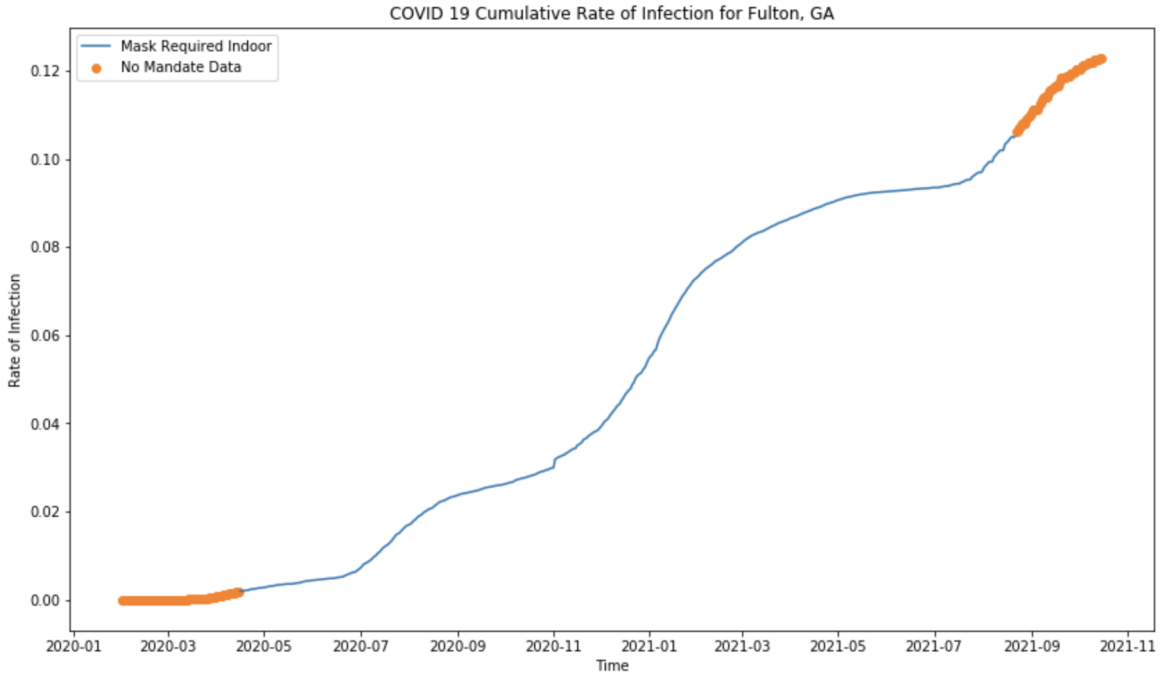
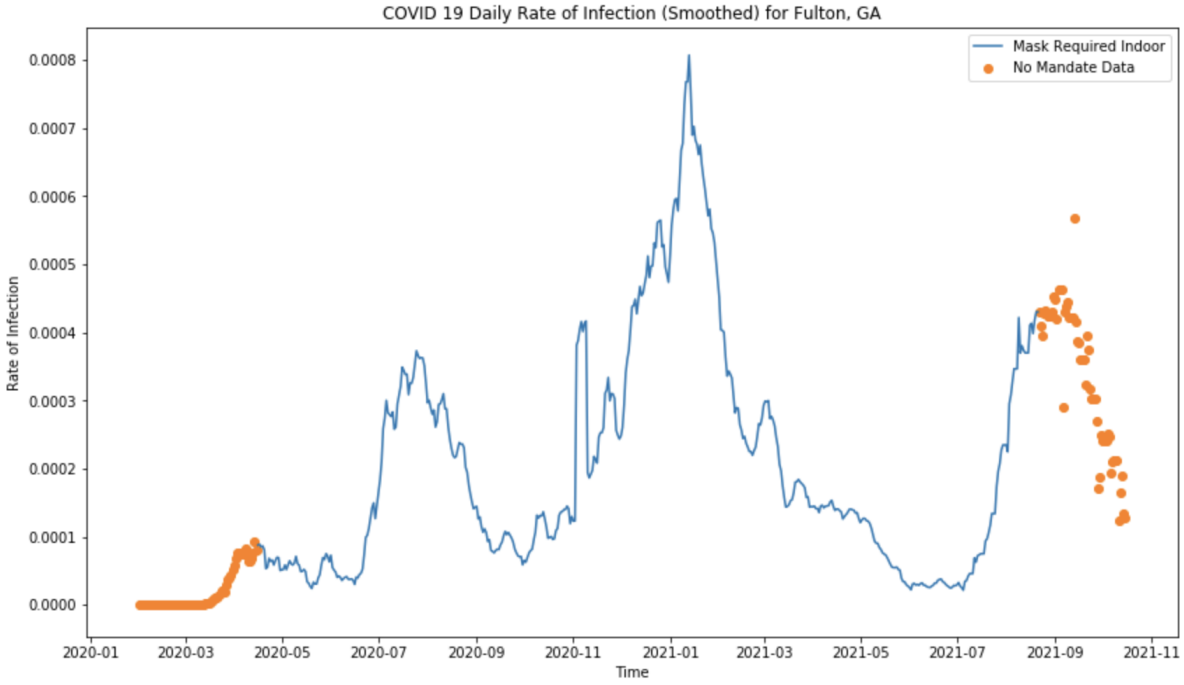
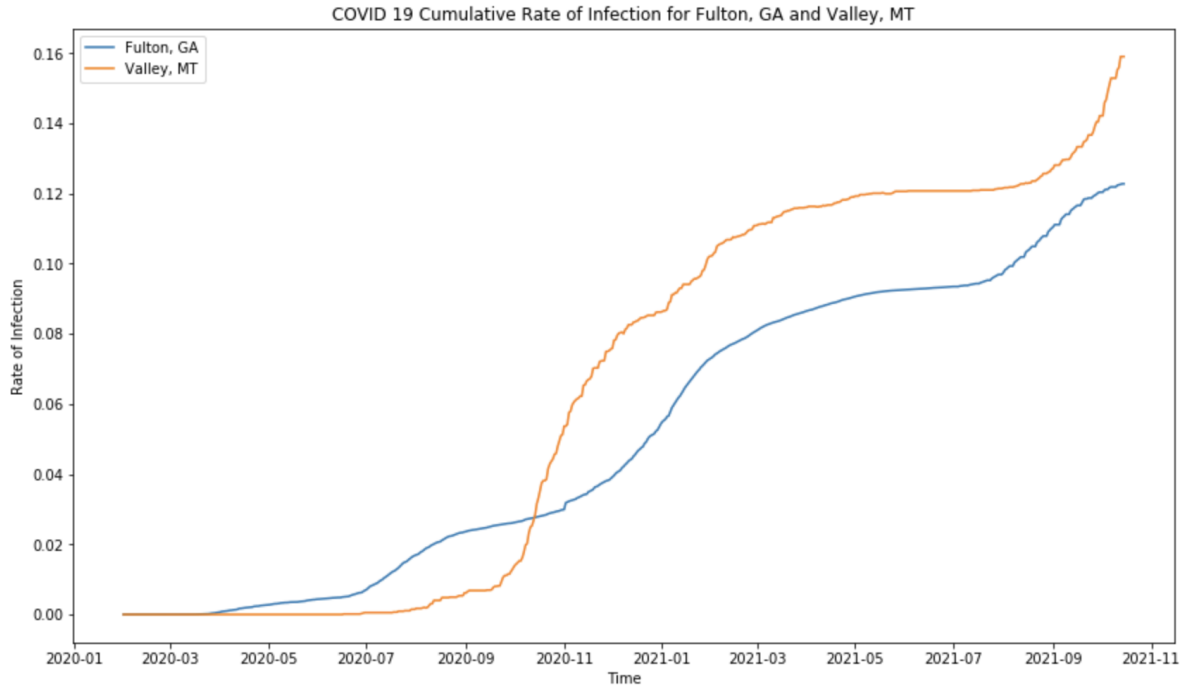
Visualization 1



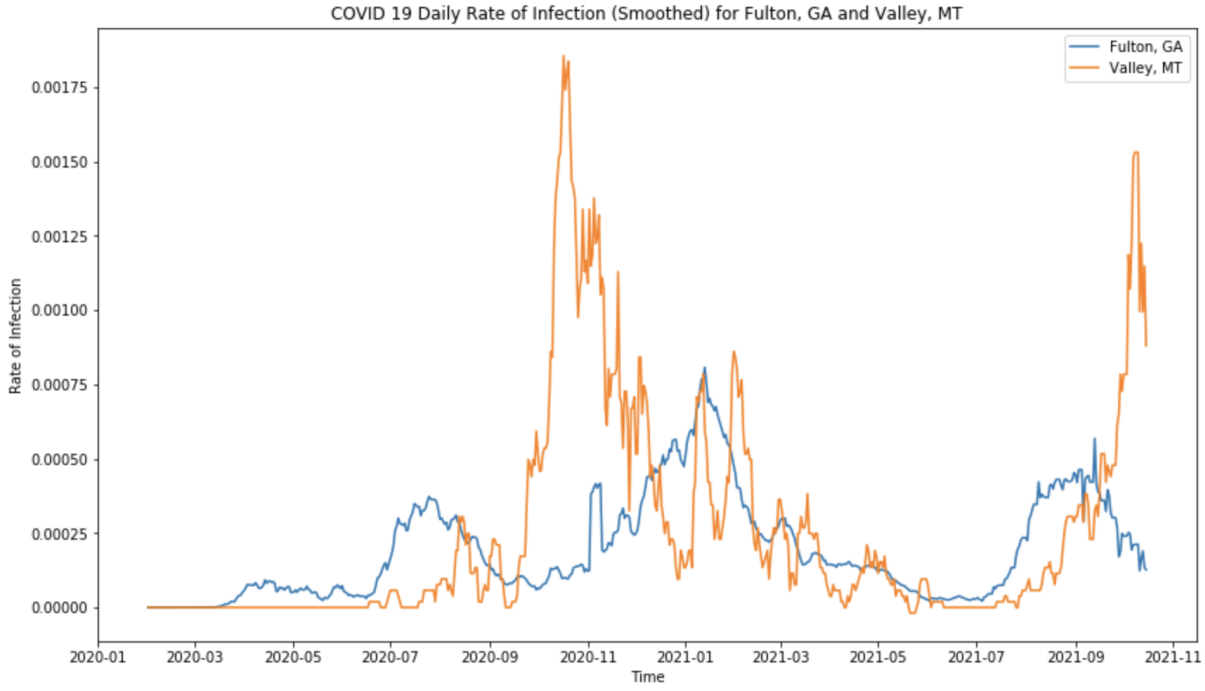
Visualization 2



Visualization 3



Visualization 4



There are four total visualizations shown above. The first two shows the cumulative and daily infection rate for Fulton County in Georgia, with or without mask mandate. Due to Fulton County always having mask mandate during the data availability, we also make a comparison with another county, Valley County in Montana, in the next two visualizations.

We use the following process to create all four visualizations. First, we retrieve The RAW_us_confirmed_cases.csv file from the Kaggle repository of John Hopkins University COVID-19 data, The CDC dataset of masking mandates by county, and The New York Times mask compliance survey data. We standardized the data to have the same date formats. Due to a delay between the time of infection and the time a case is confirmed, we delay the mask mandate dates by 6 days because average incubation time for COVID-19 is 5.6 days. Then we divide the confirmed cases by total population to estimate the rate of infection. For cumulative rate of infections (visualization 1 and 3), we simply make a scatterplot of infection rate vs. date. For daily infection rates (visualization 2 and 4), we calculate the daily incremental cases from the cumulative counts, smooth the data by taking the 7 day average to lower periodicity (due to more confirmed cases on Monday when everyone takes COVID test), and divide by total population. For visualization 1 and 2, we also use color to represent if a mask is required, or indicate if such data is unavailable.

In short, for viewing purposes of all four scatter plots, the horizontal axis represents time and the vertical axis represents infection rate, either cumulative or incremental and smoothed over 7 days, as indicated in the axis label. The color in the first two visualizations indicate mask mandate status, while the color in the last two visualizations indicate the county.

Note that I chose Valley for comparison because it has the lowest “ALWAYS” answer to the survey question in the New York Times data, as shown below.