Samuel Perebikovsky

Data 512 - A5

1. With all the COVID data available I wanted to see how the vaccination progress was influencing the number of COVID cases throughout the course of the pandemic after the vaccines were starting to roll out. I thought that it might be interesting to see if there was a proportional decrease in the number of COVID cases after the fact. For example, it makes sense that after 50% of the population was fully vaccinated, we would ideally see something close to a 50% reduction of cases. This number might vary based on the number of people who already had COVID before the vaccines were rolled out; but it is difficult to say how much it would vary by because we do not know that everyone who had COVID before the vaccine didn't get the vaccine later. It would be interesting to analyze the effectiveness of vaccines and how they have limited the spread of COVID throughout the county of my analysis (Miami-Dade). But after beginning the search for this answered I stumbled upon some information that made me pivot the focus of my project into looking at the transparency of reporting for COVID vaccinations. This is human centered because we would be able to gain a deeper understanding of how trustworthy our data is.

2. My problem statement originally was: "How have the COVID vaccines changed the number of COVID cases in the Miami-Dade County?" But after doing some research and diving deeper into the question, I found something that made it extremely difficult to answer the previous question. The reporting of the COVID vaccinations was/is likely completely messed up. There are some zip codes in the county with over 100% vaccination rates (first shot). This was cause for me to investigate further. I was not able to find official data for COVID vaccines by zip code in my county (it exists but was not available to the public) so I was stuck at the county granularity which limited the amount of data analytics I could perform. I was eventually able to find an estimated data set by the Miami Herald but only after bypassing a paywall. From seeing some data and reading various articles I was able to see a much more concerning problem. So, my question pivoted to have more of an investigative journalism flavor: "How has the lack of transparency in the reporting methodologies of COVID vaccinations in the Miami-Dade County artificially inflated the vaccination rates for the Miami-Dade County; and how has this affected our understanding of the progress of the COVID pandemic?" [With this lack of data, I am wondering if it would be okay to use different news sources and analyses performed by other sources: I learned about this problem from multiple articles but have no way of accessing the data they used. If not, I can pivot my research question to something where I actually have more data (how COVID affected unemployment rates in the county, for example.)]

3. I am using mainly the CDC's vaccination data by county found here (https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh) as well as here (https://covid.cdc.gov/covid-data-tracker/#county-view|Florida|12086|Vaccinations|Administered_Dose1_Pop_Pct) for supplemental information. In addition to these data sources, I am using some articles to aid in the

answering of my new question which can be found here (https://www.miamiherald.com/news/coronavirus/article254801602.html) and here (https://www.local10.com/news/local/2021/09/22/90-of-miami-dade-vaccinated-its-just-not-true-experts-say/), here (https://www.nytimes.com/interactive/2021/us/miami-dade-florida-covid-cases.html), and here (https://floridapolitics.com/archives/458515-miami-dade-90-vaccination-rate/). I plan on citing all the sources I end up using. The data I used is mostly from the CDC which is mostly publicly available. Using articles which are closer to the true sources should give my analysis a heavier impact. That paired with some quantitative data will hopefully be enough to convey my point of how terrible the 'official' data is because of the terrible reporting structure and lack of transparency.

4. Clearly the greatest unknown in this problem is the actual COVID vaccination rate. With reports that Florida has been doctoring the numbers in many aspects of the COVID reporting, it becomes difficult to take any of the data seriously. There is a period of time in which Florida completely stopped reporting COVID deaths because they had determined that the worst of the pandemic was over. I am also lacking data of covid vaccinations at the zip code level of granularity, which is severely undercutting my ability to provide quantitative results for my questions. There is also the consideration that qualitative sources are usually less convincing than conclusions drawn from quantitative data so I will have to ensure I cite enough sources to convince people.

5. My methodologies for answering this question are perhaps not so standard. I was originally planning to compare the number of COVID cases before and after the vaccine to observe the differences in infection rate and to see if there was a significant impact from the vaccine. However, given the strangely high numbers of vaccinations in a county and state that have been more opposed to vaccinations and COVID preventative measures than many other states I began to wonder how accurate the information was. Diving deeper into this suspicion I began to find articles that were claiming these numbers were likely highly inflated and did not accurately reflect the true vaccination numbers of the county. I scoured the internet for the official data by zip codes across the Miami-Dade County but to very little avail. So, I am left using the analysis that the various articles performed to essentially show the complete lack of transparency in the COVID reporting done by this county and how much confusion this caused. I want to use heatmaps to show just how inflated some of the data from the zip codes were both in Miami-Dade and the nearby Broward County. I also plan on using quotes from people who had access to the official data to get a perspective on how messed up the reporting for COVID vaccinations and how it also can invalidate much of the other data collected in relation to the pandemic. There is a problem between the relationship between the state governor and the leadership of Miami-Dade County and that is partially a reason for these terrible reporting structure and complete lack of transparency. I want to frame my presentation in a way that is easy and logical for humans to understand just how strange the reporting is and how those lack of good practices allows people to form conclusions (based on essentially falsified data) that has no bearing on what is actually happening in Florida with respect to the COVID pandemic.

6. A big potion of my timeline includes finding data and articles and doing some research to get a wider understanding of how the reporting works in the Miami-Dade County. There is very little data regarding my question, so I have a slightly easier time in that regard. But because of this I need to analyze as much qualitative data as I can and try to have these observations lead to some interesting conclusions about the transparency of COVID reporting in Miami-Dade County. I will be graphing the COVID cases infection rate along with the vaccination rate to look at the potential relationship between them. From there I will begin to demonstrate why any kind of analysis on the previous data will yield less than accurate results. My analysis will begin to pivot into the articles that I found essentially showing that, I will probably finish this part over the next weekend. I will include a couple choropleth map of the county and similar nearby counties by zip code to show the zip codes that have greater than 100% vaccination rate to show just how messed up the reporting for COVID is. I want to structure my presentation and report in such a way that shows the natural projection from trying to answer a question using data into an exploration of how the data was collected; and how the methods of collection and reporting lead to completely insufficient analyses using that data. How can we understand the true effectiveness of the vaccines without the data that accurately reflects 'real life'. I plan to create my presentation and work on a script for the presentation, I will collect quotes and snippets from the article to help convey my point to the audience and lead them on an abridged version of the journey I took to understand just how bad the transparency of the Miami-Dade County is in reporting COVID cases. I will probably spend about a week to a week and a half working on the basic format of my presentation and have a nice outline along with a script to go with it. From there I want to spend a few days refining and improving my presentation as well as starting on my final report. The last week before the presentation will be focused on reviewing my work and making sure my logic and presentation makes sense.