**Reflections1**

**In this step we want you to create a graph that visualizes how the course of the disease was changed by masking policies. For your county, you should create a time series showing the changes in the derivative function of the rate of infection. Your graph should indicate days where masking policies were in effect (or not) and whether the difference in the derivative function was significant. Optionally, you can add a second time series that shows the actual rate of infection.**

First, for the visualization you created in Step 2, you should write up an explanation of the visualization. Some of the important things you might need to explain include: What does the figure show? How does the viewer “read” the figure? What are the axes, and what do they represent? What is the underlying data and how was it processed? You might think of this explanation as an extended figure caption. This explanation should be no more than one written page. Making a good effort now will make it easier to write your final report for Part 4.

I have shown 2 visualizations in the end. The first visualization shows the trend in the infection rate in the Hamilton County of the Ohio state with change in time. The x-axis is the timestamps and y-axis is the number of cases. Each data point on the chart tells you the infection rate on a particular day. The infection rate is calculated by dividing the number of daily confirmed cases with the total population of the county. The number of daily confirmed cases is the rolling average of last 7 days’ daily confirmed cases and the population is assumed to be constant during the duration of the analysis. The color of the lines in the chart represents the mask mandate policy in place for Hamilton County. ‘-1’ in the legend means that the mask mandate policy is unknown, ‘1’ means mask mandate policy is ‘Yes’ and ‘0’ means mask mandate policy is ‘No’. The period used to plot this chart is from Feb,2020 to October, 2021. The red vertical lines in the chart are the change points during which the derivative of the curve changes significantly but not much inference can be drawn from the change points here.

We can see that as the mask policy changed to ‘Yes’/1 , the infection rate started soaring. Although it seems counter-intuitive this could be because the mask mandate policy was enforced as the cases started soaring. But later, the infection rate started decreasing which could be because of the continuation of the mask mandate as ‘Yes’ and new vaccination policies. Later , When the mask mandate policy was lifted, the infection rate kept decreasing for a while before it started increasing again.

In the second visualization, only the y-axis is changed from infection rate to derivative of infection rate. We can see a lot of fluctuation in the middle where the disease seems to be uncontrolled and this is the time when the mask mandate policy was enforced.

This is the period when covid was at its peak. Hence wearing masks was mandated but it did not control the spread although the survey results tells that 70% of the population in the county wore masks frequently. This could be because more tests were conducted during this period than before due the easy availability of the rapid test kits.

The data source used here is the raw\_us\_confirmed\_cases.csv from the [John Hopkins Kaggle repository](https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university). The data contains information on County level and the metadata consists of fields like Province state, Admin2, Lat, Lon, UID , FIPS and number of confirmed cases for each day since January 2020. The number of confirmed cases in the dataset was cumulative. Some data preprocessing and cleaning was done to get the daily confirmed cases ,infection rate along with other fields. This data was merged with the US mask mandate dataset to retrieve the mask mandate policy for each day.