Ningji Shen
DATA 512
December 14, 2021

A7: Project Report

I. Introduction

Over the past two years, coronavirus disease 2019 reshaped our world. The datafication of the pandemic gives us the privilege to examine this pandemic, as well as many health related topics, from potentially many different perspectives, to understand how it has changed the society, and how we can control and prevent diseases in the future. By collecting, aggregating, and re-representing many aspects of the individual toll of the pandemic, organizations such as CDC are able to provide human-centered data to the public, and enables scientists and researchers to explore and gain insight from vast datasets. By conducting a human centered data science analysis of some available COVID-19 data, that is, putting the perspectives of the human — be it the end-user, the audience of the research, or the general public who benefits from this research, we intend to understand how various factors and public health policies contributes to the trends of COVID-19. For this project, although it does not tackle an unresolved research question, I aim to discover if the correlation between the two main actions we took for this pandemic, masking and vaccination, and number of confirmed COVID-19 cases and deaths in Macomb County, Michigan can be shown from the some of the matter-related datasets.

II. Background/Related Work

Ever since the COVID-19 vaccines were developed, it became a trending topic, especially on whether vaccinations are effective. With many fully vaccinated people still getting infected with COVID (Ellyatt, 2021), people start to doubt the effectiveness of the vaccinations. Studies and analysis have been done in England and shown a waning effectiveness of the Pfizer-BioNTech shot (Ellyatt, 2021). However, this analysis also showed that after two doses the Pfizer-BioNTech Covid vaccine is highly effective against hospitalization and death (Ellyatt, 2021). Given differences in the nature of the vaccination programs in England versus the United States, as well as differences in the study dates, age groups and Covid testing regimes involved, I aim to research on the topic of "how did vaccination change the progression of COVID-19 cases and deaths in Macomb County, Michigan in 2021?", with a hypothesis that is very similar to the Ellyatt's point from the article, that is "the higher percentage of people who are fully vaccinated (have a second dose of a two-dose vaccine or one dose of a single-dose vaccine) leads to the incline in COVID-19 cases and decline in COVID-19 deaths in Macomb County, Michigan." I aim to design my analysis and present my results in a way that is very similar to the figure "Covid-19: US daily new cases and deaths" from

Johns Hopkins University shown by Ellyatt in the article, but with vaccination rate and dual axis instead, so the readers can see the changes as slopes of the derivative functions.

III. Methodology

The process of my research involves the following two parts, data gathering & cleaning, and data exploration & analysis. In order to make my analysis reproducible and understandable, I documented my analysis process, along with all my code and data.

- Data Gathering & Cleaning

  I first gathered the data for Macomb County, Michigan from all the datasets by filtering on the county names in the API. All datasets downloaded from the APIs are cleaned and ready to use, so I only did some minor cleaning, such as removing the columns where the data are the same throughout rows (e.g. County, State, FIPS, etc.). All cleaned datasets can be found in my repository. I then join all three CSVs with the date column being the unique identifier, in Tableau Public.

- Data Exploration & Analysis

  To explore my data as well as present my findings, I created graphs in Tableau that visualizes how the course of the disease was changed along with the changes in masking policies and vaccination rate. I chose to use Tableau Public instead of a Python notebook because Tableau's interactive UI allows me to choose different visualization types and switch back and forth between different variables without thinking about the methods of implementation and debugging the codes. This allows me to go deeper in my data exploration and be more organized in generating a story I wanted to tell during my result presentation.

  During this exploration process, I modeled different aspects of this problem, and made assumptions if necessary. For masking policies, I simply used color-coding over the derivative function of the rate of infection. According to Wikipedia, the infection rate is the total number of infections, as known as the total confirmed cases, divided by the population at risk. Assuming the population at risk is the population of the whole county, and also assuming the population of the whole county stays roughly the same during this whole pandemic, the derivative function of the rate of infection would look the same as the derivative function of the total confirmed cases of COVID. I also made the assumption that wearing a mask would not prevent you from death once infected, so I did not look at the correlation between masking policies and the death count.

  In order to show the changes in the derivative function of the rate of vaccination over time, and compared it with the derivative function of the rate of

death and disease transmission, I utilized multiple time-series visualizations. In order to show the slopes of death counts, which is way smaller compared to the number of confirmed cases, I used dual-axis charts, with different colors for each variable. I then analyzed the sets of slopes and peaks in values in the time series, and tested the difference in the derivative functions. In addition, my graphs also utilized the qualitative palette color-coding techniques for categorical variables, such as the Community Transmission Level Indicator. This allows the audience to tell the distinct labels instantly. All these analysis and visualization techniques allow me to tell a more effective story, one that engages my audience at an emotional level, and captures their attention quickly.

IV. Findings

I first took a look at the masking policies. Figure 1 shows how people in Macomb County, Michigan would respond to the question "How often do you wear a mask in public when you expect to be within six feet of another person?", where segments of the pie chart represent the estimated share of people in Macomb County. The data for Figure 1 was generated from a survey, and from this visualization, we can see that over 80% of the whole population in Macomb County wear masks in public frequently or always. From this analysis, we can conclude that when the masking policy was enforced, most people in Macomb County were wearing masks. Therefore, we can use the data point "if the masking policy was enforced" as a variable for my latter analysis.

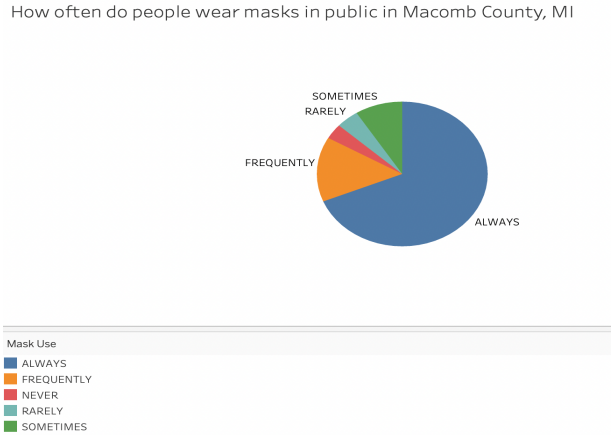

Figure 1: How Often Do People Wear Masks in Public in Macomb County

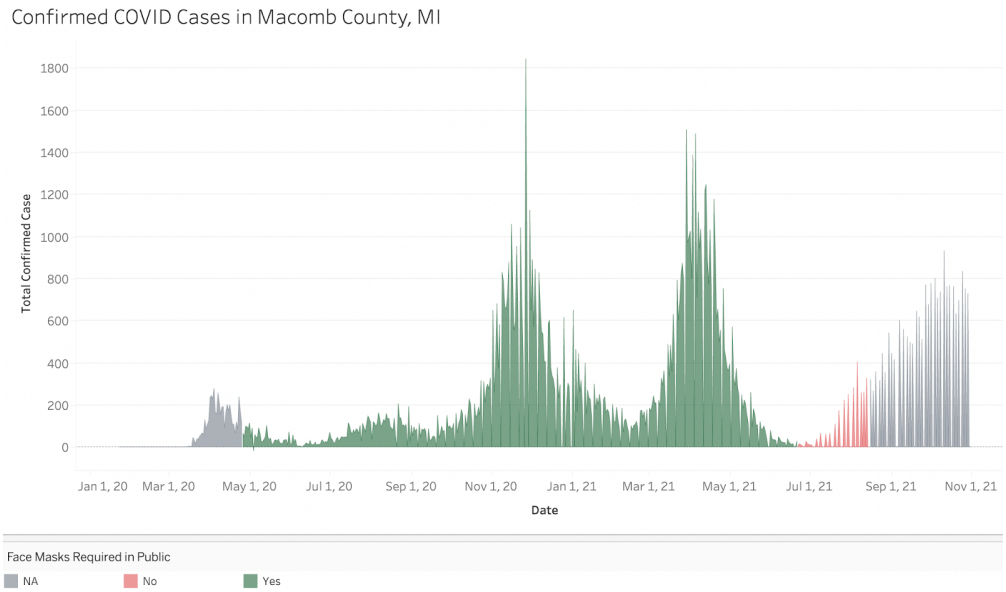

Figure 2: Confirmed COVID Cases and if Masking Policy is Enforced in Macomb County

Figure 2 shows the total number of confirmed COVID-19 cases in Macomb County, Michigan from January 1, 2020 to present. The data was processed by extracting data of Macomb County, Michigan from the combined datasets of the US confirmed COVID-19 cases from February 1, 2020 through October 15, 2021 from the Kaggle repository of John Hopkins University COVID-19 data, and Public Mask Mandates From April 10, 2020 through August 15, 2021 by County by Day from the CDC dataset. The number of confirmed cases for each day are plotted as a bar on the timeline. The bars are colored in three different colors, indicating whether wearing face masks in public is a requirement on that day. The red bars indicate that wearing a face mask in public is required, whereas the green bars indicate that wearing a face mask is not required. The gray bars in 2020 show that there is no such masking policy implemented by the official sources, likely because the pandemic just started. The gray bars in 2021 show that we do not have data showing if masks are required, but my assumption would be there is no masking policy once it gets lifted in July 2021. From Figure 2, we can see that although there are two peaks of the number of confirmed cases, the number of confirmed cases decreased twice, during the period that the masking policy was enforced. However, after the masking policy was lifted, the number of confirmed cases increased dramatically. Although there are many other factors and potential issues of what caused the number of confirmed cases to increase or decrease, we can conclude that the implementation of masking policies has a correlation with the decreasing of the confirmed COVID-19 cases from February 1, 2020 through October 15, 2021.

Obviously, besides the masking policies, there are many other factors contributing to the decrease of the number of confirmed cases, and we think that vaccination would be

one of the main contributors. Factors such as masking policies and social distance would only help with preventing some percentage of infection, but not preventing death once infected. Vaccination, on the other hand, is supposed to prevent both infection and death to some degree. From the previous analyzes, I learned the total number of confirmed COVID-19 cases increased in Macomb County, Michigan, after the masking policy was lifted. Did vaccination help with preventing that number from further increasing? In my analyses below, I visualized the correlation between the confirmed cases and the community transmission level, and the correlation between vaccination rates and the total number of confirmed COVID-19 cases and deaths in Macomb County, Michigan.

To start my second round of analyses, I first plotted both the number of confirmed cases and deaths from December 2020 to December 2021, with two distinguishable colors. From Figure 3, we can see that the trend and slopes of derivative functions of confirmed cases and deaths are very similar, with a short delay on time for death counts. Figure 3 did not show the derivative function for confirmed cases before December 2020, before the first local maximum of number of deaths. My assumption is there was a local maximum for the number of confirmed cases in fall 2020, which leads to the local maximum of the death counts in January 2021. However, how did the number of confirmed cases relate to the transmission level?

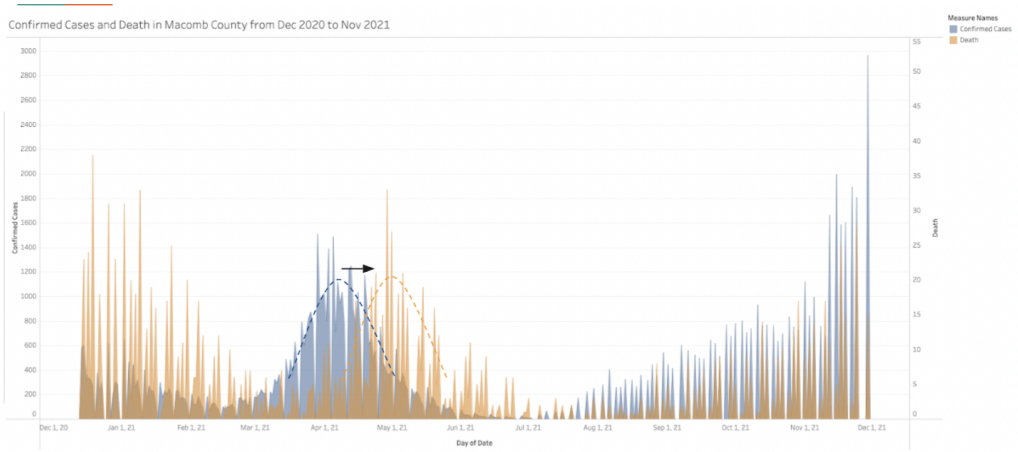

Figure 3: Confirmed Cases and Deaths in Macomb County, Dec 2020 - Dec 2021

With the similar color-coding techniques, I mapped the Community Transmission Level Indicator into the graph of the derivative function of confirmed cases. Combining Figure 4 with the definition of Community Transmission Level Indicator (e.g. “High Transmission Threshold” is being defined as: Counties with 100 or more total cases per 100,000 population in the past 7 days or a NAAT test percent positivity in the past 7 days of 10.0% or greater), we can conclude that there is a linear relationship between the the confirmed cases and the community transmission level. The higher the number of confirmed cases is, the higher the community transmission rate is.

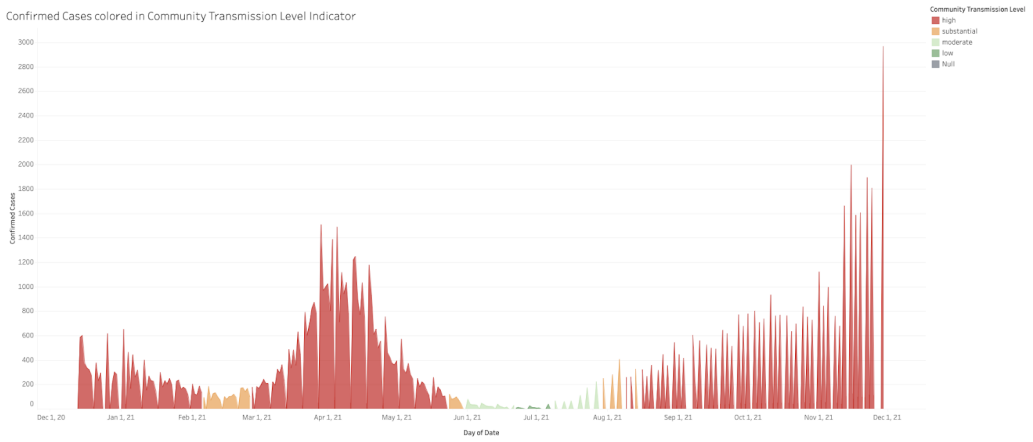

Figure 4: The Number of Confirmed Cases Implies Community Transmission Level

Now, let's take a look at if there is a correlation between community transmission rate and the vaccination rate. On top of Figure 2, I added the percent of people who lived in Macomb County and were fully vaccinated, that is, had a second dose of a two-dose vaccine or one dose of a single-dose vaccine (Figure 4). As the vaccination rate increases over time, the confirmed cases first decreased from a local maximum (April 2021 to July 2021), and then increased all the way to a historical maximum of the number of confirmed cases (December 2021). From Figure 5, we conclude that there is no visible evidence showing that there is a correlation between community transmission level and the rising vaccination rate.

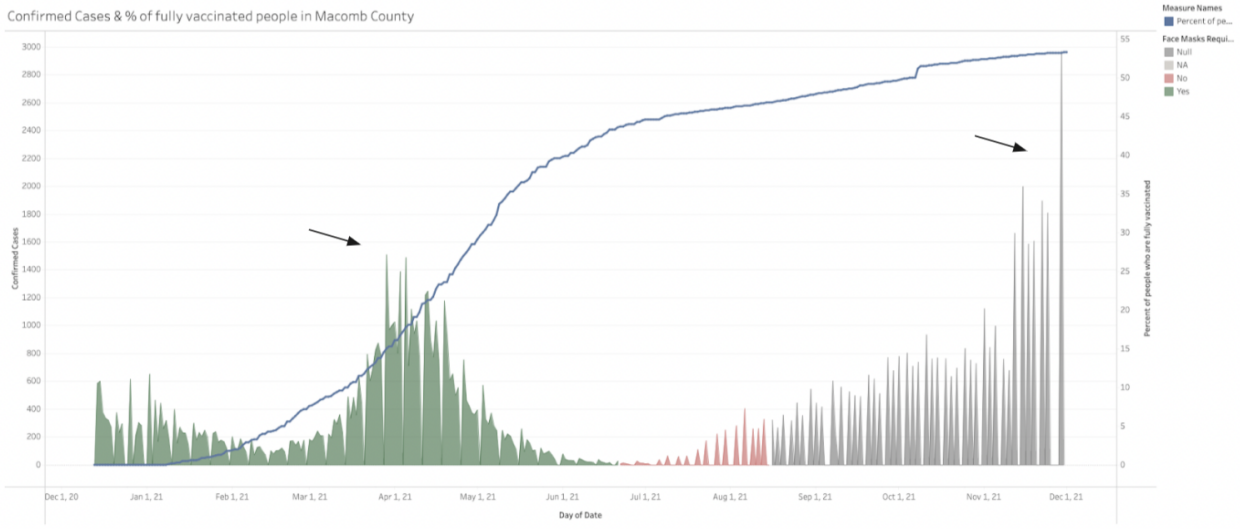

Figure 5: The Number of Confirmed Cases and the Vaccination Rate in Macomb County, Dec 2020 - Dec 2021

We did not see any correlation between the vaccination rate and the community transmission level (number of confirmed cases), but what about the correlation between the vaccination rate and the number of deaths? In Figure 6, I plotted the number of COVID deaths with the percent of people who lived in Macomb County and were fully vaccinated. Surprisingly, after more people got vaccinated in Macomb County, although there are more confirmed cases, the local maximum of death counts in December 2021 is actually lower than the local maximum in May 2021, which is lower than the first local maximum in January 2021.

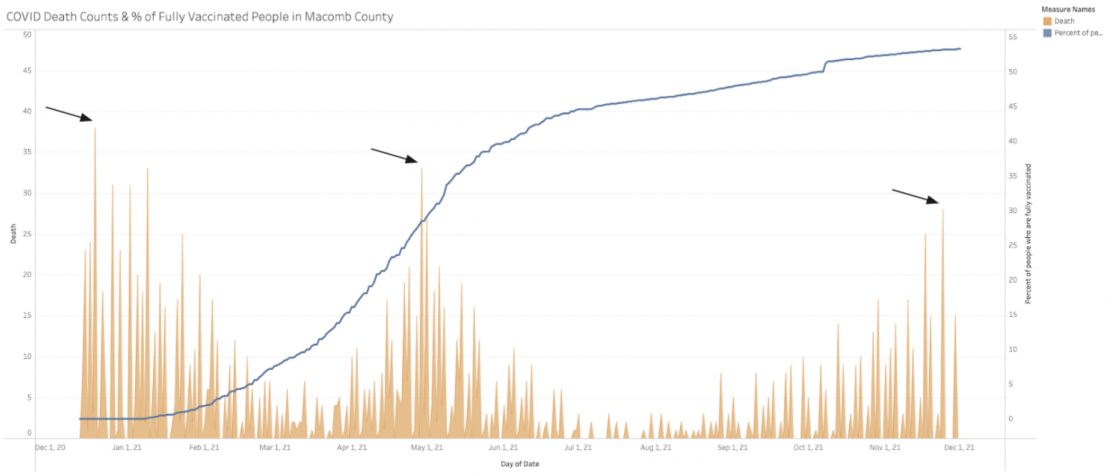

Figure 6: COVID Deaths Counts and the Vaccination Rate in Macomb County, Dec 2020 - Dec 2021

V. Discussions & Implications

For my analyses, I did not observe any obvious patterns from my plots that there is a correlation between vaccination rate and the community transmission level in Macomb County in 2021, however, if I compare the local maximums of the death counts, I observed a clear pattern that the death count was lower after more people were fully vaccinated, even with more people tested positive. Yet, I still can't conclude that vaccination works in terms of reducing the community transmission level and reducing the rate of COVID -19 death in Macomb County. Due to data limitations, I did not look at one of the major variables in the equation of the progression of this pandemic -- the virus itself. With the passage of time, the virus itself continues to evolve. Based on William Hanage, associate professor of epidemiology and a faculty member in the CCDD, "There's a widespread public perception that viruses become less dangerous over time because they evolve to avoid harming the hosts they rely on for transmission" ("What's next in the Evolution of COVID-19 Is Increasingly Difficult to Predict"). For future research, I would take the evolution of the virus into account, while analyzing the progression of the pandemic caused by this virus. From my existing research, I learned

that not all the most obvious statements, such as "vaccinations help with the spread of COVID", can be easily shown from data analysis and visualizations, without further studies on a complex list of other human-behavior related topics and factors.

VI. Limitations

There are many aspects of vaccination that are very hard to model, such as: 1) There is a delay between the time of injection and the effective date; 2) Different brands and types of vaccines have different effectiveness, and different populations within the same county may have different vaccination rates; 3) Masking is likely to impact the apparent effectiveness of vaccination. All these variables of the vaccinations are likely to have a significant impact on my results. However, for simplicity purposes, in my analysis above, I ignored all these aspects and potential issues of vaccination, and simply focused on one variable, the total number and percentage of people who are fully vaccinated.

Aside from the vaccination rate, there are many other potential factors that could be the causes of the decline in COVID-19 cases and deaths, such as the masking policies, the stay-at-home policies, and the different patterns of human interaction. There is also a great chance that although factors such as vaccination will help, vaccines alone will not be sufficient to end the pandemic. However, at this point, I do not have enough resources to research and address this issue in the time allotted.

Besides the variables mentioned above, I also made some assumptions for simplicity purposes:

- No COVID-19 vaccines available to the public in 2020, so my assumption is the vaccination rate is zero for all counties in 2020
- Assuming the population at risk is the population of the whole county, and also assuming the population of the whole county stays roughly the same during this whole pandemic, the derivative function of the rate of infection would look the same as the derivative function of the total confirmed cases of COVID.
- I assumed that wearing a mask would not prevent you from death once infected, so I did not look at the correlation between masking policies and the death count.
- After late August 2021, we do not have data showing if masks are required, but my assumption would be there is no masking policy once it gets lifted in July 2021.
- I assumed that the high (local maximum) number of death counts is caused by the high (local maximum) number of confirmed cases from weeks before.

For now, without further research to prove my assumptions to be true, I consider my above assumptions to be part of the limitations of my research.

VII. Conclusion

With my research question being "how did the progression of COVID-19 cases and deaths change in Macomb County, Michigan in 2021, after more and more people got vaccinated?", and my hypothesis of "the higher percentage of people who are fully vaccinated would lead to the decline in COVID-19 cases and deaths in Macomb County, Michigan", I rejected my hypothesis. I did not find any evidence supporting that there is a correlation between the vaccination rate and the number of confirmed cases or deaths in Macomb County. From my research, I learned that not all the most obvious statements, such as "vaccinations help with the spread of COVID", can be easily shown from data analysis and visualizations, without further studies on a complex list of other human-behavior related topics and factors.

VIII. References

"What's next in the Evolution of COVID-19 Is Increasingly Difficult to Predict." *News*, 21 Sept. 2021, www.hsph.harvard.edu/news/hsph-in-the-news/whats-next-in-the-evolution-of-covid-19-is-increasingly-difficult-to-predict/. Accessed 10 Dec. 2021.

Ellyatt, Holly. "Fully Vaccinated People Are Still Getting Infected with Covid. Experts Explain Why." CNBC, CNBC, 10 Aug. 2021, www.cnbc.com/2021/08/10/breakthrough-covid-cases-why-fully-vaccinated-people-can-get-covid.html. Accessed 14 Dec. 2021.

IX. Data Sources

The following public datasets were used. All of these abide by the ethical considerations because it does not involve any PII or PHI.

- COVID-19 data from John Hopkins University from the Kaggle.
  - License: Attribution 4.0 International (CC BY 4.0)
  - This dataset allows me to access the number of confirmed cases and deaths by US county, from January 2020 to present, by US County by Day.
- Masking mandates by county from CDC
  - License: unknown
  - This dataset allows me to assess if it is a requirement for individuals operating in a personal capacity to wear masks 1) anywhere outside their homes or 2) both in retail businesses and in restaurants/food establishments, by US County by Day.
- Mask compliance survey data from The New York Times

    - License: https://github.com/nytimes/covid-19-data/blob/master/LICENSE
    - This dataset allows me to access the answer to the question "how often do you wear a mask in public when you expect to be within six feet of another person?" obtained from 250,000 survey responses between July 2 and July 14, 2020.
- COVID-19 Vaccinations in the United States,County from CDC.
    - License: Public Domain U.S. Government
    - This dataset allows me to access the total number and percentage of people who are fully vaccinated, and the percentage of county level fully vaccinated population from January 2021 to present, by US County by Day.
    - Note: No COVID-19 vaccines available to the public in 2020, so the assumption is the vaccination rate is zero for all counties in 2020.
- United States COVID-19 County Level of Community Transmission Historical Changes from CDC.
    - License: unknown
    - This dataset allows me to access the following weekly data point, from January 2020 to present, by US County:
        - Community Transmission Level Indicator [low, moderate, substantial, high, blank]