Impact of Mobility on the spread of COVID-19

Jeremie Poisson - DATA 512 University of Washington

Introduction

The COVID-19 pandemic has changed the world considerably and it is hard for any of us to fully understand the scope and the implications that this will have over time. It is safe to say that Government officials along with the scientific community could not have potentially anticipated a pandemic of this size in this day and age. With the constant fear of getting sick and the sheer amount of contradictory information broadcasted online, information from the mainstream media along with the current political discourses in the United States, it is extremely hard for people to determine the effectiveness of the different strategies deployed by local governments in stopping the spread of COVID-19. Overall, we know that people are becoming increasingly disengaged as the pandemic progresses, which in turn probably impacts negatively the public effort in complying with these measures. In addition, the apparition of new variants means that COVID is here to stay and it is critical to be better prepared to deal with the spread of the disease from a public safety perspective from now on. The pandemic has been greatly "datafied" by the scientific community and with the sheer amount of raw data available online it becomes increasingly difficult to parse the information correctly and get a comprehensive review on the measures used to counter COVID-19 based on actual data. In this report, we are hoping to provide more of a definitive answer as to the effectiveness of the masking mandates employed by New Jersey's officials as well as analyze how the mobility and the stay-at-home orders impacted the COVID-19 infection rates throughout the pandemic. This report covers some background information, the methodology used throughout the analysis, highlights our findings and their limitations.

Background and related work

We scoped this analysis of the region of **Bergen County, NJ**, just outside of New York City. New York City is arguably one of the cities that have been the most affected by the pandemic due to its extreme population density. By scoping our analysis to a region just outside of such an affluent city, we consider it is the perfect location for analyzing how the mobility of people has impacted the course of this pandemic.

As we mentioned before, there are enormous amounts of data available publicly online about the pandemic coming from different sources. The US government has been great at publishing various public datasets online about the number of infections throughout the US daily for every county in the United States. The CDC actively published new datasets which include the current state of infections, forecasting, vaccination data, health equity data, and much more. Hospitals such as John Hopkins are also involved in the "datafication" of the pandemic. Many publish daily hospitalizations, vaccinations, and confirmed cases of COVID-19. More precisely, in this report we make great use of John Hopkin's daily confirmed COVID-19 infection dataset which provides the total number of confirmed COVID-19 cases on a per county basis for the whole nation. Big technology companies such as Google, Apple, Uber, Waze, and much more also got involved in openly publishing auxiliary data that can be used to analyze the impact of the pandemic by correlating users' activity on their platforms. More specifically, Google and Apple maintain a daily "mobility report" which measures how mobile are their users daily according to a pre-pandemic baseline. We also make great use of this resource in this report. Finally, multiple newspapers such as the New York Times are openly publishing survey data that can be used to gauge people's engagement towards the different measures in place such as masking compliance -which we make great use of in this report.

Methodology

As mentioned above, we leveraged the following datasets throughout the analysis. Below is a description of each dataset as well as it's purpose in the context of this report:

- **Official COVID-19 data from John Hopkins University**
  This dataset aggregates the total number of confirmed COVID-19 cases daily throughout the world. The data comes from different sources online and you can find a list of the official sources in the references. The data is refreshed daily and reports the confirmed cases of COVID-19, not necessarily the suspected cases or infections. We mainly use this report to get an accurate picture of whether the number of cases is increasing or worsening daily.
- **U.S. State and Territorial Public Mask Mandates**
  This reports the states and territorial executive orders, administrative orders, resolutions, and proclamations collected from the various US government websites. It reports the date at which such orders were issued and when they expired. At the time of writing, the dataset was last updated on September 10th, 2021. We used this data in the first part of our analysis to check whether the masking mandates were effective in countering the spread of COVID-19 in **New Jersey**.
- **Mask-Wearing Survey Data from the New York Times**
  This data estimates the mask use by counties in the United States. It aggregates survey

data in which each participant was asked "*How often do you wear a mask in public when you expect to be within six feet of another person?*". We use data to gauge whether masking mandates were uniformly respected by the population in **Bergen County, NJ**.

- **COVID-19 Mobility Data Aggregator (Github: ActiveConclusion/COVID19_Mobility)** This is a data scraper of mobility reports in different formats. As mentioned above, big technology companies such as Google and Apple released different "mobility reports" which aim to provide insights into what has changed in response to the work from home, shelter in place orders. In this analysis, we used data from the **Google COVID-19 Community Mobility Reports** (reports mobility trends across different categories) as well as the **Apple COVID-19 Mobility Trends Reports** (reports mobility trends in terms of driving, walking, and public transit). Those two reports provide the percentage change in mobility according to a baseline that was established in the pre-pandemic era.

Analysis of the COVID infections data

The first part of our analysis was to look at how masking mandates contributed to the "flattening of the curve" throughout the pandemic. Looking at the number of confirmed COVID-19 cases for Bergen County, NJ, we quickly realized that it was not necessarily useful to look at the cumulative active cases. Since we're focusing our attention on whether masking mandates have been effective in preventing the spread of COVID (making the situation better or worse in terms of *new infections*), looking at the daily increases in confirmed cases makes more sense. We achieved this by subtracting the cumulative number of cases from the day prior. In addition, we overlaid the periods from which a masking mandate became effective in Bergen County, NJ. This was to confirm whether we could see a decline in new daily cases after the mandate was in effect. Moreover, according to the New York Times survey data, people were mostly compliant with masking rules.

Table 1: Proportion of people who answered "*How often do you wear a mask in public when you expect to be within six feet of another person?*" (filtered down to Bergen County, NJ)

NEVER	RARELY	SOMETIMES	FREQUENTLY	ALWAYS
~2%	~1.4%	~5.3%	~15.1%	~76.2%

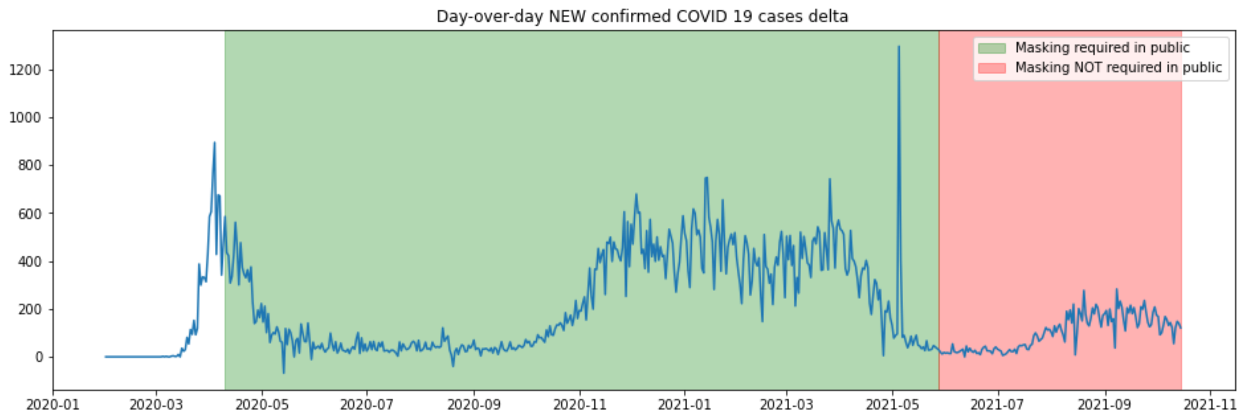

Image 1: Day-over-day NEW confirmed COVID 19 cases with masking mandate time periods

From the image above, you can see that masking (and the shelter-in-place executive orders) are likely to have played a big role in decreasing the number of new cases on a daily basis. One big issue we had to deal with when analyzing this data was to deal with the **incubation period**. Indeed, since the John Hopkins' data is reporting only "confirmed" cases of COVID-19, it means that we're likely to see the impact of masking measures a few days later after it has been implemented. This is referred to as the **incubation period**, which is the time needed for a person to develop serious symptoms and get tested for COVID-19. This will obviously offset the data by a few days. To counter that issue, we zoomed into the graph above around the time the masking mandate was issued in hope of finding the typical incubation period length through the data, by manually localizing the first significant drop in new COVID-19 infections right after the mandate was issued. Surely enough, we found that the incubation period is typically 14 days (which is in line with the official CDC guidelines published online).

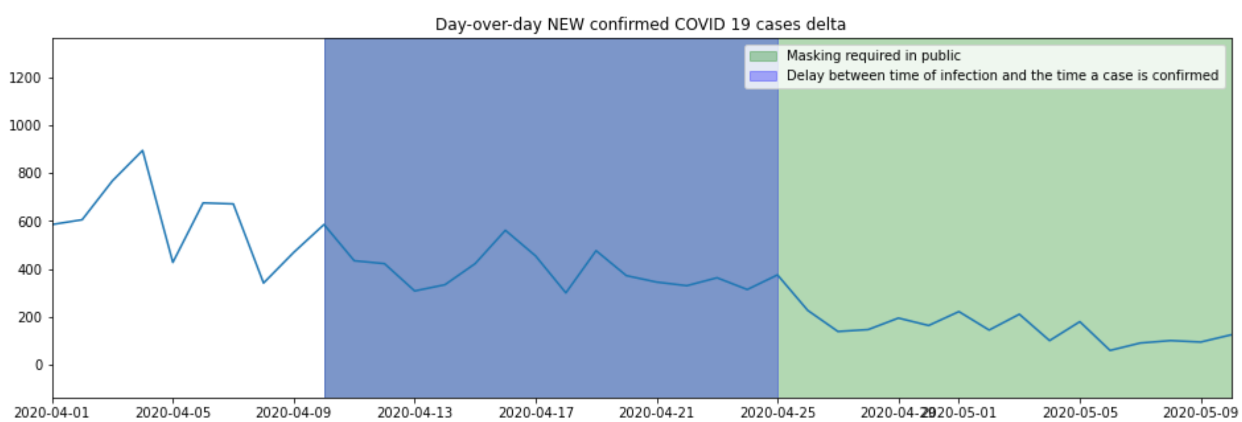

Image 2: Zoomed in version of the graph above, highlighting the first significant drop after the masking mandate was in effect

Now that we figured the incubation period length, we used **binning** to counter (or offset) the data to take into account the incubation period. This had the effect of averaging out the new infection cases over 14 days buckets which dealt with the data offset imposed by in the incubation period.

Finally, in order to get a better idea whether the situation was improving or worsening in terms of new COVID-19 infections, we derived the change in slope in the difference between the increases and decreases of new infections (getting the increase/decrease delta with the day prior).

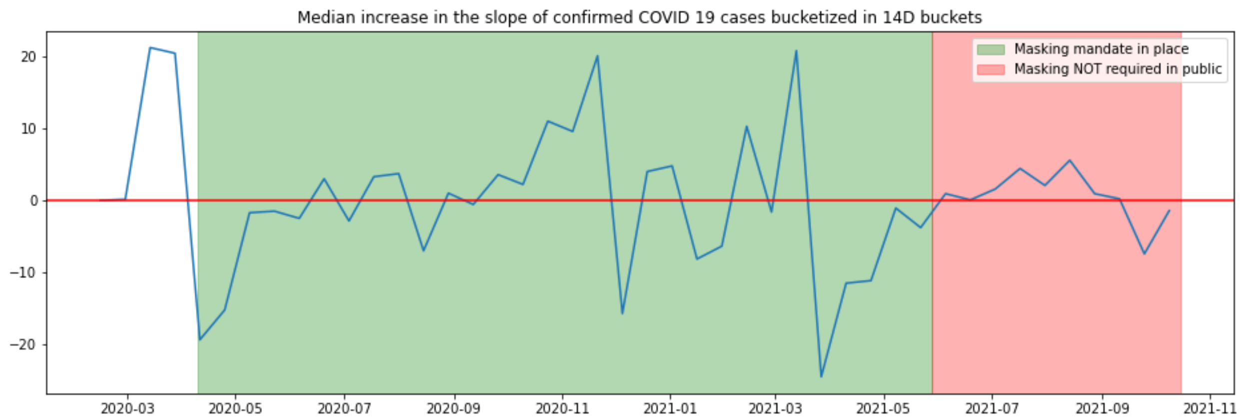

Image 3: Change in slopes in increases of COVID-19 infections

Analysis of the mobility data

Luckily, thanks to the hard work of the contributors of the ActiveConclusion/COVID19_Mobility datasets, the data was already correlated and standardized into a common format for us. Since we were interested in the mobility trends from the beginning of the pandemic to this day, we were constrained to only take the **Google** and **Apple** mobility data. This is because the **Waze** and **TomTom** data was limited and stopped in December 2020. Please see the limitations of this project for more details.

The **Google Mobility Report** data provides a percent difference against a baseline that was determined pre-pandemic. It reports insights into what has changed in terms of users' mobility in 5 broad categories: Grocery & Pharmacy, Parks, Transit Stations, Retails & Recreation, Residential, and Workplaces. This data is available on a per county basis for the United States.

The **Apple COVID-19 Mobility** data is similar to the Google Mobility Report described above, except it focuses on the transit mode used by users on their platform: Driving, Walking, and Public Transit. One caveat here is that we averaged out the percentage baseline for transit with the score given in the Google report above. Please see the limitations of this project for more details.

For the most part, we were able to simply plot the data against the COVID-19 cases above without any issue.

Findings

Research question #1

We made several findings throughout this analysis. The first part pertains to the impact of masking and the different mechanisms in place to counter the spread of COVID-19.

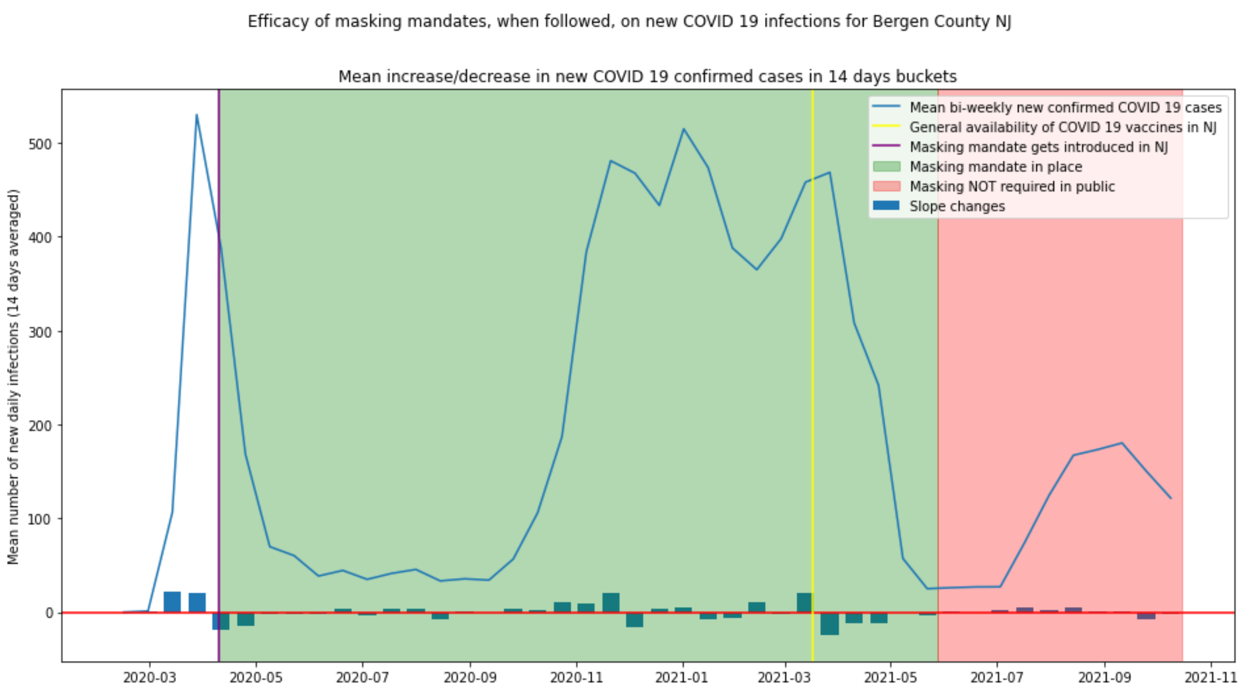

Image 4: Mean increase in new COVID 19 confirmed cases in 14 days bucket for Bergen County, NJ

The figure above shows the mean change in new COVID 19 confirmed cases in 14 days buckets for the period of 2020/02/01 through 2021/10/15 for Bergen County NJ. The Y-axis represents the mean number of daily infections (14 days averaged) whereas the X-axis represents the time since 2020/02/01. The purple line shows when the masking mandate was first introduced on 2020/04/10 and the yellow line shows when the COVID 19 vaccine first got available to the overall population roughly on 2021/03/17 as per New Jersey state ordonnance. The shaded green and red background show when the vaccine mandate was in place (people were compliant in wearing masks in public) as well and when it got removed respectively. The slope change bars in dark blue at the bottom of the graph is showing how the average mean of new COVID 19 confirmed cases is changing over time. A slope change close to zero means there is little to no change day-over-day of the mean number of new COVID 19 cases discovered over that period (roughly the same number of people end up catching COVID 19 every day, the pace at which people catch the disease is steady). An increase means that the infection rate is growing (more people end up getting sick, on average, compared to the previous period, the pace at which people catch the disease is accelerating) whereas a decrease means that the infection rate is decreasing (fewer people end up getting sick, on average, compared to the previous period, the pace at which people catch the disease is decelerating).

Research question #1: How effective were the preventive measures enforced in stopping the spread of COVID-19 ?

> You can see from the graph that the mean number of new COVID 19 cases falls rapidly after the introduction of the masking mandate as well as the general availability of COVID 19 vaccines in the state of New Jersey. Furthermore, looking at the slope change after those events is also a great way to validate this hypothesis. Another auxiliary finding here is you can see how the vaccine has greatly impacted the spread of the disease and the number of new cases plummeted almost back to zero.

Research question #2

For simplicity's sake, we grouped the different mobility areas in 4 broad categories and plotted it alongside the slope changes from the first research question. This gives us a clear picture of whether the COVID-19 infection rates were worsening or improving over time along with how the mobility of people changed in the different areas. By doing so, we can infer whether the mobility in certain areas might have played a role in spreading the disease more significantly than others. The hypothesis we had at first (and which is the scientific consensus) was that indoor activities were more conducive to spreading the disease whereas outdoor activities are more likely to not impact the spread of the disease at all.

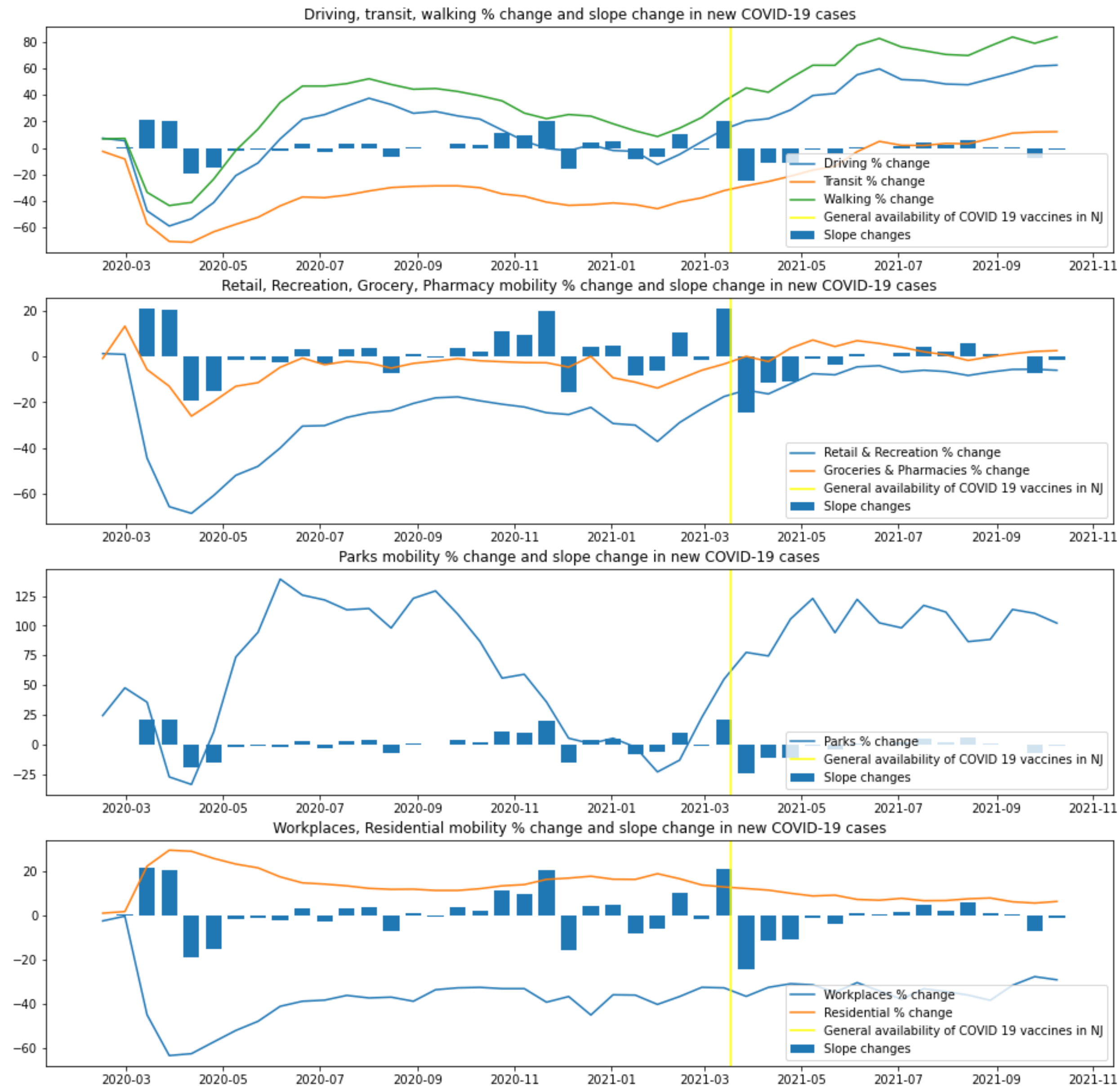

Image 5: Mean increase in new COVID-19 cases in relation to activity in mobility areas

The figure above shows the mean change in new COVID 19 confirmed cases in 14 days buckets for the period of 2020/02/01 through 2021/10/15 for Bergen County NJ. The Y-axis represents the percentage difference in mobility areas (14 days averaged) whereas the X-axis represents the time since 2020/02/01. There are 4 broad categories: **Transportation** (driving, walking, and public transportation activities)**, Indoor** (retail & recreation, groceries & pharmacies activities)**, Outdoor** (park activity)**, and Living Spaces** (residential and workplaces activities)**.** The yellow line represents the general availability of COVID 19 vaccines in New Jersey.

Research question #2: How did mobility impact the COVID-19 infection rates throughout the course of the pandemic?

Not so surprisingly, it seems like any activity involving the displacement of people and indoor activities are likely to have played a significant role in the spread of the disease. In the **Transportation** and **Indoor** mobility areas, the first section of the graph until roughly 2020/05 shows that a decrease in those mobility areas corresponds to a decrease in new COVID-19 infections (change of slope is decreasing). The subsequent increase and the relatively stable changes in slope until we reach the baseline (0%) on 2020/07 actually strengthen our first hypothesis that masking plays a big factor when dealing with indoor activities. Furthermore, you can see the impact of the new COVID variants as well as the COVID-19 vaccine implementation.

You can also notice that the **stay-at-home order** seems to also have played a key role in limiting the spread of COVID-19 as denoted by the last graph (**Living Spaces**).

One final key finding is that, as expected, **Outdoor** activities do not seem to play a significant role in the spread of COVID-19.

Limitations

There are obviously several limiting factors that were considered throughout this project. In this section, we explore some limitations in detail and explain the plausible effects they have had on the study.

- The scope of the analysis is for Bergen County, NJ. Bergen County is located just on the outskirts of New York City which has been hit historically hard throughout the pandemic. There are a lot of people commuting in and out of the city which certainly had an impact on the COVID cases and mobility analyses above.
- In this project, we surfaced a lot of correlations between the different mobility areas as well as the confirmed COVID cases. As all good statisticians know, causation does not imply correlation. The COVID pandemic is an extremely complicated situation that is being influenced by a multitude of internal and external factors which are not in the scope of this analysis. It is important to take those results with a grain of salt before making any assumptions about the spread of coronavirus.
- Using binning to offset the data with the incubation period is an imperfect strategy. It is important to understand that confirmed COVID cases are impacted by a lot of external factors such as location, seasonality, indoor vs outdoor activities, etc. By averaging it out, we believe it is possible to average out the number of cases just enough to offset

and get a clear picture of what was happening. In addition, this data has certainly been impacted by the apparition of new COVID-19 variants which have been reported in many cases to be spreading at a much faster rate. The vaccine availability also impacted the number of cases that were reported.

- Another important limitation of this study is the lack of actual statistical analysis. Although we tried to apply several methods to rationalize the results between COVID cases and mobility, we could not find any sound analysis to explain the correlation between the two. Instead, we decided to focus our attention on visualizations and the hypotheses around why those two factors might be influencing each other, leaving the statistical analysis part out of it. This is obviously a major flaw of this analysis and is left as future work.
- There are multiple ethical considerations to consider in this analysis, especially in the use of “public” mobility data. I was personally unaware that Google, Apple and alike are collecting transit and mobility data. On their website, they do mention that the data is completely anonymous and it is clear those companies intend to help the public going through this pandemic. This leaves an open-ended human-centered question about the proper use of this data. I think users should have the right to know that this kind of data can be collected and published online even if anonymous. In my opinion, it should also be clear how this setting can be disabled from my Google/Apple account from the get-go. This data should be owned by the user, and users should not be put in a place where their data is being used against their will or put in a position where their data is being collected “by default”.

Again, we must stress the importance of these limitations in this research project. We believe this is a great starting point to anyone willing to further investigate the effects of mobility of COVID-19, but we believe that it is not mature enough to make any strong conclusions on the matter.

Conclusion

In conclusion, it is clear that throughout this report we found that mobility seems to have played a role in the spread of COVID-19, especially within indoor environments. COVID-19 is an ongoing public health concern that is evolving every day and that is here to stay. Doing this analysis really impacted my understanding of human-centered data science in the sense that we made good use of data to better understand one problem that, in particular, is impacting every human on this planet. There is a serious lack of ethics in the current media cycle and this project aimed to provide an unbiased, data-driven approach of analyzing the data produced during this pandemic and to put it in good use to solve real-world problems at a larger scale.

References

COVID-19 Community Mobility Reports. (n.d.). Google. Retrieved December 13, 2021, from https://www.google.com/covid19/mobility/

COVID-19 - Mobility Trends Reports - Apple. (n.d.). Coronavirus Self-Checker. Retrieved December 13, 2021, from https://covid19.apple.com/mobility

Goldbloom, A. (2020, 11 02). *COVID-19 data from John Hopkins University*. COVID-19 data from John Hopkins University. https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv

Kassinger, C. (n.d.). *US State and Territorial Public Mask Mandates From April 10, 2020 through August 15, 2021 by County by Day | Data | Centers for Disease Control and Prevention*. CDC Data Sets. Retrieved December 13, 2021, from https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i

Liebeskind, M. (n.d.). *ActiveConclusion/COVID19_mobility: COVID-19 Mobility Data Aggregator. Scraper of Google, Apple, Waze and TomTom COVID-19 Mobility Reports* 🚶🚘🚉. GitHub. Retrieved December 13, 2021, from https://github.com/ActiveConclusion/COVID19_mobility

Management of Patients with Confirmed 2019-nCoV. (n.d.). CDC. Retrieved December 13, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html

New York Times. (n.d.). *Mask use*. GitHub. Retrieved December 13, 2021, from https://github.com/nytimes/covid-19-data/tree/master/mask-use

Data Sources

- The COVID-19 confirmed cases data is from (Goldbloom, 2020)
- The masking mandate data is from (Kassinger, n.d.)
- The mask use data is from (New York Times, n.d.)
- The Google Mobility report data is from (*COVID-19 Community Mobility Reports*, n.d.)
- The Apple Mobility report data is from (*COVID-19 - Mobility Trends Reports - Apple*, n.d.)