be missing data for a specific date (i.e., In the choropleth map provided below, states that do not have any data will not have a black border around them). Tables have been added to showcase the percent change in vaccination rate before and one month after the event took place.

Our group has identified three main categories of events to examine vaccine progress: (1) dates where initial cases of COVID variants became a concern in the United States, (2) dates where the COVID vaccine was made available for certain age brackets, and (3) date where there was a recommended pause in the use of a particular vaccine brand. Mean sentiment scores by state were calculated by averaging sentiment scores of tweets from one month before and one month after the date of interest.

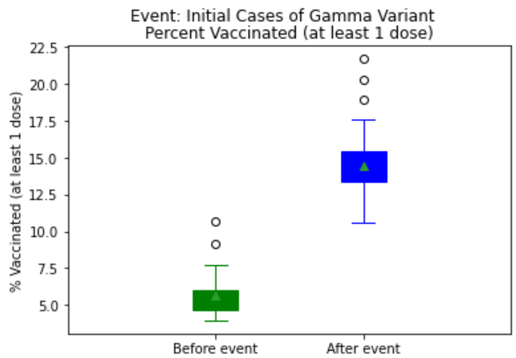
1. *COVID-19 variants of concern*

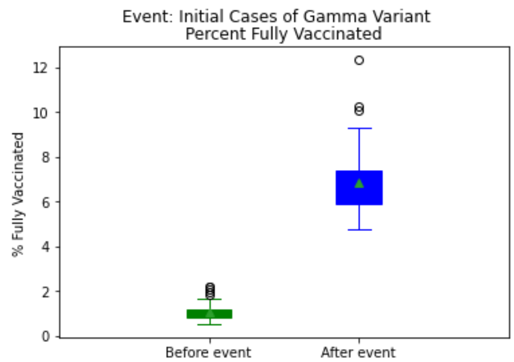
We were interested in looking at vaccine progress prior to the announcement of initial cases of COVID-19 variants in the United States. This included the following dates:

- **January 25, 2021:** First case of Gamma variant in the U.S.
- **January 28, 2021:** First case of Beta variant in the U.S.
- **July 7, 2021:** The Delta variant becomes the dominant COVID-19 strain in the U.S.
- **December 1, 2021:** The U.S. announces first cases of Omicron variant.

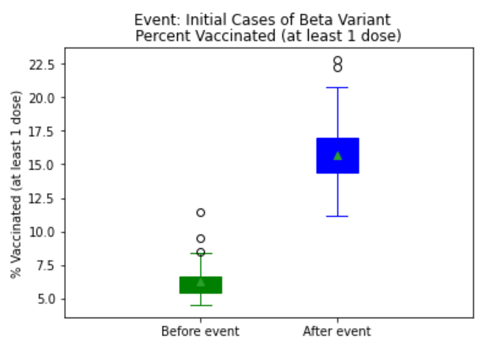
The Alpha variant was excluded from our analysis as initial cases were reported on January 4, 2021, which was not within the dates examined in our dataset. The state vaccine progress before and after these dates is depicted in Figure 5.2.

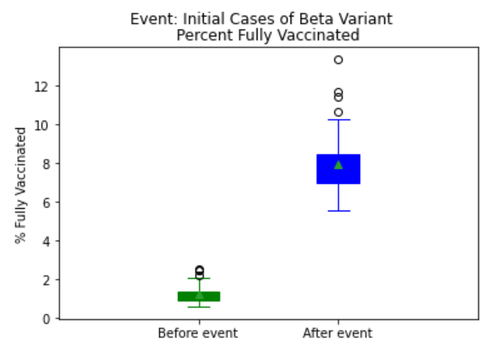
Although Virginia and Hawaii's sentiment regarding COVID vaccines decreased about 33% (0.68 to 0.45, 0.62 to 0.41), both states experienced the highest percent change in full vaccination rate and vaccination rate (at least one dose) one month after the initial cases of the Gamma variant in the U.S. was announced (1053.33%, 243.12%). Alaska had the highest fully vaccinated rate (12.33%) and vaccination (at least 1 dose) rate (21.71%) at the time, with its sentiment regarding COVID vaccines having slightly decreased (0.38 to 0.36). In the United States, the average fully vaccinated rate one month after the initial cases of the Gamma variant was announced was 6.88%, while the average vaccinated rate (at least 1 dose) was 14.46%. A total of 28 states were below the average in terms of fully vaccinated rate and 27 were below the average in terms of vaccination (at least 1 dose) rate. The average sentiment score regarding COVID vaccines was about 0.50, a 9% increase from the month before Gamma variant cases were announced.



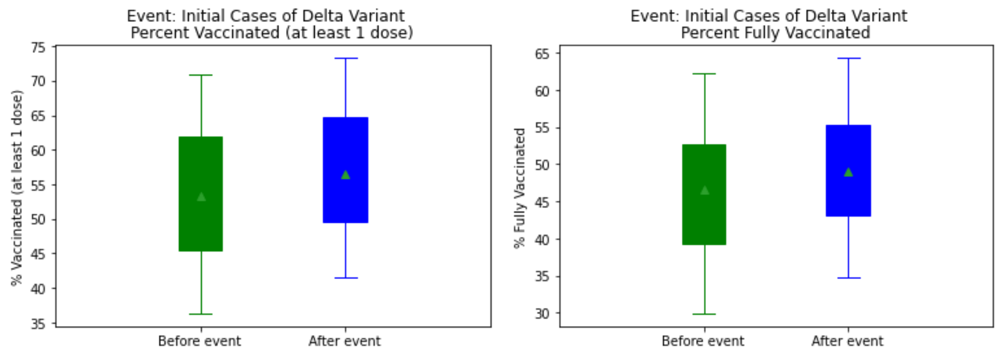


Mississippi's sentiment regarding COVID vaccines (0.25) remained unchanged one month before and one month after initial cases of the Beta variant were announced, having the highest percent change in full vaccination rate (1100%). On the other hand, Wisconsin's sentiment increased 25% (0.36 to 0.45) and it exhibited the highest percent change in vaccination rate (at least one dose) (240.79%). Alaska, with a sentiment score of 0.33 before and after the event, maintained the highest fully vaccinated rate (13.34%) and vaccination (at least 1 dose) rate (22.79%). The average fully vaccinated rate and the average vaccinated rate (at least 1 dose) in the United States rose to 7.93% and 15.69%. The average sentiment score regarding COVID vaccines was about 0.50, a 7% increase from the month before Beta variant cases were announced.

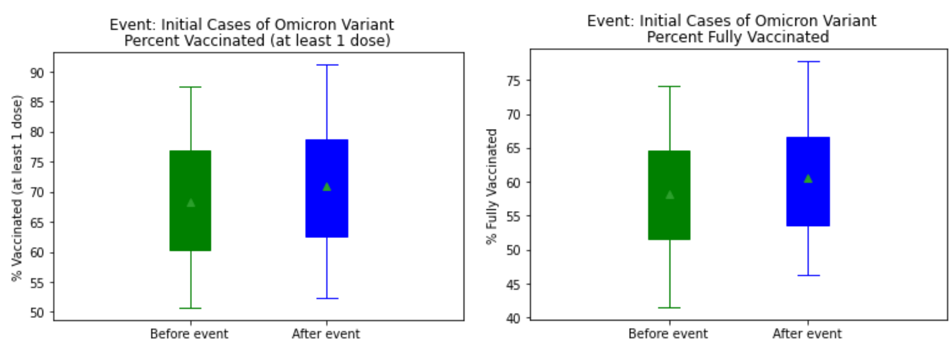




Interestingly, the state of Utah experienced the highest percent change in full vaccination rate a month after the initial cases of the Delta variant were announced, even though its sentiment score went down about 78% (0.80 to 0.18). Although Arkansas' sentiment score decreased from 1 to 0.38, it still had the highest percent change in vaccination rate (at least one dose) (15%). However, it is important to note that from the state of Utah, there were only 20 tweets regarding COVID vaccines before and only 17 tweets after this event. Additionally, Arkansas only had 3 tweets before and 8 tweets after regarding COVID vaccines. This indicates that the sentiment scores provided may not exactly be reflective of Utah's actual sentiments toward COVID vaccines at the time. Although Massachusetts' sentiment regarding COVID vaccines declined about 13% (0.58 to 0.50), Massachusetts still had the highest fully vaccinated rate (64.33%) and vaccination (at least 1 dose) rate (73.29%). The average fully vaccinated rate and the average vaccinated rate (at least 1 dose) in the United States rose to 49.06% and 56.58%. The average sentiment score regarding COVID vaccines was about 0.50, an 11% increase from the month before Delta variant cases were announced.



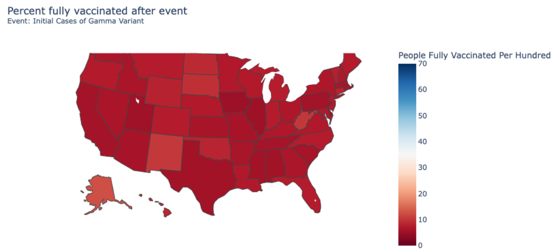
While West Virginia's sentiment on COVID vaccines one month prior to and one month following the announcement of initial cases of the Omicron variant jumped from 0.83 to 1, West Virginia also experienced the highest percent change in full vaccination rate and vaccination rate (at least one dose) (32.93%, 15.28%). Although Vermont's sentiment on COVID vaccines dropped about 40% (1 to 0.60), it had the highest fully vaccinated rate (77.75%). As mentioned previously, West Virginia also had a significantly low number of tweets regarding COVID vaccines during this time, so the sentiment score provided may not be representative of the state's sentiments toward COVID vaccines. On the other hand, Massachusetts had the highest vaccination (at least 1 dose) rate (91.17%), with a sentiment score of 0.49 (a 6% decrease one month prior to initial Omicron variant cases). The average fully vaccinated rate and the average vaccinated rate (at least 1 dose) in the United States rose to 60.65% and 71%. The average sentiment score regarding COVID vaccines was about 0.39, a 19% decrease from the month before Omicron variant cases were announced.



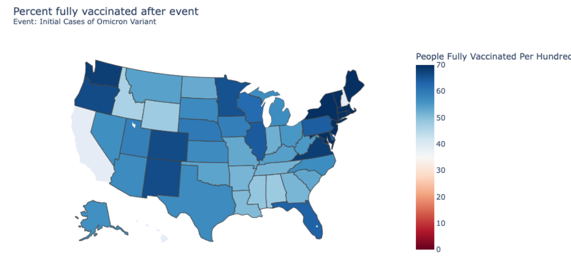
As shown in the choropleth maps below, the mid-Atlantic, west, southwest, and northeast regions of the United States seem to have had higher vaccination rates as compared to other regions.

Vaccination Progress (fully vaccinated)

1 month after initial cases of the Gamma variant

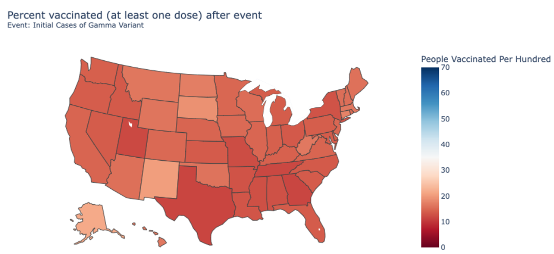


1 month after initial cases of the Omicron variant

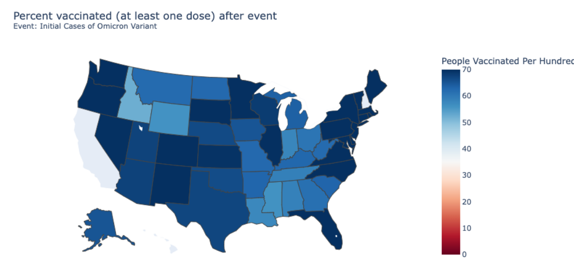


Vaccination Progress (at least one dose)

1 month after initial cases of the Gamma variant



1 month after initial cases of the Omicron variant



While the percent change in vaccination rate was highest after the report of initial cases of the Gamma and Beta variants, it might be prudent to not put too much weight in the percent change in vaccination rate after the initial cases of the Gamma and Beta variants due to the recency of the vaccine rollout.

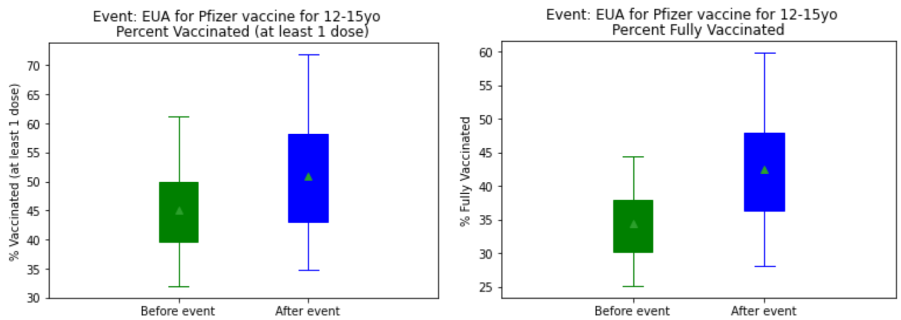
A statistical two-sample t-test was performed to assess the significance of mean sentiment across the dates observed. Using an alpha of 0.05, we fail to reject the null hypothesis that there was a statistically significant difference in mean sentiment one month before and one month after the initial cases of the Gamma variant were announced (p-value = 0.42). We fail to reject the null hypothesis that there was a statistically significant difference in mean sentiment before and after the initial cases of the Beta variant were announced (p-value = 0.58). Further, we also conclude that a significant difference in mean sentiment before and after the initial cases of the Delta variant was announced does not exist (p-value = 0.07). We do not have sufficient evidence to say that the mean sentiment before and after the initial cases of these variants was announced is different. On the other hand, we reject the null hypothesis and conclude that a significant difference in mean sentiment before and after the initial cases of the Omicron variant was announced does exist (p-value = 0.009). The significant difference in mean sentiment suggests that there may have been a potential confounder. About a month prior to the initial cases of the Omicron variant, the FDA authorized an EUA for Pfizer-BioNTech COVID-19 vaccine for children 5-11 years, which suggests that this event may have affected sentiment regarding COVID vaccines (more positive sentiment as vaccine was made available for children).

2. *COVID-19 vaccine availability by age bracket*

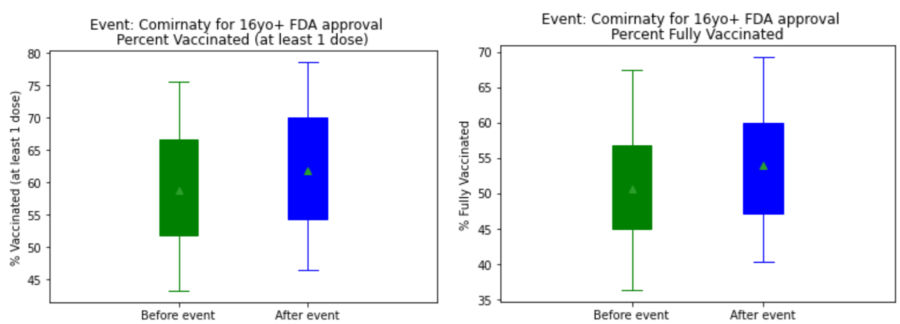
In addition to examining COVID vaccine progress by dates of initial cases of COVID variants, we wanted to look at the extent to which the availability of COVID-19 vaccines by age bracket may have influenced vaccine progress. Below are the dates that were taken into consideration for this analysis:

- **May 10, 2021**: The FDA expands the emergency use authorization of the Pfizer-BioNTech COVID-19 vaccine to include adolescents 12-15 years of age.
- **August 23, 2021**: The FDA approves the first COVID-19 vaccine Comirnaty (Pfizer-BioNTech) for individuals 16 and older. (The EUA remains in effect for individuals 12 years of age and older and a third dose for immunocompromised individuals 12 years of age and older).
- **September 22, 2021**: The FDA authorizes a booster dose of Pfizer-BioNTech COVID-19 Vaccine for aged 65 years and older, aged 18 through 64 at high risk of severe COVID-19, aged 18 through 64 who have institutional or occupational exposure to SARS-CoV-2.
- **October 29, 2021**: The FDA authorizes EUA for Pfizer-BioNTech COVID-19 vaccine for children 5-11 years

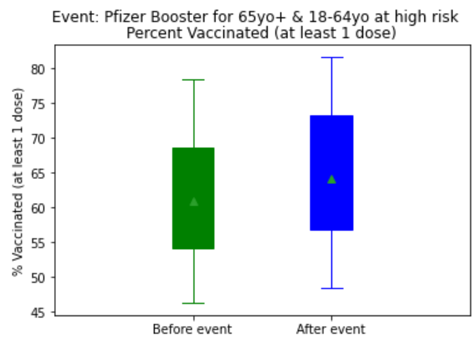
As provided in Figure 5.3, New Hampshire experienced the highest percent change in full vaccination rate one month after the FDA issued a EUA of Pfizer for those aged 12-15 (58.53%). Although it may seem as though New Hampshire had a tremendous increase in sentiment on COVID vaccines (0.13 to 0.58), there were fewer than 15 tweets regarding COVID vaccines at this time. On the other hand, Puerto Rico had the highest percent change in vaccination rate (at least one dose) one month after the announcement (25.35%). The average sentiment on COVID vaccines could not be provided for Puerto Rico as no tweets were available. Although Vermont had the highest fully vaccinated rate (59.81%) and vaccination (at least 1 dose) rate (71.84%) at the time, it also had a 47% decrease in sentiment on COVID vaccines one month prior to and one month after the EUA was issued (0.54 to 0.29). However, it should be taken into account that there was a low count of tweets regarding COVID vaccines. In the United States, the average fully vaccinated rate one month after the Pfizer EUA for 12-15-year-olds was granted was 42.48%, while the average vaccinated rate (at least 1 dose) was 50.92%. A total of 27 states were below the average in terms of fully vaccinated rate and 29 were below the average in terms of vaccination (at least 1 dose) rate. The average sentiment score regarding COVID vaccines was about 0.51, an 11% decrease from the month before the FDA issued a EUA of Pfizer for 12-15-year-olds.

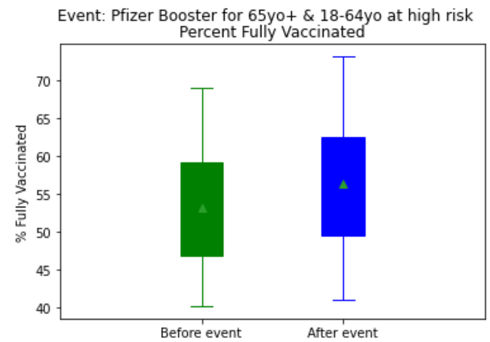


Although Mississippi experienced the highest percent change in full vaccination rate and the highest percent change in vaccination rate (at least one dose) (15.93%, 9.97%) after the FDA approved Comirnaty for those 16 and older, its sentiment regarding COVID vaccines changed from positive to negative (0.38 to -0.20). However, it should be noted that there were only 5 tweets regarding COVID vaccines after this event, so this shift in sentiment may not be reflective of the actual sentiment toward COVID vaccines at the time. Puerto Rico had the highest fully vaccinated rate (69.23%) and vaccination (at least 1 dose) rate (78.58%). The average fully vaccinated rate and the average vaccinated rate (at least 1 dose) in the United States rose to 53.89% and 61.79%. The average sentiment score regarding COVID vaccines was about 0.44, a 6% decrease from the month before the FDA approved Comirnaty for those 16 and older.

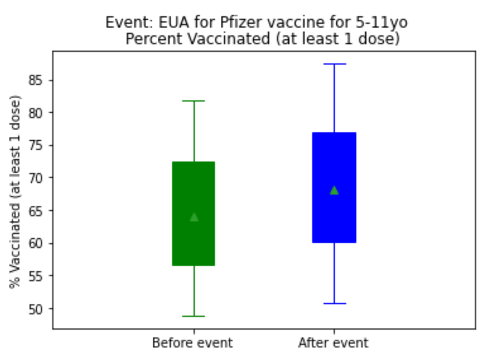


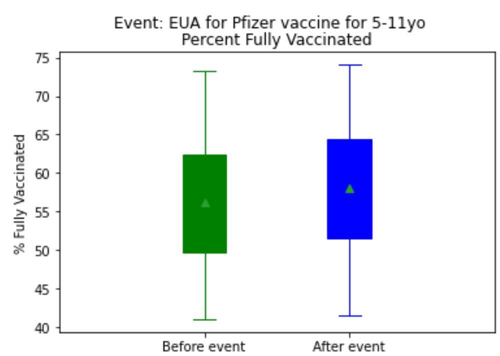
One month after the FDA authorized Pfizer booster dose for those aged 65+, age 18-64 at high risk of severe COVID-19 or who have had institutional or occupational exposure to SARS-CoV-2, Georgia (7.92%) experienced the highest percent change in full vaccination rate and its sentiment towards COVID vaccines increased from 0.30 to 0.37 (about a 24% increase). While North Carolina (7.28%) had the highest percent change in vaccination rate (at least one dose), its sentiment toward COVID vaccines dropped by about 37% (0.49 to 0.31). At this time, Puerto Rico remained with the highest fully vaccinated rate (73.12%) and vaccination (at least 1 dose) rate (81.62%). The average fully vaccinated rate and the average vaccinated rate (at least 1 dose) in the United States rose to 56.26% and 64.21%. The average sentiment score regarding COVID vaccines was about 0.37, a 19% decrease from the month before the FDA authorized Pfizer booster dose for these populations.





On the other hand, after a EUA of the Pfizer vaccine was issued for children aged 5-11, North Dakota's sentiment toward COVID vaccines increased 100% (0.50 to 1) and it had the highest percent change in full vaccination rate (6.47%). Although North Dakota's sentiment increased 100%, it is worth noting that there were less than 5 tweets regarding COVID vaccines before and after this event. New Hampshire (16.13%) had the highest percent change in vaccination rate (at least one dose). Its sentiment toward COVID vaccines remained unchanged one month prior to and one month after the EUA was issued (0.67). While Puerto Rico had the highest fully vaccinated rate (74.07%), New Hampshire had the highest vaccination (at least 1 dose) rate (87.39%). The average fully vaccinated rate and the average vaccinated rate (at least 1 dose) in the United States rose to 58.02% and 68.14%. The average sentiment score regarding COVID vaccines was about 0.48, a 25% increase from the month before the EUA of the Pfizer vaccine was issued for those 5-11 years old.





Overall, the event where the FDA provided a EUA of the Pfizer vaccine for 12-15-year-olds exhibited the greatest percent increase in vaccination rates across many states in the United States, as opposed to the other three events analyzed in this section (Figures 5.5 - 5.8). In addition, there were 8 states (located in the northeast and mid-Atlantic regions of the United States) that had over 50% of their population fully vaccinated:

1. Vermont, 59.81%
2. Maine, 57.21%
3. Massachusetts, 57.13%
4. Connecticut, 56.08%
5. Rhode Island, 53.81%
6. New Hampshire, 52.62%
7. New Jersey, 51.15%
8. Maryland, 50.50%

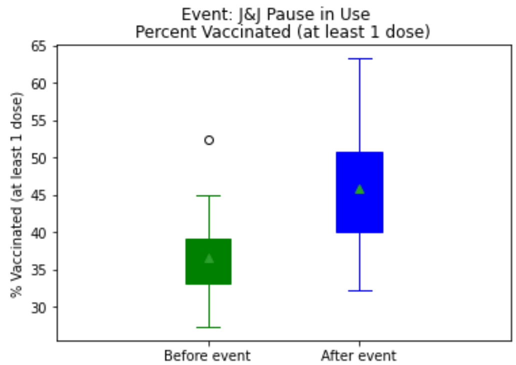
These states had an average sentiment score of 0.56 one month after the FDA provided a EUA of the Pfizer vaccine for 12-15-year-olds, which is 10% higher than the national average sentiment score.

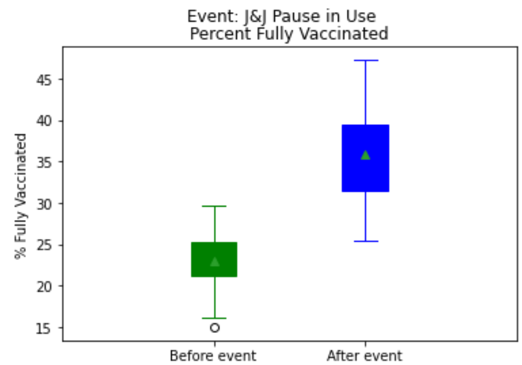
A statistical two-sample t-test was performed to assess the significance of mean sentiments across the dates observed. Using an alpha of 0.05, we fail to reject the null hypothesis that there was a statistically significant difference in mean sentiment one month before and one month after the EUA of Pfizer vaccine was issued for adolescents aged 12-15 (p-value = 0.66). We fail to reject the null hypothesis that there was a statistically significant difference in mean sentiment before and after the EUA of Comirnaty was issued for those 16+ (p-value = 0.09). Further, we also conclude that a significant difference in mean sentiment before and after the FDA approved the Pfizer Booster for those 65+ and those 18-64 at high risk does not exist (p-value = 0.43). We do not have sufficient evidence to say that the mean sentiment before and after these events is different. On the other hand, we reject the null hypothesis and conclude that a significant difference in mean sentiment before and after the Pfizer EUA issue for children 5-11 years old does exist (p-value = 0.0005). The significant difference in mean sentiment suggests that there may have been a potential confounder. About a month after the FDA provided an EUA for the Pfizer vaccine for those 5-11 years old, initial cases of the Omicron variant were announced, which suggests that this event may have affected sentiment regarding COVID vaccines (more positive sentiment as cases were rising).

3. Pause in COVID-19 vaccine use

We were also interested in seeing how the pause in COVID-19 vaccine use may have impacted vaccine progress. A particular date we focused on was April 13, 2021, when the CDC and the FDA recommended a pause in the use of the Janssen (Johnson & Johnson) COVID-19 vaccine in the U.S. out of an abundance of caution. A CDC Health Alert Network (HAN) was issued with recommendations.

As provided in Figure 5.4, Georgia experienced the highest percent change in full vaccination rate one month after the recommended pause in the use of the J&J vaccine (90.35%). Its sentiment regarding COVID vaccines increased dramatically from 0.37 to 0.63 (a 68% change). On the other hand, Hawaii had the highest percent change in vaccination rate (at least one dose) one month after the announcement (63%), with its sentiment regarding COVID vaccines remaining almost the same (0.71 to 0.70). Maine had the highest fully vaccinated rate (47.25%), despite having a 50% decrease in sentiment (1 to 0.50). However, this can be explained by the low number of tweets regarding COVID vaccines. Vermont had the highest vaccination (at least 1 dose) rate (63.23%) at the time, with a 26% increase in sentiment on COVID vaccines (0.43 to 0.54). In the United States, the average fully vaccinated rate one month after the recommended pause in the use of the J&J vaccine was 35.82%, while the average vaccinated rate (at least 1 dose) was 45.88%. A total of 26 states were below the average in terms of fully vaccinated rate and 28 were below the average in terms of vaccination (at least 1 dose) rate.





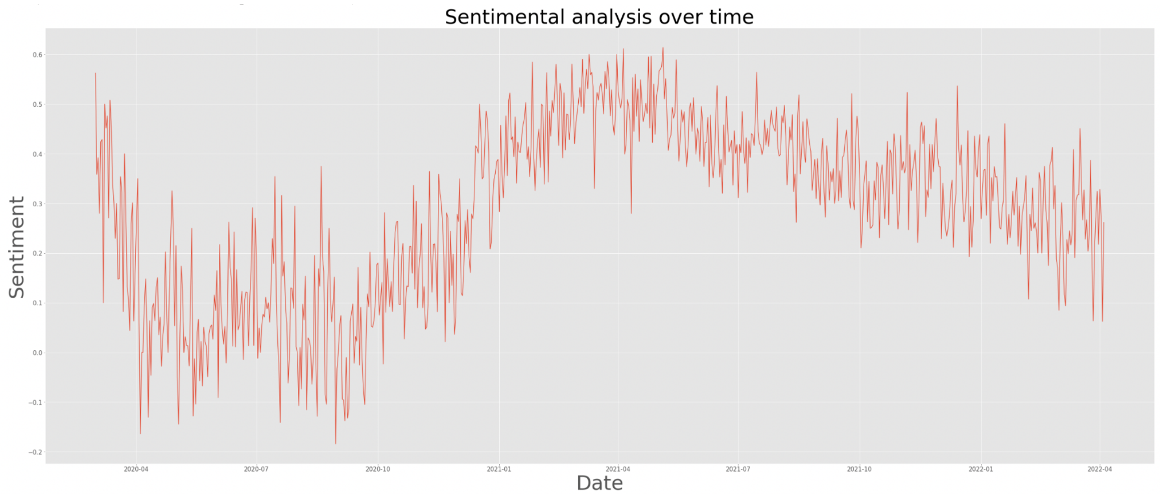
The bar charts in Figures 5.9 - 5.10 present an interesting phenomenon that seems to suggest that the pause in the Johnson & Johnson vaccine did not negatively impact people's decisions to get vaccinated. The average sentiment score regarding COVID vaccines was about 0.57, a less than a 1% decrease from the month before the pause in J&J vaccine use. The statistical two-sample t-test, using an alpha of 0.05, confirms that there was no significant difference in mean sentiment one month before and one month after the recommended pause in the use of the Johnson & Johnson vaccine (p-value = 0.11).

4. *Time Series Analysis*

Observation

The graph provided is used to observe the change in the sentiments over a period of time. The x-axis in the graph represents the date and the y-axis represents the sentiments. The timeline followed in this graph is from March 1, 2020 until April 4, 2022. It can be seen on the graph that the sentiments are relatively negative from the time April 2020 to January 2021. The variation in the graph follows the hypothesis that the sentiments were relatively negative in the initial phase of the COVID-19 pandemic and then the sentiments are relatively positive in the later phase of the COVID-19 pandemic.

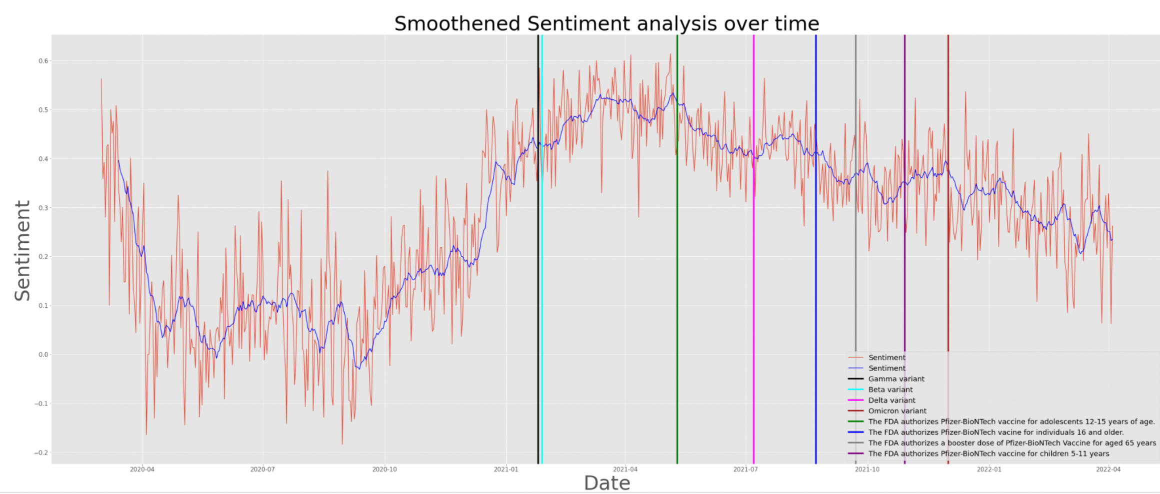
The cause of the pattern of variation in the sentiments could be due to a lack of information and treatment for COVID-19. We believe that the earlier mentioned causes could have led to negative sentiments about vaccines and COVID -19 from April 2020 to January 2021. With the rollout of the vaccine on December 14, 2020, there was a change in the pattern of sentiments seen. The sentiments are relatively positive from January 2021 as the authorization of emergency vaccine use by the FDA contributed to a change in the sentiments. The time-series graph below is further decomposed into trends, seasonality, and noise component.



Smoothing of the graph

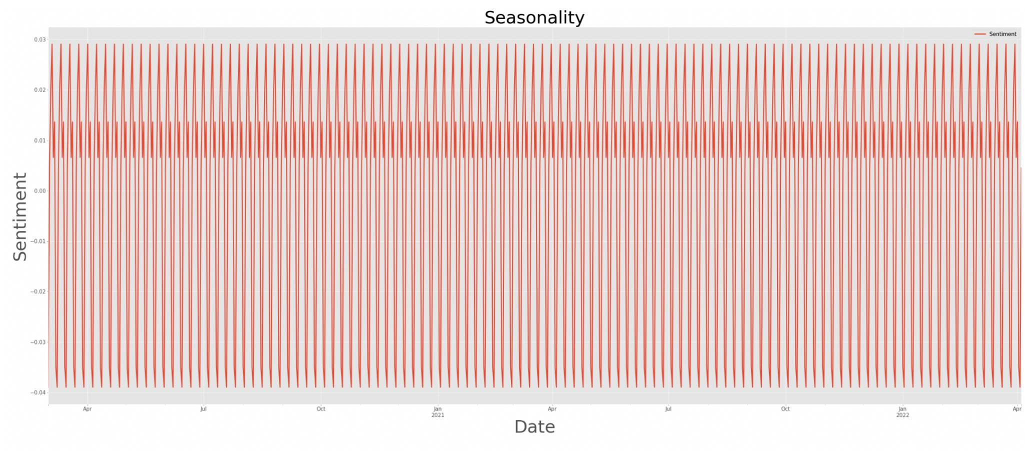
The observation graph was smoothed using a moving average or rolling window period of 14 days. This was done to remove the fine-grained variation between time steps. The blue line on the graph represents the change in sentiment over time after smoothing.

The change in sentiments is also affected by various factors like detection of the first cases of Gamma, Beta, Delta, and Omicron variants, FDA authorization of vaccines, and other external events. The following lines represent the events and the change in sentiment pattern can be analyzed.

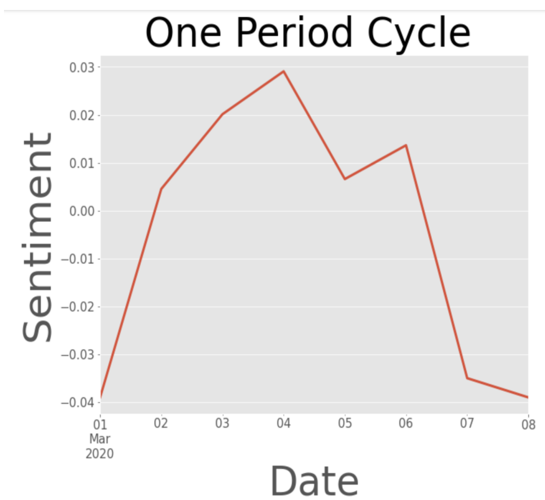


Seasonality

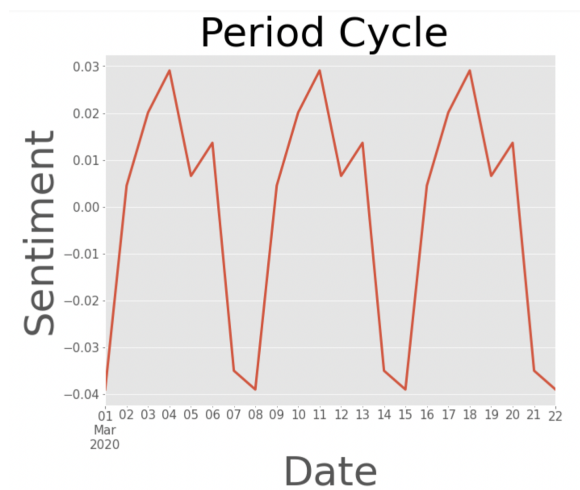
A seasonal pattern is a rise and fall in the data values that repeat regularly over the same period. Seasonal patterns always have a fixed and known period. In the graph below, the x-axis is the timeline from March 1, 2020 to April 4, 2022. The Y-axis represents sentiments ranging from -0.4 to 0.3. It can be seen from the graph that the distribution of sentiments follows a wave pattern that repeats after every 8 days. Graph 2 represents a single waveform for a period of 8 that is followed by the sentiments. Graph 3 represents a waveform for 3 cycles.



Graph 1 : Seasonality graph the timeline from March 1, 2020 to April 4, 2022.



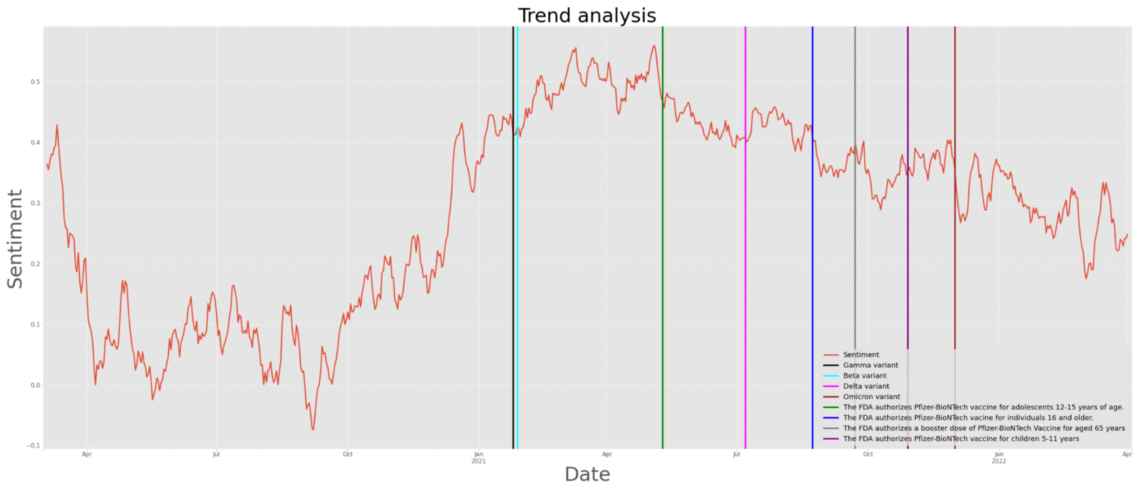
Graph 2: Waveform for 8 days window period



Graph 3: Waveform for 3 cycles

Trend Analysis

A trend analysis was done for positive and negative sentiments. The overall trend of a time series shows whether it increased, decreased, or stayed constant (flat) over a time period. The overall trend in the data showed that at the onset of the pandemic, sentiment toward a potential vaccine was generally positive. However, as time continued and sentiments became polarized, we see that a significant drop occurred and lasted until the availability of the vaccines.



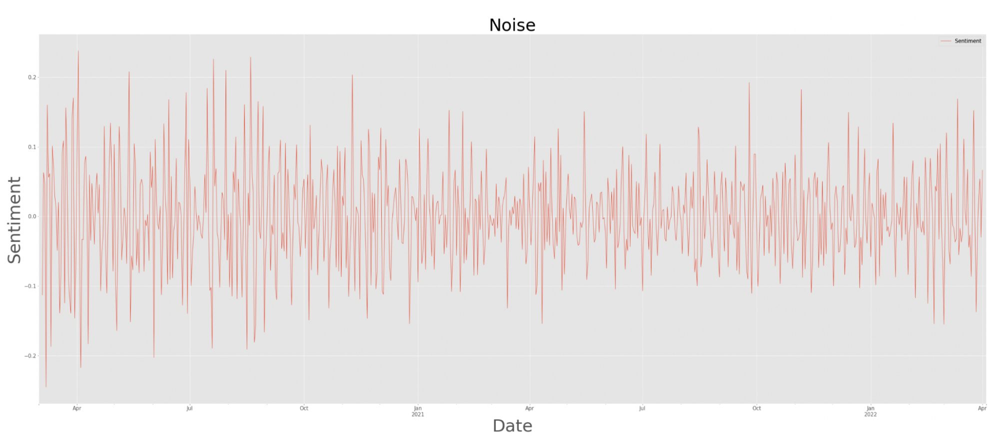
Date	Event	Color	Hypothesis	Result
January 25, 2021	First case of Gamma variant was detected in the U.S.	Blank	The detection of Gamma variants will cause a change of pattern towards negative sentiments.	There was no significant change in sentiments before and after the event. They were relatively positive.
January 28, 2021	First case of Beta variant was detected in the U.S	Cyan	The detection of Beta variant will cause a change of pattern towards negative sentiments.	There was no significant change in sentiments before and after the event. They were relatively positive.
May 10, 2021	The FDA expands the emergency use authorization of the Pfizer-BioNTech COVID-19 vaccine to include adolescents 12-15 years of age.	Green	The authorization of vaccines will cause a change of pattern towards positive sentiments.	The sentiment pattern changes and the sentiments are relatively less positive after the event.
July 7, 2021	The Delta variant becomes the dominant COVID-19 strain in the U.S.	Pink	The detection of the Delta variant will cause a change of pattern towards negative sentiments.	The sentiments are relatively positive after the event.
Aug 23, 2021	The FDA approves the first COVID-19 vaccine Comirnaty (Pfizer-BioNTech) for individuals 16 and older. (The EUA remains in effect for individuals 12 years of age and older and a third dose for immunocompromised	Blue	The authorization of vaccines will cause a change of pattern towards positive sentiments.	The sentiment pattern changes and sentiments are relatively more negative after the events.

	individuals 12 years of age and older).			
September 22, 2021	The FDA authorizes a booster dose of Pfizer-BioNTech COVID-19 Vaccine for aged 65 years and older, aged 18 through 64 at high risk of severe COVID-19, aged 18 through 64 who have institutional or occupational exposure to SARS-CoV-2.	Gray	The authorization of vaccines will cause a change of pattern towards positive sentiments.	The sentiment pattern changes and sentiments are relatively more negative after the events.
October 29, 2021	The FDA authorizes EUA for the Pfizer-BioNTech COVID-19 vaccine for children 5-11 years.	Magenta	The authorization of vaccines will cause a change of pattern towards positive sentiments.	The sentiments are relatively more positive after the event.
December 1, 2021	The U.S. announces first cases of the Omicron variant.	Brown	The detection of the Omicron variant will cause a change of pattern towards negative sentiments.	The sentiments are relatively more negative after the event.

Noise

The white noise is a part of the decomposition of time series. This graph represents the random variation of the sentiments around the mean sentiments. In the graph below the y-axis represents the variation of sentiments in the range of -0.2 to +0.2 values around the mean range of sentiments. The x-axis represents the timeline around which these variations can be observed.

The white noise in a time series analysis signifies randomness of the model and the model cannot accurately predict the future sentiments based on the historical data.



Decomposing time series helps us understand data in a structured manner. Instead of imagining a series as a value changing over time, this process helps to think of it as a distribution with a particular seasonality signal or a feature with a particular slope. This level of data understanding can be a key factor during feature engineering and modeling.

Discussion

Data Ingestion Limitations

Twitter stopped allowing misinformation about COVID-19 in March of 2020, making tweets' sentiments biased towards neutral or positive sentiments. Additionally, tweets without geolocation were filtered out - out of our 222,264 tweets, 55159 did not have any type of geolocation (location, place, or coordinates). We attempted to look at historical tweets (last 1000 tweets) from each user with no geolocation, and use the mode for the location, coordinates, and place for each of the tweets to get the most recurring location with the assumption that it would most accurately represent the user's location. However, after testing our geolocation model for 5,000 tweets, we were not able to retrieve any additional geolocation. Additionally, our tweets were filtered to only include tweets that explicitly stated were written in English.

There are also inherent limitations that arise due to our use of Twitter data. Alizadeh (2021) states that only about 22% of Americans use Twitter, and those who use it tend to be younger, more left-leaning, and more affluent. This means that Twitter users do not represent the general population. There is also an issue of the inconsistent ratio of Twitter use - some users tweet a lot, while others do not at all. According to the Pew Research Center, the most active 10% of users create about 80% of overall tweets by US adults (Wojcik et al., 2019). This leads us to believe that the tweets we are seeing are not representative of Twitter users, but instead, are the opinions of a small subset of users.

Due to our method of data ingestion, limitations arose in terms of sampling error. Our data ingestion process was a cross-section of the total subset of tweets regarding COVID-19. We collected ten tweets, per day, per phrase - however, the tweets collected were the last 10 tweets of each day. Additionally, after filtering out tweets with no geographical location, we did not have an equal number of tweets per day per phrase. Additionally, our tweets were not equally spread out by geographical location or search phrase, so some states and search phrases had a higher count of tweets than others.

Socioeconomic Status Sentiment Analysis Limitations

A limitation of the study was that we had more tweets from San Francisco than Detroit which has led to an imbalanced dataset. Therefore, further research is needed in this area with a larger, more balanced number of tweets. Our dataset had around 2,000 tweets from San Francisco and 290 tweets from Detroit. After removing irrelevant tweets, San Francisco had only 1280 tweets and Detroit 255 tweets. Michigan discontinued reporting COVID-19 vaccination data by race (Hicks, 2021). Therefore, we were unable to obtain stratified daily COVID-19 vaccination progress for all races from Detroit city.

State Vaccine Progress Limitations

A limitation with the state vaccine progress analysis is that the Kaggle dataset that was used to perform this analysis did not include vaccine progress data from December 20, 2020, when the FDA issued EUAs for Pfizer and Moderna vaccines. In addition to that, some dates did not have any vaccine progress data. Thus, if we wanted to look at a specific date, the data would not be available. A workaround for this was to locate a date closest to the desired date where the data was available. Another limitation is that there were instances where vaccine progress data was available for most states but not all states. For the states that were missing data, the percent change in vaccine progress rate was not calculated.

The reasoning behind analyzing vaccine progress one month after a specified event was specifically for the events relating to vaccine EUAs that were issued. We thought it would be appropriate to assess vaccine progress across the United States one month after an event has taken place as we expected there to be a lag in vaccine rollout. In addition to that, we realize that other events may have taken place or there may have been other factors that may have influenced people's decision to get vaccinated besides the event that was being analyzed.

Model Limitations

In an ideal situation, we would've been able to create a fully-custom model and architecture trained on our topic of interest. However, we were limited to building on top of existing models with varying sets of training data. The relative sentiment and topics of conversation for COVID vaccinations have fluctuated throughout the pandemic. A majority of the models that were used, including that on which the final set of prediction data was created using, were trained on tweets from 2020 till mid-2021. The additional year of conversations and topics could've had a large impact on the true sentiment of tweets, but unfortunately, we would not be able to capture this.

Nevertheless, the model performed relatively well and in line with conventional expectations regarding the COVID-19 vaccine. However, we weren't able to properly train our model on the total number of iterations as we would've liked to. This was the result of hardware limitations on the cloud computing platforms we used. We were only able to reasonably train models on 4 epochs before running out of dedicated GPU memory. This is compared to the 10-15 epochs that we would've liked to train each of our models on, especially **model_ctb_v2**.

Further Research

There is scope for further research by verifying the tweets and Twitter users. Further, a daily count of tweets and retweets on the vaccine can be used for further analysis. Other vaccines developed around the world like Covidshield, Covaxine, and Sputnik can be used as search parameters to analyze the sentiments of people towards those vaccines around the world.

Conclusion

The study aimed to analyze sentiments towards COVID-19 vaccines in the United States according to positive, neutral, and negative polarities. Overall, the majority of the tweets (53.90%) in the United States were positive, while 20.30% were neutral and 15.90% displayed negative sentiment.

The data was preprocessed using several NLP techniques, and a classifier model was successfully selected against a total set of 7 different models. Using the COVID tweet-BERT v2 model, we were able to develop an architecture that resulted in an 82% weighted F1 score.

Sentiment analysis in Red versus Blue states revealed that political affiliations may have little to no effect on the public's sentiment toward COVID vaccines. Additionally, sentiments towards COVID vaccines may not be affected by the socioeconomic status of the geographic location and a lower rate of COVID vaccination. Analysis of state-wide vaccination progress revealed that although there was generally a decline in positive sentiment regarding COVID vaccines as the FDA issued EUA approval of COVID vaccines, there was an increase in vaccination rates. It was also found that the pause in the use of a particular vaccine brand (J & J vaccine) did not negatively impact sentiment regarding COVID vaccines, nor did it slow down vaccination rates across the United States.

Lexicon-based Twitter sentiment analysis is a useful and simple approach for tracking public sentiment regarding COVID-19 vaccinations. High vaccination uptake is critical for ending the epidemic and identifying events that influence vaccine sentiment allows for improved planning and implementation of specific interventions.

Contribution Summary

Task (in chronological order)	Person
Parsing Tweets	Yena, Elizabeth
Codebook (for labeling tweets)	Elizabeth
Labeling of Tweets (Sentiment & Location)	Yena, Arjun, Elizabeth, Sunny, Kashmira
Inter-rater reliability of tweet labels	Elizabeth
Geographical Cleaning	Arjun
Tweet Cleaning	Sunny
Sentiment Analysis Model	Sunny
Red/Blue States Data Analysis	Yena
Socio-economic Status Data Analysis for SF and Detroit	Arjun
State Vaccine Progress Data Analysis	Elizabeth
Statistical Tests	Yena, Arjun, Elizabeth, Kashmira, Sunny
Visualizations for general trend on sentiment	Sunny, Kashmira
Powerpoint	Yena, Arjun, Elizabeth, Sunny, Kashmira
Final Report	Yena, Arjun, Elizabeth, Sunny, Kashmira

Appendix

Figure 1.1 Overview of codes

Overall Theme	Code	Description	Example
Positive	1	Tweet was considered positive if it mentioned research discoveries related to the COVID-19 vaccine trial; if a person is seeking information on how to volunteer for the clinical trial, or if the tweet reflected a positive attitude toward the COVID-19 vaccination.	"Don't worry, we will return to normal with vaccines worldwide. #NoNewNormal #ScienceWillWin
Negative	-1	Tweet was considered as negative if it mentioned vaccines as unnecessary, provided misinformation, conspiracy theories, or raised safety and effectiveness concerns about vaccines against coronavirus.	"So, the plan is still to wait for an untested vaccine that was produced in less than a year when it normally takes 10 years?"
Neutral	0	Tweet was considered neutral if the tweet's overall sentiment was neutral; if tweets provided information that did not pertain to emotional words or mentioned a COVID-19 vaccine clinical trial in general.	"Phase 3 clinical trial of investigational #vaccine for #COVID19 begins #Moderna"
Irrelevant	Flag for removal	Tweets were considered irrelevant if the tweets were not related to COVID-19 vaccine or if the context is not understood. Posts with pictures or URLs only without providing a context in the tweet were also considered as irrelevant.	💕 COURTESY POST🐾 🚨 🚨 LOCAL KITTEN NEEDS DONATIONS IN NEWBERRY, SC 🚨 🚨 PetCare Animal Hospital of Newberry has offered to remove her eye for $350.00, as well as spay and vaccinate her. Please call in donations to PetCare's @ 1 (803) 276-2498 Many THANKS! 😻 https://t.co/lYY2m7ZshB

Figure 1.2 Themes and sub-themes of tweets with positive sentiment

Code Theme	Code Subtheme	Description
Protection	Protect family from getting infected with COVID-19	Exhibiting desire to protect family and friends from getting infected
	Protect oneself	Exhibiting desire to protect oneself from getting infected
	Identify as part of a high-risk group	Exhibiting desire to protect oneself from getting infected due to identifying as higher risk for serious illness from COVID-19 (for example, those who are older adults, those with heart disease/diabetes/brain and nervous system conditions/cancer/blood disorders/weakened immune system/chronic kidney or liver disease/etc.)
Return to normal life	Want to return to normal familial and social life	Exhibiting desire to return to pre-COVID times, due to the impact COVID-19 has caused in contributing to one's decreased familial/social interaction
	Want to resume normal professional activities	Exhibiting desire to return to pre-COVID times, due to the impact COVID-19 has caused such as transitioning to in-person to remote work, furloughs, layoffs
Trust in recommendations by experts/knowledgeable resources	Recommended by health protection agencies	Motivated by reecommendations posed by the CDC, World Health Organization (WHO), Ministry of Public Health, etc.
	Recommended by health care workers	Motivated by recommendations made by physicians and pharmacists
	Recommendations from other entities	Sources of information include television, social media platforms, friends and family
	Consulting with the scientific community	Motivated by recommendations made by scientists and relying on scientific releases
Vaccine mandate	May become mandatory for jobs	Motivated to get vaccine due to job security
	May become mandatory for traveling abroad	Motivated to get vaccine due to desire to travel having undergone/experienced quarantine or isolation period
Confidence in vaccine safety	Vaccine will prevent the severity and complications of COVID-19	Displays confidence in vaccine safety in terms of serious or life-threatening side effects, general adverse reactions, and long-term safety
	Belief that vaccine will outweigh its side effects	There are more benefits to receiving a vaccination than to not and the vaccine will prevent transmission, symptomatic disease, and death

Figure 1.3 Themes and sub-themes of tweets with negative sentiment

Code Theme	Code Subtheme	Description
Safety and Effectiveness	Vaccine will produce side effects	Vaccines against coronavirus is harmful/will cause injuries/adverse side effects
	Fast-paced vaccine development	Concerns that the first COVID-19 vaccines have been developed at a rushed speed compared to the traditional timeframe for vaccine development, therefore questioning the safety and efficacy of vaccine
	Vaccine will not be effective	A potential COVID-19 vaccine will create only temporary/ineffective immunity; vaccine will not be potent enough to protect against infections; preference for natural immunity
Conspiracy theories	Profit from developing a COVID-19 vaccine	The production of a potential COVID-19 vaccine is motivated only by a quest for profit by large pharmacological companies/health care services/big corporations
	Other unusual/unique theories	Unique theories about the vaccine against COVID-19
Misinformation and falsehoods	Falsehoods	Unsupported/misleading statements made regarding COVID-19 vaccines that were not substantiated by evidence
	Vaccines are untested	Potential COVID-19 vaccines will not be adequately tested for safety; questioning the credibility of the vaccine developed
	Vaccines cause illnesses	Vaccine will cause autism or other illnesses
	Vaccination is unnecessary	Vaccination is not necessary to protect against COVID-19
Mistrust	Mistrust of scientists and vaccine advocates	Mistrust of science, scientists, and vaccine advocates involved in financing, development, and distribution of potential COVID-19 vaccines
	Mistrust in the government	Mistrust in the governmental institutions pushing for COVID-19 vaccination
Lack of intent to get the COVID-19 vaccine	Individual unwillingness to get a COVID-19 vaccine	No intention of getting vaccination/expressing future noncompliance or discouraging/demotivating others from getting vaccinated
	Discouraging others to get vaccinated against coronavirus	
Freedom of choice	Violation of individual rights	Civil liberties will be violated by taking away an individual's right to choose in getting a potential COVID-19 vaccination
	Totalitarianism	Accusations of totalitarianism were made, which included warnings that citizens would be forced to comply with a mandatory vaccination program
Religious beliefs	Religious concerns	Religious concerns about the ingredients used in developing the vaccine or about the vaccines in general

Figure 2.1 Word cloud of tweets stratified by sentiment

Figure 2.2 Themes and sub-themes of tweets with negative sentiment

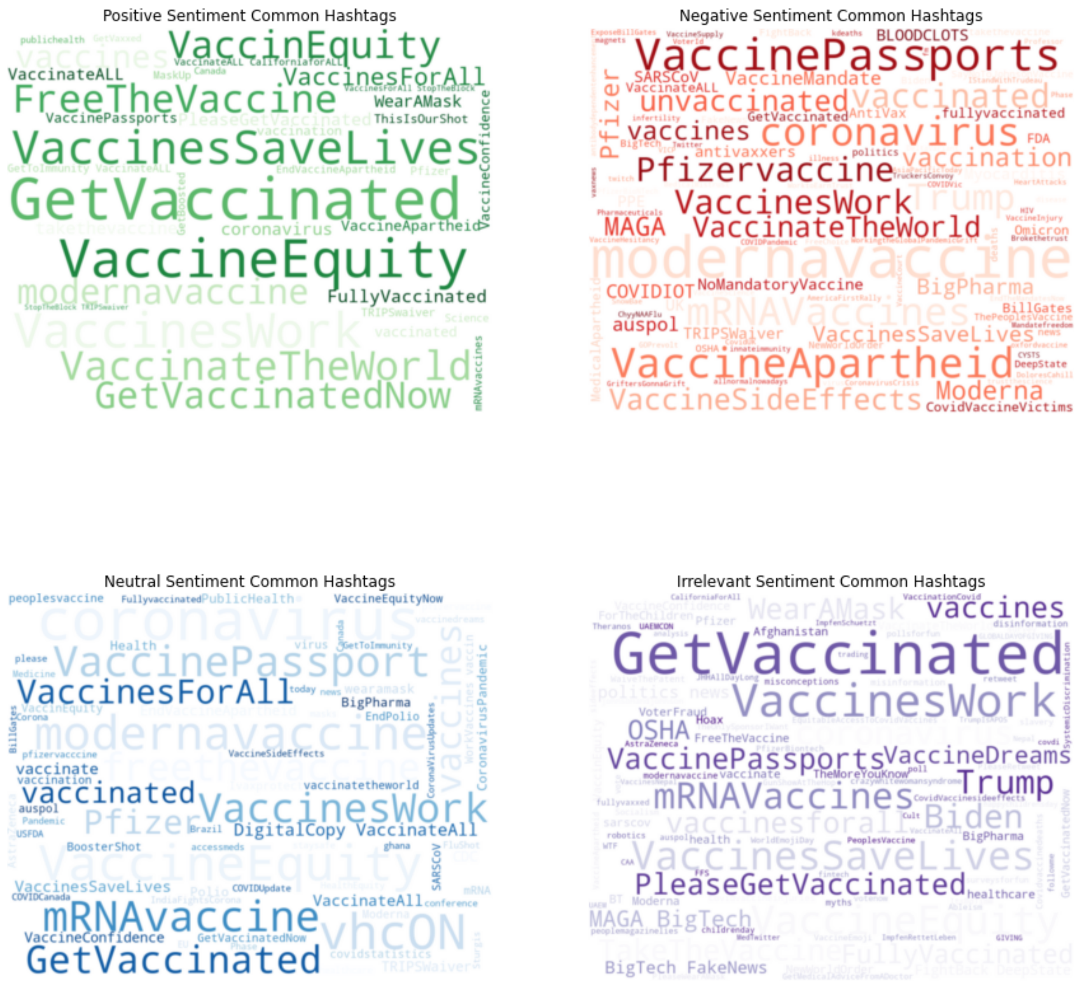


Figure 2.3 Model Architectures

```
==========================================================================================================================================
Layer (type:depth-idx)                     Param #(model_ctb_v2)        Param #(model_bcb_u)         Param #(model_vct)           Param #(model_msc)
==========================================================================================================================================
BertForSequenceClassification              --                           --                           --                           --
├─BertModel: 1-1                           --                           --                           --                           --
│    └─BertEmbeddings: 2-1                 --                           --                           --                           --
│    │    └─Embedding: 3-1                 31,254,528                   49,152,768                   49,152,768                   192,001,536
│    │    └─Embedding: 3-2                 524,288                      99,840                       99,840                       394,752
│    │    └─Embedding: 3-3                 2,048                        768                          768                          768
│    │    └─LayerNorm: 3-4                 2,048                        1,536                        1,536                        1,536
│    │    └─Dropout: 3-5                   --                           --                           --                           --
│    └─BertEncoder: 2-2                    --                           --                           --                           --
│    │    └─ModuleList: 3-6                302,309,376                  85,054,464                   85,054,464                   85,054,464
│    └─BertPooler: 2-3                     --                           --                           --                           --
│    │    └─Linear: 3-7                    1,049,600                    590,592                      590,592                      590,592
│    │    └─Tanh: 3-8                      --                           --                           --                           --
├─Dropout: 1-2                             --                           --                           --                           --
├─Linear: 1-3                              4,100                        3,076                        3,076                        3,076
==========================================================================================================================================
Total params:
        model_ctb_v2          335,145,988
        model_bcb_u                   134,903,044
        model_vct                     134,903,044
        model_msc                     278,046,724
Trainable params:
        model_ctb_v2          335,145,988
        model_bcb_u                   134,903,044
        model_vct                     134,903,044
        model_msc                     278,046,724
Non-trainable params:
        model_ctb_v2          0
        model_bcb_u                   0
        model_vct                     0
        model_msc                     0
==========================================================================================================================================
```

Figure 2.4 Confusion Matrix (Training Set)

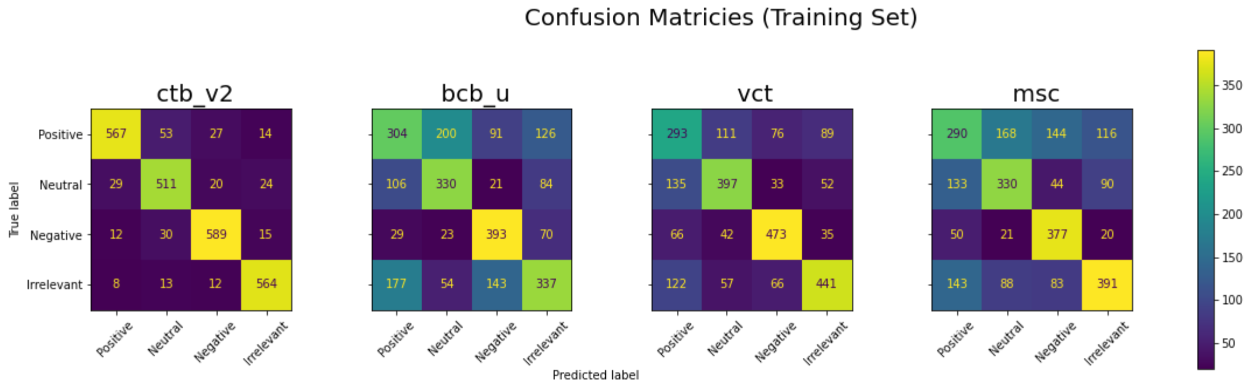


Figure 2.5 Confusion Matrix (Validation Set)

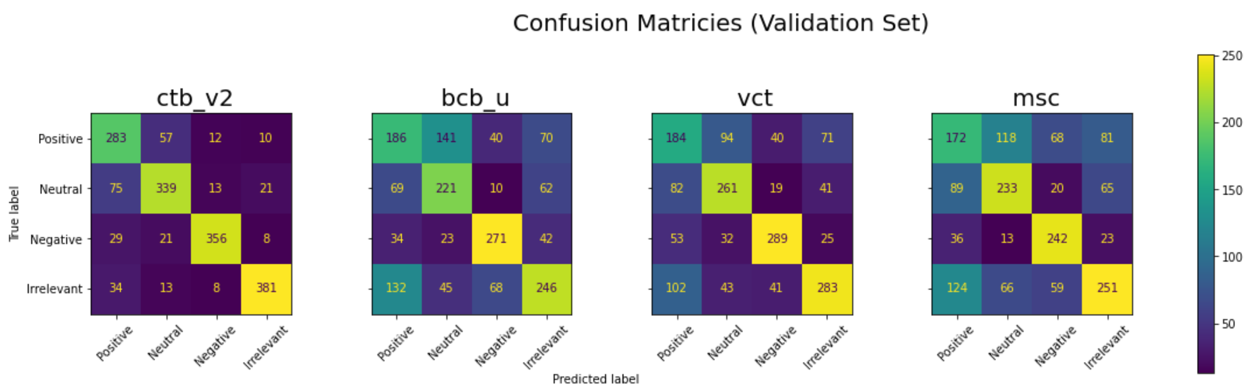


Table 2.1 BERT Models

Model Name	Model Source	Model Description
model_ctb_v1	digitalepidemiologylab/covid-twitter-bert	This model was trained on 160M tweets collected between January 12 and April 16, 2020 containing at least one of the keywords "wuhan", "ncov", "coronavirus", "covid", or "sars-cov-2".
model_ctb_v2	digitalepidemiologylab/covid-twitter-bert-v2	97M unique tweets (1.2B training examples) collected between January 12 and July 5, 2020 containing at least one of the keywords "wuhan", "ncov", "coronavirus", "covid", or "sars-cov-2".
model_bcb_c	vinai/bertweet-covid19-base-cased	23M COVID-19 English Tweets (cased)
model_bcb_u	vinai/bertweet-covid19-base-uncased	23M COVID-19 English Tweets (uncased)
model_bb	vinai/bertweet-base	850M English Tweets (cased)
model_vct	ans/vaccinating-covid-tweets	850M General English Tweets (Jan 2012 to Aug 2019), with 23M COVID-19 English Tweets, pre-trained on COVID-19/vaccined related tweets using a masked language modeling (MLM) objective starting from BERTweet
model_msc	clampert/multilingual-sentiment-covid19	trained on a large-scale (18437530 examples) dataset of multi-lingual tweets that was collected between March 2020 and November 2021 using Twitter's Streaming API with varying COVID19-related keywords

Table 2.2 Weighted F1 Score Across Models and Training Epochs

Model	Epoch 1	Epoch 2	Epoch 3	Epoch 4
model_ctb_v2	0.67	0.77	0.81	0.82
model_bcb_u	0.18	0.38	0.53	0.53
model_vct	0.11	0.32	0.6	0.6
model_msc	0.13	0.37	0.53	0.53

*Green box represents model chosen for final predictions

Figure 3.1 Count and Sentiment by Presidential Election Political Affiliation

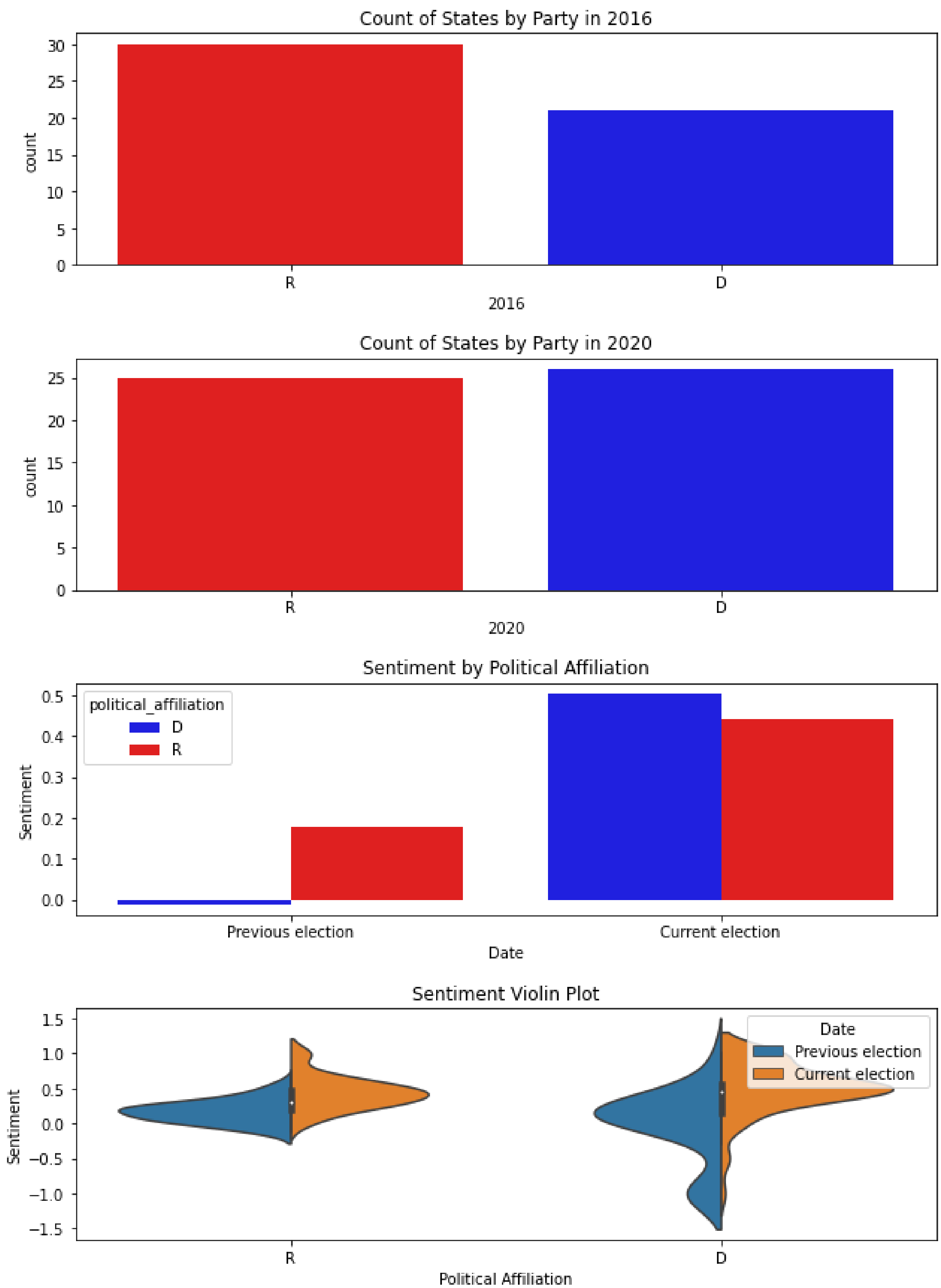


Figure 3.2 Count and Sentiment by Governor Political Affiliation

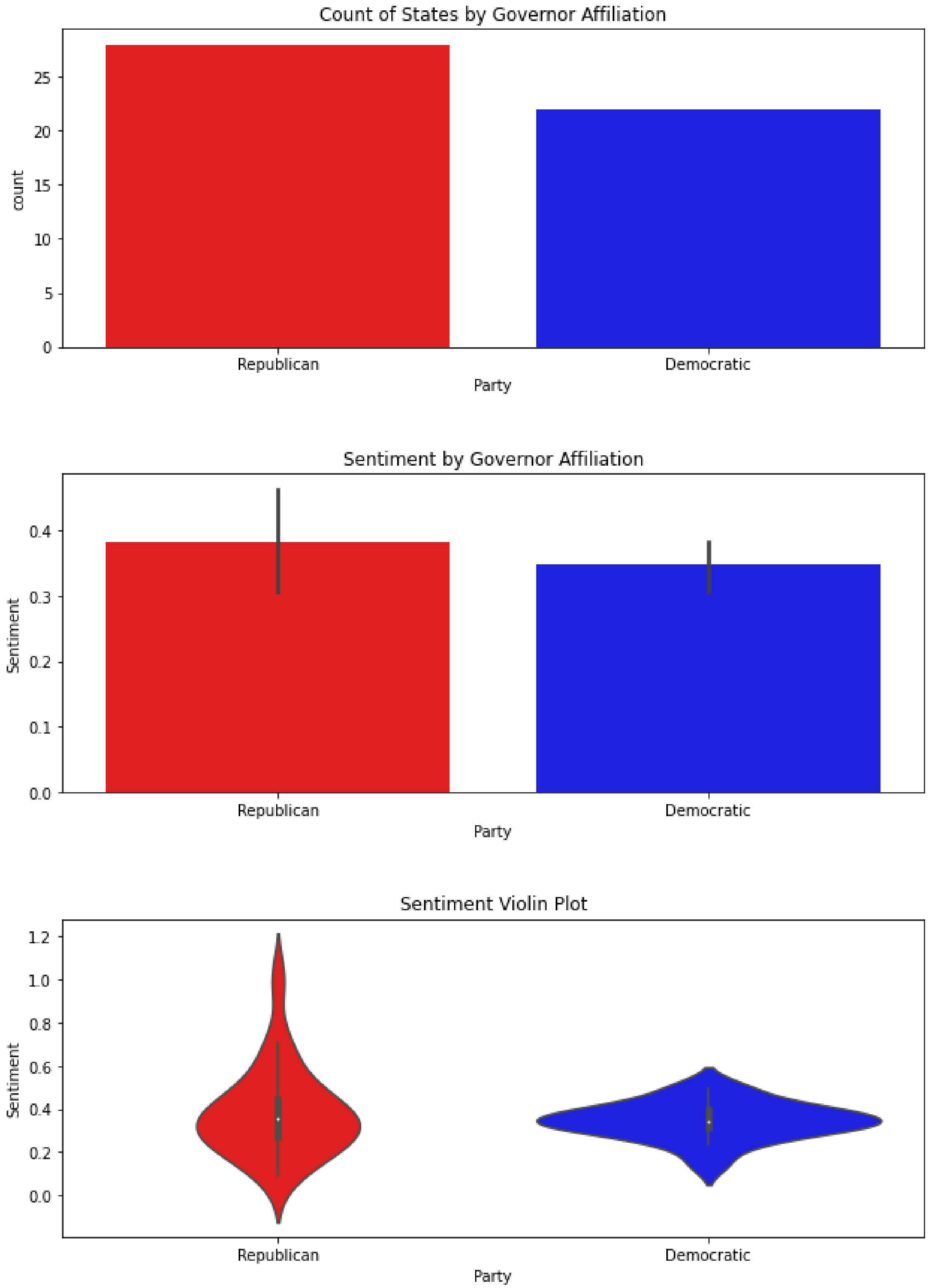


Figure 3.3 Count and Sentiment by State Legislature Political Affiliation

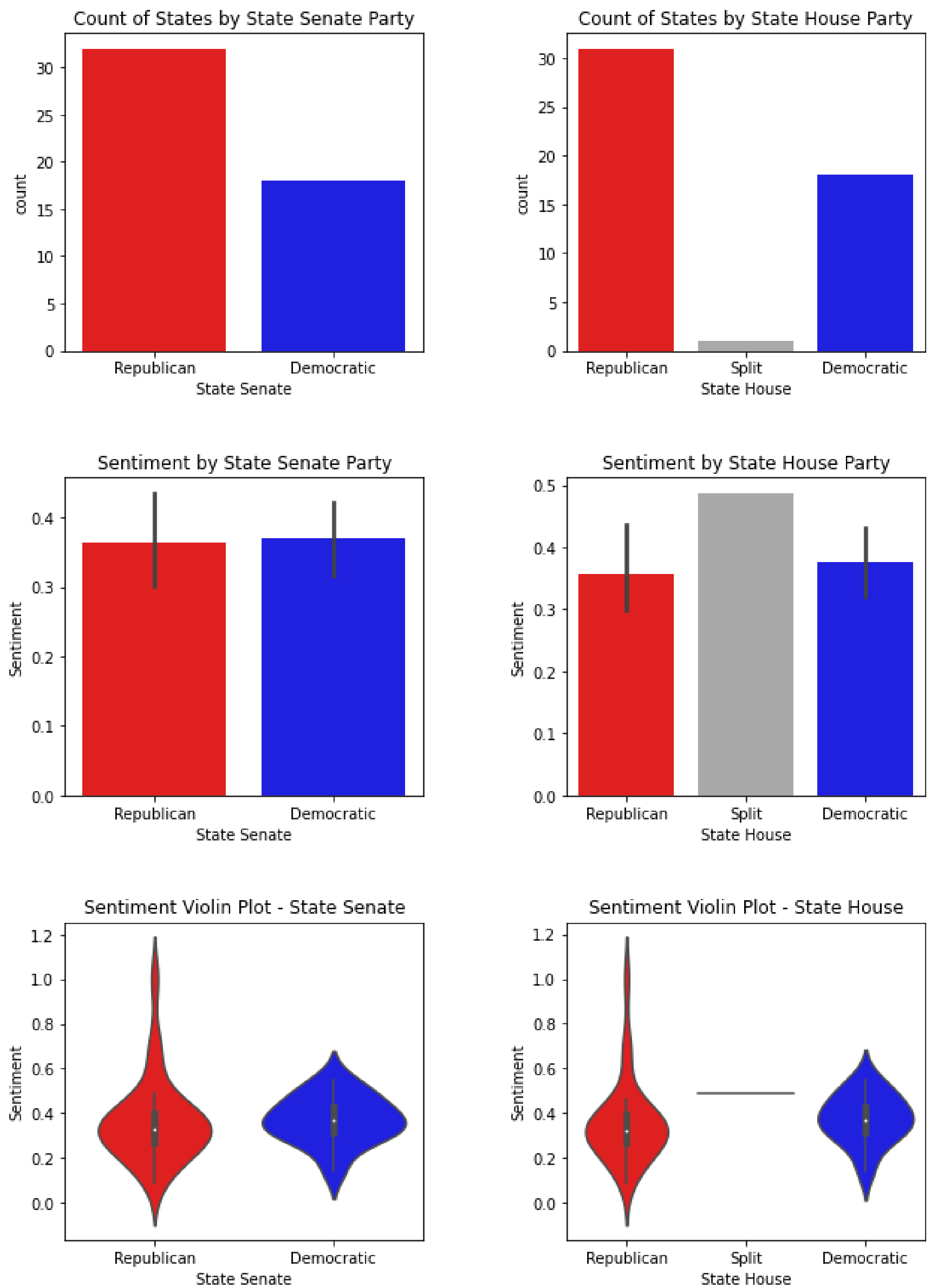


Figure 4.1: Sub themes for Social Vulnerability Index calculation

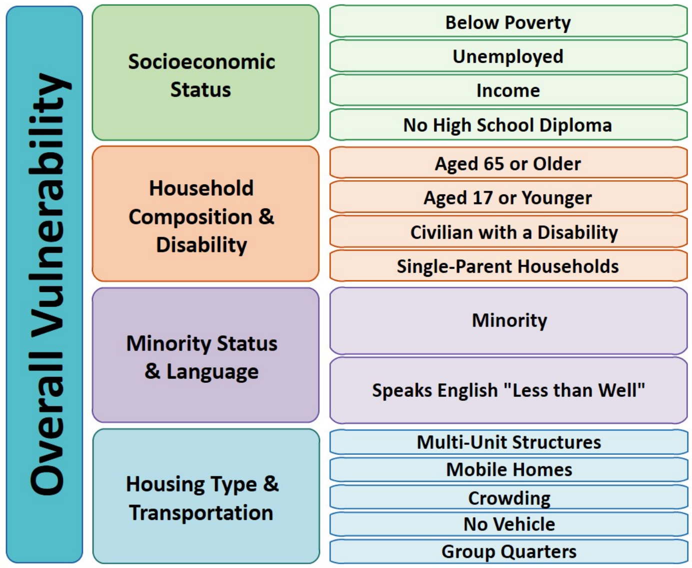
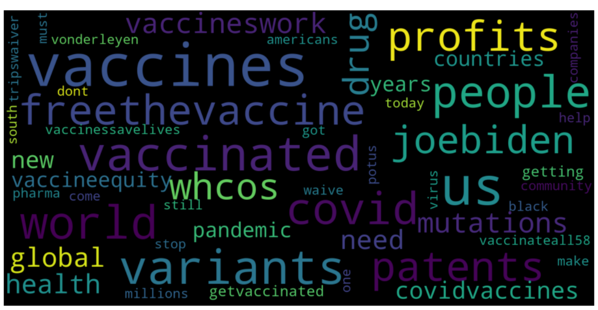
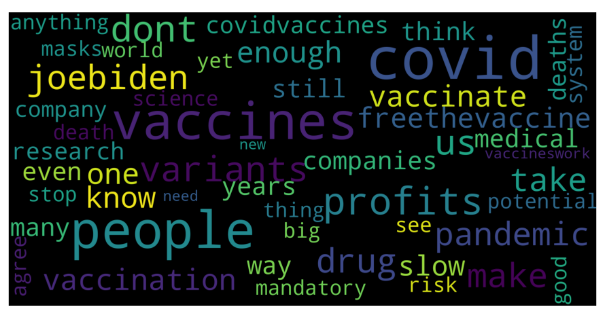


Figure 4.2: Word Cloud for positive and negative tweets from San Francisco and Detroit

San Francisco, California:

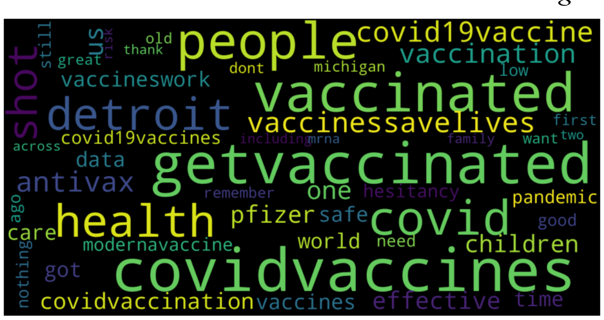
Positive Tweets *Negative Tweets*





Detroit, Michigan

Positive Tweets *Negative Tweets*



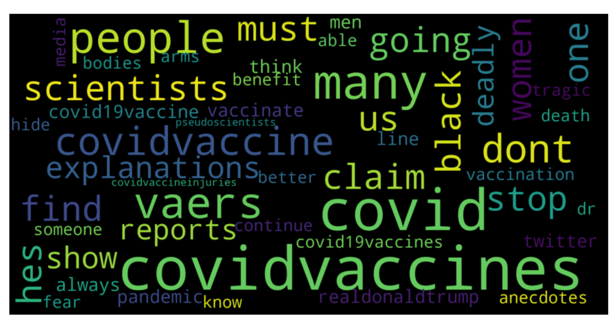


Figure 5.1 Date Reference for Fig 5.2-5.4

Event Category	Events	Event Details	Day Before Event Start Date	Event Start Date	1 month after event*
COVID-19 variants	Gamma Variant	First case of the Gamma variant in the U.S.	1/24/2021	1/25/2021	2/25/2021
	Beta Variant	First case of Beta variant in the U.S.	1/27/2021	1/28/2021	2/28/2021
	Delta Variant	The Delta variant becomes the dominant COVID-19 strain in the U.S.	7/6/2021	7/7/2021	8/7/2021
	Omicron Variant	The U.S. announces first cases of Omicron variant.	11/30/2021	12/1/2021	1/4/2022
COVID-19 vaccine availability by age bracket	EUA for Pfizer vaccine for 12-15yo	The FDA expands the emergency use authorization of Pfizer-BioNTech COVID-19 vaccine to include adolescents 12-15 years of age.	5/9/2021	5/10/2021	6/10/2021
	Comirnaty for 16yo+ FDA approval	The FDA approves first COVID-19 vaccine Comirnaty (Pfizer-BioNTech) for individuals 16 and older. (The EUA remains in effect for individuals 12 years of age and older and for a third dose for immunocompromised individuals 12 years of age and older).	8/22/2021	8/23/2021	9/23/2021
	Pfizer Booster for 65yo+ & 18-64 at high risk	The FDA authorizes booster dose of Pfizer-BioNTech COVID-19 Vaccine for aged 65 years and older, aged 18 through 64 at high risk of severe COVID-19, aged 18 through 64 who have institutional or occupational exposure to SARS-CoV-2.	9/21/2021	9/22/2021	10/22/2021
	EUA for Pfizer vaccine for 5-11yo	The FDA authorizes EUA for Pfizer-BioNTech COVID-19 vaccine for children 5-11 years	10/28/2021	10/29/2021	11/29/2021
Pause in Use	Johnson & Johnson vaccine pause	CDC and FDA recommended a pause in use of the Janssen (Johnson & Johnson) COVID-19 vaccine in the US out of an abundance of caution. A CDC Health Alert Network (HAN) was issued with recommendations	4/12/2021	4/13/2021	5/13/2021

*Note: The date provided may not be exactly 1 month after the event as data may have not been available for that particular date.

Figure 5.2 State Vaccine Progress before and after initial cases of COVID-19 variants

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
Arkansas	Gamma Variant	1.08%	6.12%	5.94%	12.77%	466.67%	114.98%
	Beta Variant	1.36%	7.54%	6.94%	13.90%	454.41%	100.29%
	Delta Variant	34.57%	37.43%	42.47%	48.84%	8.27%	15.00%
	Omicron Variant	49.32%	51.40%	60.46%	62.98%	4.22%	4.17%
Mississippi	Gamma Variant	0.53%	5.48%	5.43%	12.40%	933.96%	128.36%
	Beta Variant	0.59%	7.08%	5.80%	13.91%	1100.00%	139.83%
	Delta Variant	29.89%	35.06%	36.30%	41.61%	17.30%	14.63%
	Omicron Variant	46.90%	48.69%	53.71%	56.17%	3.82%	4.58%
Alabama	Gamma Variant	0.61%	5.48%	4.36%	12.54%	798.36%	187.61%
	Beta Variant	0.69%	6.39%	4.74%	13.25%	826.09%	179.54%
	Delta Variant	32.99%	34.83%	40.31%	45.02%	5.58%	11.68%
	Omicron Variant	46.18%	47.84%	56.54%	58.84%	3.59%	4.07%
Louisiana	Gamma Variant	0.91%	7.01%	6.26%	12.94%	670.33%	106.71%
	Beta Variant	1.01%	8.06%	6.69%	14.29%	698.02%	113.60%
	Delta Variant	35.44%	37.44%	38.86%	44.43%	5.64%	14.33%
	Omicron Variant	48.89%	50.44%	55.58%	57.70%	3.17%	3.81%
Oklahoma	Gamma Variant	1.04%	8.20%	7.06%	15.94%	688.46%	125.78%
	Beta Variant	1.18%	9.93%	7.49%	17.40%	741.53%	132.31%
	Delta Variant	38.77%	40.77%	45.20%	49.21%	5.16%	8.87%
	Omicron Variant	51.53%	53.68%	62.83%	66.40%	4.17%	5.68%
Florida	Gamma Variant	NA	7.30%	NA	13.63%	NA	NA
	Beta Variant	0.82%	8.20%	6.76%	14.73%	900.00%	117.90%
	Delta Variant	46.48%	49.50%	54.29%	59.38%	6.50%	9.38%
	Omicron Variant	61.39%	63.55%	71.81%	74.86%	3.52%	4.25%
Puerto Rico	Gamma Variant	1.22%	5.39%	4.51%	10.58%	341.80%	134.59%
	Beta Variant	1.33%	6.09%	4.89%	11.12%	357.89%	127.40%
	Delta Variant	56.40%	60.48%	65.54%	69.45%	7.23%	5.97%
	Omicron Variant	74.08%	77.20%	84.86%	89.32%	4.21%	5.26%
Texas	Gamma Variant	0.89%	5.51%	5.18%	11.40%	519.10%	120.08%
	Beta Variant	1.12%	6.22%	5.69%	12.18%	455.36%	114.06%
	Delta Variant	41.70%	44.44%	48.63%	52.95%	6.57%	8.88%
	Omicron Variant	54.70%	57.17%	64.09%	67.16%	4.52%	4.79%
Missouri	Gamma Variant	1.19%	5.93%	3.96%	12.14%	398.32%	206.57%
	Beta Variant	1.30%	6.73%	4.51%	13.15%	417.69%	191.57%
	Delta Variant	39.43%	41.96%	45.42%	49.77%	6.42%	9.58%
	Omicron Variant	50.93%	53.10%	59.72%	62.48%	4.26%	4.62%
South Carolina	Gamma Variant	0.85%	5.83%	4.53%	13.53%	585.88%	198.68%
	Beta Variant	0.97%	6.71%	5.35%	14.33%	591.75%	167.85%
	Delta Variant	39.13%	41.04%	44.67%	47.81%	4.88%	7.03%
	Omicron Variant	51.40%	53.32%	60.27%	63.12%	3.74%	4.73%
Tennessee	Gamma Variant	1.41%	5.81%	5.19%	11.64%	312.06%	124.28%
	Beta Variant	1.65%	6.71%	5.48%	12.87%	306.67%	134.85%
	Delta Variant	37.67%	39.52%	42.63%	45.78%	4.91%	7.39%
	Omicron Variant	49.55%	51.53%	56.84%	58.99%	4.00%	3.78%
Georgia	Gamma Variant	0.56%	6.36%	5.03%	11.48%	1035.71%	128.23%
	Beta Variant	0.71%	7.00%	5.87%	11.87%	885.92%	102.21%
	Delta Variant	36.98%	39.06%	43.84%	47.21%	5.62%	7.69%
	Omicron Variant	NA	51.23%	NA	61.48%	NA	NA

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
Nevada	Gamma Variant	0.72%	6.25%	4.20%	13.59%	768.06%	223.57%
	Beta Variant	0.77%	7.33%	4.85%	14.54%	851.95%	199.79%
	Delta Variant	42.50%	45.05%	50.39%	54.81%	6.00%	8.77%
	Omicron Variant	54.50%	56.66%	66.44%	69.93%	3.96%	5.25%
Arizona	Gamma Variant	0.80%	5.94%	4.56%	15.61%	642.50%	242.32%
	Beta Variant	0.95%	7.22%	5.35%	16.77%	660.00%	213.46%
	Delta Variant	43.58%	45.71%	50.93%	53.94%	4.89%	5.91%
	Omicron Variant	54.67%	57.25%	64.30%	67.65%	4.72%	5.21%
North Carolina	Gamma Variant	0.80%	7.36%	5.26%	13.42%	820.00%	155.13%
	Beta Variant	0.84%	8.19%	5.99%	15.01%	875.00%	150.58%
	Delta Variant	42.18%	44.18%	49.03%	52.22%	4.74%	6.51%
	Omicron Variant	54.21%	56.94%	70.28%	77.09%	5.04%	9.69%
Kansas	Gamma Variant	0.78%	6.11%	4.37%	13.53%	683.33%	209.61%
	Beta Variant	0.91%	7.19%	4.83%	14.77%	690.11%	205.80%
	Delta Variant	42.33%	45.73%	49.55%	54.34%	8.03%	9.67%
	Omicron Variant	54.65%	57.21%	66.22%	69.66%	4.68%	5.19%
Wyoming	Gamma Variant	0.85%	7.99%	6.20%	16.11%	840.00%	159.84%
	Beta Variant	0.94%	9.39%	6.56%	17.25%	898.94%	162.96%
	Delta Variant	35.36%	36.97%	39.92%	42.34%	4.55%	6.06%
	Omicron Variant	45.57%	47.69%	53.65%	56.08%	4.65%	4.53%
Utah	Gamma Variant	0.65%	5.26%	5.47%	11.79%	709.23%	115.54%
	Beta Variant	0.94%	5.55%	6.27%	12.22%	490.43%	94.90%
	Delta Variant	37.48%	45.43%	48.87%	53.10%	21.21%	8.66%
	Omicron Variant	55.30%	58.98%	63.89%	67.59%	6.65%	5.79%
New York	Gamma Variant	0.91%	6.64%	6.03%	12.76%	629.67%	111.61%
	Beta Variant	1.09%	7.35%	6.61%	13.83%	574.31%	109.23%
	Delta Variant	54.73%	57.69%	60.56%	64.06%	5.41%	5.78%
	Omicron Variant	68.50%	72.04%	78.24%	84.61%	5.17%	8.14%
California	Gamma Variant	0.90%	5.72%	4.65%	14.52%	535.56%	212.26%
	Beta Variant	1.06%	6.60%	5.46%	15.54%	522.64%	184.62%
	Delta Variant	50.49%	53.49%	62.05%	65.80%	5.94%	6.04%
	Omicron Variant	63.17%	NA	78.83%	NA	NA	NA
Idaho	Gamma Variant	0.79%	6.21%	4.05%	13.38%	686.08%	230.37%
	Beta Variant	0.86%	7.17%	4.52%	14.37%	733.72%	217.92%
	Delta Variant	36.37%	37.68%	39.83%	41.75%	3.60%	4.82%
	Omicron Variant	45.22%	46.32%	50.73%	52.28%	2.43%	3.06%
Rhode Island	Gamma Variant	1.25%	6.25%	5.12%	15.19%	400.00%	196.68%
	Beta Variant	1.55%	6.98%	5.65%	17.30%	350.32%	206.19%
	Delta Variant	59.44%	62.04%	65.04%	68.31%	4.37%	5.03%
	Omicron Variant	72.58%	76.76%	83.20%	89.63%	5.76%	7.73%
New Jersey	Gamma Variant	0.73%	7.11%	5.28%	14.75%	873.97%	179.36%
	Beta Variant	0.90%	8.13%	5.97%	16.03%	803.33%	168.51%
	Delta Variant	55.79%	58.96%	63.36%	66.93%	5.68%	5.63%
	Omicron Variant	67.82%	70.72%	79.13%	84.19%	4.28%	6.39%
Nebraska	Gamma Variant	1.15%	7.05%	5.68%	14.53%	513.04%	155.81%
	Beta Variant	1.27%	8.21%	5.90%	16.02%	546.46%	171.53%
	Delta Variant	48.01%	49.93%	51.99%	54.83%	4.00%	5.46%
	Omicron Variant	57.41%	60.07%	64.13%	66.67%	4.63%	3.96%

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
Virginia	Gamma Variant	0.60%	6.92%	4.61%	14.53%	1053.33%	215.18%
	Beta Variant	0.93%	8.45%	6.14%	15.81%	808.60%	157.49%
	Delta Variant	52.47%	55.01%	59.56%	62.54%	4.84%	5.00%
	Omicron Variant	64.94%	68.23%	75.58%	79.50%	5.07%	5.19%
Pennsylvania	Gamma Variant	1.03%	5.62%	4.78%	13.80%	445.63%	188.70%
	Beta Variant	1.22%	6.51%	5.42%	14.93%	433.61%	175.46%
	Delta Variant	50.21%	52.96%	63.28%	66.45%	5.48%	5.01%
	Omicron Variant	NA	64.09%	NA	78.77%	NA	NA
New Mexico	Gamma Variant	1.66%	10.23%	7.76%	20.33%	516.27%	161.98%
	Beta Variant	1.83%	11.70%	8.40%	22.30%	539.34%	165.48%
	Delta Variant	55.20%	57.71%	63.39%	66.45%	4.55%	4.83%
	Omicron Variant	63.52%	66.39%	76.57%	80.95%	4.52%	5.72%
North Dakota	Gamma Variant	2.05%	8.71%	7.37%	17.01%	324.88%	130.80%
	Beta Variant	2.22%	10.23%	7.70%	19.30%	360.81%	150.65%
	Delta Variant	39.12%	40.37%	44.14%	46.06%	3.20%	4.35%
	Omicron Variant	48.76%	52.76%	57.95%	62.48%	8.20%	7.82%
South Dakota	Gamma Variant	1.95%	9.30%	7.13%	18.93%	376.92%	165.50%
	Beta Variant	2.47%	10.62%	7.35%	20.80%	329.96%	182.99%
	Delta Variant	45.65%	47.39%	50.79%	53.64%	3.81%	5.61%
	Omicron Variant	54.64%	57.34%	66.99%	71.32%	4.94%	6.46%
Colorado	Gamma Variant	1.34%	7.40%	6.17%	14.93%	452.24%	141.98%
	Beta Variant	1.57%	8.02%	6.63%	15.65%	410.83%	136.05%
	Delta Variant	52.41%	54.91%	58.35%	61.04%	4.77%	4.61%
	Omicron Variant	63.25%	66.45%	71.48%	74.88%	5.06%	4.76%
Wisconsin	Gamma Variant	0.78%	7.09%	4.48%	15.31%	808.97%	241.74%
	Beta Variant	0.90%	8.59%	4.83%	16.46%	854.44%	240.79%
	Delta Variant	50.34%	52.15%	54.04%	56.34%	3.60%	4.26%
	Omicron Variant	59.50%	62.15%	65.82%	68.50%	4.45%	4.07%
Montana	Gamma Variant	1.32%	7.31%	5.85%	16.30%	453.79%	178.63%
	Beta Variant	1.44%	8.68%	6.15%	17.66%	502.78%	187.15%
	Delta Variant	43.21%	44.63%	48.02%	50.07%	3.29%	4.27%
	Omicron Variant	51.86%	54.11%	59.78%	62.19%	4.34%	4.03%
Delaware	Gamma Variant	1.18%	5.64%	5.40%	13.90%	377.97%	157.41%
	Beta Variant	1.27%	7.12%	6.67%	14.80%	460.63%	121.89%
	Delta Variant	50.68%	53.21%	58.66%	61.50%	4.99%	4.84%
	Omicron Variant	61.60%	64.41%	73.07%	77.11%	4.56%	5.53%
Alaska	Gamma Variant	2.21%	12.33%	10.65%	21.71%	457.92%	103.85%
	Beta Variant	2.47%	13.34%	11.40%	22.79%	440.08%	99.91%
	Delta Variant	44.15%	45.87%	50.06%	52.15%	3.90%	4.17%
	Omicron Variant	54.23%	56.46%	62.80%	65.21%	4.11%	3.84%
Indiana	Gamma Variant	1.16%	7.39%	5.47%	14.05%	537.07%	156.86%
	Beta Variant	1.36%	8.44%	6.43%	14.92%	520.59%	132.04%
	Delta Variant	42.71%	44.60%	45.51%	47.92%	4.43%	5.30%
	Omicron Variant	50.58%	52.07%	55.67%	58.03%	2.95%	4.24%
Maryland	Gamma Variant	0.62%	6.86%	4.97%	13.65%	1006.45%	174.65%
	Beta Variant	0.80%	8.07%	5.63%	14.63%	908.75%	159.86%
	Delta Variant	56.62%	59.39%	62.40%	65.58%	4.89%	5.10%
	Omicron Variant	67.56%	70.58%	76.94%	80.83%	4.47%	5.06%

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
District of Columbia	Gamma Variant	1.85%	4.79%	6.98%	10.60%	158.92%	51.86%
	Beta Variant	2.05%	5.68%	7.59%	11.86%	177.07%	56.26%
	Delta Variant	52.98%	55.48%	61.82%	64.78%	4.72%	4.79%
	Omicron Variant	64.52%	67.84%	81.60%	89.17%	5.15%	9.28%
Iowa	Gamma Variant	0.98%	5.01%	5.00%	15.01%	411.22%	200.20%
	Beta Variant	1.14%	5.66%	5.24%	16.95%	396.49%	223.47%
	Delta Variant	48.39%	50.01%	51.67%	54.03%	3.35%	4.57%
	Omicron Variant	56.73%	59.15%	62.62%	65.15%	4.27%	4.04%
Minnesota	Gamma Variant	1.20%	6.93%	4.48%	14.72%	477.50%	228.57%
	Beta Variant	1.34%	8.08%	5.30%	16.70%	502.99%	215.09%
	Delta Variant	52.32%	54.17%	57.32%	59.61%	3.54%	4.00%
	Omicron Variant	62.48%	65.65%	69.03%	71.70%	5.07%	3.87%
Connecticut	Gamma Variant	1.14%	8.04%	7.72%	17.59%	605.26%	127.85%
	Beta Variant	1.46%	8.45%	8.49%	19.37%	478.77%	128.15%
	Delta Variant	61.26%	63.77%	67.51%	70.73%	4.10%	4.77%
	Omicron Variant	72.00%	74.90%	84.09%	89.34%	4.03%	6.24%
Oregon	Gamma Variant	0.96%	7.17%	5.80%	13.71%	646.88%	136.38%
	Beta Variant	1.14%	8.37%	6.55%	14.99%	634.21%	128.85%
	Delta Variant	54.29%	56.36%	59.17%	61.34%	3.81%	3.67%
	Omicron Variant	64.01%	66.54%	71.46%	74.15%	3.95%	3.76%
Hawaii	Gamma Variant	1.19%	8.31%	4.80%	16.47%	598.32%	243.12%
	Beta Variant	1.31%	9.67%	5.59%	17.52%	638.17%	213.42%
	Delta Variant	52.21%	53.85%	70.14%	71.99%	3.14%	2.64%
	Omicron Variant	61.01%	NA	82.09%	NA	NA	NA
Michigan	Gamma Variant	1.11%	7.53%	5.74%	13.66%	578.38%	137.98%
	Beta Variant	1.31%	8.44%	6.30%	14.93%	544.27%	136.98%
	Delta Variant	47.35%	49.17%	51.62%	53.65%	3.84%	3.93%
	Omicron Variant	54.59%	56.98%	61.23%	63.70%	4.38%	4.03%
Maine	Gamma Variant	1.42%	7.19%	5.90%	15.77%	406.34%	167.29%
	Beta Variant	1.61%	8.09%	6.41%	16.94%	402.48%	164.27%
	Delta Variant	61.94%	64.22%	66.73%	69.05%	3.68%	3.48%
	Omicron Variant	72.30%	76.05%	82.02%	86.28%	5.19%	5.19%
Massachusetts	Gamma Variant	0.92%	6.16%	5.17%	16.45%	569.57%	218.18%
	Beta Variant	1.06%	7.20%	5.67%	17.71%	579.25%	212.35%
	Delta Variant	62.21%	64.33%	70.86%	73.29%	3.41%	3.43%
	Omicron Variant	71.18%	74.82%	85.78%	91.17%	5.11%	6.28%
West Virginia	Gamma Variant	2.18%	10.10%	9.14%	16.45%	363.30%	79.98%
	Beta Variant	2.52%	11.44%	9.47%	17.69%	353.97%	86.80%
	Delta Variant	37.47%	39.17%	43.90%	46.25%	4.54%	5.35%
	Omicron Variant	41.54%	55.22%	53.92%	62.16%	32.93%	15.28%
Vermont	Gamma Variant	1.58%	7.72%	6.72%	15.44%	388.61%	129.76%
	Beta Variant	1.73%	8.68%	7.13%	16.76%	401.73%	135.06%
	Delta Variant	NA	NA	NA	NA	NA	NA
	Omicron Variant	72.88%	77.75%	85.45%	89.75%	6.68%	5.03%
Washington	Gamma Variant	0.90%	6.39%	4.86%	13.97%	610.00%	187.45%
	Beta Variant	1.07%	7.65%	5.80%	14.90%	614.95%	156.90%
	Delta Variant	55.24%	58.16%	61.84%	64.74%	5.29%	4.69%
	Omicron Variant	65.04%	68.09%	72.91%	75.93%	4.69%	4.14%

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
New Hampshire	Gamma Variant	0.98%	6.74%	5.88%	15.44%	587.76%	162.59%
	Beta Variant	1.34%	7.35%	6.24%	17.17%	448.51%	175.16%
	Delta Variant	56.74%	58.59%	62.95%	65.29%	3.26%	3.72%
	Omicron Variant	64.78%	NA	87.60%	NA	NA	NA
Kentucky	Gamma Variant	0.66%	6.74%	6.01%	14.15%	921.21%	135.44%
	Beta Variant	0.81%	7.81%	6.39%	15.59%	864.20%	143.97%
	Delta Variant	43.97%	46.18%	49.87%	53.39%	5.03%	7.06%
	Omicron Variant	52.13%	54.39%	60.25%	62.73%	4.34%	4.12%
Illinois	Gamma Variant	1.07%	5.15%	4.39%	14.65%	381.31%	233.71%
	Beta Variant	1.17%	6.36%	4.82%	15.82%	443.59%	228.22%
	Delta Variant	46.74%	49.04%	59.98%	63.22%	4.92%	5.40%
	Omicron Variant	NA	64.41%	NA	72.62%	NA	NA
Ohio	Gamma Variant	0.61%	6.63%	4.92%	13.36%	986.89%	171.54%
	Beta Variant	0.76%	7.67%	5.62%	14.43%	909.21%	156.76%
	Delta Variant	45.18%	46.86%	48.49%	50.46%	3.72%	4.06%
	Omicron Variant	53.04%	55.44%	58.38%	60.71%	4.52%	3.99%

Figure 5.3 State Vaccine Progress before and after COVID-19 vaccine availability

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
Arkansas	EUA for Pfizer vaccine for 12-15yo	27.84%	32.14%	36.70%	40.38%	15.45%	10.03%
	Comirnaty for 16yo+ FDA approval	39.88%	45.01%	51.97%	55.33%	12.86%	6.47%
	Pfizer Booster for 65yo+ & 18-64 at high risk	44.73%	47.52%	55.15%	57.47%	6.24%	4.21%
	EUA for Pfizer vaccine for 5-11yo	47.81%	49.30%	57.78%	60.43%	3.12%	4.59%
Mississippi	EUA for Pfizer vaccine for 12-15yo	25.13%	28.05%	31.98%	34.88%	11.62%	9.07%
	Comirnaty for 16yo+ FDA approval	36.78%	42.64%	45.12%	49.62%	15.93%	9.97%
	Pfizer Booster for 65yo+ & 18-64 at high risk	42.53%	45.31%	49.52%	51.68%	6.54%	4.36%
	EUA for Pfizer vaccine for 5-11yo	45.55%	46.90%	51.91%	53.71%	2.96%	3.47%
Alabama	EUA for Pfizer vaccine for 12-15yo	25.92%	29.78%	33.74%	36.45%	14.89%	8.03%
	Comirnaty for 16yo+ FDA approval	36.34%	41.71%	47.85%	52.13%	14.78%	8.94%
	Pfizer Booster for 65yo+ & 18-64 at high risk	41.58%	44.26%	52.05%	53.93%	6.45%	3.61%
	EUA for Pfizer vaccine for 5-11yo	44.61%	46.14%	54.33%	56.44%	3.43%	3.88%
Louisiana	EUA for Pfizer vaccine for 12-15yo	28.69%	32.24%	33.42%	36.61%	12.37%	9.55%
	Comirnaty for 16yo+ FDA approval	39.86%	44.77%	48.23%	51.31%	12.32%	6.39%
	Pfizer Booster for 65yo+ & 18-64 at high risk	44.52%	47.21%	51.16%	53.40%	6.04%	4.38%
	EUA for Pfizer vaccine for 5-11yo	47.51%	48.87%	53.70%	55.56%	2.86%	3.46%
Oklahoma	EUA for Pfizer vaccine for 12-15yo	31.39%	34.50%	39.45%	42.20%	9.91%	6.97%
	Comirnaty for 16yo+ FDA approval	42.38%	46.89%	51.79%	55.99%	10.64%	8.11%
	Pfizer Booster for 65yo+ & 18-64 at high risk	46.63%	49.51%	55.80%	58.64%	6.18%	5.09%
	EUA for Pfizer vaccine for 5-11yo	49.84%	51.48%	59.19%	62.79%	3.29%	6.08%
Florida	EUA for Pfizer vaccine for 12-15yo	32.68%	41.04%	44.05%	50.75%	25.58%	15.21%
	Comirnaty for 16yo+ FDA approval	51.56%	56.48%	62.58%	66.42%	9.54%	6.14%
	Pfizer Booster for 65yo+ & 18-64 at high risk	56.25%	59.23%	66.27%	68.56%	5.30%	3.46%
	EUA for Pfizer vaccine for 5-11yo	59.57%	61.33%	68.91%	71.76%	2.95%	4.14%
Puerto Rico	EUA for Pfizer vaccine for 12-15yo	28.50%	39.84%	42.48%	53.25%	39.79%	25.35%
	Comirnaty for 16yo+ FDA approval	61.92%	69.23%	72.23%	78.58%	11.81%	8.79%
	Pfizer Booster for 65yo+ & 18-64 at high risk	69.03%	73.12%	78.48%	81.62%	5.92%	4.00%
	EUA for Pfizer vaccine for 5-11yo	73.33%	74.07%	81.86%	84.77%	1.01%	3.55%
Texas	EUA for Pfizer vaccine for 12-15yo	30.15%	37.20%	39.69%	45.61%	23.38%	14.92%
	Comirnaty for 16yo+ FDA approval	46.13%	NA	55.77%	NA	NA	NA
	Pfizer Booster for 65yo+ & 18-64 at high risk	50.34%	52.81%	59.27%	60.83%	4.91%	2.63%
	EUA for Pfizer vaccine for 5-11yo	53.22%	54.64%	61.25%	64.02%	2.67%	4.52%
Missouri	EUA for Pfizer vaccine for 12-15yo	30.36%	35.56%	38.92%	43.08%	17.13%	10.69%
	Comirnaty for 16yo+ FDA approval	44.00%	47.29%	51.81%	54.38%	7.48%	4.96%
	Pfizer Booster for 65yo+ & 18-64 at high risk	47.12%	49.34%	54.26%	56.29%	4.71%	3.74%
	EUA for Pfizer vaccine for 5-11yo	49.60%	50.84%	56.64%	59.53%	2.50%	5.10%
South Carolina	EUA for Pfizer vaccine for 12-15yo	30.02%	35.32%	37.82%	41.85%	17.65%	10.66%
	Comirnaty for 16yo+ FDA approval	42.29%	46.43%	50.29%	54.60%	9.79%	8.57%
	Pfizer Booster for 65yo+ & 18-64 at high risk	46.17%	49.39%	54.41%	56.89%	6.97%	4.56%
	EUA for Pfizer vaccine for 5-11yo	49.80%	51.36%	57.35%	60.22%	3.13%	5.00%
Tennessee	EUA for Pfizer vaccine for 12-15yo	27.80%	32.69%	35.79%	39.90%	17.59%	11.48%
	Comirnaty for 16yo+ FDA approval	40.79%	44.72%	48.17%	52.22%	9.63%	8.41%
	Pfizer Booster for 65yo+ & 18-64 at high risk	44.48%	47.22%	52.05%	54.10%	6.16%	3.94%
	EUA for Pfizer vaccine for 5-11yo	47.54%	49.49%	54.40%	56.71%	4.10%	4.25%
Georgia	EUA for Pfizer vaccine for 12-15yo	27.60%	33.82%	36.44%	41.28%	22.54%	13.28%
	Comirnaty for 16yo+ FDA approval	40.43%	44.62%	49.89%	54.35%	10.36%	8.94%
	Pfizer Booster for 65yo+ & 18-64 at high risk	44.08%	47.57%	53.98%	56.40%	7.92%	4.48%
	EUA for Pfizer vaccine for 5-11yo	47.96%	NA	56.70%	NA	NA	NA
Nevada	EUA for Pfizer vaccine for 12-15yo	31.46%	38.21%	41.29%	46.73%	21.46%	13.18%
	Comirnaty for 16yo+ FDA approval	46.65%	50.21%	56.88%	60.21%	7.63%	5.85%
	Pfizer Booster for 65yo+ & 18-64 at high risk	50.02%	52.37%	60.06%	62.62%	4.70%	4.26%
	EUA for Pfizer vaccine for 5-11yo	52.77%	54.43%	63.11%	66.39%	3.15%	5.20%
Arizona	EUA for Pfizer vaccine for 12-15yo	31.56%	37.22%	42.29%	47.34%	17.93%	11.94%
	Comirnaty for 16yo+ FDA approval	47.03%	50.61%	55.74%	58.85%	7.61%	5.58%
	Pfizer Booster for 65yo+ & 18-64 at high risk	50.44%	52.61%	58.72%	60.76%	4.30%	3.47%
	EUA for Pfizer vaccine for 5-11yo	52.91%	54.64%	61.05%	64.24%	3.27%	5.23%
North Carolina	EUA for Pfizer vaccine for 12-15yo	32.51%	37.19%	40.51%	44.07%	14.40%	8.79%
	Comirnaty for 16yo+ FDA approval	45.37%	49.05%	54.35%	58.70%	8.11%	8.00%
	Pfizer Booster for 65yo+ & 18-64 at high risk	48.82%	52.16%	58.50%	62.76%	6.84%	7.28%
	EUA for Pfizer vaccine for 5-11yo	52.48%	54.07%	63.82%	69.96%	3.03%	9.62%

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
Kansas	EUA for Pfizer vaccine for 12-15yo	34.94%	39.43%	43.77%	47.62%	12.85%	8.80%
	Comirnaty for 16yo+ FDA approval	47.19%	50.54%	56.31%	59.33%	7.10%	5.36%
	Pfizer Booster for 65yo+ & 18-64 at high risk	50.35%	52.71%	59.18%	61.69%	4.69%	4.24%
	EUA for Pfizer vaccine for 5-11yo	53.03%	54.47%	62.20%	65.87%	2.72%	5.90%
Wyoming	EUA for Pfizer vaccine for 12-15yo	29.39%	32.61%	34.75%	37.86%	10.96%	8.95%
	Comirnaty for 16yo+ FDA approval	37.91%	41.01%	44.11%	47.93%	8.18%	8.66%
	Pfizer Booster for 65yo+ & 18-64 at high risk	40.83%	43.40%	47.72%	50.22%	6.29%	5.24%
	EUA for Pfizer vaccine for 5-11yo	43.79%	45.52%	50.67%	53.55%	3.95%	5.68%
Utah	EUA for Pfizer vaccine for 12-15yo	26.26%	33.64%	40.92%	46.04%	28.10%	12.51%
	Comirnaty for 16yo+ FDA approval	46.71%	50.00%	55.27%	58.49%	7.04%	5.83%
	Pfizer Booster for 65yo+ & 18-64 at high risk	49.75%	52.97%	58.29%	60.62%	6.47%	4.00%
	EUA for Pfizer vaccine for 5-11yo	53.26%	55.30%	60.87%	63.89%	3.83%	4.96%
New York	EUA for Pfizer vaccine for 12-15yo	39.41%	48.95%	50.11%	57.19%	24.21%	14.13%
	Comirnaty for 16yo+ FDA approval	59.11%	62.99%	66.23%	70.45%	6.56%	6.37%
	Pfizer Booster for 65yo+ & 18-64 at high risk	62.73%	65.96%	70.19%	73.42%	5.15%	4.60%
	EUA for Pfizer vaccine for 5-11yo	66.44%	68.44%	73.97%	78.13%	3.01%	5.62%
California	EUA for Pfizer vaccine for 12-15yo	34.92%	45.31%	51.34%	58.31%	29.75%	13.58%
	Comirnaty for 16yo+ FDA approval	54.93%	58.45%	67.83%	71.21%	6.41%	4.98%
	Pfizer Booster for 65yo+ & 18-64 at high risk	58.08%	60.68%	70.86%	73.78%	4.48%	4.12%
	EUA for Pfizer vaccine for 5-11yo	61.06%	63.12%	74.28%	78.78%	3.37%	6.06%
Idaho	EUA for Pfizer vaccine for 12-15yo	29.05%	33.73%	35.03%	38.33%	16.11%	9.42%
	Comirnaty for 16yo+ FDA approval	38.52%	41.09%	43.29%	46.47%	6.67%	7.35%
	Pfizer Booster for 65yo+ & 18-64 at high risk	40.90%	43.22%	46.29%	48.50%	5.67%	4.77%
	EUA for Pfizer vaccine for 5-11yo	43.69%	45.20%	48.86%	50.70%	3.46%	3.77%
Rhode Island	EUA for Pfizer vaccine for 12-15yo	41.20%	53.81%	54.95%	62.26%	30.61%	13.30%
	Comirnaty for 16yo+ FDA approval	63.74%	67.51%	70.45%	74.56%	5.91%	5.83%
	Pfizer Booster for 65yo+ & 18-64 at high risk	67.12%	70.31%	74.20%	77.28%	4.75%	4.15%
	EUA for Pfizer vaccine for 5-11yo	70.70%	72.47%	77.84%	82.96%	2.50%	6.58%
New Jersey	EUA for Pfizer vaccine for 12-15yo	41.02%	51.15%	54.20%	61.97%	24.70%	14.34%
	Comirnaty for 16yo+ FDA approval	60.42%	63.77%	68.86%	71.95%	5.54%	4.49%
	Pfizer Booster for 65yo+ & 18-64 at high risk	NA	65.88%	NA	74.27%	NA	NA
	EUA for Pfizer vaccine for 5-11yo	66.22%	67.75%	74.80%	79.01%	2.31%	5.63%
Nebraska	EUA for Pfizer vaccine for 12-15yo	37.06%	43.78%	44.91%	49.39%	18.13%	9.98%
	Comirnaty for 16yo+ FDA approval	51.11%	54.17%	56.59%	59.03%	5.99%	4.31%
	Pfizer Booster for 65yo+ & 18-64 at high risk	54.00%	55.90%	58.92%	60.56%	3.52%	2.78%
	EUA for Pfizer vaccine for 5-11yo	56.14%	57.20%	60.85%	63.72%	1.89%	4.72%
Virginia	EUA for Pfizer vaccine for 12-15yo	37.18%	47.25%	49.61%	56.65%	27.08%	14.19%
	Comirnaty for 16yo+ FDA approval	56.31%	59.83%	64.39%	67.79%	6.25%	5.28%
	Pfizer Booster for 65yo+ & 18-64 at high risk	59.75%	62.48%	67.75%	70.34%	4.57%	3.82%
	EUA for Pfizer vaccine for 5-11yo	62.94%	64.87%	70.91%	75.50%	3.07%	6.47%
Pennsylvania	EUA for Pfizer vaccine for 12-15yo	36.16%	45.92%	52.55%	60.08%	26.99%	14.33%
	Comirnaty for 16yo+ FDA approval	54.27%	57.33%	68.32%	71.73%	5.64%	4.99%
	Pfizer Booster for 65yo+ & 18-64 at high risk	57.04%	59.94%	71.43%	76.08%	5.08%	6.51%
	EUA for Pfizer vaccine for 5-11yo	NA	NA	NA	NA	NA	NA
New Mexico	EUA for Pfizer vaccine for 12-15yo	NA	49.23%	NA	58.79%	NA	NA
	Comirnaty for 16yo+ FDA approval	58.93%	NA	68.54%	NA	NA	NA
	Pfizer Booster for 65yo+ & 18-64 at high risk	NA	NA	NA	NA	NA	NA
	EUA for Pfizer vaccine for 5-11yo	61.98%	63.48%	71.88%	76.51%	2.42%	6.44%
North Dakota	EUA for Pfizer vaccine for 12-15yo	34.36%	37.15%	40.10%	42.72%	8.12%	6.53%
	Comirnaty for 16yo+ FDA approval	41.15%	43.54%	47.47%	50.49%	5.81%	6.36%
	Pfizer Booster for 65yo+ & 18-64 at high risk	43.36%	NA	50.30%	NA	NA	NA
	EUA for Pfizer vaccine for 5-11yo	45.74%	48.70%	53.02%	57.85%	6.47%	9.11%
South Dakota	EUA for Pfizer vaccine for 12-15yo	39.70%	43.34%	45.60%	48.97%	9.17%	7.39%
	Comirnaty for 16yo+ FDA approval	48.40%	51.13%	55.20%	58.53%	5.64%	6.03%
	Pfizer Booster for 65yo+ & 18-64 at high risk	50.96%	52.13%	58.37%	60.98%	2.30%	4.47%
	EUA for Pfizer vaccine for 5-11yo	52.46%	54.42%	61.70%	66.64%	3.74%	8.01%
Colorado	EUA for Pfizer vaccine for 12-15yo	37.72%	47.38%	48.71%	55.62%	25.61%	14.19%
	Comirnaty for 16yo+ FDA approval	56.12%	58.85%	62.60%	65.13%	4.86%	4.04%
	Pfizer Booster for 65yo+ & 18-64 at high risk	58.69%	61.04%	64.99%	67.05%	4.00%	3.17%
	EUA for Pfizer vaccine for 5-11yo	61.43%	63.16%	67.45%	71.38%	2.82%	5.83%

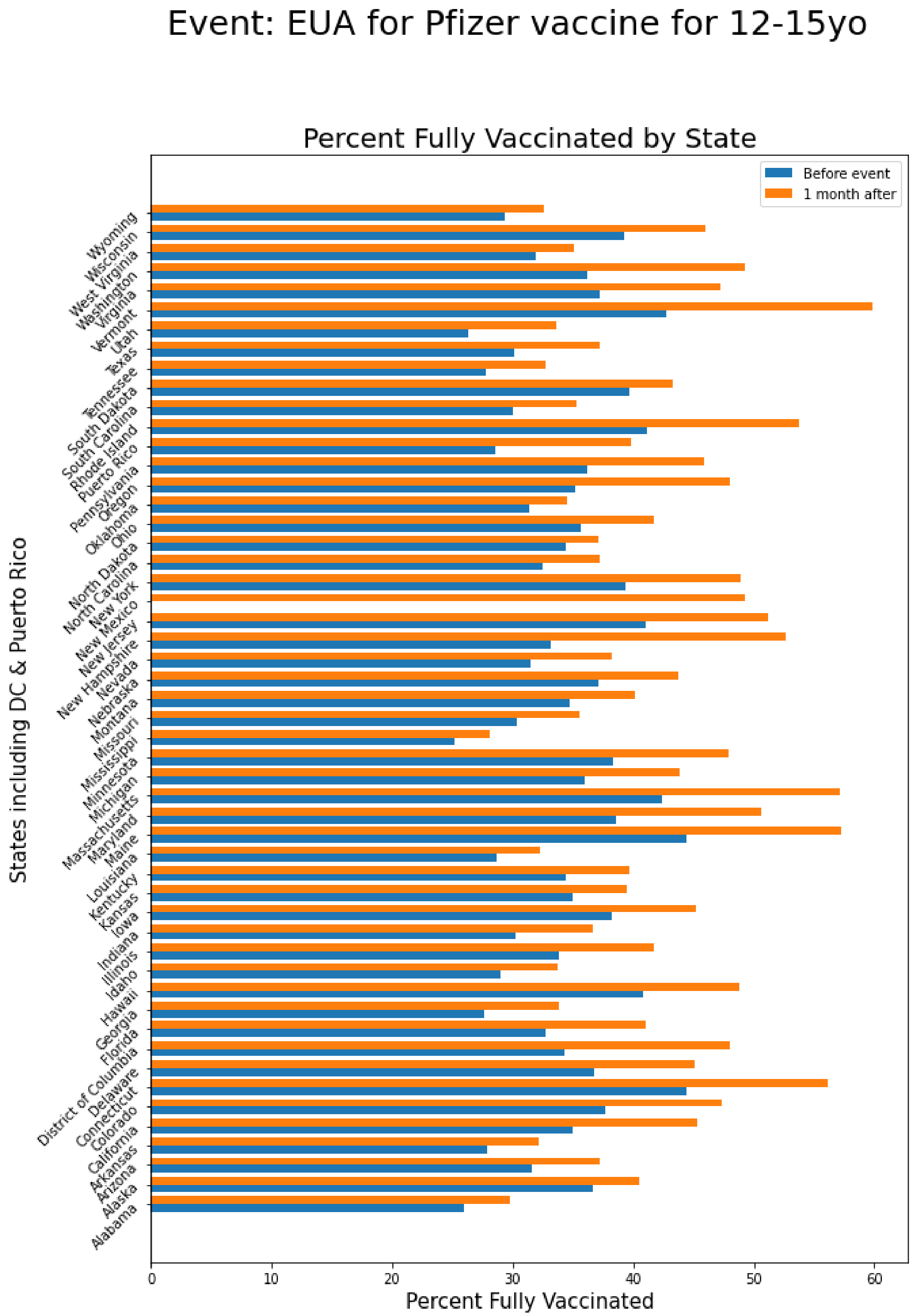
State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
Wisconsin	EUA for Pfizer vaccine for 12-15yo	39.21%	46.03%	47.17%	52.11%	17.39%	10.47%
	Comirnaty for 16yo+ FDA approval	53.13%	55.91%	57.78%	60.57%	5.23%	4.83%
	Pfizer Booster for 65yo+ & 18-64 at high risk	55.75%	57.88%	60.45%	62.22%	3.82%	2.93%
	EUA for Pfizer vaccine for 5-11yo	58.13%	59.47%	62.51%	65.79%	2.31%	5.25%
Montana	EUA for Pfizer vaccine for 12-15yo	34.69%	40.20%	42.06%	46.36%	15.88%	10.22%
	Comirnaty for 16yo+ FDA approval	45.42%	48.07%	51.39%	54.32%	5.83%	5.70%
	Pfizer Booster for 65yo+ & 18-64 at high risk	47.90%	49.93%	54.16%	56.26%	4.24%	3.88%
	EUA for Pfizer vaccine for 5-11yo	50.33%	51.82%	56.69%	59.69%	2.96%	5.29%
Delaware	EUA for Pfizer vaccine for 12-15yo	36.80%	45.07%	49.12%	55.93%	22.47%	13.86%
	Comirnaty for 16yo+ FDA approval	54.32%	57.10%	63.09%	65.76%	5.12%	4.23%
	Pfizer Booster for 65yo+ & 18-64 at high risk	56.80%	59.39%	65.50%	68.37%	4.56%	4.38%
	EUA for Pfizer vaccine for 5-11yo	59.70%	60.95%	68.79%	71.44%	2.09%	3.85%
Alaska	EUA for Pfizer vaccine for 12-15yo	36.69%	40.49%	42.72%	46.99%	10.36%	10.00%
	Comirnaty for 16yo+ FDA approval	46.62%	49.73%	53.26%	56.79%	6.67%	6.63%
	Pfizer Booster for 65yo+ & 18-64 at high risk	49.31%	52.16%	56.60%	58.84%	5.78%	3.96%
	EUA for Pfizer vaccine for 5-11yo	52.51%	54.07%	59.23%	62.46%	2.97%	5.45%
Indiana	EUA for Pfizer vaccine for 12-15yo	30.19%	36.67%	37.92%	42.64%	21.46%	12.45%
	Comirnaty for 16yo+ FDA approval	45.61%	47.98%	49.30%	51.83%	5.20%	5.13%
	Pfizer Booster for 65yo+ & 18-64 at high risk	47.82%	49.51%	51.71%	53.34%	3.53%	3.15%
	EUA for Pfizer vaccine for 5-11yo	49.69%	50.57%	53.56%	55.64%	1.77%	3.88%
Maryland	EUA for Pfizer vaccine for 12-15yo	38.57%	50.60%	51.19%	58.85%	31.19%	14.96%
	Comirnaty for 16yo+ FDA approval	60.61%	63.60%	67.24%	70.14%	4.93%	4.31%
	Pfizer Booster for 65yo+ & 18-64 at high risk	63.42%	65.69%	69.98%	72.38%	3.58%	3.43%
	EUA for Pfizer vaccine for 5-11yo	NA	67.47%	NA	76.83%	NA	NA
District of Columbia	EUA for Pfizer vaccine for 12-15yo	34.24%	48.01%	51.14%	58.24%	40.22%	13.88%
	Comirnaty for 16yo+ FDA approval	56.65%	59.46%	66.56%	69.94%	4.96%	5.08%
	Pfizer Booster for 65yo+ & 18-64 at high risk	59.28%	61.85%	69.72%	73.28%	4.34%	5.11%
	EUA for Pfizer vaccine for 5-11yo	62.21%	64.44%	74.01%	81.46%	3.58%	10.07%
Iowa	EUA for Pfizer vaccine for 12-15yo	38.19%	45.26%	45.64%	50.07%	18.51%	9.71%
	Comirnaty for 16yo+ FDA approval	50.90%	53.55%	55.46%	57.74%	5.21%	4.11%
	Pfizer Booster for 65yo+ & 18-64 at high risk	53.42%	55.16%	57.63%	59.18%	3.26%	2.69%
	EUA for Pfizer vaccine for 5-11yo	55.41%	56.51%	59.50%	62.23%	1.99%	4.59%
Minnesota	EUA for Pfizer vaccine for 12-15yo	38.38%	47.92%	49.44%	55.41%	24.86%	12.08%
	Comirnaty for 16yo+ FDA approval	55.15%	57.75%	61.16%	63.19%	4.71%	3.32%
	Pfizer Booster for 65yo+ & 18-64 at high risk	57.61%	59.41%	63.09%	64.67%	3.12%	2.50%
	EUA for Pfizer vaccine for 5-11yo	59.70%	62.26%	65.00%	68.73%	4.29%	5.74%
Connecticut	EUA for Pfizer vaccine for 12-15yo	44.37%	56.08%	57.31%	64.70%	26.39%	12.89%
	Comirnaty for 16yo+ FDA approval	65.13%	68.17%	72.57%	75.57%	4.67%	4.13%
	Pfizer Booster for 65yo+ & 18-64 at high risk	68.01%	70.25%	75.44%	78.15%	3.29%	3.59%
	EUA for Pfizer vaccine for 5-11yo	70.56%	71.89%	78.71%	83.90%	1.88%	6.59%
Oregon	EUA for Pfizer vaccine for 12-15yo	35.19%	47.98%	48.14%	56.24%	36.35%	16.83%
	Comirnaty for 16yo+ FDA approval	57.36%	60.15%	62.96%	66.11%	4.86%	5.00%
	Pfizer Booster for 65yo+ & 18-64 at high risk	60.02%	62.39%	66.07%	67.89%	3.95%	2.75%
	EUA for Pfizer vaccine for 5-11yo	62.71%	63.97%	68.22%	71.41%	2.01%	4.68%
Hawaii	EUA for Pfizer vaccine for 12-15yo	40.80%	48.86%	58.81%	67.95%	19.75%	15.54%
	Comirnaty for 16yo+ FDA approval	54.66%	57.27%	73.38%	76.28%	4.77%	3.95%
	Pfizer Booster for 65yo+ & 18-64 at high risk	NA	59.40%	NA	78.20%	NA	NA
	EUA for Pfizer vaccine for 5-11yo	59.66%	60.89%	78.57%	81.98%	2.06%	4.34%
Michigan	EUA for Pfizer vaccine for 12-15yo	35.93%	43.82%	44.44%	49.78%	21.96%	12.02%
	Comirnaty for 16yo+ FDA approval	49.96%	51.95%	54.69%	56.63%	3.98%	3.55%
	Pfizer Booster for 65yo+ & 18-64 at high risk	51.83%	53.08%	56.52%	57.65%	2.41%	2.00%
	EUA for Pfizer vaccine for 5-11yo	53.40%	54.46%	58.06%	60.94%	1.99%	4.96%
Maine	EUA for Pfizer vaccine for 12-15yo	44.44%	57.21%	57.15%	64.67%	28.74%	13.16%
	Comirnaty for 16yo+ FDA approval	65.07%	67.97%	70.44%	73.60%	4.46%	4.49%
	Pfizer Booster for 65yo+ & 18-64 at high risk	67.77%	70.03%	73.45%	75.72%	3.33%	3.09%
	EUA for Pfizer vaccine for 5-11yo	70.40%	72.14%	76.33%	81.70%	2.47%	7.04%
Massachusetts	EUA for Pfizer vaccine for 12-15yo	42.36%	57.13%	60.09%	68.18%	34.87%	13.46%
	Comirnaty for 16yo+ FDA approval	65.29%	67.52%	74.60%	76.97%	3.42%	3.18%
	Pfizer Booster for 65yo+ & 18-64 at high risk	67.37%	69.22%	76.83%	79.31%	2.75%	3.23%
	EUA for Pfizer vaccine for 5-11yo	69.51%	71.11%	79.88%	85.69%	2.30%	7.27%

State (including DC and Puerto Rico)	Event	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
West Virginia	EUA for Pfizer vaccine for 12-15yo	31.98%	35.05%	36.68%	41.76%	9.60%	13.85%
	Comirnaty for 16yo+ FDA approval	39.46%	40.24%	46.75%	47.94%	1.98%	2.55%
	Pfizer Booster for 65yo+ & 18-64 at high risk	40.18%	40.94%	47.87%	48.78%	1.89%	1.90%
	EUA for Pfizer vaccine for 5-11yo	41.01%	41.54%	48.91%	53.91%	1.29%	10.22%
Vermont	EUA for Pfizer vaccine for 12-15yo	42.79%	59.81%	61.26%	71.84%	39.78%	17.27%
	Comirnaty for 16yo+ FDA approval	67.42%	69.11%	75.57%	77.33%	2.51%	2.33%
	Pfizer Booster for 65yo+ & 18-64 at high risk	69.03%	70.75%	77.26%	79.03%	2.49%	2.29%
	EUA for Pfizer vaccine for 5-11yo	71.03%	72.87%	79.41%	85.41%	2.59%	7.56%
Washington	EUA for Pfizer vaccine for 12-15yo	36.25%	49.26%	49.54%	58.27%	35.89%	17.62%
	Comirnaty for 16yo+ FDA approval	59.37%	60.16%	66.62%	66.69%	1.33%	0.11%
	Pfizer Booster for 65yo+ & 18-64 at high risk	NA	62.92%	NA	68.80%	NA	NA
	EUA for Pfizer vaccine for 5-11yo	63.35%	64.87%	69.22%	72.49%	2.40%	4.72%
New Hampshire	EUA for Pfizer vaccine for 12-15yo	33.13%	52.62%	57.40%	60.90%	58.83%	6.10%
	Comirnaty for 16yo+ FDA approval	59.27%	61.22%	66.43%	68.97%	3.29%	3.82%
	Pfizer Booster for 65yo+ & 18-64 at high risk	61.08%	62.59%	68.80%	73.69%	2.47%	7.11%
	EUA for Pfizer vaccine for 5-11yo	62.75%	64.75%	75.25%	87.39%	3.19%	16.13%
Kentucky	EUA for Pfizer vaccine for 12-15yo	34.36%	39.75%	42.38%	47.52%	15.69%	12.13%
	Comirnaty for 16yo+ FDA approval	47.56%	NA	55.75%	NA	NA	NA
	Pfizer Booster for 65yo+ & 18-64 at high risk	NA	NA	NA	NA	NA	NA
	EUA for Pfizer vaccine for 5-11yo	NA	52.09%	NA	60.22%	NA	NA
Illinois	EUA for Pfizer vaccine for 12-15yo	33.84%	41.75%	48.80%	56.58%	23.37%	15.94%
	Comirnaty for 16yo+ FDA approval	50.35%	NA	64.99%	NA	NA	NA
	Pfizer Booster for 65yo+ & 18-64 at high risk	NA	NA	NA	NA	NA	NA
	EUA for Pfizer vaccine for 5-11yo	NA	NA	NA	NA	NA	NA
Ohio	EUA for Pfizer vaccine for 12-15yo	35.62%	41.69%	41.83%	46.86%	17.04%	12.02%
	Comirnaty for 16yo+ FDA approval	47.76%	49.78%	51.79%	53.81%	4.23%	3.90%
	Pfizer Booster for 65yo+ & 18-64 at high risk	49.73%	51.48%	53.73%	55.42%	3.52%	3.15%
	EUA for Pfizer vaccine for 5-11yo	51.73%	53.01%	55.68%	58.35%	2.47%	4.80%

Figure 5.4 State Vaccine Progress before and after the J & J vaccine pause in use

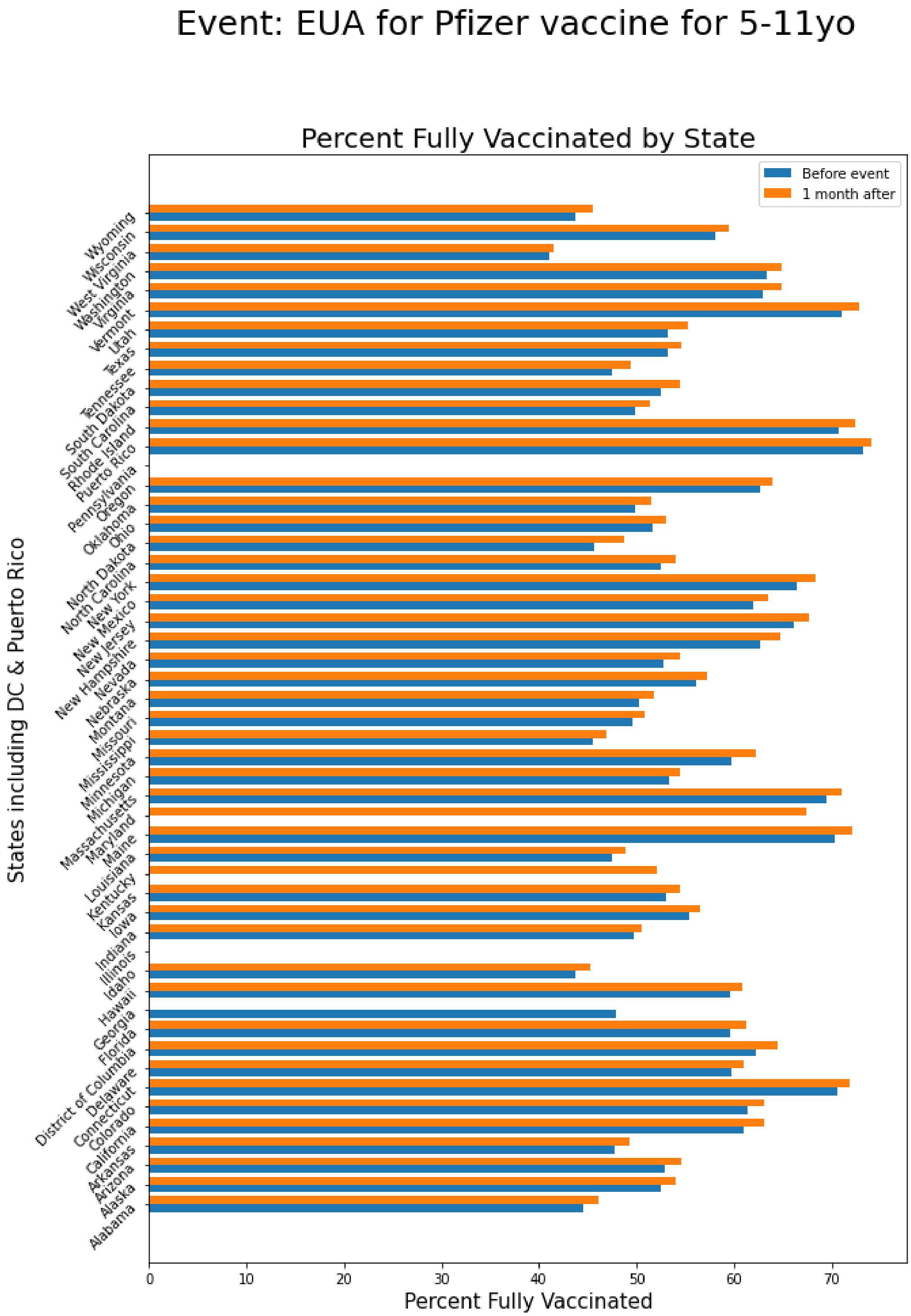
State (including DC and Puerto Rico)	% Fully Vaccinated before event	% Fully Vaccinated 1 month after event	% Vaccinated (at least one dose) before event	% Vaccinated (at least one dose) 1 month after event	% change in vaccine progress (Fully Vaccinated)	% change in vaccine progress (at least one dose)
Arkansas	18.98%	28.47%	31.49%	37.10%	50.00%	17.82%
Mississippi	19.21%	25.51%	27.90%	32.26%	32.80%	15.63%
Alabama	16.81%	26.81%	28.34%	34.23%	59.49%	20.78%
Louisiana	21.06%	29.18%	29.85%	33.77%	38.56%	13.13%
Oklahoma	23.70%	31.85%	35.64%	39.73%	34.39%	11.48%
Florida	21.44%	34.27%	34.81%	44.96%	59.84%	29.16%
Puerto Rico	16.34%	29.79%	27.29%	44.11%	82.31%	61.63%
Texas	19.86%	31.24%	32.42%	40.23%	57.30%	24.09%
Missouri	20.59%	31.61%	31.47%	39.43%	53.52%	25.29%
South Carolina	20.58%	30.92%	32.66%	38.31%	50.24%	17.30%
Tennessee	18.20%	28.60%	30.29%	36.33%	57.14%	19.94%
Georgia	15.03%	28.61%	28.84%	37.07%	90.35%	28.54%
Nevada	21.51%	32.71%	33.95%	41.96%	52.07%	23.59%
Arizona	22.01%	32.70%	35.15%	43.05%	48.57%	22.48%
North Carolina	21.68%	33.29%	34.04%	40.83%	53.55%	19.95%
Kansas	22.88%	35.73%	37.67%	44.21%	56.16%	17.36%
Wyoming	22.81%	29.95%	31.15%	35.05%	31.30%	12.52%
Utah	16.12%	28.59%	31.81%	42.04%	77.36%	32.16%
New York	25.35%	41.18%	39.23%	50.88%	62.45%	29.70%
California	21.41%	36.62%	38.78%	52.20%	71.04%	34.61%
Idaho	NA	30.12%	NA	35.59%	NA	NA
Rhode Island	28.71%	43.42%	40.54%	55.82%	51.24%	37.69%
New Jersey	26.29%	42.75%	42.09%	55.13%	62.61%	30.98%
Nebraska	23.88%	38.77%	37.18%	45.37%	62.35%	22.03%
Virginia	22.97%	38.96%	39.00%	50.52%	69.61%	29.54%
Pennsylvania	22.74%	37.83%	39.52%	53.64%	66.36%	35.73%
New Mexico	29.65%	NA	44.97%	NA	NA	NA
North Dakota	27.13%	34.70%	36.82%	40.28%	27.90%	9.40%
South Dakota	28.49%	40.31%	41.24%	45.88%	41.49%	11.25%
Colorado	22.78%	39.45%	38.09%	49.45%	73.18%	29.82%
Wisconsin	25.76%	40.38%	39.70%	47.61%	56.75%	19.92%
Montana	24.91%	35.37%	36.52%	42.43%	41.99%	16.18%
Delaware	21.81%	38.24%	38.34%	49.85%	75.33%	30.02%
Alaska	27.58%	37.36%	37.21%	43.18%	35.46%	16.04%
Indiana	20.87%	31.29%	31.08%	38.60%	49.93%	24.20%
Maryland	23.76%	40.80%	38.48%	52.15%	71.72%	35.52%
District of Columbia	19.76%	37.00%	36.29%	52.60%	87.25%	44.94%
Iowa	25.25%	39.34%	38.11%	46.05%	55.80%	20.83%
Minnesota	26.04%	39.99%	39.35%	50.13%	53.57%	27.40%
Connecticut	28.10%	46.46%	44.42%	58.12%	65.34%	30.84%
Oregon	22.34%	37.29%	35.48%	49.42%	66.92%	39.29%
Hawaii	25.21%	41.86%	36.65%	59.74%	66.05%	63.00%
Michigan	23.08%	37.29%	35.48%	45.11%	61.57%	27.14%
Maine	27.66%	47.25%	42.95%	58.21%	70.82%	35.53%
Massachusetts	25.82%	44.79%	43.84%	61.31%	73.47%	39.85%
West Virginia	24.19%	32.41%	33.57%	37.41%	33.98%	11.44%
Vermont	27.21%	44.81%	42.46%	63.23%	64.68%	48.92%
Washington	24.05%	38.37%	37.17%	50.75%	59.54%	36.53%
New Hampshire	24.99%	35.18%	52.31%	59.32%	40.78%	13.40%
Kentucky	24.01%	35.37%	36.33%	42.99%	47.31%	18.33%
Illinois	22.27%	35.29%	38.75%	49.75%	58.46%	28.39%
Ohio	22.93%	36.82%	35.65%	42.41%	60.58%	18.96%

Figure 5.5 State Vaccine Progress before and after the EUA for Pfizer Vaccine for 12-15 yo



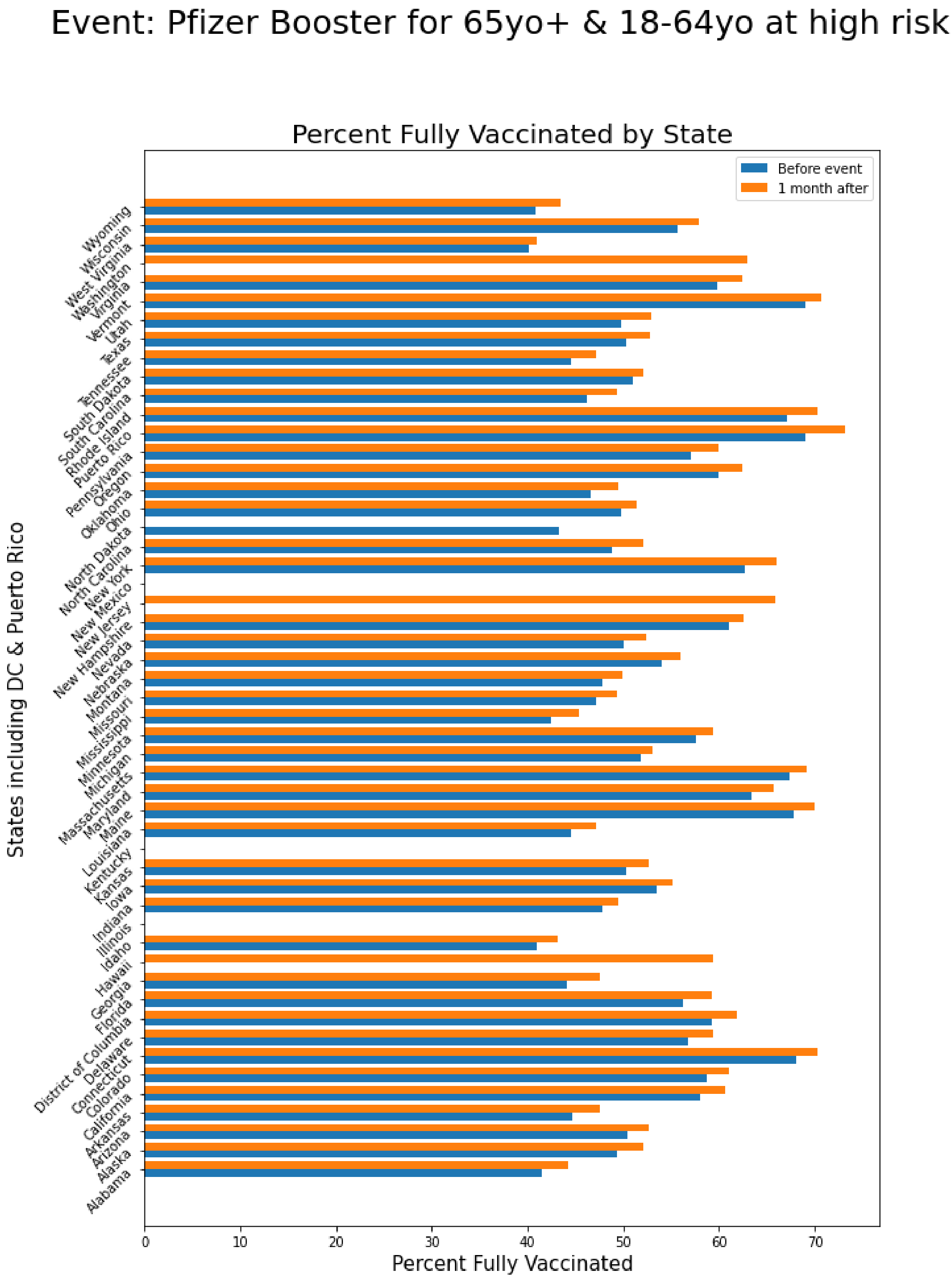
*Note: Some states will have missing data.

Figure 5.6 State Vaccine Progress before and after the EUA for Pfizer Vaccine for 12-15 yo



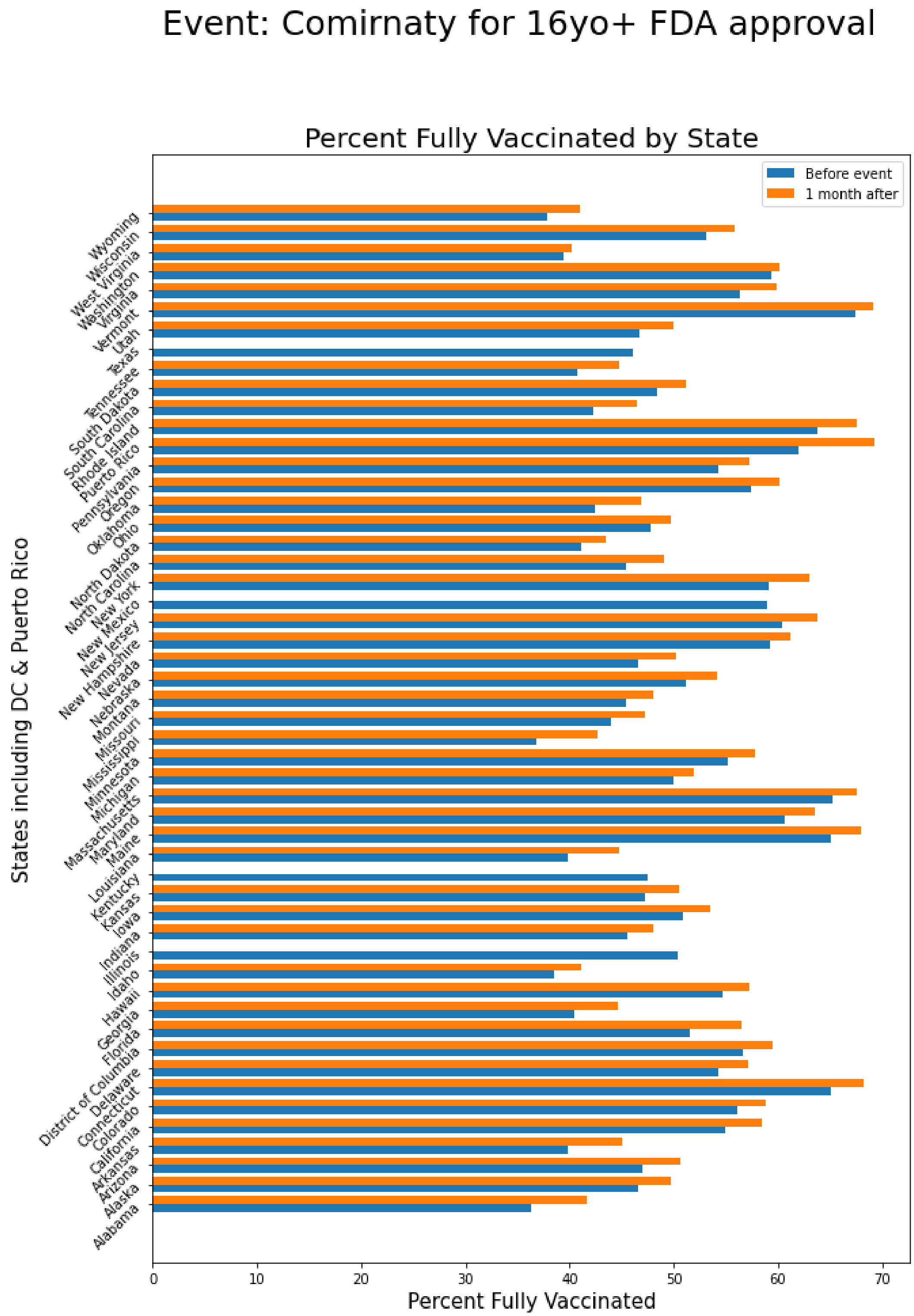
*Note: Some states will have missing data.

Figure 5.7 State Vaccine Progress before and after the FDA authorized Pfizer booster dose for those aged 65+ and age 18-64 at high risk



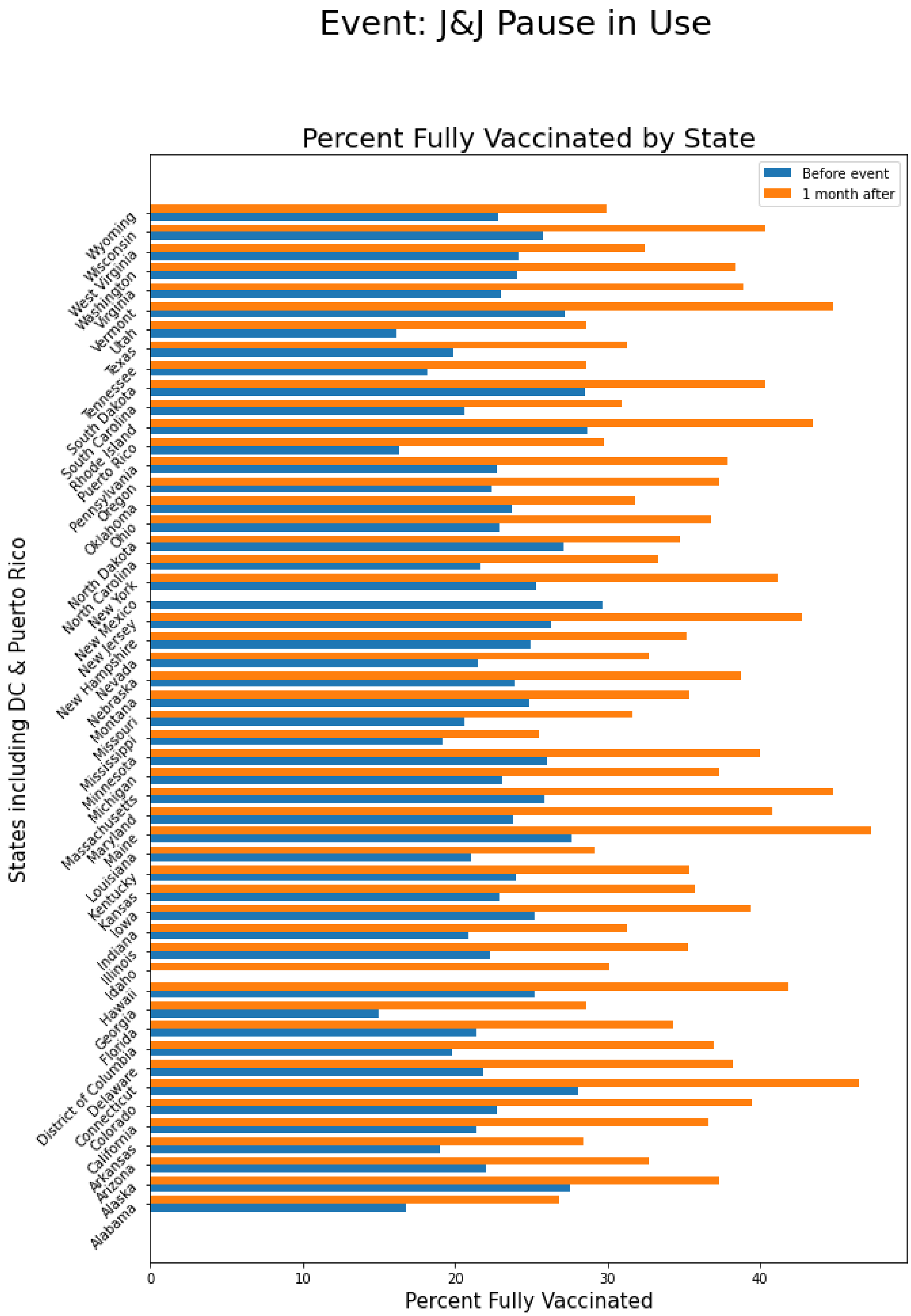
*Note: Some states will have missing data.

Figure 5.8 State Vaccine Progress before and after the FDA approved Comirnaty for those 16+



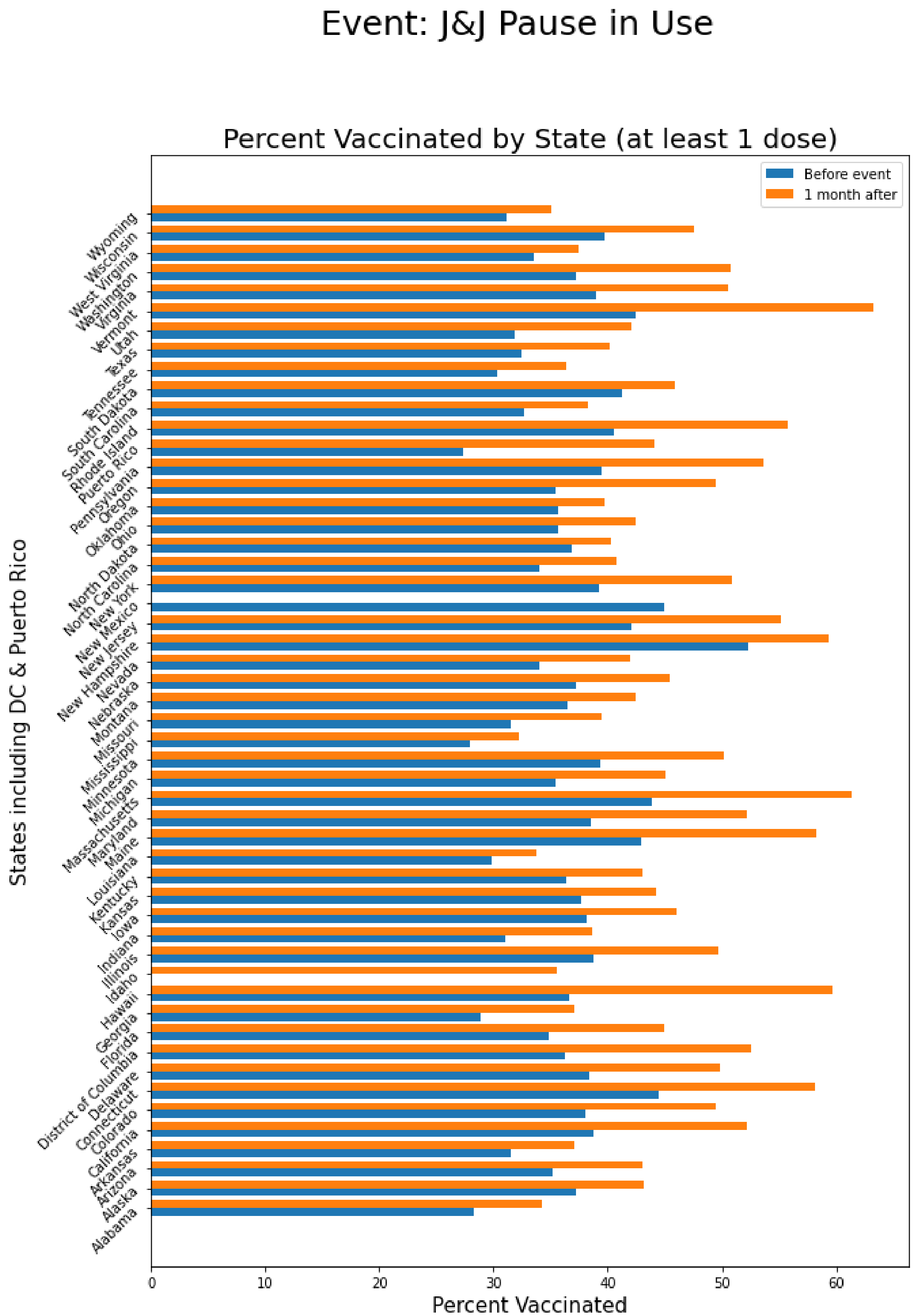
*Note: Some states will have missing data.

Figure 5.9 State Vaccine Progress before and after J & J Pause in Use (Full Vaccination Rate)



*Note: Some states will have missing data.

Figure 5.10 State Vaccine Progress before and after J & J Pause in Use (Vaccination Rate)



*Note: Some states will have missing data.

Works Cited

Agarwal, R., Dugas, M., Ramaprasad, J., Luo, J., Li, G., & Gao, G. G. (2021). Socioeconomic privilege and political ideology are associated with racial disparity in COVID-19 vaccination. Proceedings of the National Academy of Sciences, 118(33). https://doi.org/10.1073/pnas.2107873118

Albrecht, D. (2022). Vaccination, politics and COVID-19 impacts. *BMC Public Health*, *22*(1), 96. https://doi.org/10.1186/s12889-021-12432-x

Alizadeh, K. (2021, April 2). *Limitations of Twitter Data*. Medium. https://towardsdatascience.com/limitations-of-twitter-data-94954850cacf

Beusekom, M. V. (2021, June 18). *Poorer US counties have lower COVID-19 vaccine uptake*. CIDRAP. https://www.cidrap.umn.edu/news-perspective/2021/06/poorer-us-counties-have-lower-covid-19-vaccine-uptake

Centers for Disease Control and Prevention/ Agency for Toxic Substances and Disease Registry/ Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index [2018] Database [California]. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html

Centers for Disease Control and Prevention/ Agency for Toxic Substances and Disease Registry/ Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index [2018] Database [Michigan]. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html

Cristóbal Colón-Ruiz, I. (2020). Comparing deep learning architectures for sentiment analysis on drug reviews, Journal of Biomedical Informatics https://doi.org/10.1016/j.jbi.2020.103539

Concern about safety is main reason many Detroiters are not getting vaccinated, U-M survey finds. (2021, July 22). Gerald R. Ford School of Public Policy. https://fordschool.umich.edu/news/2021/concern-about-safety-main-reason-many-detroiters-are-not-getting-vaccinated-u-m-survey

CDC. (2020, March 28). COVID Data Tracker. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#datatracker-home

COVID-19 basics. Harvard Health Publishing. Published January 26, 2022. https://www.health.harvard.edu/diseases-and-conditions/covid-19-basics

COVID-19 misleading information policy. Twitter. Published December 2021. https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy

DeJesus J. Cleaning Location Data with GeoPy. Medium. Published March 4, 2019. https://towardsdatascience.com/cleaning-location-data-with-geopy-51613bcd3c89

Election results, 2021: Party control of state legislatures. (n.d.). Ballotpedia. Retrieved April 23, 2022, from https://ballotpedia.org/Election_results,_2021:_Party_control_of_state_legislatures

Ethnic and Racial Minorities & Socioeconomic Status. (2017, July). APA. https://www.apa.org/pi/ses/resources/publications/minorities

Holder J. Tracking Coronavirus Vaccinations Around the World. *The New York Times*. https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html. Published January 29, 2021.

Justin P. Hicks, (2021, February 5). Michigan isn't reporting COVID-19 vaccination data by race, but initial reports suggest disparity. Mlive. https://www.mlive.com/public-interest/2021/02/michigan-isnt-reporting-covid-19-vaccination-data-by-race-but-initial-reports-suggest-disparity.html

Kates, J., Tolbert, J., & Rouw, A. (2022, January 19). The Red/Blue Divide in COVID-19 Vaccination Rates Continues: An Update. *KFF.* https://www.kff.org/policy-watch/the-red-blue-divide-in-covid-19-vaccination-rates-continues-an-update/

KFF COVID-19 Vaccine Monitor Dashboard. KFF. Published March 1, 2022. https://www.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/

Kirzinger, A., Kearney, A., Hamel, L., & Brodie, M. (2021, November 16). KFF COVID-19 Vaccine Monitor: The Increasing Importance of Partisanship in Predicting COVID-19 Vaccination Status. KFF. https://www.kff.org/coronavirus-covid-19/poll-finding/importance-of-partisanship-predicting-vaccination-status/

List of current United States governors. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_current_United_States_governors&oldid=1081486822

List of United States presidential election results by state. In: Wikipedia. ; 2021. https://en.wikipedia.org/w/index.php?title=List_of_United_States_presidential_election_results_by_state&oldid=1040690723

Mooney P. USA COVID-19 Vaccinations. Kaggle. https://www.kaggle.com/paultimothymooney/usa-covid19-vaccinations

Mooney P. Latitude and Longitude for Every Country and State. Kaggle. https://www.kaggle.com/paultimothymooney/latitude-and-longitude-for-every-country-and-state

More Than 11.4 Billion Shots Given: Covid-19 Tracker. *Bloomberg.com*. https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/

San Francisco to Expand COVID-19 Vaccinations to People with Disabilities and Severe Underlying Conditions and those in High-Risk Congregate Settings on March 15 | Office of the Mayor. (2021, March 12). Office of the Mayor |. Retrieved April 24, 2022, from https://sfmayor.org/article/san-francisco-expand-covid-19-vaccinations-people-disabilities-and-severe-underlying

Shacham O. USA States to region. Kaggle. https://www.kaggle.com/omer2040/usa-states-to-region

Timeline of the COVID-19 pandemic in the United States (2021). In: Wikipedia. ; 2022. https://en.wikipedia.org/w/index.php?title=Timeline_of_the_COVID-19_pandemic_in_the_United_States_(2021)&oldid=1072395501

Vaccine Timeline. Immunize. https://www.immunize.org/timeline/

Wojcik, S., & Hughes, A. (2019, April 24). Sizing Up Twitter Users. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

Wood, D., & Brumfiel, G. (2021, December 5). Pro-Trump counties now have far higher COVID death rates. Misinformation is to blame. *NPR*. https://www.npr.org/sections/health-shots/2021/12/05/1059828993/data-vaccine-misinformation-trump-counties-covid-death-rate

U.S. Census Bureau QuickFacts: Detroit city, Michigan; San Francisco County, California. (n.d.). Census Bureau QuickFacts. Retrieved April 20, 2022, from https://www.census.gov/quickfacts/fact/table/detroitcitymichigan,sanfranciscocountycalifornia/AGE135220#AGE135220

U.S. Census Bureau (2020). Poverty Status in the past 12 months, 2015-2020 American Community Survey 5-year estimates. Retrieved from https://data.census.gov/cedsci/table?t=Income%20and%20Poverty%3APoverty&g=0600000US2616322000_1600000US0667000,2622000&y=2020.

U.S. Census Bureau (2020). Education Attainment, 2015-2020 American Community Survey 5-year estimates. Retrieved from https://data.census.gov/cedsci/table?t=Educational%20Attainment&g=0600000US2616322000_1600000US0667000,2622000&y=2020.

U.S. Census Bureau (2020). Mean income in the past 12 months (in 2020 inflation-adjusted dollars), 2015-2020 American Community Survey 5-year estimates. Retrieved from https://data.census.gov/cedsci/table?t=Income%20and%20Poverty&g=0600000US2616322000_1600000US0667000,2622000&y=2020.