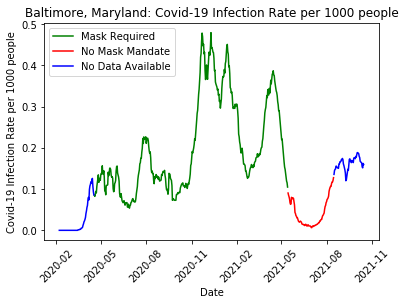
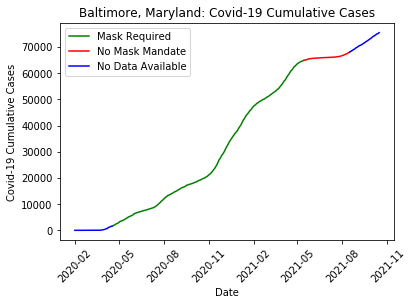
A4 - Common Analysis

Corina Geier

11/4/2021



### 

For my first visualization I showed the Cumulative Covid-19 Cases in my county of Baltimore, Maryland from February 1, 2020 through October 15, 2021. The second visualization shows Covid-19 Infection rate per 1000 people. For both charts, when the line is green, it signifies that there was a masking mandate in place, red signifies no masking mandate, and where the line is blue, there is no data available on masking mandates. The x-axis shows the month and year date. In the Cumulative Covid-19 Cases chart, the y-axis shows the number of cumulative Covid-19 cases, ranging from 10,000 to around 80,000. In the Infection Rate per 1000 people visual, the y-axis depicts the infection rate, ranging from 0 to .5 roughly. The lines on both charts do not show the exact number of each point but are meant to show the viewer the overall trend in cases or infection rate as time goes on. For instance, the viewer can see that as cases started picking up in April of 2020 a mask mandate was imposed, and then when cases seemed to plateau around March of 2021 the mask mandate was dropped.

It is worth noting that mask mandate data was only available from April, 2020 through August, 2021. The reader might assume that prior to April of 2020 there was no masking mandate in place, but after August, 2021 it is unclear.

The masking mandate data was found on the [CDC Website](https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i) and was filtered to only Baltimore county in Maryland, the column “Face\_Masks\_Required\_in\_Public” was used to determine whether or not a masking mandate was in place. Probability of wearing a mask and vaccination status was not considered in this analysis.

The information on Covid Cases was found on the [Kaggle repository of John Hopkins University COVID-19 data](https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university). It was filtered for dates between 2/1/2020 and 10/15/2021 for Baltimore county in Maryland. Given daily cases, cumulative cases were calculated by taking the cumulative summation of daily cases. The infection rate per 1000 people was calculated as daily cases with a rolling average of 7 days, divided by the population at risk all multiplied by 1000. The daily cases rolling average was used instead of the daily cases to account for delay in time of infection and time of confirmation. The population at risk was defined as the total population of the county minus the cumulative cases.

These two graphs together show both the overall picture of how many cases there are total at a given time, as well as how many new people are getting infected each day and how both of these measures are affected by masking mandates.

1. The collaborative nature of this assignment really helped me to think about possibilities I wouldn’t have thought of otherwise. When thinking of the question posed as to how to deal with the delay from the time of infection to the time of confirmation I found the discussion on slack between Grant Savage and Kevin Sweet to be helpful. Grant suggested using a 7-day moving average and provided the below code which I used to account for the delay.

* df.column.rolling(window=7).mean().round()

When considering how to calculate Infection rate I found [this informational document](https://healthcentricadvisors.org/wp-content/uploads/2017/03/Cal_Inf_Rates.pdf) posted by Emily Lineberger to be very useful in getting a working definition to use in my calculation.

For the visualization portion of the assignment I found the graphs posted by Patrick Peng and Apoorv Sharma to be helpful guides in considering how to display the information. I was considering how best to show whether or not there was a masking mandate and found their method of using color to show the masking mandate status very helpful.

Something that I did not use as it did not apply to my specific data set but that I found interesting, was the discussion people were having over what to do with counties that didn’t have any masking information. I saw some discussions around using nearby state or county information, or even scraping tweets to conduct a sentiment analysis on the use of masks in a particular area. It will be very interesting to see the kinds of outside data people bring in to attempt to answer this question.

A related issue that I will need to explore later is the question of how many people are actually wearing masks. For instance, some people might not wear a mask even given a masking mandate and some people might wear a mask even if there is no mandate. While there was some discussion of this problem on the class slack, I think people are still considering how to address this problem, and it has also caused me to think deeper about how I want to consider this information.

Regarding modeling, the class discussions have also helped me to do some outside research for the next parts of the projects. I saw some discussion around the SIR model by Grant Savage, Ryan Williams, and Kevin Sweet, which prompted me to start thinking about what I could do with the data I have and what data I would need. I did some more research into the SIR model which requires information on the susceptible, infectious, and recovered populations. Therefore, this model would require some outside data on recovered people, as currently I do not take into account those people that have gotten Covid-19 and recovered from it.

Overall, the discussion with my peers around how to tackle the question of how masking policies changed the progression of confirmed Covid-19 cases from 2/1/2020 to 10/15/2021 was very informative and helped me think of ideas I would not have considered on my own.