

Reflection Statement on Collaboration:

I found the collaborative activities in this assignment to be very helpful in understanding how to effectively communicate data visualization information and what are the best approaches we can take to solve this problem. I collaborated with my friend Harry McNinson, and we exchanged several ideas about the steps we should take to reach our final solution. I also found it helpful in getting feedback from him on my visualization. Also, the collaborative activities helped me understand the different ways to approach a problem and the different ways to think about a problem. It was also helpful to see how other people approached the same problem by checking the discord group for this class.

Based on the research questions posed in this assignment, the first thing that I learned from this research is that there is often a delay between the time of infection and the time a case is confirmed. This delay can be caused by several factors, including the incubation period of the disease, the time it takes for symptoms to appear, and the lag time between infection and the availability of testing results. This delay can have important implications for the spread of the disease, so it is important to take it into account when modeling the pandemic. The second thing that I learned from the assignment is that masking may simply make it longer to get infected or it may prevent some percentage of infection and we should consider the effect of a mask when making decisions about whether to wear one.

The possibility of collaboration on this part helped me to think more deeply about the problem and to consider different perspectives. It also helped me to see the problem from a different angle and to better understand the potential solutions. Also, it allowed me to see how other colleagues might think about the same issue and what their perspectives might be. This helped me to come up with the steps to reach the final solution. The details for the same have been given below:

- I started with understanding the column names for all the 3 datasets
- I did the visual inspection of the data and after that I loaded the data using pandas in my Jupyter notebook and started exploration of the data.
- With exploration, I came to know that there is a lot of missing data present in the data frame. For the data with the infection numbers, I used the back fill and forward fill methodology to deal with the missing data. This type of methodology is generally used when we work on time series data.
- I used the same methodology (back fill and forward fill) for the mask mandate data set.
- With the concatenation of mask mandate and infection rates data frame, I started understanding the effect of the interruption in the time series data.
- I started with the OLS methodology which is a very Naïve approach and thereafter used the ARIMA model for my analysis.
- ARIMA model helped me to understand the effect of mask mandate on the rate of infections. Also, by plotting contrafactual, I understood about the effect if masks would have not been placed as a mandate and where the infection would have led without it.

- Although with the help of mask the infection rates do not seem to decrease at a rapid phase but with the help of contrafactual we saw that the infection would have spread linearly among the population if the mandate was not put in place. With the help of the mandate, we can see that the infection curve does not grow exponentially and rather flattens once the mask mandate has been put in place.