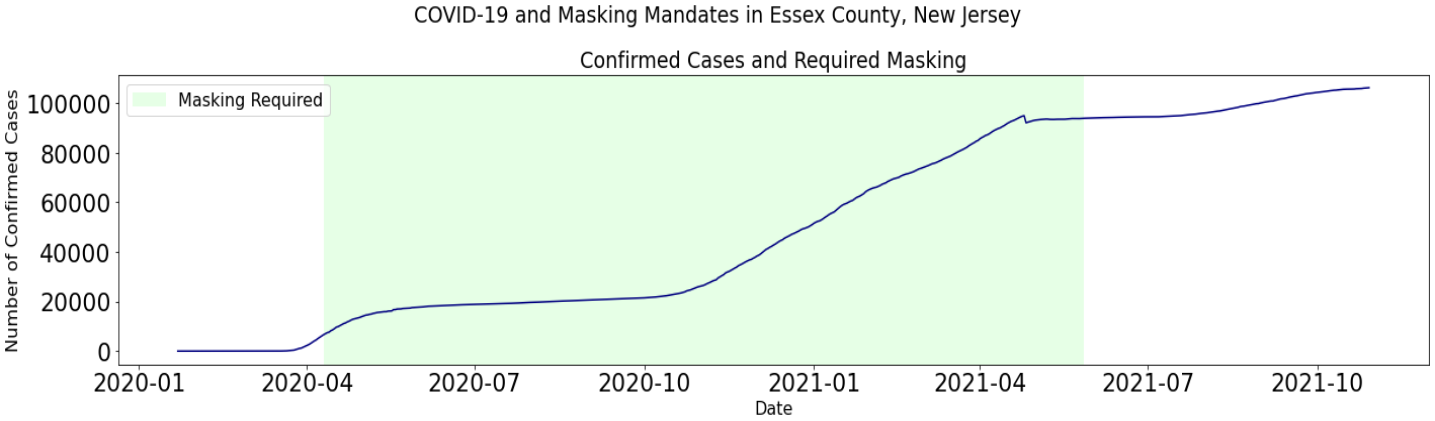
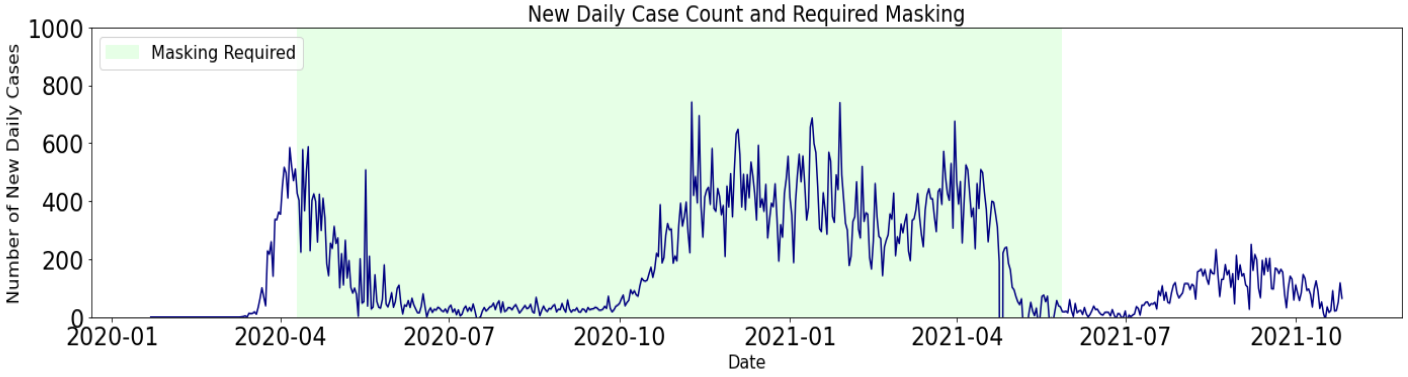
A4 Common Analysis: The Course Project

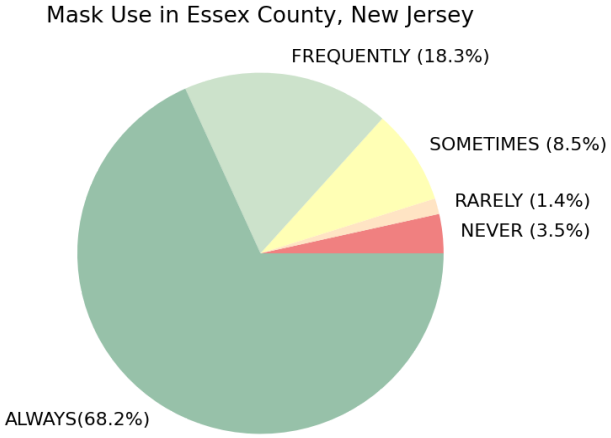
By Aaliyah Hänni

11/4/2021

Research Question: How did masking policies change the progression of confirmed COVID-19 cases from February 1, 2020 through October 15, 2021?







Reflection Part 1: Extended Figure Caption

In order to analyze the research question of how masking mandates impacted COVID-19, I decided to create multiple visualizations. The topmost visualization shows the cumulative count of all confirmed number of COVID-19 cases in Essex County, New Jersey. The y-axis displays the number of people that live in Essex County and have or had COVID-19, while the x-axis shows the date of the confirmed case, in month-year format, spanning from January, 2020 through October, 2021. The cumulative confirmed case counts for where gathered from the Kaggle repository of John Hopkins University COVID-19 raw United States confirmed cases dataset (source: https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university).

Highlighted in green shows the date range in which there was a masking mandate in place for Essex County. The data for masking mandates was sourced from the CDC dataset of masking mandates by county (source: https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i). The CDC states that county masking regulations are, "...defined to mean either (1) anywhere outside the home or (2) both in retail businesses and in restaurants/food establishments."

To better show the rate of the number of confirmed cases, the graph below the first one shows the number of new daily cases in Essex County, New Jersey. On the y-axis is the number of new daily cases (calculated as the change in the number of confirmed cases), and along the x-axis is the date in month-year format, spanning from January, 2020 through October, 2021. Again, the dates in which a masking mandate is in effect is highlighted in green. This graph is useful in visualizing the change of number of new cases from day-to-day.

The last visualization is a simple pie chart to show the masking utilization among residents in Essex County, New Jersey. The data from this survey was gathered from a large number of online surveys conducted by the survey firm Dynata where participants answered about how often they wear a mask in public when within six feet of another person, selecting from the following options: always, frequently, sometimes, rarely, and never. The survey includes responses from 250,000 global users in July 2020 (source: https://github.com/nytimes/covid-19-data/tree/master/mask-use).

Reflection Part 2: Reflection Statement

Originally, I was unsure of how to approach visualizing the rate of infection for COVID-19. I corresponded with Andrew Zhou, a fellow classmate, who suggested plotting the daily new case count to show the change in infection rate over time. This led me to calculate the number of new daily confirmed cases of COVID-19, by using the cumulative number of cases and then calculated the difference between each day, to get the difference. This yielded the total number of new cases for every day between January 2020 to October 2021.

When it came to figuring out the best way to display the masking mandate, I compared techniques with Emily Linebarger, Patrick Peng, and Apoorv Sharma. Patrick gave me the idea to display multiple visualizations, in order to capture the different angles in which our research question can be analyzed, while Emily helped me conclude that highlighting the area of the graph when the masking mandate was in effect was the best way to communicate this to users. To learn how to highlight my visualization for when masking mandate was in effect I referenced the article "Data Viz with Python and R" (source: https://datavizpyr.com/highlight-a-time-range-in-time-series-plot-in-python-with-matplotlib/)

From this assignment I learned that although the question of how masking impacts COVID-19 may sound simple, there are a multitude of variables to consider. An example of how complex this issue is, is that when visualizing the masking mandate with the rate of new cases, it is clear that there are also external factors at play. It appears that the masking mandate significantly impacts the rate of the COVID-19 transmission, until winter 2020 when there is a large spike in the number of new daily cases. There are many factors that could cause this increase in new cases, such as the introduction of the delta variant, more people being close together indoors due to the change in weather, or other factors that are difficult to guess or measure. Slightly before the release of the mask mandate in Essex County, New Jersey, around June 2021, the rate of cases appears to slow.

Another interesting thing that I gathered from the visualization was that approximately a month after the mask mandate is lifted, there appeared to be another slight rise in cases, which could be caused by the lag of the time it takes for people to get the news about the change in policy, the spread of the disease, and the time it takes to show symptoms and get tested. Given the timing, it is possible that the slight uptake in cases from July to October 2021 could be from the lifting of the masking mandate. I found the visualizations that I created to be useful when creating hypotheses and identifying correlations between COVID-19 cases and masking regulations.