

A4 – Common Analysis

Visualization Explanation

The visualization I produced in Step 2 has two subplots each showing a different look at the COVID-19 infection rate in Marion County, IN from February 1, 2020 – October 15, 2021. Both plots have a yellow backdrop which indicates when a public mask mandate was in effect for the county. They have a shared x-axis that represents date with a tick for each month but have different y-axes. The top plot represents the change in infection rate by day while the bottom plot shows the actual infection rate (smoothed using a 7-day rolling average). Infection rate was calculated as a percentage of active cases divided by the at-risk population. Daily change in infection rate was calculated by finding the difference in infection rate from one day to the next.

This figure shows that infection rate continued to rise at an increasing rate when the mask mandate was first implemented. Infection rate then peaked (aside from some fluctuations likely due to data reporting inconsistencies) before decreasing at a similar rate. From the data available, it is unclear how much of this can be attributed to public masking or to some seasonality from the infection tearing its way through the vulnerable population in the area. It was interesting to see that the infection rate decrease was steeper for a sustained period of time during the mask mandate period than anywhere else in the date range of concern.

This visualization aggregates data from a Kaggle [repository](#) of John Hopkins University COVID-19 confirmed cases, mask mandates by county as aggregated by the [CDC](#), and a snapshot of mask compliance data taken by the [NYTimes](#) in July 2020. These datasets were filtered down to show Marion County, IN and joined using the FIPS code (a unique US-based identifier for state & county level geography).

In addition to processing this data, some simplifying assumptions were made. It is assumed that all cases were properly identified by a confirmed positive COVID-19 test result (i.e. no false positives or false negatives and all individuals infected got tested) and that all cases were equal in duration. Individuals were assumed to have “active cases” for 3 days prior to and 10 days following a confirmed test – this aligns with CDC guidelines to get tested 3-5 days post exposure and remain in quarantine for 10 days post diagnosis. I also assumed all test results were reported on the same day; we know this is not the case because there are some days with negative confirmed cases which indicates some data reporting error. At-risk population was calculated by subtracting active cases from the total population (2020 Census Bureau data). This definition of at-risk population assumes that every patient recovers (i.e. total population remains constant) and that reinfection is possible immediately upon falling out of the active case pool. It also ignores any impact from vaccinations and assumes that mask compliance remains constant throughout.