Final Report - Vaccination Status

ABSTRACT

Over the course of the COVID-19 Pandemic, 570 million doses of the vaccine were given across the United States. However, only two-thirds of Americans have been fully vaccinated against the virus, a vaccination rate significantly less than many other countries across the globe.[1] The objective of this project was to try to learn more about the factors which influenced vaccination status in the United States throughout the COVID-19 pandemic. As a result of the gravity of the situation and the pressing need to learn more about the virus, hundreds of raw datasets are available which describe many characteristics of vaccinated Americans. However, the majority of this data has not been standardized in terms of the United States population, which leads to extraneous variability and prevents any conclusions from accurately depicting the current situation. The lack of standardization, alongside the absence of a clear portrait of the typical vaccinated American, has created a major gap in the current research which must be addressed. To further past research on the study of the vaccination status in the US, analysis on vaccine hesitancy was done based on social vulnerability, ethnicities, presence of children under the age of 18, marital status of households, and other household characteristics for each US state to narrow the scope of the project rather the US overall. The work in this study identifies demographic characteristics of the vaccinated population, compares characteristics of vaccinated individuals in different states, investigates the role of social hesitancy and vaccine distribution in vaccination status, and identifies factors which influence the COVID-19 recovery rate. Datasets from a variety of sources were brought together to achieve a thorough understanding of the research questions; these sources included the Centers for Disease Control and Prevention COVID Data Tracker and the Household Pulse Survey. At the end of the analysis, it was discovered that women are more likely to get vaccinated than men. The most vaccinated age group was 75+ years, and the most vaccinated race/ethnicity was American Indian/Alaska Native Non-Hispanic. Positive vaccine attitudes were associated with income levels of less than $25,000, educational attainment, and an unmarried status. A correlation was found between social vulnerability and level of concern for vaccination rollout. Another relationship was discovered between higher recovery rate and liberal political orientation in a state. While some clear patterns emerged as a result of this study, future research is needed to further understand whether these traits of the vaccinated and unvaccinated populations hold true.

[1] Centers for Disease Control and Prevention. (n.d.). CDC Covid Data tracker. Centers for Disease Control and Prevention. Retrieved May 2, 2022, from https://covid.cdc.gov/covid-data-tracker/#vaccination-demographic

INTRODUCTION

Throughout the COVID-19 pandemic, millions of Americans were faced with the decision of whether or not to receive the COVID-19 vaccine. Across the entire United States, 219 individuals are now fully vaccinated, a statistic which constitutes 66.5% of the population. Globally, the United States ranks 54th in terms of countries most vaccinated against COVID.[2] While vaccination efforts and campaigns spurred increasing willingness to receive the vaccine, many researchers and scientists remained surprised by Americans' hesitancy when it came to the vaccination. In an attempt to better understand and establish the characteristics which vaccinated Americans tend to have in common, this study investigated the factors which influenced vaccination status in the US during the COVID-19 pandemic.

At the peak of the pandemic, many individuals across the globe were testing positive for COVID-19 daily; a large subsection of the positive population ended up losing their lives to the virus. Just within the United States, 80.9 million Americans tested positive for COVID-19, and nearly a million died as a result.[3] Due to the intense nature of the pandemic and the urgency on the part of both the United States government and international organizations to help decrease the number of confirmed cases, there is an enormous amount of data which has been collected. Many of these datasets related directly to descriptors associated with the populations most likely and least likely to receive a vaccine. Much work has been completed which relates common demographic traits to vaccination status; the challenge faced in this project was to differentiate the analysis from the large body of research already conducted on the subject. After examining a broad range of studies, the greatest gap identified was within the United States-specific data, which exhibited a tendency to lack standardization procedures in the visualizations. While other work asked relatively similar questions to those posed here, the absence of standardization led to graphics which did not accurately portray a comprehensive portrait of the vaccinated American. Using prior art and expanding from it, the goal of this particular project hence became to establish clarity and standardize effectively to uncover key traits which related directly to vaccination status.

The scope of the analysis reached across the United States, and the overarching topic examined was which factors influenced vaccination status during the COVID-19 pandemic. Four specific questions were created which shared a connection to the general topic, and each question was investigated separately. The first of the questions examined how demographic information impacts vaccination status. Demographic factors which were examined included gender of the individual (male or female), age range into which the individual falls, and race/ethnicity. Each of these factors was plotted in its original form, standardized, and subsequently compared to the original visualization. The second question asked what characteristic influences vaccination status the most in unvaccinated households between different states. Household characteristics such as the presence of children in the household, educational attainment, marital status, employment status, and income level were explored for each state. Each characteristic was categorized and the number of unvaccinated households that matched that characteristic was counted, then standardized based on the state's total population of unvaccinated households. The

[2] Holder, J. (2021, January 29). Tracking coronavirus vaccinations around the world. The New York Times. Retrieved May 2, 2022, from https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html

[3] *Covid-19 data explorer*. Our World in Data. (n.d.). Retrieved May 2, 2022, from https://ourworldindata.org/explorers/coronavirus-data-explorer

data was visualized using map models and stacked/facet wrapped bar plots to correlate a state's vaccine hesitancy to their household characteristics. The third question took a slightly different approach and looked into what elements played a role in both vaccine distribution and social hesitancy. The third question investigated the different factors that influence social hesitancy to the COVID-19 vaccination. Along with analyzing social hesitancy in regards to the COVID-19 vaccine across the United States, multiple data sets were used to try to discover a relationship between the distribution of the first dose of the vaccine and different demographic categories. The fourth and final question investigated recovery rate from COVID-19 and how it appeared to change based on factors such as vaccination rate, political party, and education level by state. Each of these factors was plotted on the U.S map using standardized data, and subsequently compared to the original visualization.

When examining the findings of the study, clear differences in proportions emerged for the demographic analysis. The proportion of women receiving a vaccination was higher than that of men, but only by a small margin. Older age groups tended to get vaccinated at higher rates, with the 75+ age range being the most vaccinated. The vaccination rates of different races/ethnicities varied widely, with the most vaccinated being American Indian/Alaska Native Non-Hispanic and the least vaccinated being Black Non-Hispanic. Looking at vaccine hesitancy, more negative attitudes toward the vaccine seemed to correlate to an absence of children in unvaccinated households. Also within unvaccinated households, positive vaccine attitudes were directly tied to income levels less than $25,000, overall educational attainment, and unmarried status. A strong correlation was observed between average Social Vulnerability Index (SVI) and average CVAC level of concern for vaccination rollout in each state as well. However, not enough additional evidence was present to draw a concrete conclusion about this SVI analysis. Finally, an examination of recovery rates yielded the finding that states appearing to have higher recovery rates shared a tendency toward political liberalism, high proportions of residents with college education, and high vaccination rates.

RELATED WORK

In the wake of the COVID-19 pandemic, many organizations researched vaccination status in an attempt to determine ways to convince more individuals to receive a COVID vaccination. Demographic information was collected by the Centers for Disease Control and Prevention COVID Data Tracker and was depicted in various visualizations.[4] However, this data was not standardized, meaning that the provided plot did not depict the relation in terms of the proportion within a studied demographic population. For example, the plot below shows the percentage vaccinated and the percentage of the US population as two separate bars and does not merge them.[5] The problem of what relationship arises from a standardized dataset was left unsolved and was a gap within the research that our study aims to fill.

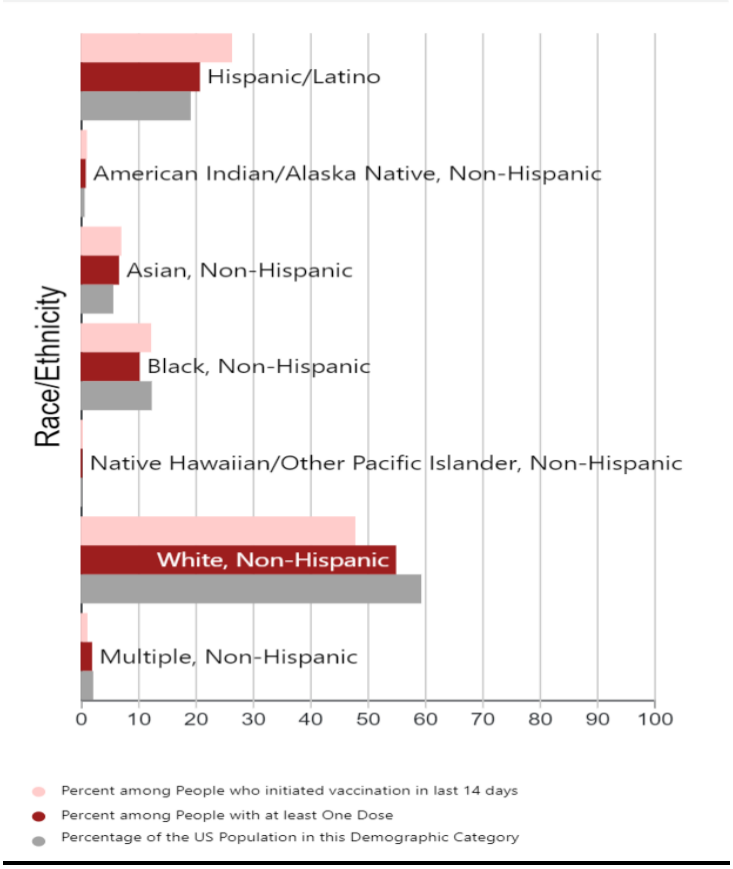


Figure 1 - Racial Vaccination Analysis Conducted by CDC

[4] Centers for Disease Control and Prevention. (n.d.). Home Page - CDC Covid Data tracker. Centers for Disease Control and Prevention. Retrieved May 8, 2022, from https://covid.cdc.gov/covid-data-tracker/#datatracker-home

[5] Centers for Disease Control and Prevention. (n.d.). CDC Covid Data tracker. Centers for Disease Control and Prevention. Retrieved May 2, 2022, from https://covid.cdc.gov/covid-data-tracker/#vaccination-demographic

These figures show a timeline of COVID-19 developments, including important events and vaccination timeline. The data is aggregated to a weekly level and is without any control variables.The solid line in each figure is a fitted linear curve between the growth rate of total cases/hospitalizations and vaccination rate.[6] This study is different because it is comparing the vaccination rate to the recovery rate (new cases/total cases) and other variables like political affiliation and education level. The data collected includes the infection and hospitalization data in the U.S. from 12 October 2020 to 7 March 202, which is older and findings might have changed with the resurgence of covid variants like omicron.

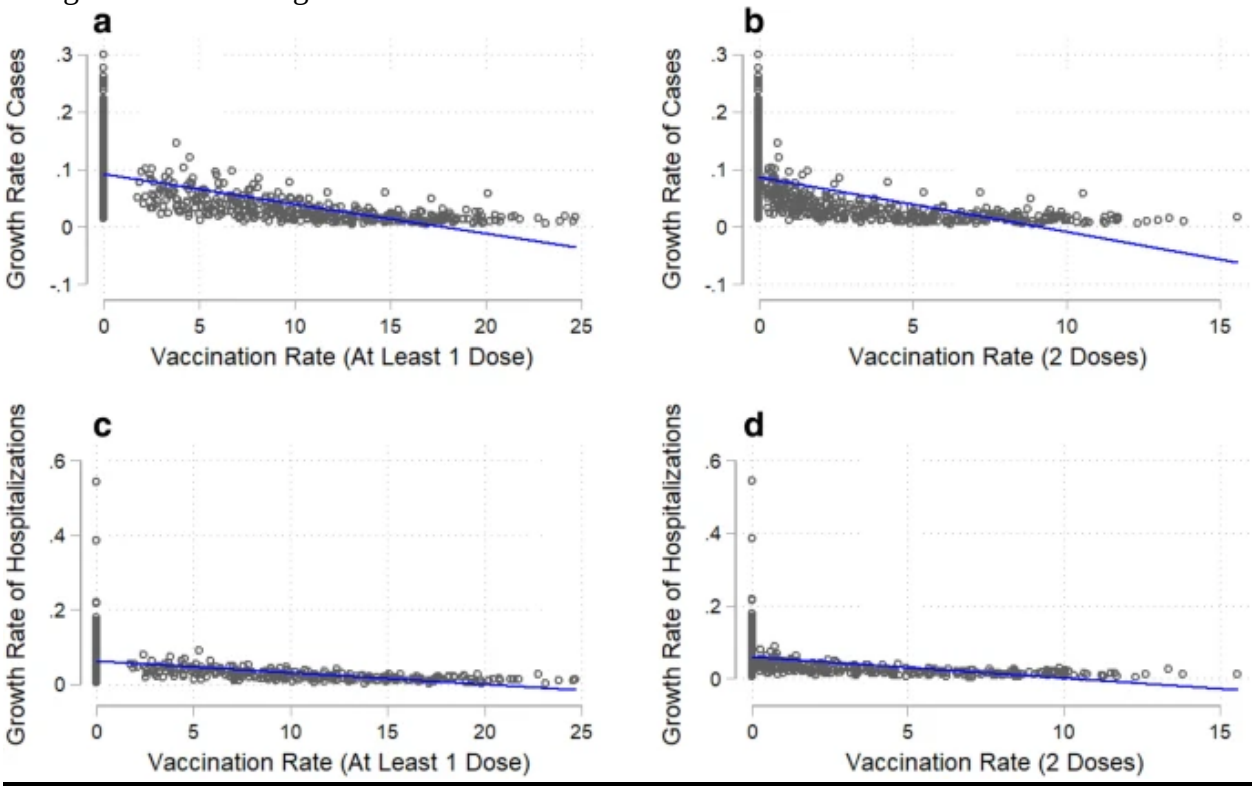


Figure 2 - These plots show the negative correlation between the vaccination rate and the growth rates of total cases and hospitalizations.

Past research suggests that there is a decline of vaccine attitudes and intentions where participants in the survey that identified as Republicans were the driving force behind this decline. Evidence depicts that differential exposure to media channels and social networks could explain the polarization between Democrats and Republicans.[7] However, the extent of this study only reached the general US and not each state within the US as well as it only studied the characteristic of political affiliation, not specifics of the participant's education, income, marital status, and employment status. Another research studied vaccine hesitancy amongst Americans

[6] Chen, X., Huang, H., Ju, J., Sun, R., & Zhang, J. (2022, January 28). Impact of vaccination on the COVID-19 pandemic in U.S. states. Nature News. Retrieved May 8, 2022, from https://www.nature.com/articles/s41598-022-05498-z

[7] Fridman, A., Gershon, R., & Gneezy, A. (n.d.). *Covid-19 and vaccine hesitancy: A longitudinal study*. PLOS ONE. Retrieved May 2, 2022, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0250123

and stated that vaccine efficacy was more influential on vaccine acceptance among whites than among blacks. Additionally, the vaccine had a stronger adverse effect on willingness to vaccinate older Americans and women. For political affiliation, Democrats were generally more sensitive to vaccine efficacy than Republicans, but both groups responded differently to endorsements of the vaccine. Attitudes toward general vaccine safety, past flu vaccination history, personal contact with severe cases of Covid could explain the variation within each group in vaccination hesitancy.[8] The difference between this study and the analysis of the subquestion "what characteristics influence vaccination status the most in unvaccinated households between different states?" is that it focuses mainly on determining the shared characteristics between those that are unvaccinated, not those already vaccinated and unvaccinated.

[8] Kreps, S. E., & Kriner, D. L. (2021, June 2). *Factors influencing covid-19 vaccine acceptance across subgroups in the United States: Evidence from a conjoint experiment*. Vaccine. Retrieved May 2, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8064867/

ANALYSIS DESIGN

A series of four questions were created, each of which covered a different component of the overall topic. Together, these questions form a comprehensive analysis of vaccination status in the United States.

The first question was as follows: How does demographic information impact vaccination status? As the term "demographic" can include several subfactors, addressing this question consisted of identifying key demographic characteristics for which data was commonly available. It was also imperative that this demographic information was more specifically accessible with regard to individuals who received the COVID-19 vaccine. Datasets were acquired which had raw data reporting race/ethnicity, age range, and gender (either male or female) for individuals who had received at least one dose of the COVID-19 vaccine. As a result, these three demographic characteristics were selected for the analysis. Each characteristic was plotted and an initial qualitative analysis was conducted to observe any general trends. From there, the data for the characteristics were standardized within R, and updated visualizations were coded. Based on these standardized analyses, more concrete conclusions were made.

The second question was as follows: What characteristic influences vaccination status the most in unvaccinated households between different states? The main goal of this study is to analyze which factors influenced vaccination status in the US during the COVID-19 pandemic and this subquestion aims to determine what characteristics do unvaccinated households share amongst those with a positive vaccine hesitancy and those with a negative vaccine hesitancy. Map visualizations of those unvaccinated that answered a survey question asking about their intention on getting vaccinated as of March 14, 2022 were made to find the vaccine hesitancy per state. Each state was then categorized by those with positive vaccine hesitancy defined as those unvaccinated that would definitely get the vaccine and negative vaccine hesitancy defined as those unvaccinated that would definitely not get the vaccine. Similarities in household statistics were compared to see if those characteristics had an effect on the vaccine hesitancy such as whether or not there were children present in the household, education level, marital status, employment status, and income level. The goal was to determine which states have the greatest positive and negative vaccine hesitancy and what characteristics they share and don't share to see if there is a correlation.

The third question was as follows: What factors influence social hesitancy regarding the COVID-19 vaccine? The main goal of this study is to mainly find relationships between different factors that might influence social hesitancy towards the vaccine, as well as determining what factors influence the distribution of the COVID-19 vaccine. I began asking myself whether different demographic categories such as age, gender, and ethnicity affected when an individual received their first dose of the vaccine. First, I plotted the distribution of the total count of people with at least one dose of the vaccine against the month-year date. I then looked at when each demographic group, age, gender, and ethnicity, received the first dose of the vaccine. After analyzing the vaccination distribution with different demographic categories, I gained background knowledge on the COVID-19 vaccine regarding when the most vaccinations took place in the United States. By analyzing data with a large scope, such as the United States, my goal was to find a correlation between social hesitancy regarding the vaccine in each state. Within each state, I began trying to find a relationship between the states whose population was strongly hesitant about the vaccine. I later analyzed and plotted the United States COVID-19 Vaccine Advisory Committee (CVAC) Level of Concern for Vaccination Rollout within each

state. Along with plotting the CVAC Level of Concern and vaccine hesitancy on a United States map, I also found the Social Vulnerability Index (SVI) and percent of adults fully vaccinated for each state.

The fourth question was as follows: How does recovery rate change with vaccination rate, political party and education level by state? The term recovery rate was defined as the number of new covid cases divided by the total number of covid cases. The objective of the question is to find a correlation between the vaccination rate and the other variables. The following requires datasets for each variable which data is organized by the U.S states for a certain period of time. Data will need to be standardized using the population for each state. Each variable will be studied separately then compared to the recovery rate to find a relevant conclusion.

ANALYSIS

Question 1: How does demographic information impact vaccination status?

For this question, three different datasets were used which all originated from the Centers for Disease Control and Prevention (CDC) COVID Data Tracker. The datasets were titled *race ethnicity of people with at least one dose administered, sex of people with at least one dose administered,* and *age groups of people with at least one dose administered*; the name of each dataset indicates what it examines. The race/ethnicity dataset was broken down into seven distinct race categories, the sex dataset into two categories (male or female), and the age dataset into eight age ranges spanning from 5-11 years of age to 75 and older. These datasets in particular are relevant to the first question of the analysis because they provide a comprehensive examination of key demographic characteristics which may differentiate the typical vaccinated individual from the unvaccinated individual. As a continuation of the discussion of the datasets, the relevant variables for investigation were drawn directly from the same demographic traits; these variables consisted of age range, race/ethnicity, gender, proportion of individuals receiving at least one dose within the overall vaccinated population, and proportion of individuals receiving at least one dose within their respective age/gender/race. It was determined that these variables were the most applicable to a demographic analysis after doing preliminary research on which demographic characteristics are the most insightful when conducting a demographic-based study. Gender, age, and race/ethnicity appeared most frequently and were prominent across a wide variety of demographic research. Through a detailed investigation of these three key factors, it was hoped that the role of demographics in vaccination status would be revealed.

None of the datasets used to answer this question had any missing values, so it was not necessary to use the na.rm = TRUE command. However, it was suspected that some of the data points could be outliers given the skew of the data. For instance, all of the proportions vaccinated in the age range data were above 0.65 with the exception of one age group (5-11 years) which was well below. While a lower proportion vaccinated was expected for this specific age range, the enormity of the difference in proportions suggested there could still be an outlier present. To examine outliers in each of the datasets, the *quantile()* command was used to record the values for quartile 1, quartile 3, and the interquartile range (quartile 3 - quartile 1). The following formulas were used to calculate the lower and upper limits upon which to base conclusions about outliers:

*Lower Outlier Bound = Quartile 1 - (1.5 * Interquartile Range)*
*Upper Outlier Bound = Quartile 3 + (1.5 * Interquartile Range)*

It was observed that the proportion recorded for the 5-11 years age group was in fact an outlier according to these formulas. There were no outliers within the race/ethnicity data, nor were there any within the gender data. To explore the distribution of each variable, it was decided that scatter plots made the most sense, as they allow for a clear comparison between a categorical (age range, gender, race/ethnicity) and continuous (proportion receiving at least one dose) variable. With the scatter plots, it was easy to decipher which proportions were the largest and smallest. For the gender data, while a scatter plot was created, it was decided that the table was more insightful since gender only had two different categories.

The primary strategy used to transform the data was to create new, standardized variables from the originally unstandardized variables within the three principal datasets. Standardization is an essential transformation because without it, the results gained will neither be accurate or indicative of any greater conclusion.

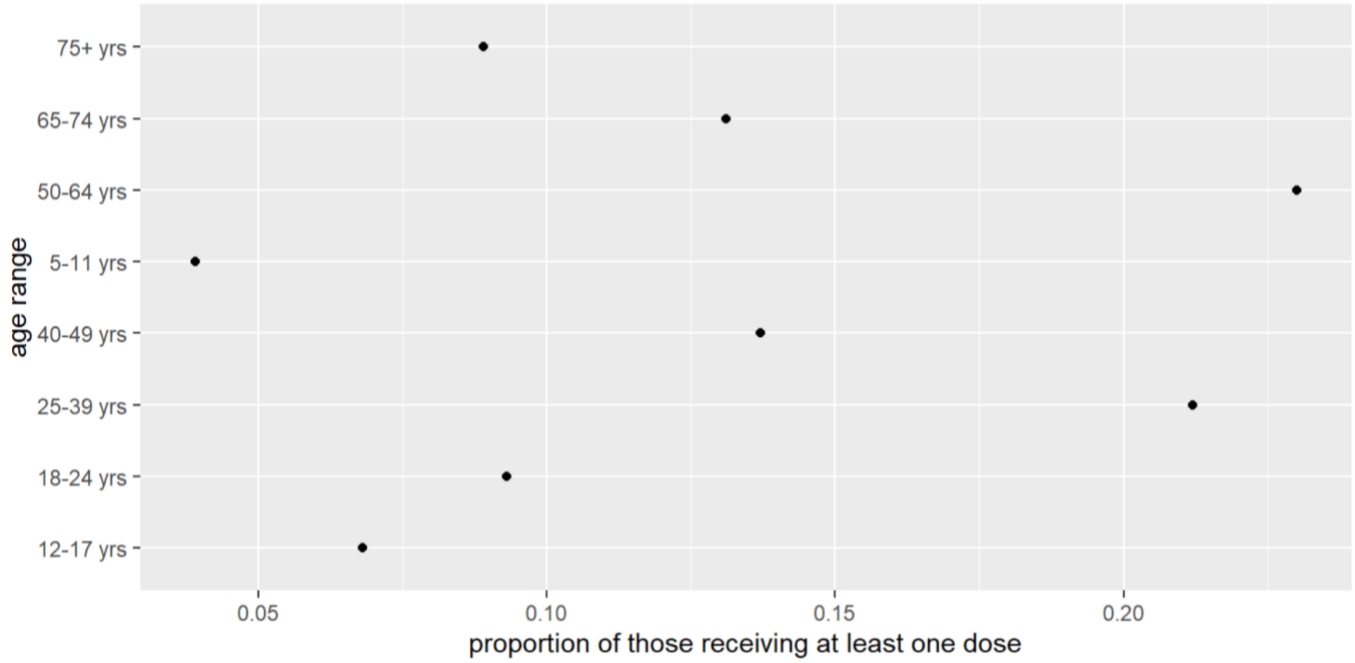


Figure 3 - Distribution of the Proportion of People with at Least One Dose by the Age Range

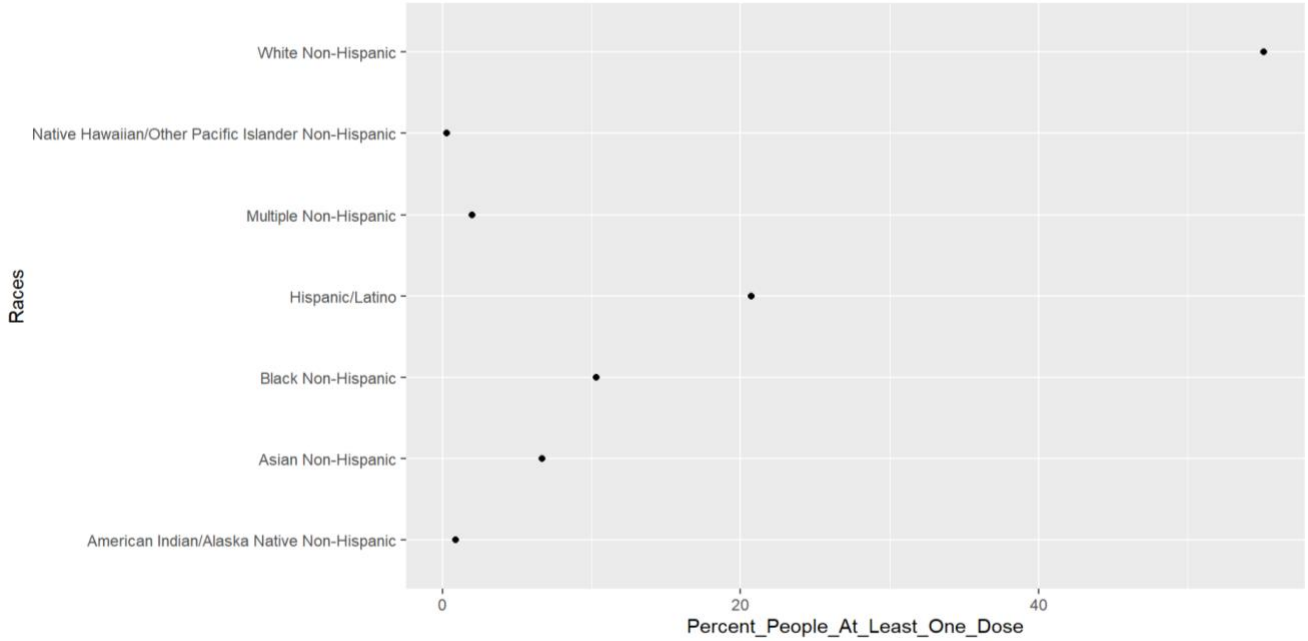


Figure 4 - Distribution of the Proportion of People with at Least One Dose by the Race

RStudio: Notebook Output

	0.478	0.522
Female	0	1
Male	1	0

Figure 5 - Table Demonstrating Proportion of Vaccinated Population Belonging to Each Gender

In order to standardize the data, new variables were created which found the proportions within each individual specific racial group, gender and age range; these normalized proportions differed significantly from the proportions within the number of individuals receiving at least one dose of the COVID-19 vaccine. The total number of people in the United States was obtained from the World Bank, and was then multiplied by the percentage which each age group/gender/race comprises of the total US population. "% US Population" was a variable available within each of the three datasets for every age range, gender, and race/ethnicity. In this way, the number of individuals from each specific category within the larger United States population was found. The number of individuals from each specific category who received at least one dose was available in each dataset as "# People at least One Dose". The variable "# People at least One Dose" was divided by the total number within the United States population for each demographic trait. In this way, all data was standardized.

Once standardization was complete, the exploratory analysis was continued by creating updated visualizations representing models of the newly coded variables.

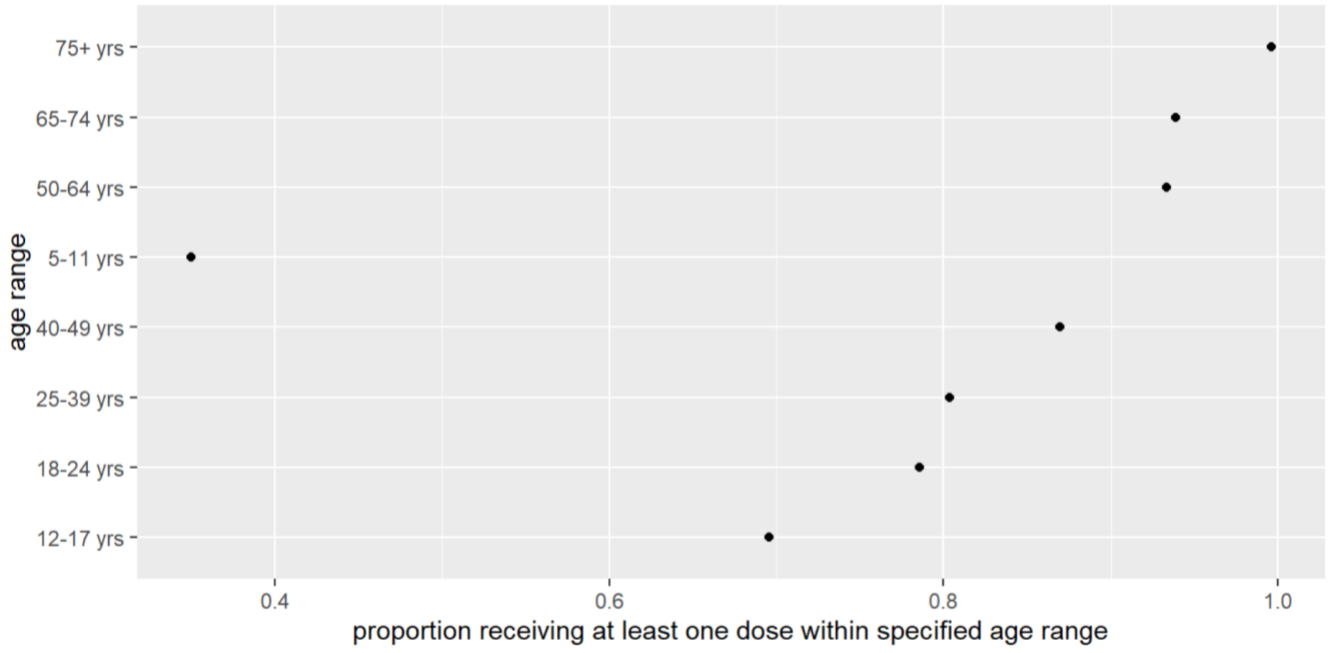


Figure 6 - Distribution of the Proportion of People with at Least One Dose by Age Range

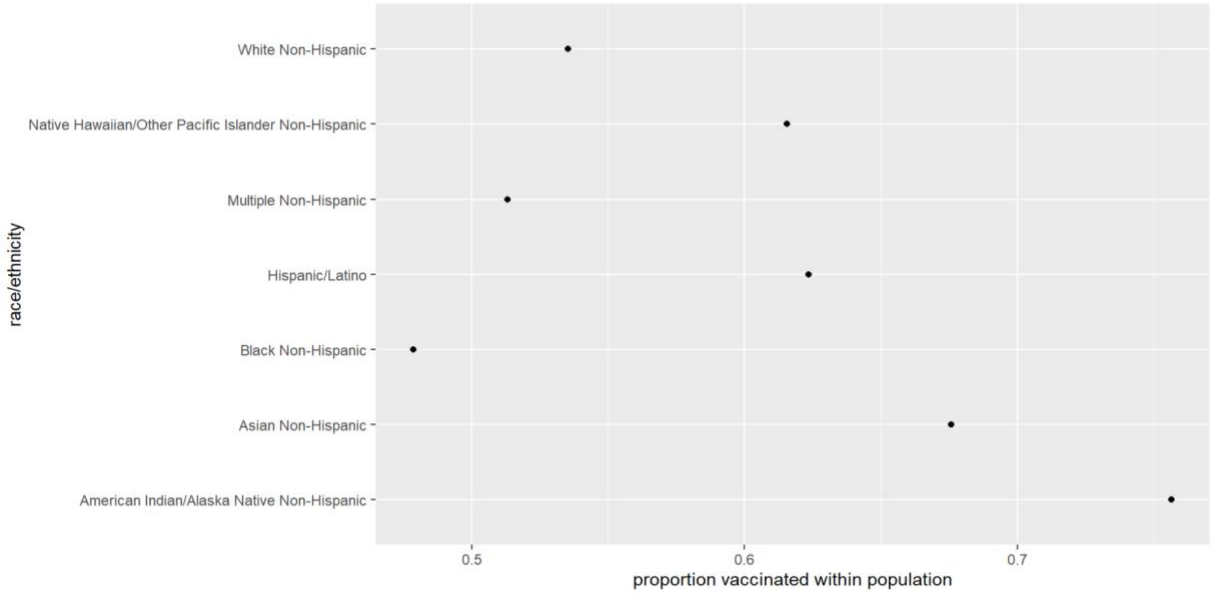


Figure 7 - Distribution of the Proportion of People with at Least One Dose by Race

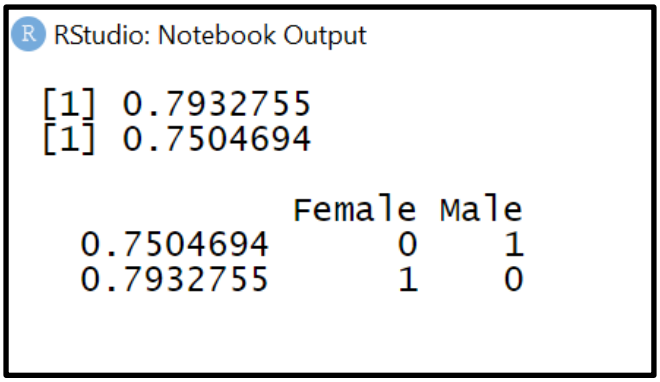


Figure 8 - Table Demonstrating Proportion Vaccinated Within Each Gender

Question 2: What characteristic influences vaccination status the most in households between different states?

The main dataset used for this question was sourced from the United States Census Bureau in the Household Pulse Survey released weekly from May 5, 2020 to the present day. The public use file from the Phase 3.4 PUF Releases Week 43 Household Pulse Survey: March 2 - March 14 was utilized because within the dataset there are variables that ask the surveyee the state they live in, their intention on getting a vaccine, the number of people under 18 years-old living in their household, their highest degree or level of school they've completed, their marital status, employment status for the last seven days, and total household income before taxes. The dataset was chosen because it was the most recent data retrieved at the time analysis began and it

has been over a year since the vaccine first came out in December 11, 2020, so it was assumed that there would be a clear reason as to why there still existed unvaccinated households in the US other than inaccessibility to the vaccine. Each variable accounts for the characteristics of the unvaccinated household the question aims to study and provides insight on the vaccine hesitancy per state.

The variables that were determined to be relevant were *EST_ST*: state, *GETVACRV*: intention on getting the vaccine, *THHLD_NUMKID*: total number of people under 18 years-old in the household, *EEDUC*: educational attainment, *MS*: marital status, *ANYWORK*: employment status for the last seven days, and *INCOME*: total household income. The miscellaneous values were -88 (missing/did not report) and -99 (question seen but category not selected) which was handled by replacing those values with NA. Additionally, the District of Columbia was removed from analysis as the usmap package did not support modeling the 50 states and the District of Columbia.

Variable	**Variable Categorization**
GETVACRV	1) Definitely get a vaccine 2) Probably get a vaccine 3) Be unsure about getting a vaccine 4) Probably NOT get a vaccine 5) Definitely NOT get a vaccine
THHLD_NUMKID	Number of people under 18 (whole number)
EEDUC	1) Less than high school 2) Some high school 3) High school graduate or equivalent (for example GED) 4) Some college, but degree not received or is in progress 5) Associate's degree (for example AA, AS) 6) Bachelor's degree (for example BA, BS, AB) 7) Graduate degree (for example master's, professional, doctorate)
MS	1) Now married 2) Widowed 3) Divorced 4) Separated 5) Never married
ANYWORK	1) Yes 2) No
INCOME	1) Less than $25,000 2) $25,000 - $34,999 3) $35,000 - $49,999 4) $50,000 - $74,999 5) $75,000 - $99,999 6) $100,000 - $149,999 7) $150,000 - $199,999 8) $200,000 and above

Figure 9 - Table showing all of the variables and their categorization

In the R environment, the libraries tidyverse, dplyr, and usmap were installed and the Household Pulse Survey of Week 43 was read in as the *week_43* variable. The state variable column labeled each state using the numerical system and had to be renamed to each state. The names of each state and the number it corresponded to was found in the data dictionary for the survey. Other organizations included creating a vector of the state names as well as a vector of the total unvaccinated household population in each state. The initial analysis covered the overall vaccine hesitancy of each state and the values of the *GETVACRV* variable varied from 1-5 (Figure 10) with 1 representing the surveyee's intention on definitely getting the vaccine if it was available and 5 representing definitely not getting the vaccine. The variable was recategorized down to two levels of positive vaccine hesitancy that contained levels 1 & 2 and negative vaccine hesitancy that contained levels 4 & 5. The number of unvaccinated households that fit in each category was counted separately for each state, then this count was standardized by dividing the count by the total unvaccinated population in each state. The data was standardized this way in order to consider the percentage of those with a characteristic amongst the unvaccinated households as this is the aim of this question. A new data frame was created to clean the data down to the columns that were needed to visualize through a map model using the usmap package (Figures 10 and 11 found below). The scale fill was continuous with the lower percentage on the lighter end of the color scale and the higher on the darker end.

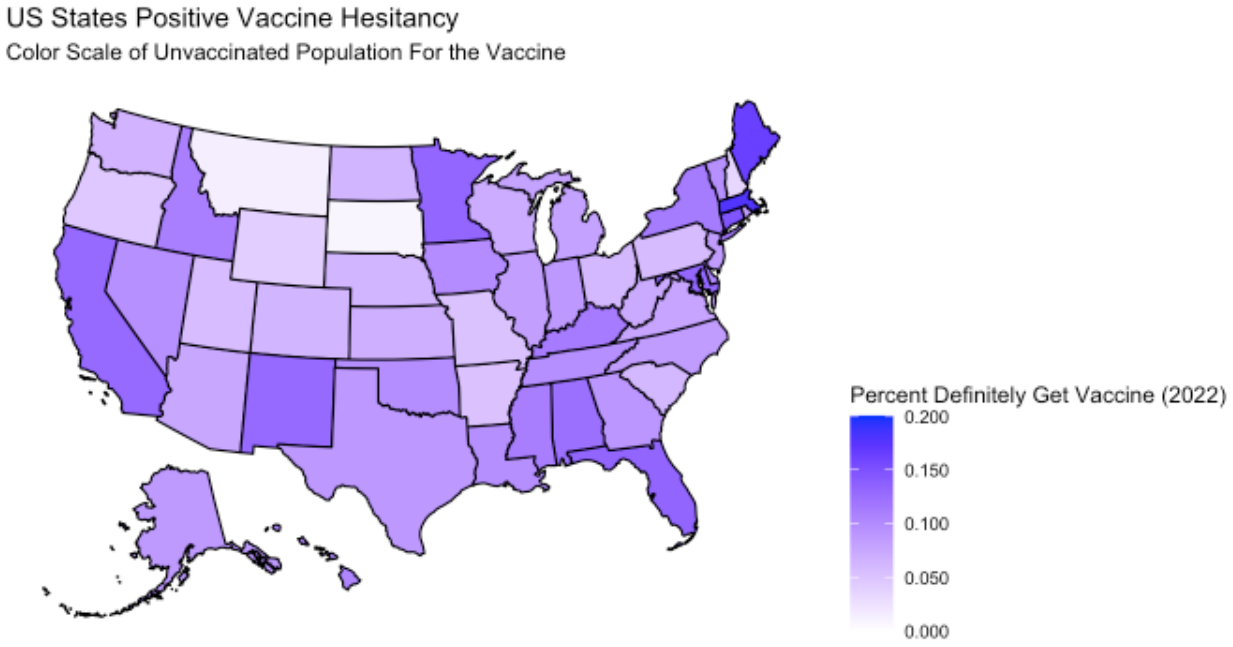


Figure 10 - U.S Map showing the positive hesitancy of taking the vaccine

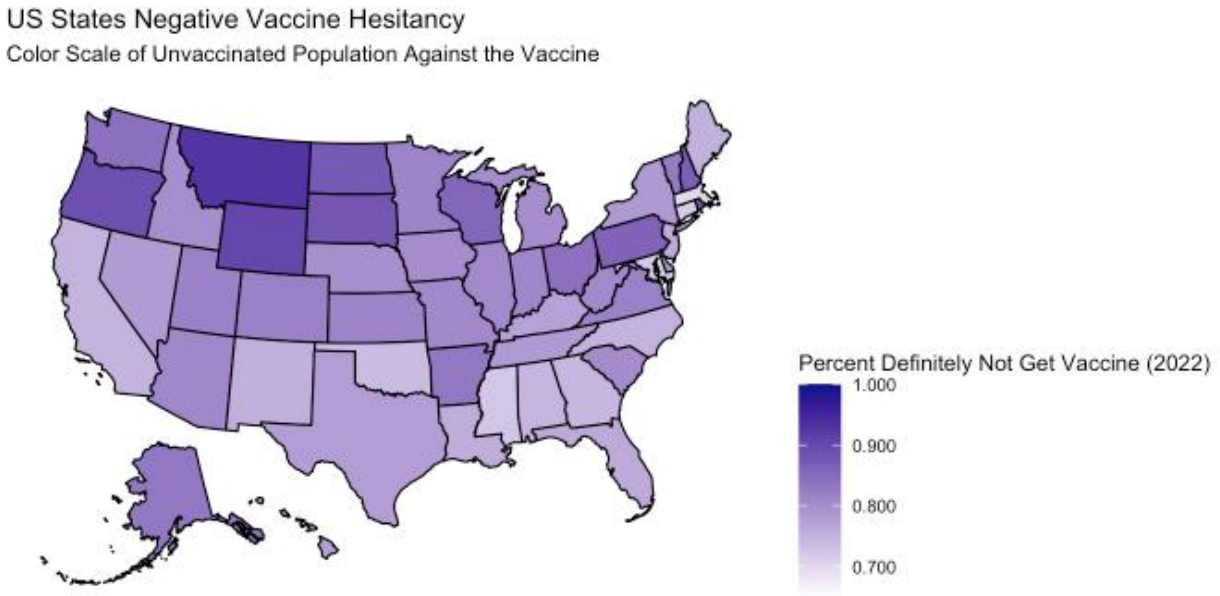


Figure 11 - U.S Map showing the negative hesitancy of taking the vaccine

The secondary analysis covered the presence of people under 18 years of age in the household characteristic. The number of unvaccinated households with kids present was counted for each state through a conditional sum function, and then the count was standardized. A new data frame was created to plot this analysis on a map visualization and the color scale fill portrayed high percentages as a darker red and the lower percentages on the lower end of the color scale (Figure 12 found below).

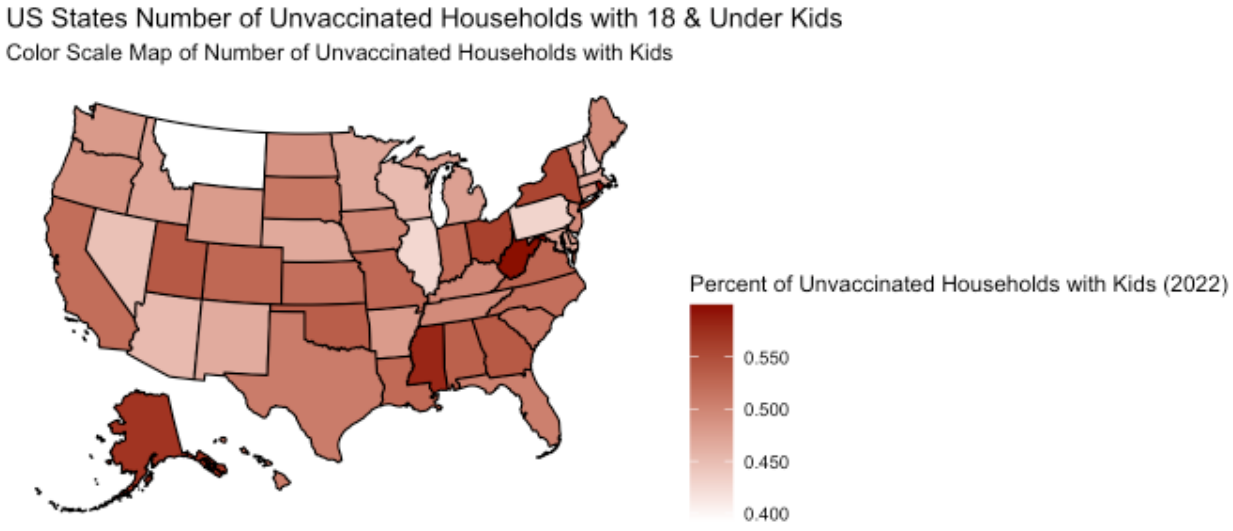
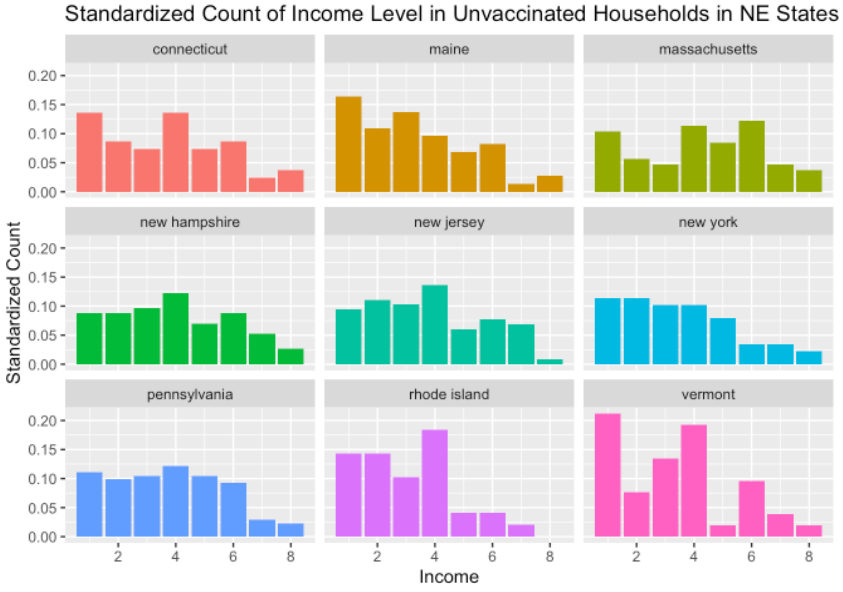
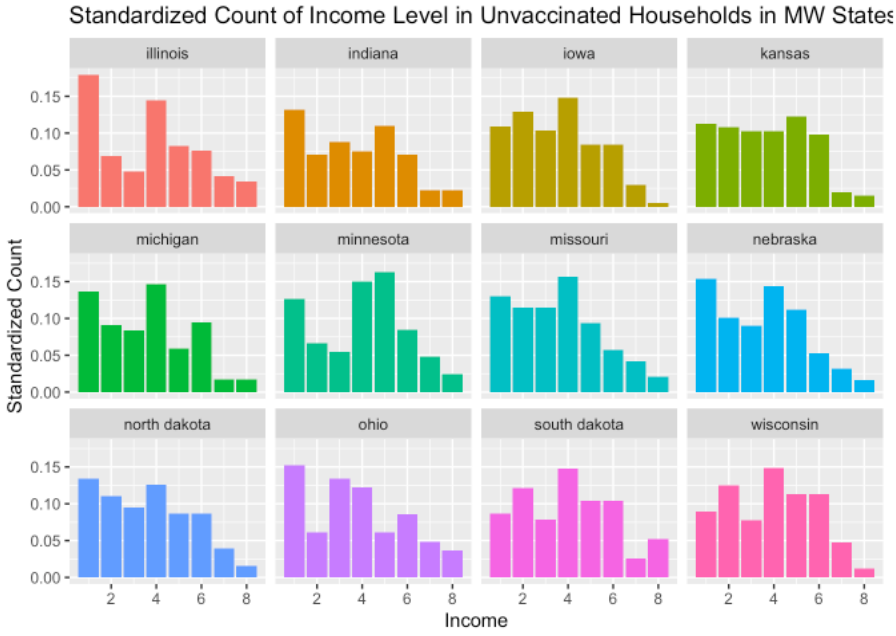


Figure 12 - U.S Map showing the number of unvaccinated households with 18 & under kids

Tertiary analysis focused on the household characteristics of educational attainment, marital status, employment status, and income level within each state. For the tertiary analysis, the data was cleaned in a different strategy than compared to the initial and secondary analyses by subsetting the *week_43* dataset to include the households that answered the intention on

getting the vaccine survey question as those who answered this question did not get the vaccine. The outliers were determined to be the vaccinated households and were removed by subsetting the data. To summarize the subset for each characteristic, a new variable was initialized to edit the subset four times and a column was added to each subset called population to include the total unvaccinated population for each state. Using the count and population column, the standardized count was created by dividing the count column by the population column. For educational attainment, the levels of education scaled from 1-7 (Figure 13.1-13.4) and the education subset was grouped by state and education level to summarize them by count. The count was then standardized into a new column and the ggplot package was used to create a facet wrap of the standardized count of the education level in each state based on their general region of Northeast, MidWest, South, and West.





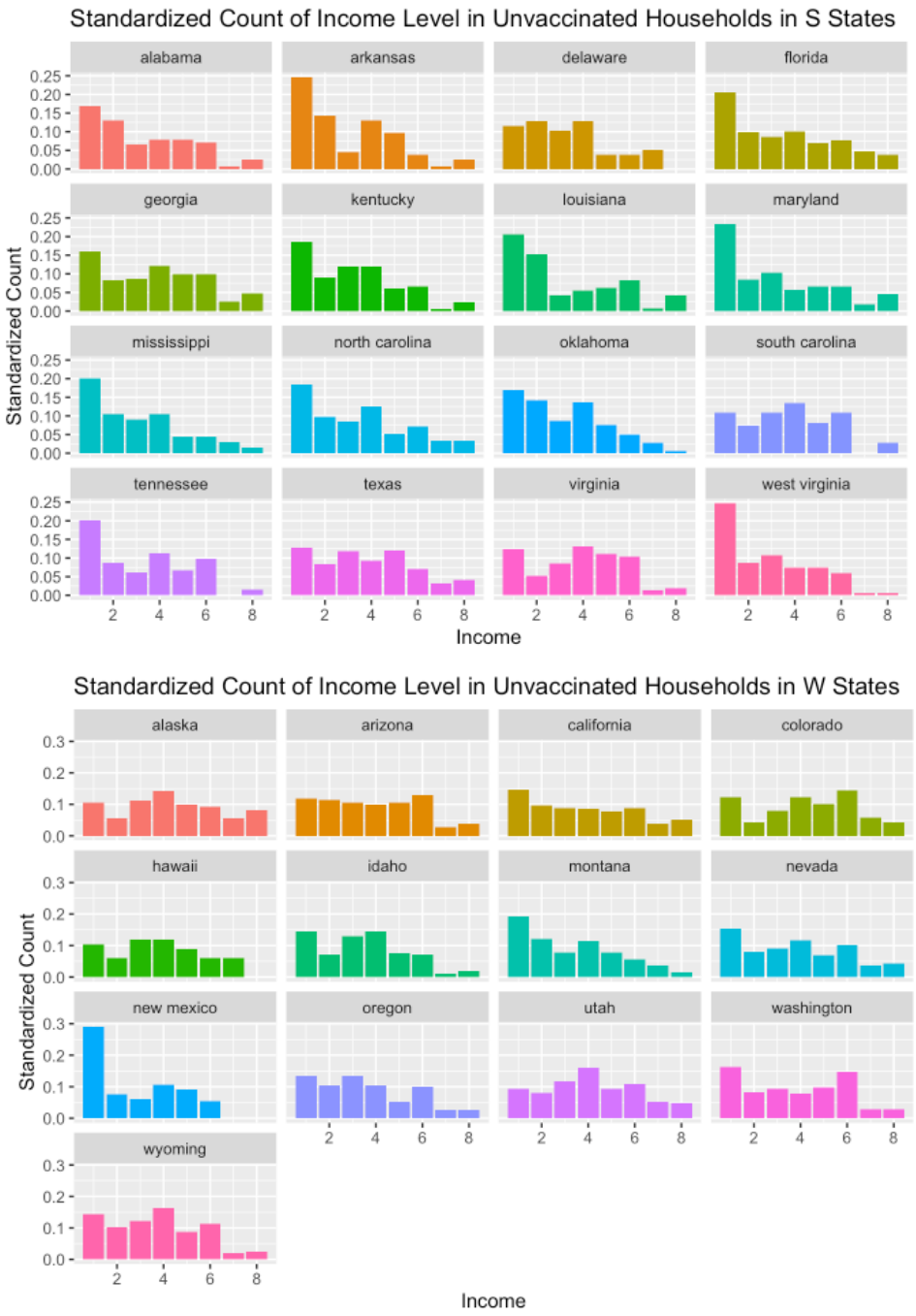


Figure 13.1-13.4 - Standardized count of educational attainment of each US region (NE: Northeast, MW: Midwest, S: South, W: West) in unvaccinated households.

For marital status, the levels of marital status was categorized from 1-5, however the data was cleaned to include married, previously married, and not married households which was visualized through a stacked barplot (Figure 14).

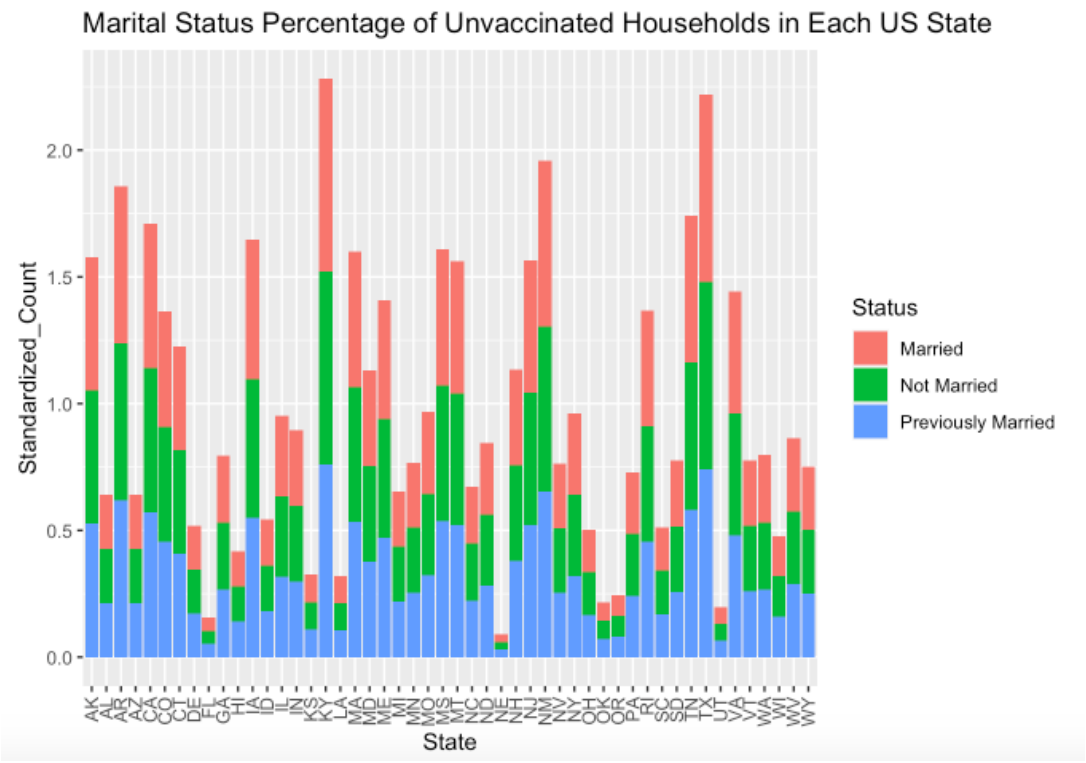


Figure 14 - Bar graph showing the marital status of unvaccinated households by state

For employment status, a vector was created that counted the number of unvaccinated households that were employed in the last seven days which was standardized by dividing by the total unvaccinated households in each state. A data frame was created to plot the map visualization using usmap with a continuous color scale fill with the lower percentage on the lighter end of the color scale and the higher on the darker end (Figure 15).

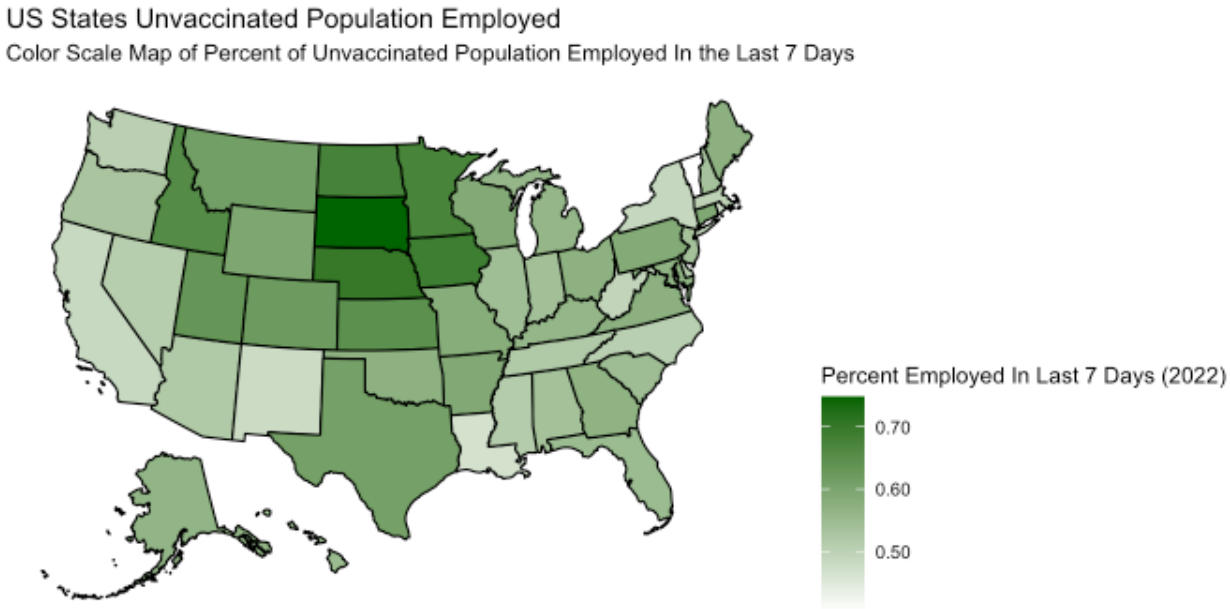
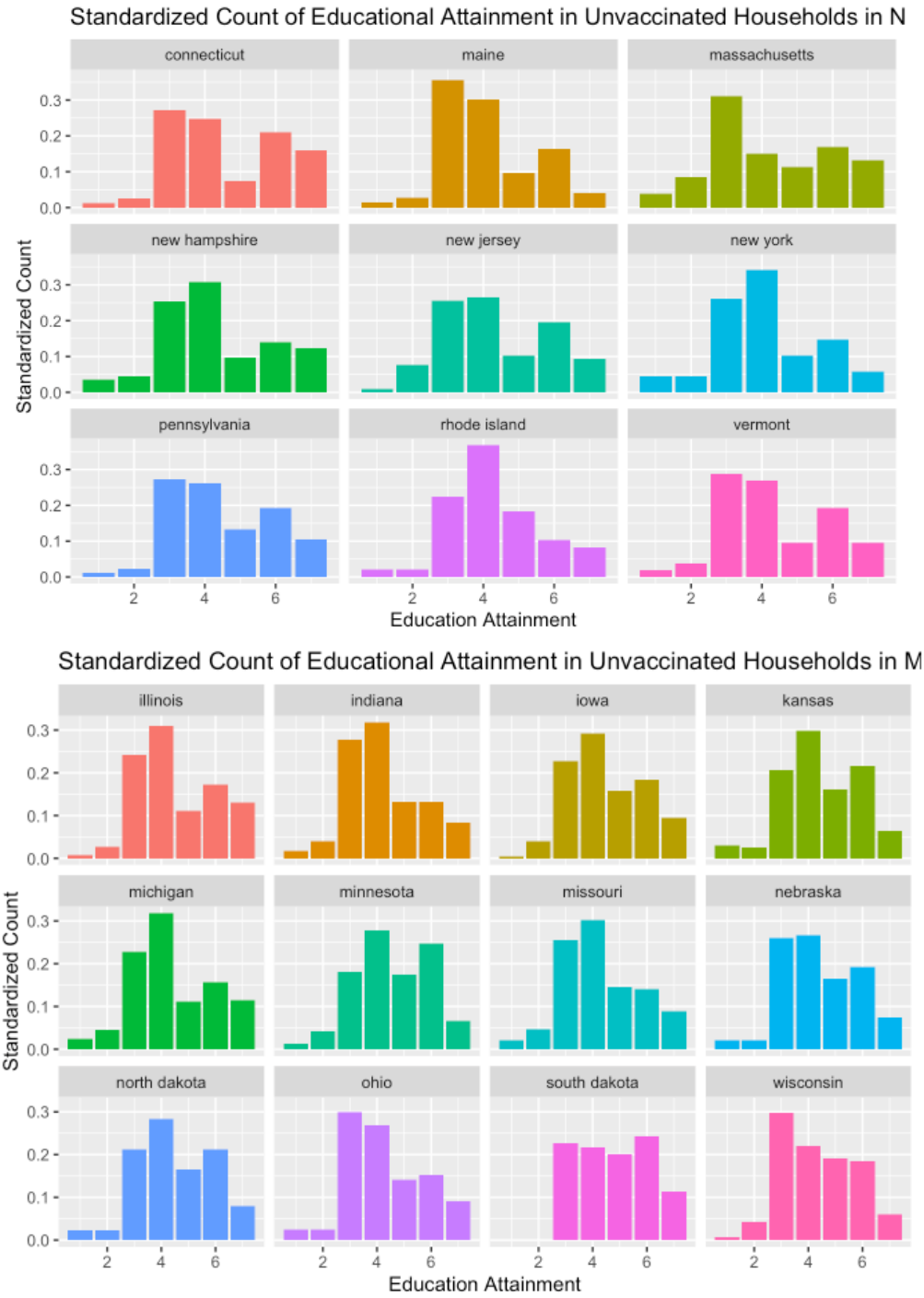


Figure 15 - U.S Map showing the unvaccinated population employed by state

For income level, the income levels varied from $25,000 to $200,000 which was categorized into eight groups. The subset was grouped by state and income level and summarized by count. The count was then mutated into a new standardized count column and the ggplot package was used to create a facet wrap of the standardized count of the income level in each state based on their general region of Northeast, MidWest, South, and West (Figure 16.1-16.4).



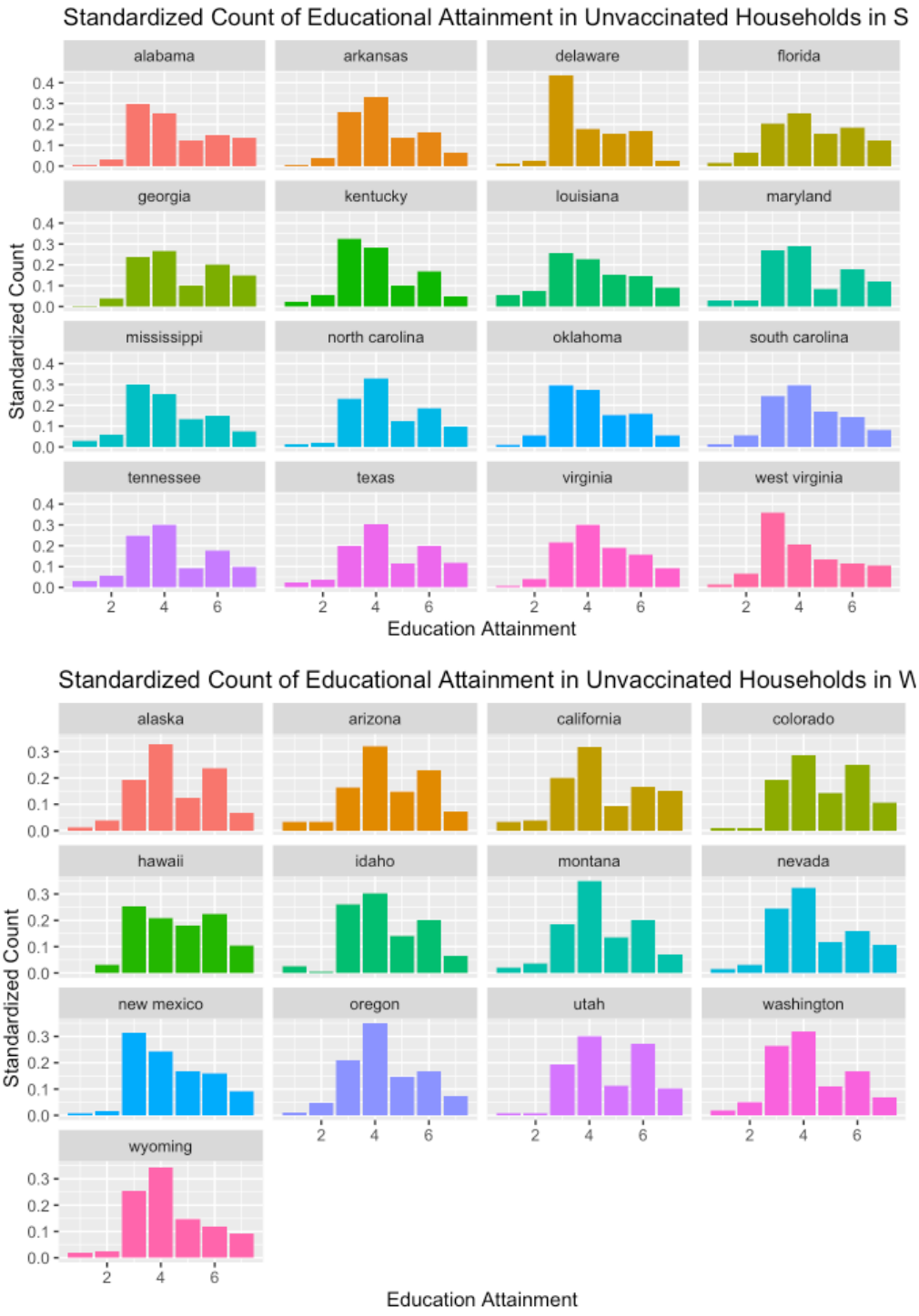


Figure 16.1-16.4 - Standardized count of income level of each US region (NE: Northeast, MW: Midwest, S: South, W: West) in unvaccinated households.

Question 3: What factors influence social hesitancy regarding the COVID-19 vaccine?

The first dataset used describes the overall demographic characteristics of people receiving COVID-19 vaccinations in the United States, with all vaccine partners represented within the data. This data set from the Centers for Disease Control and Prevention (CDC), was evaluated to see when different demographic groups got vaccinated between 2020 to 2022. The goal of this analysis is to determine if there was a difference between when each demographic group, such as age, gender, and ethnicity, received their first dose of the COVID-19 vaccine.

The next dataset used during the analysis process was again from the CDC, where data was presented for each county and state describing individuals intention to vaccinate between April 14-26, 2021. As this data set was created closer to the beginning of the COVID-19 pandemic, the goal of the analysis of this data set was to see the beginning of COVID-19 hesitancy trends throughout each state.

The dataset was reorganized to graph the dates when the first dose of vaccination was given. A formatted data column was added to the dataset to better graph the dates of vaccinations. A new dataset was created with the newly formatted dates, as well as two new variables created using the mutate() function. The dataset was then grouped by the two new variables, month and year. With the new dataset "month_year_data", the total count of people with at least one dose of the COVID-19 vaccination was plotted against the month they were vaccinated. The plot was facet wrapped in respect to the year they were vaccinated (Figure 17).

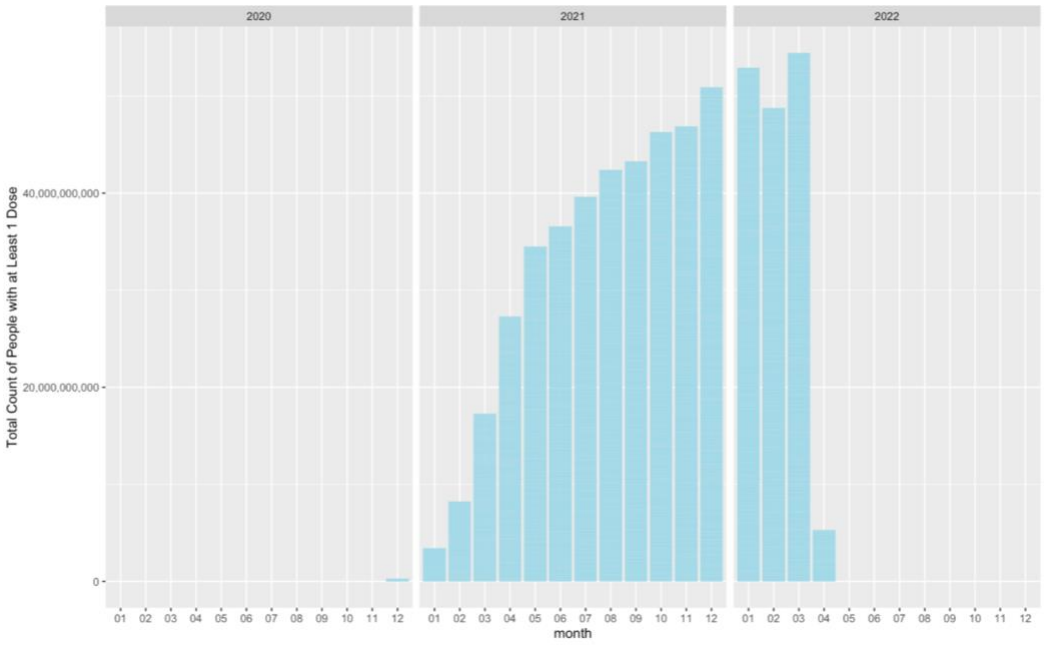


Figure 17 - Graph showing the distribution of the total count of people in the US with at least one COVID-19 vaccination dose between 2020 and 2022

Through Figure 17, it is apparent that the number of people vaccinated with at least one dose increased over time, with the majority of vaccinations occurring between late 2021 to 2022. Based on the graph, it seems there are unusual data values during April of 2022. As this dataset is updated regularly, the last data was from the beginning of April 2022. Therefore, not all vaccinations during April of 2022 were reported as the analysis process began during early April 2022. The trends of this distribution show that the total count of people with at least one dose of the COVID-19 vaccine should be closer to 50,000,000 people.

In order to find the distribution of each demographic group. A new data set was made for gender, age, and ethnicity. For gender, the dataset "gender_data" was created and the droplevels() function was used to keep the variables 'Sex_Female', 'Sex_male', and 'Sex_unknown'. The graph below shows the total count of people with at least one dose of the COVID-19 vaccine against the month each person was vaccinated, facet wrapped with the year of vaccination. The distribution of vaccinations among females and males do not have any apparent differences between when each gender was vaccinated.

Figure 18 - Graph showing the total count of people in the US with at least one COVID-19 vaccine dose between 2020 and 2022 in respect to gender

As explained earlier, a new data set was also made for age and ethnicity. For age, the droplevels() function was used to eliminate demographic categories that didn't relate to age. Only specific demographic categories that represented age were used in this dataset, as there were different variables that included age ranges that were already represented in other variables. Based on Figure 19, it is apparent that those who are 12 years and younger received the first dose of vaccination later than older age groups.

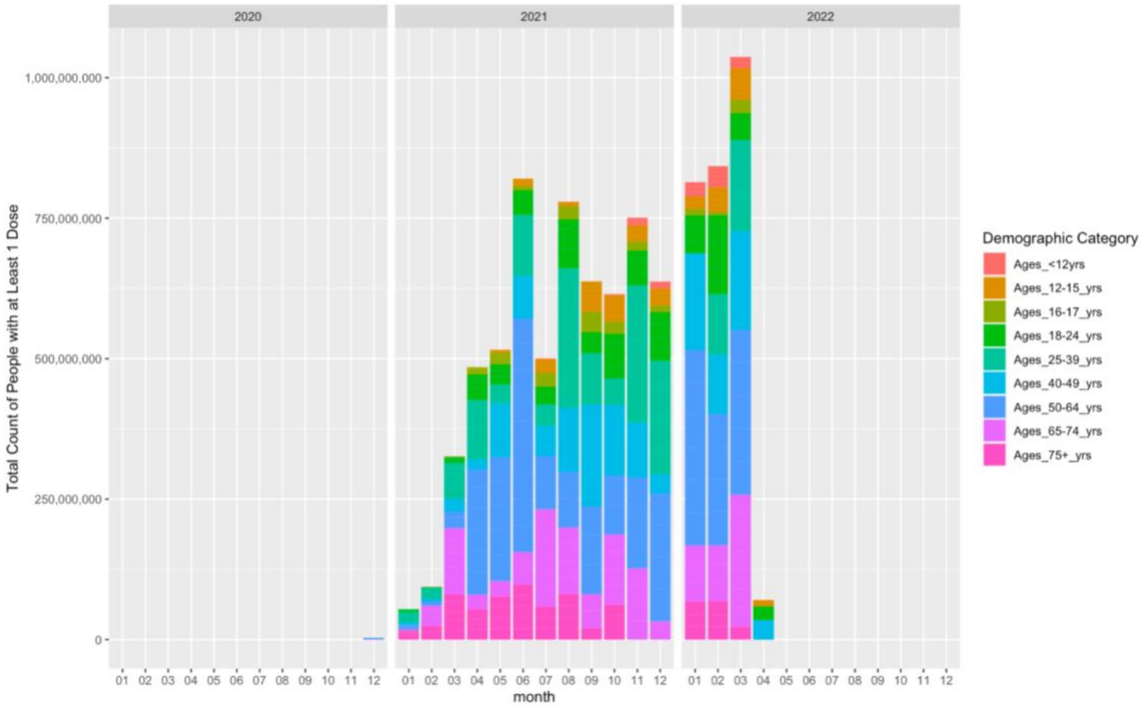


Figure 19 - Graph showing the total count of people in the US with at least one COVID-19 vaccine dose between 2020 and 2022 in respect to age

Looking at the distribution of the total count of people in the US with at least one dose of the vaccine with different ethnicities, such as Hispanic, Asian, Black, and White Americans, it is more difficult to see a correlation. The population of White Americans is much greater than Hispanic, Asian, and Black Americans, which you can see in Figure 20. It is difficult to draw a conclusion based on the graph as each ethnicity showed a significant amount of vaccinations. However, it is surprising to see that individuals that identify as Hispanic did not have vaccinations during June of 2021.

Figure 20 - Graph showing the total count of people in the US with at least one COVID-19 vaccine dose between 2020 and 2022 in respect to ethnicity

Taking this analysis a step further, a second dataset was used from the CDC that involved American's hesitancy towards the vaccine. First, a new dataset was created where the data was grouped by each state in the US. The summarize() function was used to create new variables finding the average estimated hesitancy, using the mean() function, where missing values were removed by setting na.rm equal to TRUE. Again, using the summarize() function, the mean Social Vulnerability Index (SVI) and the mean CVAC level of concern for vaccination rollout were calculated. For reference in regards to the CVAC level of concern, 0 is the lowest concern and 1 is the highest concern.

As stated earlier, this dataset depicted not only the hesitancy regarding vaccination, but it also provided data to see the percent of adults fully vaccinated against COVID-19 as of June 10, 2021 in each county of the United States. Along with this data, the CDC also provided the percentage of fully vaccinated adults within different ethnicities. The first data set helped better display the distribution and timeline of the vaccination process in the US. However, with this new data, it will be more helpful to find correlations and patterns between hesitancy, vaccination rollout, and vaccination status among ethnicities at the beginning of the pandemic.

To find the percentage of individuals who were strongly hesitant to take the vaccine, the mean() function was used for 50 variables representing each state. Because the dataset was bound by county rather than state, this was necessary in order to plot those who were strongly hesitant on a graph of the United States. After making 50 new variables, a new data frame was created with the averages of each state's population who were strongly hesitant to take the vaccine and the corresponding states. Each average percent was rounded to 3 decimal places to help better visualize and read the graph below (Figure 21). Unfortunately, New Mexico had no

data in regards to vaccine hesitancy in the case of those who were strongly hesitant. However, it is evident that states such as Montana, Alaska, and Wyoming had a large percent of individuals who were strongly hesitant in regards to the vaccine.

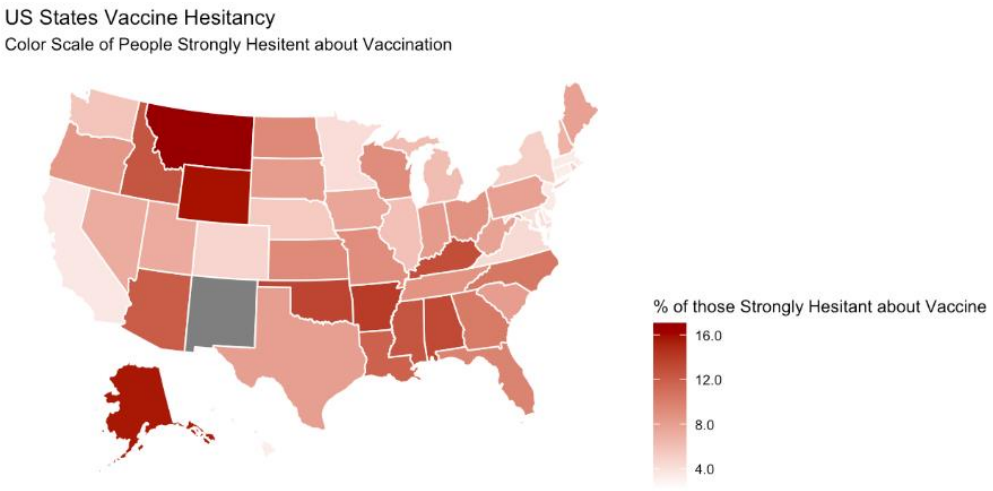


Figure 21 - Color scale US map showing the average percentage of individuals who are strongly hesitant in regards to the COVID-19 vaccine by state

The same process was followed by creating 50 new variables to find the average CVAC level of concern for vaccination rollout for each state. Again, a new data frame was created with each state and its corresponding CVAC average level of concern. Based on the graph pictured below, southern states such as Texas, Mississippi, and Arizona have high CVAC levels of concern for vaccination rollout. With New Mexico having missing data, the CVAC level of concern can not be pictured in the figure.

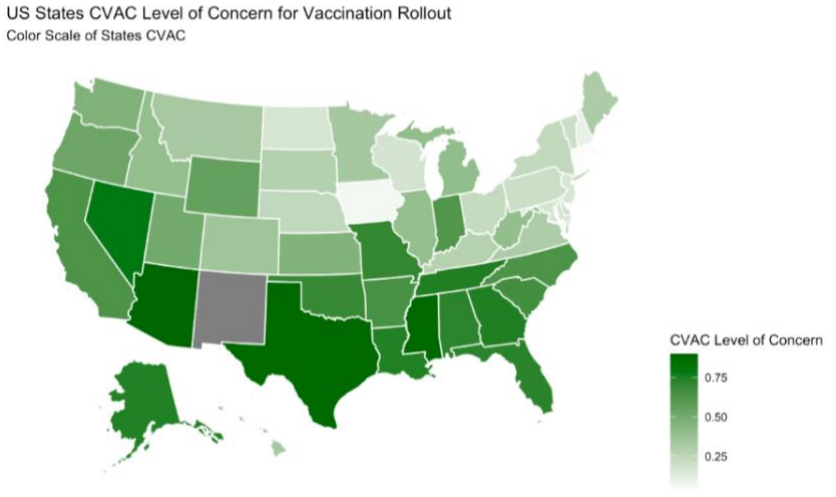


Figure 22 - Color scale US map showing the average CVAC level of concern for vaccination rollout in regards to the COVID-19 vaccine by state

The same process was again repeated, where 50 new variables were made to find the social vulnerability averages for each state. A new data frame was created with the states and their corresponding SVI averages. Here, when the SVI is closest to 0, the state is least vulnerable to potential negative effects caused by external stresses on human health. Whereas states like Texas, Mississippi, and Arizona have a high SVI, meaning they are most vulnerable to potential negative effects.

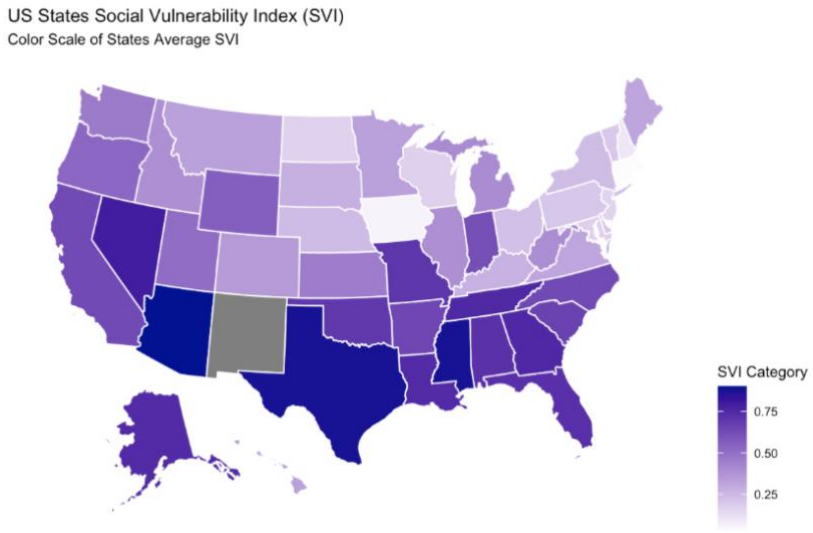


Figure 23 - Color scale US map showing the average Social Vulnerability Index (SVI) by state

After plotting US States average SVI, average CVAC Level of Concern for Vaccination Rollout, and average Vaccine Hesitancy, it was an essential part of this analysis to find a correlation between these different variables. The correlation between "avg_estimated_strongly.hesitant" and "avg_SVI" was 0.307. This means that there is a slight correlation between the average Social Vulnerability Index and average percent of those who are strongly hesitant regarding the vaccine in each state. Again, looking at the average percent of those who are strongly hesitant about taking the vaccine, there was a 0.536 correlation between those strongly hesitant and average CVAC level of concern for vaccination rollout. The correlation between "avg_SVI" and "avg_CVAC" was 0.759, meaning there is a strong, direct relationship between the average SVI and average CVAC level of concern for vaccination rollout in each state.

Looking at the overall vaccination distribution, it was intuitive to look into the percent of fully vaccinated individuals in each state. To do this, a new dataset was created, titled "new_hes_df". In this new data set, the data was grouped by state, missing values were dropped by using drop_na(), and the average of fully vaccinated adults in the US as of June 10, 2021 was found using the summarize() function. The percent average of fully vaccinated adults within each ethnicity in each state was also found. The averages found were for percent Hispanic, percent American Indian/Alaska Native, percent Asian, percent Black, percent Native Hawaaiin/Pacific Islander, and percent White. As the percent of each ethnicity was for county population, it was necessary to average the percentages to find vaccination values for each state. This data did not need to be standardized due to the fact that the CDC dataset for each county was based on percentages, not population. Six new columns were also created in this dataset by using the cut()

function. These columns contained the range of percentages of those who were fully vaccinated in each state.

In order to tidy the data, the first 8 columns of the "new_hes_df" data frame were selected, missing values were dropped, and two datasets were combined using the pivot_longer() function to create 4 new columns: state; the percent of adults fully vaccinated as of June 10, 2021; ethnicities; the percent of fully vaccinated individuals based on the ethnicity. With this new data frame, the percentage of those fully vaccinated for ethnicity in each state is depicted in the graph below (Figure 24). This figure shows how each State was graphed against the percent of fully vaccinated in terms of ethnicities. In order to create this visualization, the geom_col() function was used, the categorical variable, "ethnicities", filled the plot, and facet_wrapped was used to plot six separate distributions in one image. Based on the figure below, it is evident that the largest percent of American Indians that were vaccinated were in Arkansas with over 30% of American Indians fully vaccinated. In New Mexico, over 45% of Hispanics were fully vaccinated, meaning that the Hispanic population makes up for almost half of all fully vaccinated adults in New Mexico. The Pacific Islander population makes up only 0.5% of all fully vaccinated adults in Arkansas.

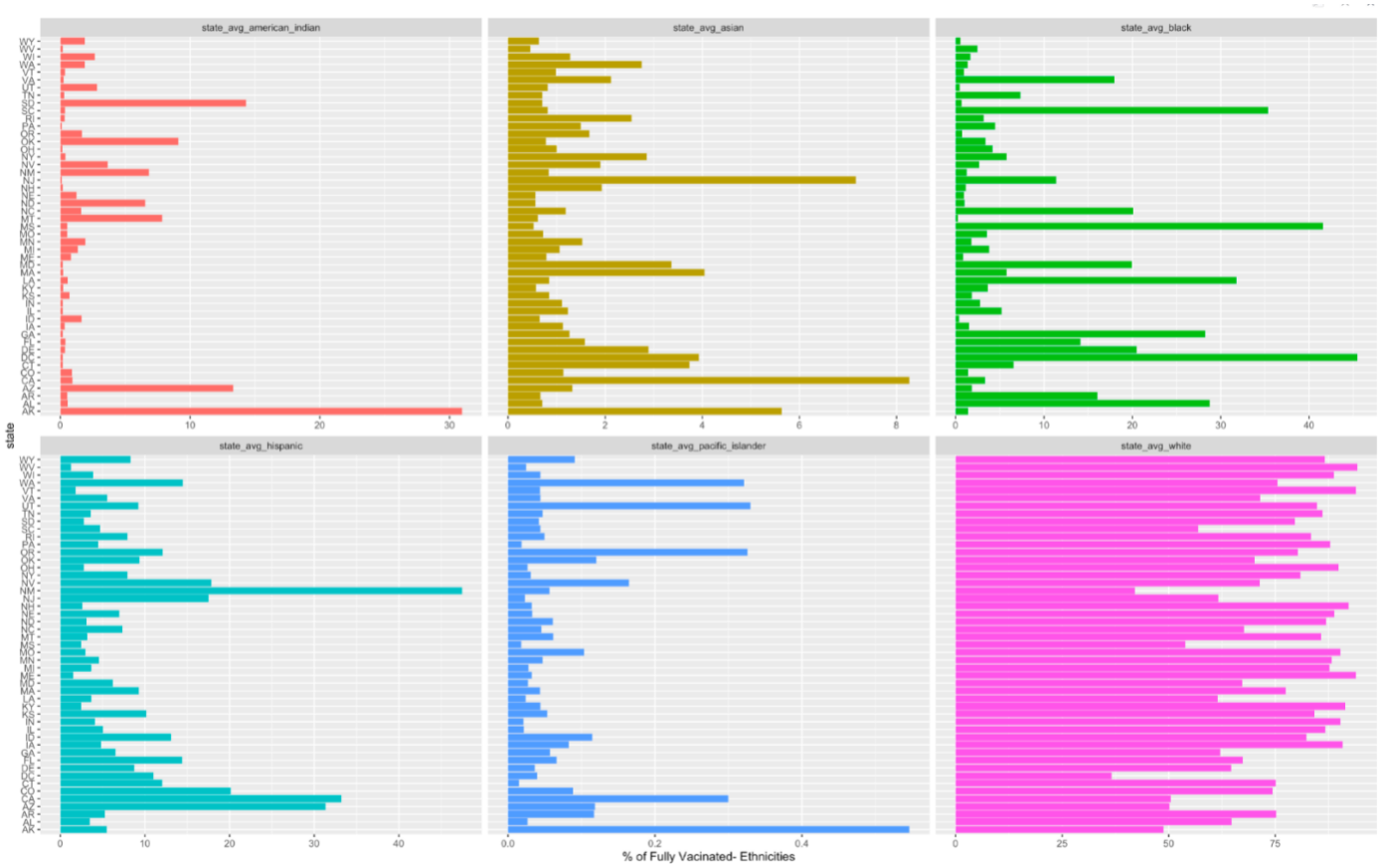


Figure 24 - Facet wrapped plot showing the percent of fully vaccinated adults in the US in each state, in terms of ethnicity

To compare and try to find a correlation between percentages of fully vaccinated adults based on ethnicities, the percentage of fully vaccinated adults in the US with each state was plotted using geom_col(). As seen in the figure below, the states with the largest percentage of adults fully vaccinated were in Rhode Island, Connecticut, and Maine, which are all North

Eastern states who typically identify as “liberal” states. This is not surprising as the majority of the Democratic party supported the COVID-19 vaccine. In fact, according to a Gallup survey released on September 29, 2021, 40% of Republicans “don't plan” to get vaccinated, while only 3% of Democrats “don’t plan” to get vaccinated. The states with the lowest percentage of fully vaccinated adults appear to be Georgia, Virginia, and West Virginia. According to the graph above, although Georgia had a low percentage of adults fully vaccinated, around 30% of those who were vaccinated in Georgia identified as Black and around 60% of those who were fully vaccinated in Georgia were White.

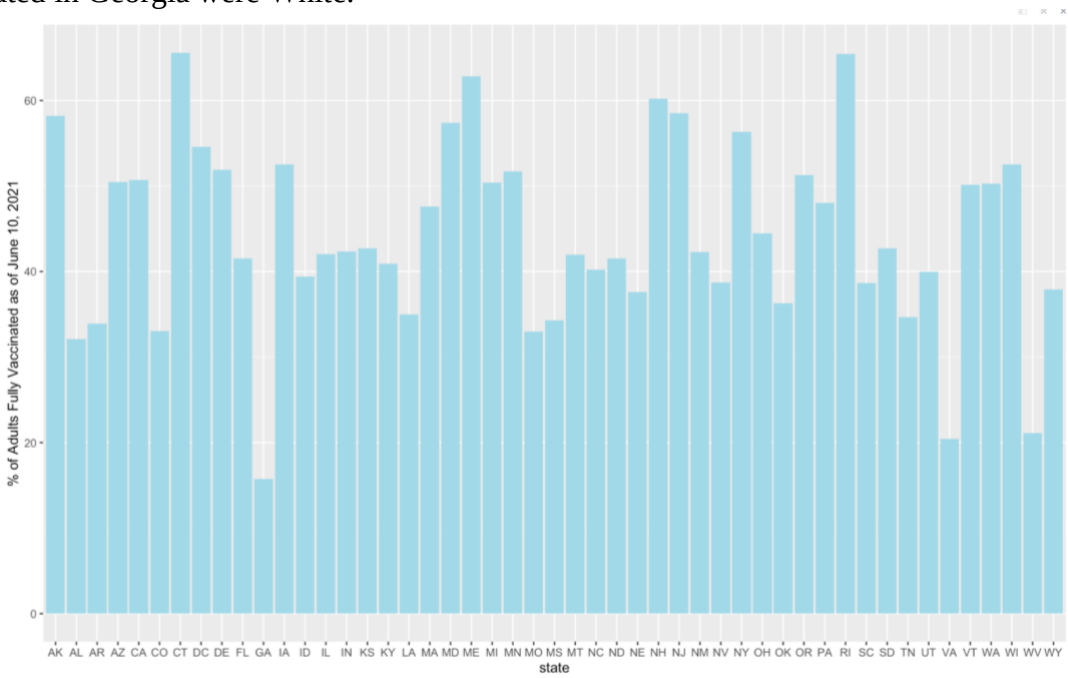


Figure 25 - Graph showing the percent of adults fully vaccinated against COVID-19 as of June 10, 2021 in each state

For the final part of this analysis, correlations between the percentage of adults who are fully vaccinated as of June 10, 2021, the average SVI, average CVAC level of vaccination rollout, and vaccination hesitancy were all found. The correlation between the average of individuals who were strongly hesitant about the vaccine and the average percent of fully vaccinated adults was -0.353. It is difficult to draw conclusions based on a small correlation value, but as these two variables have a negative correlation, this could mean that as the percentage of those who are strongly hesitant about the vaccine rises in each state, the average percentage of fully vaccinated adults would decrease. In terms of the correlation between the average percent of fully vaccinated adults and the average SVAC level of vaccination rollout, there is also a negative correlation of -0.5155. Although this isn’t a strong correlation, this could hint at the fact that as the average CVAC level of vaccination rollout increases, the average percent of fully vaccinated adults decreases. With a greater concern of obtaining the COVID-19 vaccine, it makes sense that the vaccination rate decreases. This goes along with the correlation between the average SVI and average percent of fully vaccinated adults as the correlation values is -0.362. With a greater social vulnerability index, there is a greater potential for negative effects on each state that is caused by external stresses on human health in respect to the COVID-19 pandemic.

Question 4: How does recovery rate change with vaccination rate, political party and education level by state?

For this question, four datasets were used. The first one, is the data related to covid cases which was imported from the CDC, it included a lot of variables, but most importantly the number of new covid cases per month by state and the total number of cases. It's time frame was from April 2021 to March 2022. The second one was a table with the political affiliation of each state whether republican or democratic in March 2022. The third included the covid vaccination rate whether it was two doses of Moderna or Pfizer or one dose of Johnson and Johnson of each state in March 2022. The last one includes the percentage of people with bachelor degrees or more in each state in March 2022. The recovery rate is the independent variable while the rest are dependent. To find the recovery rate, the number of covid new covid cases per month was divided by the total number of covid cases at the time for each state for one year. Thus the main dataset is necessary to collect the data about the recovery rate, while the others are also required to answer the question. Upon the first analysis, some data were found to be missing while plotting the U.S map, thus they had to be omitted. The data also needed to be standardized to the overall population, because there was discrepancy between the population number of each state. To find the new rates, the old recovery, bachelor and vaccination rates were each individually multiplied by the state's population then divided by the total population of the country. An exception was made, Washington D.C not a state as added to the study because it is a relevant area of interest, while the U.S territories were omitted from certain datasets while selecting the data from the website source. In this question, U.S maps were firstly used to help better visualize each variable separately.

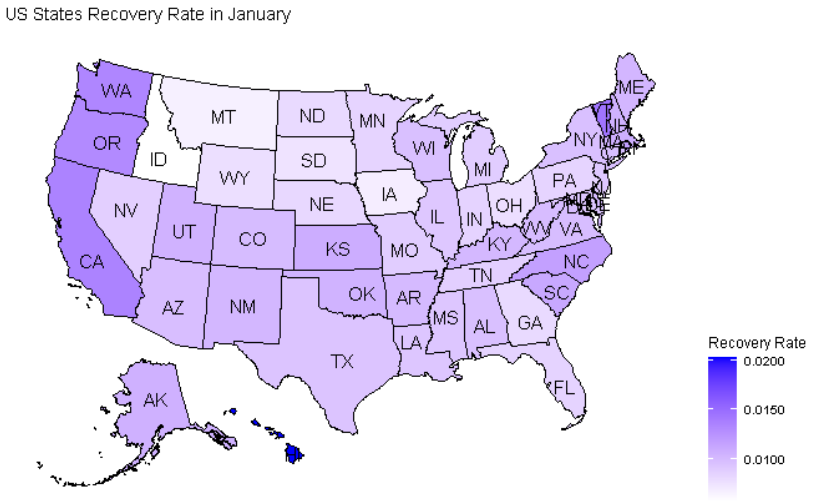


Figure 26 - U.S map showing the covid recovery rate from January 2021 by state

The figure above shows the U.S map with the covid recovery rates of each state in January. The west coast has a high number of recovery rates compared to the other regions.

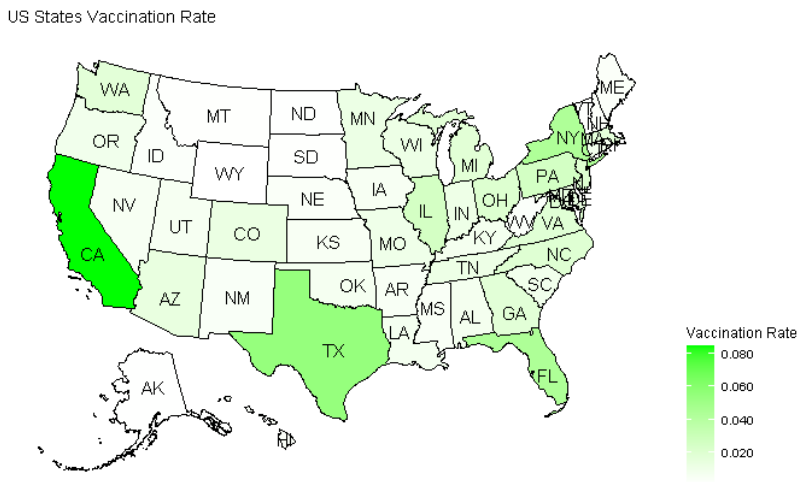


Figure 27 - U.S map showing the covid vaccination rate by state on March 2022

The figure above shows the U.S map with the covid vaccination rates of each state. Comparing it to the previous map, a trend can be noticed for certain states who had a high recovery rate and high vaccination rates.

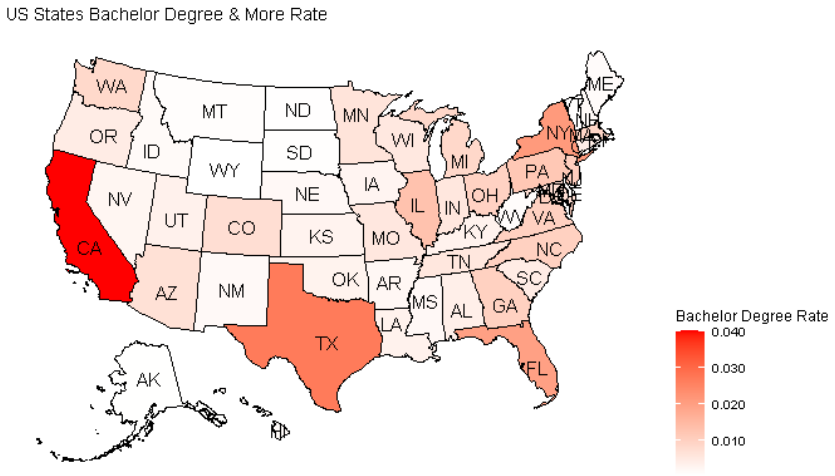


Figure 28 - U.S map showing the percentage of people with bachelor degrees or more by state on March 2022

The figure above shows the U.S map with the percentage of people with bachelor degrees or more in each state. Comparing it to the previous map, there seems to exist a very direct relationship between states who have a high percentage of bachelor degree holders and high covid vaccination rates.

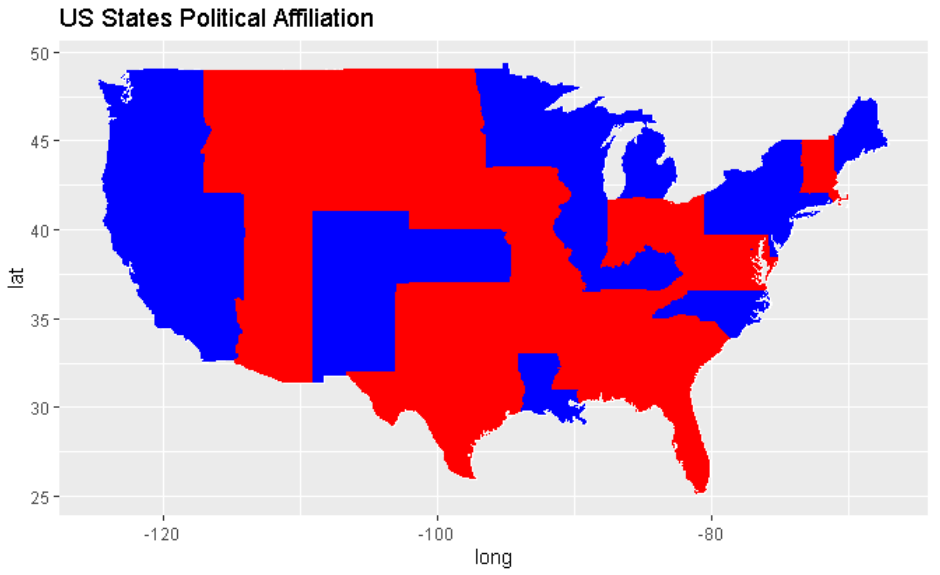


Figure 29 - U.S map showing the political affiliation of each state in 2022

The figure above shows the U.S map with the political affiliation of each state. Comparing it to the first map, it is very obvious that democratic states in blue tend to have a higher percentage of covid recovery rate compared to the republicans in red.

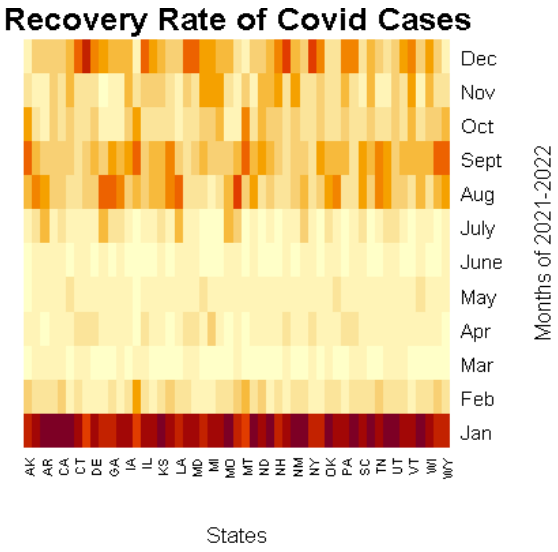


Figure 30 - Heat map showing the covid recovery rate of each state for the last year

The figure above is a heatmap with the 51 states and the 12 months on each axis. Each rectangle represents the covid recovery rate in each phase in time. The darker the color the higher the rate is, thus it is easy to conclude that in January the highest recovery rate occurred.

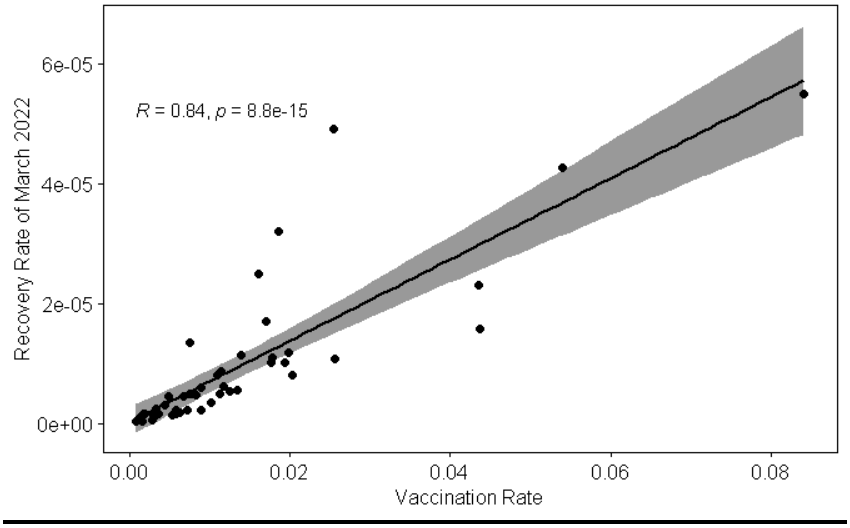


Figure 31 - Scatterplot between the covid vaccination rate and recovery rate

The figure above is a correlation scatterplot between the recovery rates and the vaccination rate in March 2022.

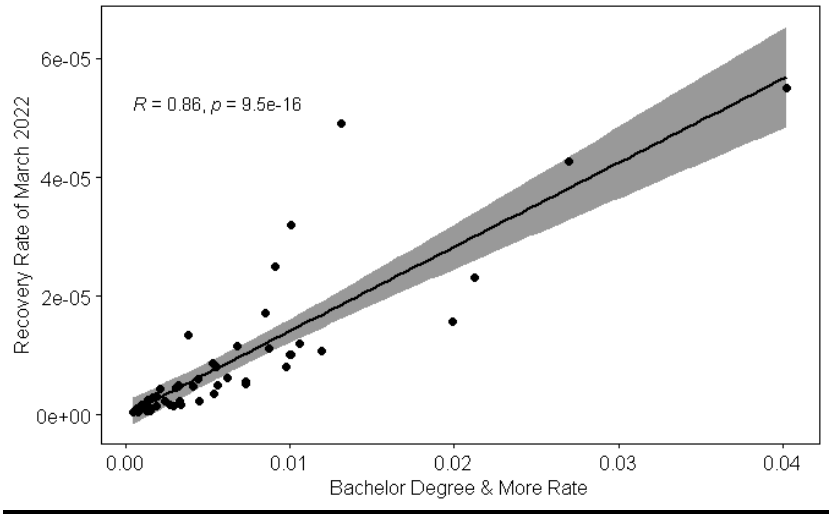


Figure 32- Scatterplot showing the recovery rate and the percentage of people with bachelor degrees or more

The figure above is a correlation scatterplot between the recovery rates and the bachelor degree rate in March 2022.

DISCUSSION

During this study, we evaluated and analyzed multiple datasets in order to find a correlation and pattern between different factors that influenced vaccination status in the United States during the COVID-19 pandemic. Through this process, our analysis was divided among four sub questions in order to extend our research and understanding into multiple case studies: "How does demographic information impact vaccination status?", "What characteristics influence vaccination status the most between different states?", What factors influence social hesitancy regarding the COVID-19 vaccine?", and "How does the recovery rate change with the vaccination rate by state?."

The first finding discovered was that the most highly vaccinated age group was those 75 years of age and older (Figure 6). This conclusion was not surprising, as it is well known that the elderly are more at risk of contracting COVID and therefore are strongly encouraged to receive a vaccine. These findings are backed up by how the vaccine rollout system was established. The timeline of vaccine rollouts for Virginia involved two phases, with Phase 1 divided into 3 sections. Phase 1a targeted health care workers, long-term care facilities, and first responders. For age groups specifically: Phase 1b included seniors 75+; Phase 1c was 65-74 years of age and people ages 16-64 with high-risk medical conditions; and Phase 2 was people ages 16-64 without high-risk medical conditions. It was also found that the most highly vaccinated race/ethnicity was the American Indian/Alaska Native Non Hispanic population (Figure 7). Generally, there was a greater number of White Non-Hispanics that were vaccinated, but looking at the proportion vaccinated within the population, White Non-Hispanic ethnicity group placed third to last amongst American Indian/Alaska Native Non-Hispanic, Asian, Hispanic/Latino, Native Hawaiian/Other Pacific-Islander, Multiple Ethnicities, and Black. The last main discovery from the first subquestion of this analysis was the fact that females were more likely to get vaccinated than males (Figure 8). This is seen through the analysis where the difference between the standardized proportions of females to males was 0.043 meaning that the difference is not statistically significant to reach a final conclusion. However, there is a pattern of females getting vaccinated more than males. This finding helps explain the overall problem of a lack of vaccination, as the failure of men to get vaccinated may contribute to the positive cases and deaths across the country. These three main findings show the significance between demographic information, such as gender, age, and ethnicity, and vaccination status.

Looking at the second sub-question of this analysis, general patterns regarding different characteristics that might affect vaccination status were found through analysis of each state. A general pattern found was that there were negative vaccine attitudes among individuals, which related to an absence of children in unvaccinated households (Figure 12). This was a surprising part of the analysis. However, the fact that adults have negative attitudes toward the vaccine when they have no kids in their house makes sense as there is less concern of spreading the virus to those who can't access the vaccine. It was also discovered that there could be a relationship between positive vaccine attitudes related to income levels less than $25,000, overall education attainment, and unmarried status in unvaccinated households (Figures 13-16). Although these findings are too vague to reach a final conclusion, a trend in income level was expected. It was also expected that education attainment would relate to those with positive vaccine attitudes. However, during the beginning of this analysis, it was not expected that unmarried status would relate to a positive vaccine attitude. As these general patterns are results of looking at the United States as a whole, vaccine hesitancy, income level, education attainment, and marital status was

also analyzed by state. Throughout this analysis, it was expected that the greatest positive vaccine attitudes would be in the North East states as these states are generally more liberal. However, it was not expected that the greatest negative vaccine attitudes would be in North Western states. Overall, looking at specific characteristics that might affect vaccination status in each state helps further analyze factors that might influence vaccination as a nation.

The third question gears more towards social hesitancy in regards to the COVID-19 vaccine. After analysis, it was discovered that there was a strong correlation of 0.759 between the average Social Vulnerability Index (SVI) and average CVAC level of concern for vaccination rollout in each state. There is also a small, negative correlation between those who were strongly hesitant about receiving the COVID-19 vaccine and the average percentage of adults that were fully vaccinated as of June 10, 2021. This could mean that as the percentage of those who are strongly hesitant about the vaccine rises in each state, the average percentage of fully vaccinated adults would decrease. Another slightly negative correlation found throughout this analysis was between the average CVAC level and the average percentage that are fully vaccinated. As the average CVAC level of vaccination rollout increases, the average percentage of fully vaccinated adults decreases. With a greater concern of obtaining the COVID-19 vaccine, it makes sense that the vaccination rate decreases. Again, looking at the average SVI in each state, there was a small, negative correlation between the SVI and the average percentage of adults who were fully vaccinated. With a greater social vulnerability index, there is a greater potential for negative effects on each state that is caused by the external stresses on human health with respect to the COVID-19 pandemic. As some of the correlations found were insignificant or small, it is difficult to make valid conclusions. However, some trends can be seen throughout this analysis.

The fourth and final question analyzed revolved around different factors that might affect recovery rate from COVID-19. In particular, the specific factors analyzed were vaccination rate, political party affiliation, and education level by state. After the analysis process, it was found that the states that had remarkably higher COVID-19 recovery rates were politically liberal (Figure 29). States that had a high percentage of people with bachelor degrees and high vaccination rates also had much higher COVID-19 recovery rates (Figures 31 and 32). Each of these factors have been shown to have influenced the recovery rate of each state.

Based on the findings of each analysis, it is evident that there are many factors that influence vaccination status in the US during the COVID-19 pandemic. These factors include gender, age, race/ethnicity, employment status, income level, educational attainment, marital status, social vulnerability, vaccine hesitancy, and political affiliation. Through this analysis, it was discovered that these factors strongly correspond to the likelihood of vaccination than others. In fact, future research is necessary to confirm whether results are applicable across a larger population, even on a global scale.

CONCLUSION

The goal of this study was to try to discover what factors influenced vaccination status of Americans during the COVID-19 pandemic. Our four part analysis covered factors of age, race/ethnicity, gender, presence of children in the household, marital status, income, employment, political affiliation, education, recovery rate, vaccine hesitancy, and social vulnerability.

Age, race/ethnicity, and gender encompassed the first part of analysis. We found that the most highly vaccinated age group was 75+ years old. As for race/ethnicity, the most highly vaccinated group was American Indian/Alaska Native Non-Hispanic. For the gender factor, females are more likely to get vaccinated than males.

Presence of children in the household, marital status, income, education, and employment were the next factors to be analyzed. The top five US states that hold the highest proportion of negative vaccine attitudes in unvaccinated households were Montana, Wyoming, Oregon, South Dakota, and Rhode Island. The top five US states that hold the highest proportion of positive vaccine attitudes in unvaccinated households were Massachusetts, Maine, Connecticut, Florida, and Minnesota. The top ten were recorded for each to find general patterns, but we found that North-Western US states held negative vaccine attitudes while the North-East held positive vaccine attitudes. General patterns conclude that negative vaccine attitudes relate to an absence of children in unvaccinated households, while positive attitudes relate to income levels less than $25,000, overall educational attainment, and an unmarried status. Employment status had similar effects on vaccine hesitancy in unvaccinated households. As for income levels, states with positive vaccine attitudes topped the proportions of income levels of $25k - $50k, $50k - $100k, and $100k - $200k, however states with negative vaccine attitudes made up most of the top ten states that had those income levels. As for states with income levels above $200k, there was no conclusive evidence to relate high income to vaccine attitude.

The third part studied vaccine hesitancy and social vulnerability through the Social Vulnerability Index (SVI) and the average CVAC level of concern for vaccination rollout in each state. The correlation between the two was 0.759 and within the specifics, those who were strongly hesitant, there was little correlation to the average SVI which is difficult to draw conclusions upon and a slight correlation between those who were strongly hesitant and the average CVAC level of concern for vaccination rollout. There is a small negative correlation between those who were strongly hesitant and the average percentage that were fully vaccinated. There is a slightly negative correlation between the average CVAC level and the average percentage that are fully vaccinated. There is a small negative correlation between the average SVI and the average percentage that are fully vaccinated.

The final part included political affiliation, educational attainment of a Bachelor's degree and more, and recovery rate in each state. There is a positive, direct correlation between vaccination rate and the attainment of a Bachelor's degree and more. For the recovery rate as of April 2022, it directly correlates to vaccination rate. Politically liberal states along the east and west coasts had higher recovery rates, higher Bachelor's Degree attainment rates, and higher vaccination rates.

A further step we could take in our study would be to look into the same factors in our datasets at a certain time period of the pandemic. The vaccination status varies greatly based on the time of year since the Pfizer vaccine was issued an Emergency Use Authorization by the

Food and Drug Administration on December 11, 2020. The possible time sections that could be analyzed are a month after the release of the Pfizer vaccine where the U.S. COVID-19 death toll surpasses 400,000 on January 18, 2021, FDA approval of Johnson and Johnson one shot COVD-19 vaccine, the Delta variant becomes the dominant variant in the U.S. on June 1, 2021, and the CDC recommends that everyone over 18 years old receive a booster shot 6 months after they are fully vaccinated on November 29, 2021. For possible future research, we could expand our analysis globally and look into factors or characteristics of countries with low and high vaccination/recovery rates. This could open up questions of how economic status, cultural differences, world population, and education systems affect how different countries approached the vaccine rollout during the pandemic. We could also take another step into a different pandemic like the HIV and AIDS in the 1980s to present day, Whooping cough in 2010 and 2014, or the H1N1 flu in 2009 given that we have datasets with similar variables that we could use to compare to the COVID-19 pandemic and reveal if similar characteristics/factors affected the population's decision to get vaccinated for the flu or find preventative measures to avoid getting ill.

REFERENCES

1) Centers for Disease Control and Prevention. (n.d.). *CDC Covid Data tracker*. Centers for Disease Control and Prevention. Retrieved May 2, 2022, from https://covid.cdc.gov/covid-data-tracker/#vaccination-demographic
2) Holder, J. (2021, January 29). *Tracking coronavirus vaccinations around the world*. The New York Times. Retrieved May 2, 2022, from https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html
3) *Covid-19 data explorer*. Our World in Data. (n.d.). Retrieved May 2, 2022, from https://ourworldindata.org/explorers/coronavirus-data-explorer
4) Centers for Disease Control and Prevention. (n.d.). *Home Page - CDC Covid Data tracker*. Centers for Disease Control and Prevention. Retrieved May 8, 2022, from https://covid.cdc.gov/covid-data-tracker/#datatracker-home
5) Fridman, A., Gershon, R., & Gneezy, A. (n.d.). *Covid-19 and vaccine hesitancy: A longitudinal study*. PLOS ONE. Retrieved May 2, 2022, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0250123
6) Chen, X., Huang, H., Ju, J., Sun, R., & Zhang, J. (2022, January 28). *Impact of vaccination on the COVID-19 pandemic in U.S. states*. Nature News. Retrieved May 8, 2022, from https://www.nature.com/articles/s41598-022-05498-z
7) Kreps, S. E., & Kriner, D. L. (2021, June 2). *Factors influencing covid-19 vaccine acceptance across subgroups in the United States: Evidence from a conjoint experiment*. Vaccine. Retrieved May 2, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8064867/