Visual Analysis of Effects of Key Demographics on Covid-19 Vaccine Hesitancy in the US

Tina M. Kirk
Department of Computer Science and Electrical Engineering
Marshall University
Huntington, WV
kirk41@marshall.edu

I. Introduction

Throughout history, the world has been through sevearal epidemics or pandemics [1]. In 1916 it was polio. The vaccine, developed in 1954, has largely eradicated the disease. There were approximately 6000 deaths from polio in the United States compared to 900,000 from Covid, but people were not hesitant to get the polio vaccine. They did not doubt that polio was real as they could see its disabling effects on themselves or loved ones, mainly children. They voluntarily isolated because they did not know what caused it [2]. "Historians say…in 1955, many Americans had an especially deep respect for science [2]." Today, the rapid roll-out of the Covid vaccine paired with the massive disinformation campaign has led many to hesitate or refuse it altogether. There are no crippled limbs from Covid, so they explain away side-effects as having another cause. Some people do not believe the scientific information presented to them about the disease, some believe they are being deceived about death rates due to Covid, and some believe it is a way for the government to exert control. It seems no amount of scientific evidence is likely to sway their opinions. The job of those trying to impart statistically significant, scientifically based information to the world has become much more difficult because so many have already "researched" it for themselves on social media. They will undoubtedly continue to fall into the trap of confirmation bias, such as choosing sources of news that fit what they already believe [3]. Why should the scientific community keep trying to get through to people? Because disinformation is, at its best, useless, and at its worst, detrimental to society.

With the power of visualization, a story can be shared in a way others can engage and understand [4]. "Humans can process images 60,000 times faster than text [4]." As exploratory analysis of data is carried out, interesting questions can be generated. Looking at the structure of the data can determine patterns and characteristics, outliers and anomalies, and assumptions can be tested [5]. There is little doubt that disinformation and misinformation is here to stay, so it is important to use the tools necessary to make data available to people in an interesting, straightforward way to aid them in making informed decisions.

The purpose of this project is to analyze how Covid 19 has progressed in the United States, looking at rates of vaccination and patterns of vaccine hesitancy from early 2020 to the present, and to provide visual representation of those findings. In analyzing the data, I hope to discover patterns and find answers to interesting questions such as 1) How did the virus spread over time across the United States? 2) Are there certain areas of the United States that were hit harder than others? If so, what do these places have in common and how are they different? 3) Are there patterns that have changed over the course of the pandemic? 4) What characteristics of people seem to feature most in their decision to be vaccinated or not? Kaggle.com [6] has many available datasets that are interesting for this purpose. For one of the datasets, us_states_covid19_daily.csv, there is a Tableau public dashboard people can use, as seen in Fig. 1 [7]. I will work to create interactive visualizations such as these to help guide use exploration and decision making amid the ongoing pandemic because "visualizing data in graphs, charts, and maps helps users…develop actionable insights [8]."

In the analysis stage, it will be interesting to see if any unique ways of looking at the situation emerge. To check for accuracy of the visualizations in this project, they will be compared to those in Fig. 1 [7] as well as visualizations provided on the sites with the data sources [6].

Fig. 1: Screenshot of public dashboard created from the us_states_covid19_daily.csv dataset [7]

II. RELATED WORKS

Covid 19 is merely the latest in a long history of global health crises. Jarus [1] compiled a list of the 20 deadliest epidemics and pandemics dating as far back as 3000 B.C. All of them horrifying with mass casualties, but a few had more similarities with the Covid pandemic than others, namely anger over masks/quarantines and the spread of misinformation. The Russian Plague of 1770-1772 had the first mention of citizens erupting into violence over quarantines, even murdering Archbishop Ambrosius for encouraging crowds not to gather for worship [1]. During the Philadelphia Yellow Fever Epidemic of 1793, it was mistakenly thought that slaves were immune and had people of African descent working as nurses for the sick [1]. The Spanish Flu of 1918-1920 got its name because Spain was a neutral country and had a freer press that published early accounts of the disease. Because of that it was thought that the disease originated in Spain and the name stuck [1].

Brink described public attitudes during the polio epidemic [2]. When a vaccine was finally developed in 1955, the disease had been raging for almost 40 years. The public was overwhelmingly in support of it and had "an especially deep respect for science [2]." They were also terrified of the visible effects of polio. With Covid, it has been a struggle to convince some people that the massive worldwide casualties were even caused by the illness. The 24/7 access to information on the internet has not helped the situation. Research shows that people will use sources of news that fit their pre-esisting views [3]. Hsu noted that it seems people more confident in their views are more likely to also look at opposing viewpoints [3]. He said a big question that remains is whether consuming all this news affects views or simply solidifies the original beliefs [3].

With a freely available dataset, Taha demonstrated analyses in Python using Plotly, Seaborn and pandas [4]. Many different types of graphs and charts are given along with code included. At the end of the paper, the link for the dataset is provided along with an invitation to "collect the data [from the link]…and play with it [4]" and an admonition that the results will be different because more data will have surely been added. Hillier described some projects that would be beneficial to include in a data analytics portfolio [5]. He makes great points in saying that "80% of all data analytics tasks involve preparing data for analysis [5]." Some of the things Hillier mentions are scraping the web for data, carrying out exploratory analyses, cleaning untidy datasets, and communicating results with visualizations. He then gives a description of each category and suggestions for projects, including a link to a tutorial to help visualize Covid-19 data using R, Shiny, and Plotly.

A list of best practices for visual analytics is listed on the website "Visual Analytics: What it is, why it matters, and best practices [8]." First, "define goals [8]" and decide what questions you want to answer. Next, "integrate and manage the data [8]," which means your data should be transformed into standard formats and made accessible to the business that needs it. Lastly, "simplify visualizations [8]."

Robinson, Chartier, Shaak, and Bovij studied a vaccine tweet dataset to determine people's overall feelings including its distribution [9]. The authors had two research questions: 1) "How can you use descriptive or exploratory methods for social media data anlytics?" and 2) "What impact has the process in distribution and receiving the Covid vaccine had on people's opinions via Twitter [9]?" This study calculated "sentiment scores [9]" for tweets and gave each state a "mean sentiment score [9]." The further negative the score was, the more negative the sentiment toward vaccination for that state, and the further positive the score was, the more positive the sentiment for that state. This can easily be compared to maps with certain demographics highlighted, such as red and blue states.

Vaccine hesitancy is defined as a "delay in acceptance or refusal of vaccines despite availability [15]." Nguyen, Joshi, Drew et al. studied "vaccine uptake [10]" (taking the vaccine) in different ethnic groups in the United States and United Kingdom. Lower levels of vaccine uptake were discovered in Black participants in the US than in the UK. The authors surmise those levels may be partially attributable to challenges in access in the US because they found that uptake rates were low even for individuals that had a self-reported willingness to get vaccinated [10].

Distrust with the health system is another supposed reason that Covid has disproportionately affected individuals in the Black community [11]. According to Padamsee, Bond, Dixon, Hovick et al., as of January 2021, few studies had examined changes in vaccine hesitancy of the Black community over time [11]. They set out to try to examine mechanisms that might have contributed to the differences. Their results showed that in December of 2020 [11], when the new vaccines were on the horizon, Black and White people had a similar level of sentiment toward them. As time went on, the differences in sentiment changed more among Blacks than Whites, with the former having a larger favorable increase [11]. While vaccination rates as of January, 2022 were still lower for Blacks than Whites, the study results support their initial thoughts that factors other than vaccine hesitancy might be more to blame [11]. They concluded that the main reason for the rapid increase in overcoming vaccine hesitancy in the Black community seemed to be the rise in belief that vaccines were a necessary way to protect themselves and their communities [11].

Vaccine hesitancy rates differ between people with opposing political views as well, as we have seen all over the news and social media. Kates, Tolbert, and Orgera reported that, as of September 2021, 52.8% of "people in counties [in the US] that avoted for Biden were fully vaccinated compared to 39.9% of Trump counties [13]." This was a very short article but did include links to the

CDC's "COVID-19 Integrated County View' dataset at https://covid.cdc.gov/covid-data-tracker/#county-view as well as the 2020 Presidential election results by county at https://github.com/tonmcg/US_County_Level_Election_Re s ults_8-20 [13]. Albrecht's goal was to improve understanding of factors related to COVID-19 vaccination decisions and consequences [14]. He focused his study on the role of political views in such decisions, specifically the relationship between political views, vaccinations, and per capita COVID deaths in US counties [14]. He concluded that the percent voting for Trump was "strongly and inversely related to percent vaccinated [14]." Additionally, in Trump leaning counties, COVID-19 deaths were more prevalent.

III. PROPOSED SOLUTION

First it was necessary to narrow down the many datasets that show the spread of Covid cases and vaccinations from early 2020 to the present. Some had data only from 2020 and some had only data from a later period. County level results for the 2016 and 2020 presidential elections were easier to find. The plan was to create visualizations to see if anything interesting emerges. I did expect to see differences between different demographics, especially political, but to what degree I was not sure.

As Albrecht described, vaccines became readily available in March of 2021 and the six month period that followed was long enough to see the consequences of vaccine use or lack thereof [14]. Although it makes sense to include this time period in my research on the topic of vaccination rates and patterns of vaccine hesitancy, the time right before the vaccines were available is when people were predetermining if they would take it or not. It is of great interest to see if patterns emerge that can help explain the enthusiasm for the vaccine or lack thereof, ahead of its availability.

Taking some direction from the Albrecht study [14], I included the number of Covid cases by county from May of 2020, March of 2021, and six months later in September of 2021. I started with those dates of interest. This will show the progression of the disease over time. Since the pandemic happened during one of the most politically polarizing times perhaps we have had as a nation, there is no surprise that previous studies have shown distinctions between vaccination rates of those with differing political views. Similar studies have also shown differeneces in vaccine hesitancy between people of different races. Usable race demographics datasets were not as easy to find for this paper, so that will be saved for future works.

Since the nation seems more divided than ever with the current and former administrations, and since the pandemic started right in the middle of it all, data regarding party affiliation from 2016 and 2020 presidential elections would seem more useful than simply party affiliation from any other time. Wearing a mask to protect from a deadly disease somehow turned into a blaring signal that you did not support Donald Trump. There is reason to suspect that political views played a role in social behavior amid the pandemic.

The metric for each county will be percent of Trump votes in that county. Those metrics will be compared to vaccination rates at different times throughout the pandemic.

Working with some datasets thus far, it was a challenge to find ones that can blend elements. Fig. 2 is an example of a preliminary attempt at merging two datasets. By color, it shows states where Joe Biden won in the 2020 presidential election. By size, it shows the total number fully vaccinated between December, 2020 and end of February, 2021. As my research progressed, showing percent fully vaccinated in each county along with whether the county swayed toward Trump was one of the goals.

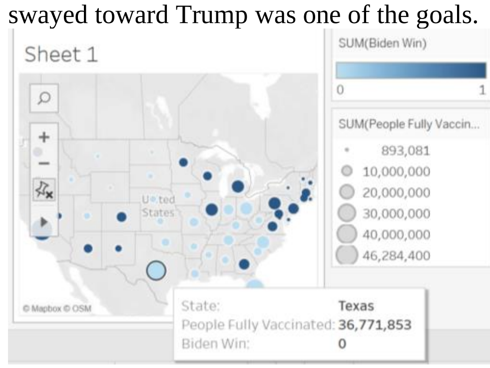

Fig. 2: Total fully vaccinated individuals (by size) in states won by Joe Biden (by color)

IV. EVALUATION

Both graphs in Fig.3 below show a weak, negative correlation between percentage of votes for Trump in the last two presidential elections by county, and percent fully vaccinated in each US county. The p-value is low, so there is significant evidence of an association, albeit a weak one. According to the data, approximately 21.7 million more people voted in 2020 than in 2016.

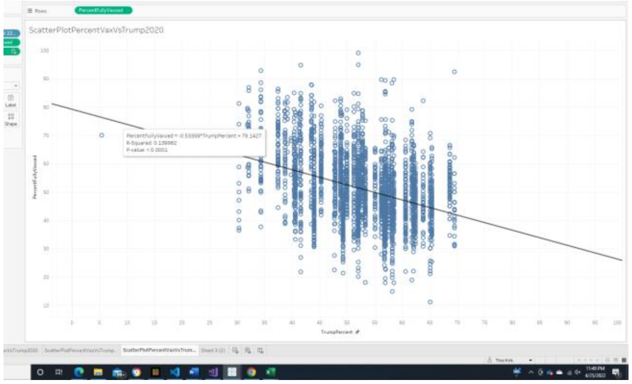

Fig. 3: Percent fully vaccinated by county as of 4/22/22 vs. percent voting for Trump for president in 2016 and 2020, respectively.

Fig. 4 below shows a subset of the percentage of votes for Trump for each state in 2020 with colors of orange to blue, with darker orange meaning lower vaccination rates and dark blue being the highest. As you get into the higher Trump percentages you can see a trend toward lower vaccination rates. This is showing the total vaccinated per hundred in each state as of April 21, 2022. There seems to be an obvious connection between vaccination rates and choice of president.

Fig. 4: Percentage of votes per state for Trump in 2020 color-coded by vaccination rates (lower-dark orange, higher-dark blue)

V. CONCLUSIONS

I have learned that a good, complete dataset is difficult to find. If you want multiple datasets that can work together, that is even harder.

With the data that I was able to find and use in the time frame, through the power of visualization and analysis here we can see that there is a trend toward vaccine hesitancy in the "red" states. It is not as strong a correlation as I had expected to find, but it is there. If the pandemic had struck at a different time in history that was not so politically polarized, I presume that there would be little distinction between vaccination rates and politics. I strongly suspect other factors at play that contributed to the trend seen here.

VI. FUTURE WORKS

Many of the sources cited in this paper gave information about the datasets used. My research will also include some exploration of those datasets with the goal of reproducing some of their results. One such example is in the Padamsee et al. study, Fig. 1, which shows mean vaccination intention over time by race [11]. It shows while black participants had lower intentions than white participants at the start of the vaccine roll out, around mid-February of 2021 the roles reversed with the greatest difference in April, and then the intention levels appear to come together by June of 2021 [11]. Another very interesting visual to reproduce would be the US States Covid-19 vaccine sentiment by color shade in the Robinson et al. study. This type of chart could also be compared with the same demographic information previously mentioned.

I was not able to create a dashboard for the data at the time of this publication. I will continue to work toward that goal as well.

REFERENCES

[1] O. Jarus, "20 of the worst epidemics and pandemics in history," November 15, 2021 https://www.livescience.com/worst-epidemicsand-pandemics-inhistory.html

[2] S.Brink, "Can't Help Falling In Love With A Vaccine: How Polio Campaign Beat Vaccine Hesitancy," West Virgina Public Broadcasting (NPR), May 3, 2021, https://www.npr.org/sections/healthshots/2021/05/03/988756973/ca nt-help-falling-in-love-with-avaccine-how-polio-campaign-beatvaccine-hesitan

[3] J. Hsu, "People Choose News That Fits Their Views," June 7, 2009, https://www.livescience.com/3640-people-choose-newsfitsviews.html

[4] B. Taha. "Covid-19 Vaccination: Data Analysis & Visualization." May 18, 2021. https://www.analyticsvidhya.com/blog/2021/05/covid19vaccination-data-analysis-visualization/

[5] Career Foundry. https://careerfoundry.com/en/blog/dataanalyticsportfolio-project-ideas/

[6] "Your Machine Learning and Data Science Community." Kaggle. Accessed April 26, 2022. http://www.kaggle.com/.

[7] "Covid19 In USA, Based on Data Collected by the Covid19 Tracking Project (@covid19tracking) #Coronavirus #Coronavirusdata #COVID19." Tableau Software. Accessed April 26, 2022. https://public.tableau.com/views/COVID19inUSA/COVID-19spreadinUSA?%3Adisplay_count=y&%3Aorigin=viz_share_link

[8] “Visual Analytics: What It Is and Why It's Important.” Qlik. Accessed April 26, 2022. https://www.qlik.com/us/data-visualization/visual-analytics.
[9] Robinson, E. Chartier, R. Shaak, and L. Bovij. "Data Analytics and The Impact on Distribution Of The Covid Vaccine."
[10] L.H. Nguyen., A.D. Joshi, D.A. Drew et al. Self-reported COVID19 vaccine hesitancy and uptake among participants from different racial and ethnic groups in the United States and United Kingdom. Nat Commun 13, 636 (2022). https://doi.org/10.1038/s41467-022-28200-3
[11] TJ. Padamsee, RM Bond, GN Dixon, et al. Changes in COVID-19 Vaccine Hesitancy Among Black and White Individuals in the US. JAMA Netw Open. 2022;5(1):e2144470. doi:10.1001/jamanetworkopen.2021.44470
[12] N. Ndugga, L. Hill, S. Artiga, S. Haldar, “Latest Data on COVID-19 Vaccinations by Race/Ethnicity,” Feb. 2, 2022, https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-oncovid-19-vaccinations-by-race-ethnicity/
[13] “The Red/Blue Divide in Covid-19 Vaccination Rates Continues: An Update.” KFF, January 31, 2022. https://www.kff.org/policy-watch/the-red-blue-divide-in-covid-19-vaccination-rates-continues-an-update/.

[14] Albrecht, D. Vaccination, politics and COVID-19 impacts. BMC Public Health 22, 96 (2022). https://doi.org/10.1186/s12889-021-12432-x
[15] D. Olivera Mesa, A.B. Hogan, O.J. Watson. et al. Modelling the impact of vaccine hesitancy in prolonging the need for NonPharmaceutical Interventions to control the COVID-19 pandemic. Commun Med 2, 14 (2022).
[16] MIT Election Data and Science Lab, 2018, "County Presidential Election Returns 2000-2020", https://doi.org/10.7910/DVN/VOQCHQ, Harvard Dataverse, V10, UNF:6:pVAMya52q7VM1Pl7EZMW0Q== [fileUNF]