Reflection Statement

In Data Science initiatives and human-centered design tasks, the collaborative model is beneficial as it plays a crucial role in providing diverse and continuous feedback to improve the overall design. Especially for this task, given the anonymity in the virus dynamics and limited knowledge about the nuances in the study, it is critical to give and take inputs in a collaborative cohort setting. Discussing multiple concept designs, approaches, and implementations to obtain a robust solution and trying out multiple algorithms and metrics were all highly beneficial in performing the detailed analysis and obtaining the final visualization.

Talking about the learnings from the assignment, the first key takeaway is that the data collected and used for any analysis needs to be cleaned thoroughly. Due to collaboration, I was able to know the downside impacts of these data gaps/outliers, and I was able to remove them from my county data, where deaths are reported negative for a few days, and cases are undercounted during the weekends. As the professor mentioned in the week five lecture, there can be side effects of datafication where few nuances can be overlooked because of the scope/considerations made while collecting the data. Given the data is time series, I cleaned the data by smoothening the outliers on a few dates using moving averages.

Another key learning is that the data we use is not always complete. We should not confuse correlation with causation. Though we concluded from our analysis that masking policy impacts the infection rate, we can't attribute the increase or decline solely to one factor (i.e., mask mandate). When discussing the change points in the data, we were drawing on the potential reasons for intermittent spikes in cases though there is a mask mandate. We then considered hospitalizations, recovery periods, time for testing and vaccinations, and other implicit factors and how they will play critical roles in the modeling process.

Impact of vaccination: Both visuals presented in the analysis don't consider the vaccination state for calculations. To simulate the impact of vaccination on the state of the disease, we came to know that we can leverage the SIRV model with a valid vaccination rate (p) and reinfection rate (alpha). This slightly changes the calculations of infection rates, but the story doesn't change much.

Also, while discussing, we all agreed that the impact is not always imminent. The change can be seen in the data after a few time intervals, and we tried to understand how it is different for other counties with stringent/no masking policies. One interesting finding is that though most counties had a masking mandate during the December 2020, and the initial phase of vaccination was started; there was an increase in cases. The reason can be that people travel during Christmas and New Year (holidays), and the impact of masks is less than the infection spread due to high mobility. In general, we can see that the infection rate slowly started to drop after the initial vaccination happened in December and the holiday season was over.

Initially, I tried various derivative analysis techniques to understand the change in virus spread. Still, with the help of Charles Reinertson, I got familiar with the concept of change point detection. I did my research and discovered ruptures, a package that uses regression splines to fit the data and identify the pivotal points where there is a change in derivative/slope. I then shared it with the cohort as I felt it was very interpretable and discussed how these change points could be mapped to the mask mandate policy. I also discussed the SIR modeling with Hriday Bhagar, where we tried to model the recovery state and reinfection rate into our model. It is super productive how this collaborative exercise brought together people with common backgrounds and research interests while simultaneously sharing the diverse school of thought to build a robust solution.

Overall, brainstorming with the cohort helped me refresh my understanding of epidemiological modeling and share new ideas about the study. I could share my thoughts more freely and adopt a few techniques that I felt were interesting. This collaboration freedom helped me accelerate my analysis and focus more on understanding and incorporating the nuances of the study in my model.