Baltimore County: Covid, Crime, and Unemployment
Corina Geier

Introduction

I had three areas I set out to perform an analysis of. First, I looked into how masking policies changed the progression of confirmed Covid-19 cases from February 1, 2020 to October 15, 2021. This is an interesting question as it can help explain the benefits of masking for reduction of infection rate, even at a time when vaccines weren't readily available. Especially with the large resistance to mask mandates it is important to be able to explain the relationship with infection rate to support masking policies.

Next, I wanted to understand the relationship between Covid-19 infection rate and the unemployment rate in Baltimore County. My interest behind this analysis was to understand the effect Covid-19 had on the labor industry, as employment can have a big impact on poverty and the economy, as well as social factors such as social unrest or conflict (1). I hoped to learn how well Baltimore County managed the employment rate while dealing with a pandemic. I also looked into the insured unemployment rate in the state of Maryland to expand upon my analysis and support my results. This is an interesting case study as another pandemic occurring is fairly likely, according to a study done by the Proceedings of the National Academy of Sciences (2). Understanding how the pandemic effected people's jobs is very important to ensuring people have better job security in the future. In this way my question is very human centered as it explores how Covid-19 shaped people's employments and lives.

The second area I investigated were crime reports in Baltimore County to see if there was any relationship between Covid-19 infection rate and the number of crime reports in a given month. My initial hypothesis was that crime reports and unemployment rate were positively correlated due to a study that showed positive effects of unemployment on property crime rates (3). This question is human centered as a correlation between the pandemic and crime would have very real consequences for people and if the problem is properly understood, measures could be taken to prevent similar consequences from happening in the future.

Background/Related Work

There have been many studies regarding the association between masking mandates and Covid-19 cases. One in particular that I found interesting was published on the CDC that examined mask mandates and Covid-19 cases in the United States on a county-level (4). They found that mask mandates were associated with a significant decrease in daily Covid-19 cases after 20 days of the mask mandate implementation. This coincided with what I expected from my analysis – Covid-19 cases would fall given a masking policy in Baltimore County.

While my second research question was "Is the crime rate correlated with the number of Covid-19 cases in Balitmore County?" it was heavily tied into a hypothesis that crime reports and unemployment rate were positively correlated. In support of this, the journal of law and economics published a paper by researchers from Berkley that looked at crime reports from across the whole United States (3). They did not consider Covid-19 as a potential confounding factor but did control for other factors such as state-level demographic and economic factors. The researchers didn't find anything significant for the relationship between unemployment and violent crime but did find a significant positive effect of unemployment on property crime reports. The results from this study led me to my initial hypothesis that there would be a positive correlation between crime reports and unemployment rate in Baltimore County.

My last research question was "how was the economy in Baltimore County influenced by the number of Covid-19 cases in the area?". Looking into unemployment rates during the pandemic, it is quite clear that unemployment drastically went up at the start of the pandemic. The Center on Budget and Policy Priorities noted that the US unemployment rate jumped up to levels that haven't been seen for 90 years in April of 2020 (5). It is also worth mentioning that this rise in unemployment rate has had a much larger effect on people working in industries that play a low average wage, as well as on specific demographics such as people of color. Due to the effects of government shutdowns on a state and national level, I expected to see a positive correlation between Covid-19 infection rate and unemployment rate. One thing to note is that I did not split up unemployment rate based on demographics to examine this factor as well, as my data only had information on county and state level unemployment rates.

Methodology

For my analysis I used Pearson correlation, cross correlation, and graphical representation. Pearson correlation measures the strength of a linear relationship between two variables. The correlation coefficient can fall between -1 and 1 with a number closer to 1 or -1 representing a stronger positive or negative correlation. I used this for my initial measurement of the relationship between infection rate, crime reports, and unemployment rate to see if there was a statistically significant relationship between any of them, using a p-value of less than 0.05 as a measure of statistical significance. To compute this, I used the pearsonr() function in python from the scipy.stats package. This was also a helpful measure to have to see if applying a cross correlation resulted in a change in correlation coefficient or statistical significance.

Next, I calculated multiple cross correlations for infection rate and insured unemployment rate in Maryland. Cross correlation measures the linear relationship between time series where one series is displaced in relation to the other. I chose this methodology to compute a correlation in case there was a delay in effect between infection rate and insured unemployment rate. Using cross correlation I found an optimal lag time, which produces the strongest correlation, and then looked at the correlation with this optimal lag time to examine both the correlation and the statistical significance of this correlation.

I also employed the use of graphical representations of my data to convey information. For instance, I created a line graph of the infection rate in Baltimore County with different colors corresponding to different masking mandate policies to answer the question of how masking policies changed Covid-19 infection rate. I also graphed all the time series of my crime reports, Baltimore County unemployment rate, and Maryland insured unemployment rate against Covid-19 infection rate in Baltimore County. One ethical consideration I had when creating these was to ensure my graphs had clear labels so at to not mislead the viewer. A bias in the crime data is that only reported crimes are recorded, and as certain populations might be less likely to report incidents to the police, it is important to stress that this analysis only deals with reported crimes, and not all crime incidents. Additionally, the Maryland unemployment data only recorded those on unemployment benefits, so I made sure to differentiate the wording between this and the calculated "true" unemployment rate for Baltimore County.

Findings

My first research question was how masking policies impacted the course of the pandemic. I created the below chart which showed how infection rate per 1000 people-days changed with masking policies.

Baltimore County, Maryland: Covid-19 Infection Rate per 1000 person-days
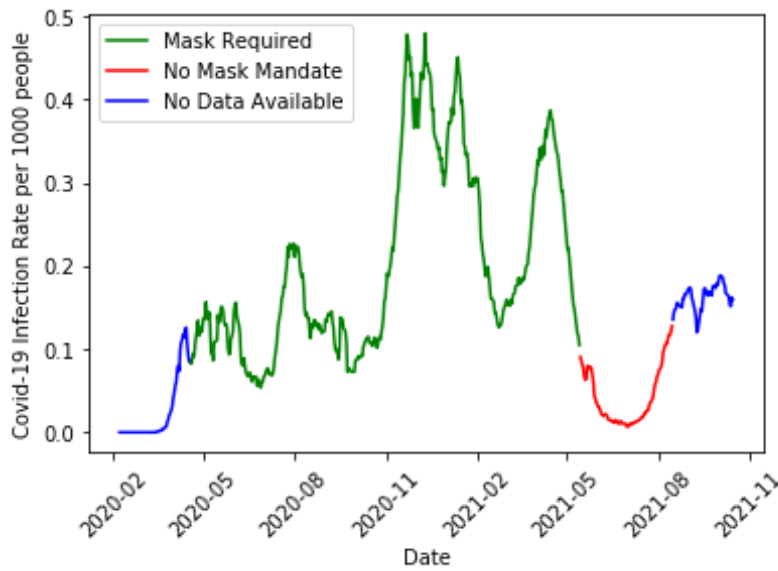
It is worth noting that mask mandate data was only available from April 2020 through August 2021. The reader might assume that prior to April of 2020 there was no masking mandate in place, but after August 2021 it is unclear. The masking mandate data was found on the CDC Website and was filtered to only 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 outside the scope of this analysis. The information on Covid-19 cases was found on the Kaggle repository of John Hopkins University COVID-19 data. It was filtered for dates between 2/1/2020 and 10/15/2021 for Baltimore County in Maryland. The infection rate per 1000 people-days was calculated as daily cases with a rolling average of 7 days, divided by the population at risk all multiplied by 1000. Since this chart is on a daily level, 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.
As seen in the color-coded lines, as infection rate started to rise in April, a masking policy was put into effect and remained in effect until May when cases started to lower. A couple months after this, infection rate again began to rise. In this way the graph conveys the overall message to the user of how infection rate per 1000 people-days varied and its relationship with masking policies.

My next question was that of the correlation between crime reports and Covid-19 infection rate. I used data from the Baltimore County Government for information on monthly crime reports which comes in two separate files, one for 2020 data and one for 2021 data. The data needed to

be separated into the two years, as in 2020 only the most serious offense in an incident was reported, whereas in 2021, all offenses in an incident were reported. Due to this discrepancy in reporting, crime report counts were much higher in 2021 than in 2020 and no statistical comparisons could be made between the two years. The crime counts were separated by crime type, but for my analysis I aggregated this into a total crime count to answer my general question of how infection rate and crime reports were related. I aggregated the crime types of Arson, Assault, Burglary, Homicide, Motor Vehicle Theft, Sexual Assault, and Robbery for both years to get this total crime report number. I used this dataset as it is publicly available for use and since it comes from the Baltimore County government, I can trust that they have recorded all reported crimes from their own internal databases, as police departments are government agencies.

A possible ethical consideration for using this dataset is that as it is coming from the Baltimore County government and police department, it only records crimes that have been reported. This could lead to some bias in the results, as people might be less likely to report something like domestic violence, and communities that have a greater mistrust of the police might also be less likely to report a crime. Therefore, it is important to note that this analysis only accounts for reported crimes and not all crimes.
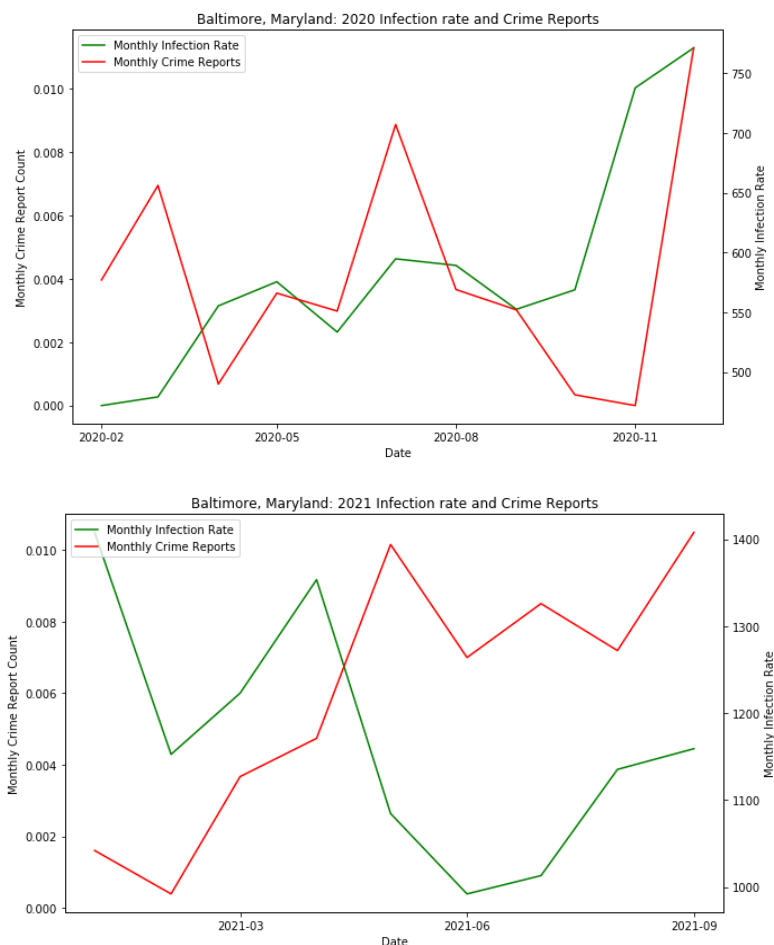
I created the above graphs and calculated a Pearson correlation and found a correlation coefficient of .21 with a p-value of .52 for the year 2020 and a correlation coefficient of -.57 with a p-value of .10 for the year 2021. Thus, I found nothing statistically significant for either year. Due to the lack of data points, I was also unable to perform a meaningful cross correlation with this data.

After looking into infection rate and crime reports, I wanted to look at crime reports and unemployment rate to check my earlier hypothesis that crime and unemployment might be positively correlated.
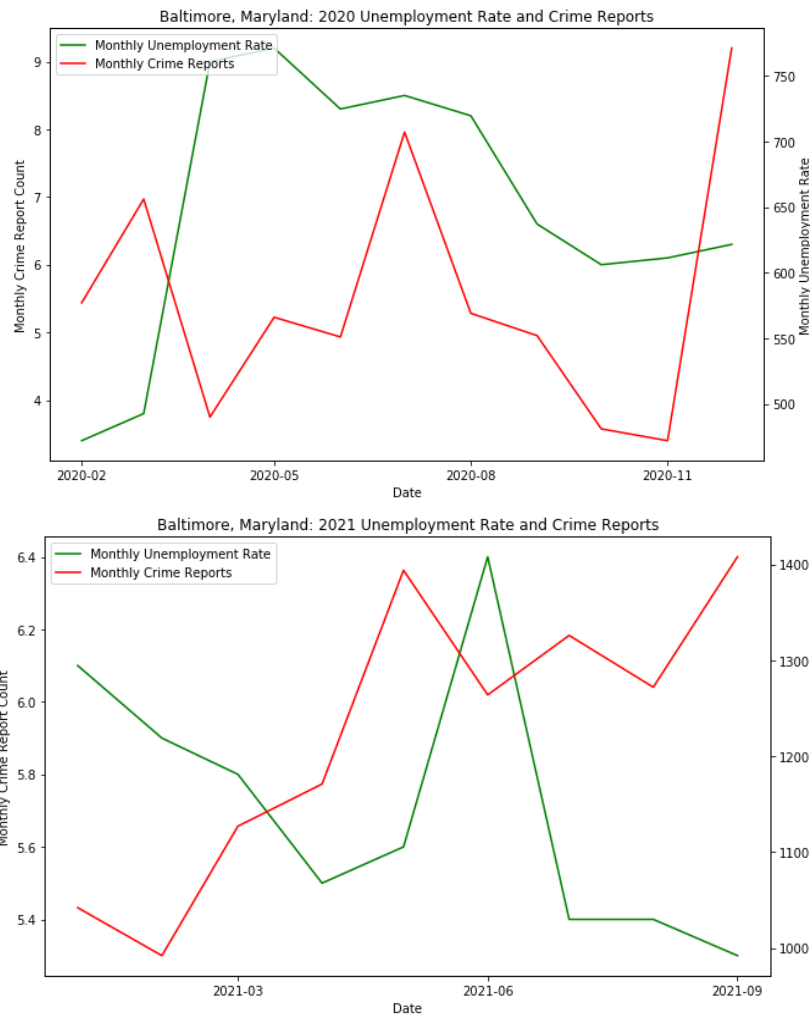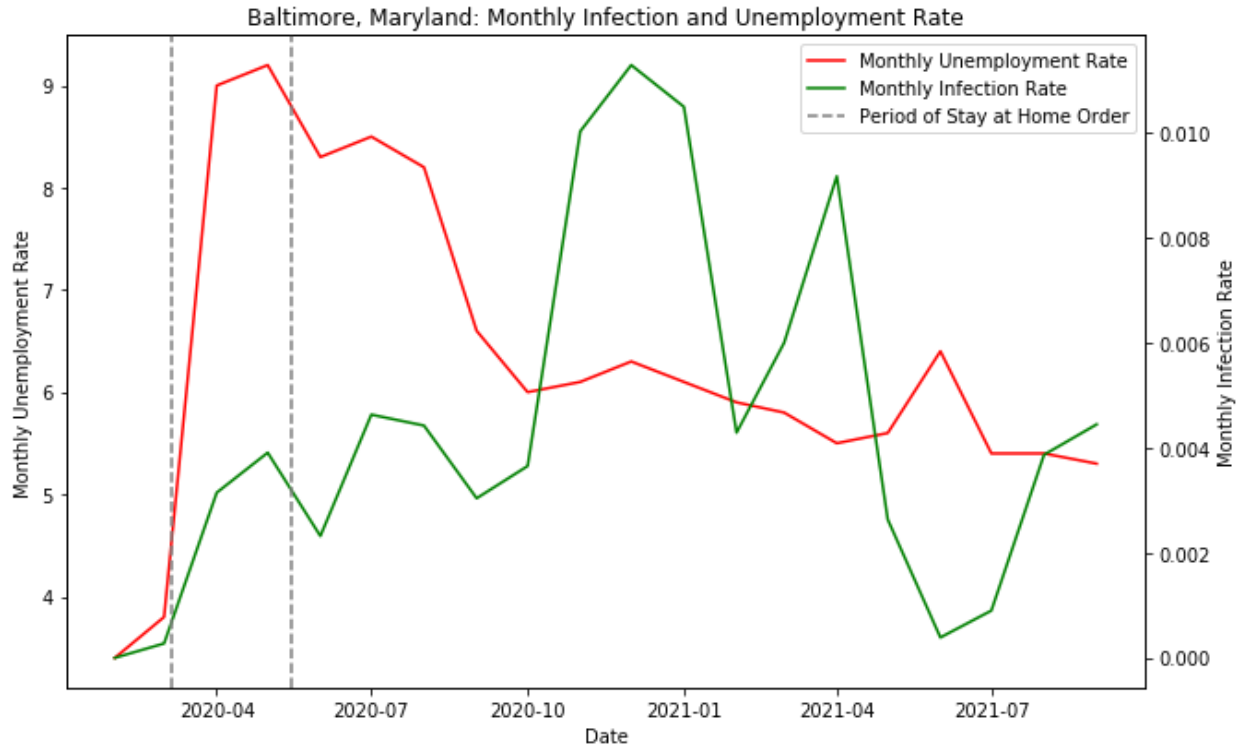

I again found no statistically significant results from this analysis. The correlation was -.12 with a p-value of .72 for 2020 and -.51 with a p-value of .16 for 2021. Therefore, I was unable to reject the null hypotheses that there is no correlation between unemployment rate and crime reports as well as that there was no correlation between crime reports and Covid-19 infection rate.

Finally, I wanted to look at unemployment rate and Covid-19 infection rate. I gathered my data from two sources, the first being the Bureau of Labor statistics (BLS) for unemployment data in Baltimore County. The unemployment dataset comes in a csv format and contains monthly information on the count of the labor force, employment, unemployment, and unemployment rate from January 2019 to September 2021. The data is collected from a survey run on a sample size of the population which is then used to extrapolate county level unemployment rates.

The second place I sourced data from was the United States Department of Labor to get information on unemployment rate in the state of Maryland on a weekly level. This data also came in a csv format. This does not represent the true unemployment rate but rather the insured unemployment rate, or those on unemployment benefits. This is one limitation as some unemployed people might not qualify for unemployment benefits or may never file a claim. Both unemployment rate data sources are in the public domain. I used these sources as the Bureau of Labor Statistics is the "principal fact-finding agency for the U.S. government in the broad field of labor economics and statistics" according to the Wikipedia page (6), and the United States Department of Labor is a credible government department responsible for gathering data on unemployment among other labor related topics. However, some possible ethical considerations when using these datasets is that they are coming from a government agency, which means it might not be accounting for people not documented by the government. Additionally, the BLS data may have a selection bias as the data is collected using a computerized questionnaire.

I first created a graph of the unemployment rate vs Covid-19 infection rate in Baltimore County.

I noticed that there was again no statistically significant correlation with a p-value of .62 and a correlation coefficient of .12, however I did notice that the initial rise in unemployment rate came in April, one month after Maryland states stay-at-home order (6). This signified to me that government shutdowns could be a better indicator of unemployment rate than Covid-19 infection rate itself. This defies an earlier assumption I had that higher Covid-19 infection rate would lead to government shutdowns.

Again, as this data was on a monthly basis, I wasn't able to apply a meaningful cross correlation, so I investigated the weekly data I had from the state of Maryland.

Before applying a cross correlation, I had a correlation coefficient of -.14 with a p-value of .28, signifying no interesting correlation. The weekly insured unemployed rate looked very similar to the monthly rate obtained for Baltimore County with a huge rise in rate in April of 2020 after the state stay-at-home order.
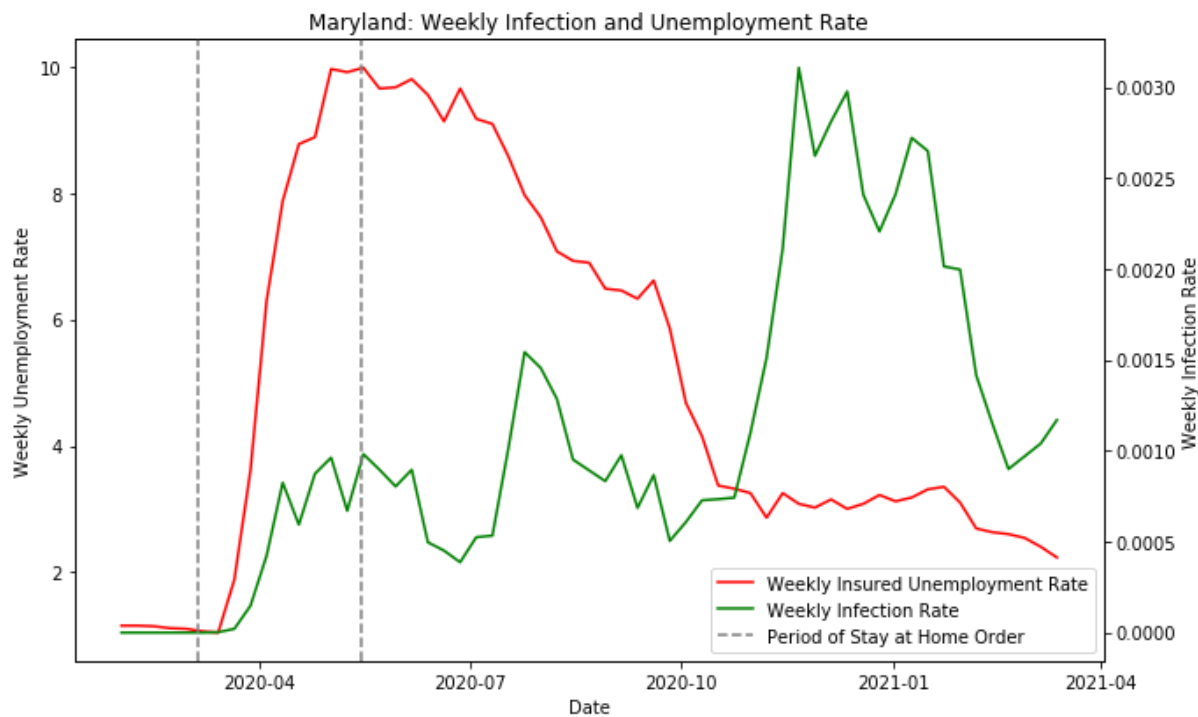
Applying a cross correlation, I found an optimal lag time of -30 weeks or approximately 7 months. This was interesting as it signified that unemployment rate actually led the interaction with Covid-19 infection rate as opposed to the other way around.

Maryland: Weekly Infection and Unemployment Rate Lagged
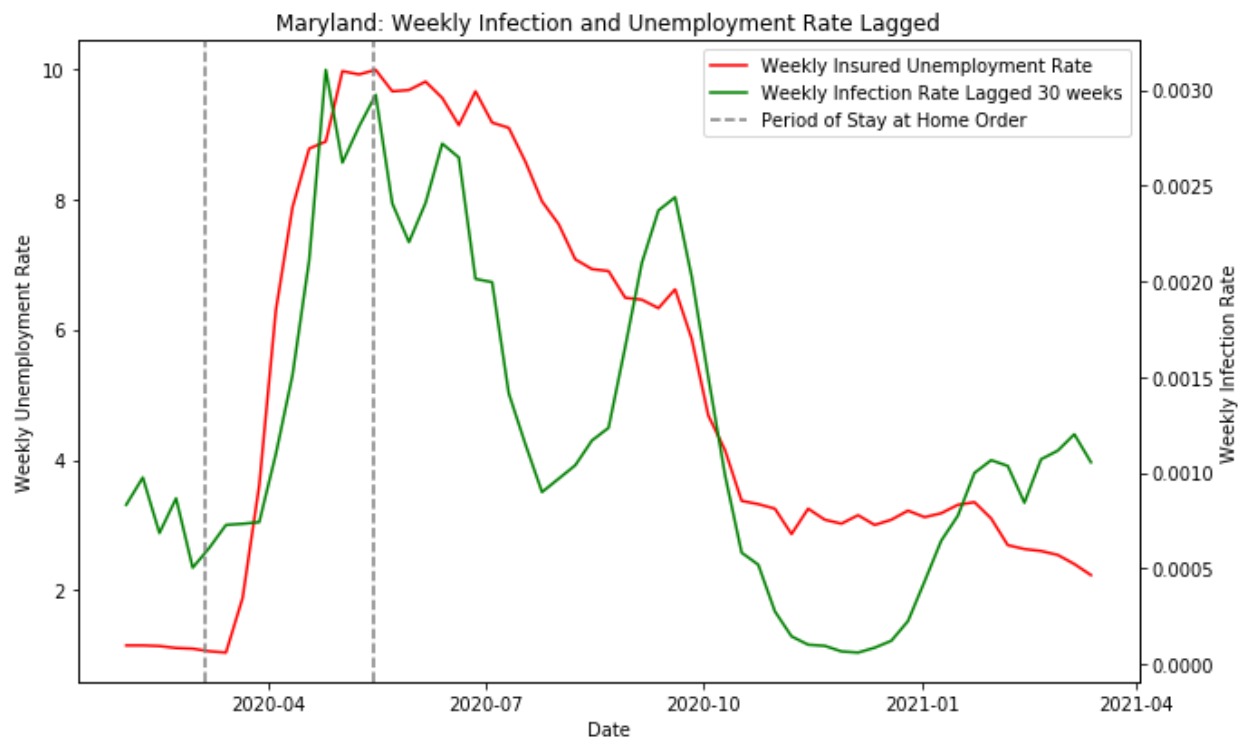
Applying this cross correlation did yield a statistically significant strong positive correlation with a p-value of <.001 and a correlation coefficient of .79. However, from a qualitative point of view this wasn't a meaningful result. Again, it seemed that the stay-at-home order was more indicative of unemployment rate rises than infection rate.

Discussion

I ended up finding no statistically significant correlation between Covid-19 infection rate, unemployment rate and crime reports in Baltimore County and failed to reject the null hypotheses. This is interesting as other studies have show that Covid-19 and unemployment rate are associated, and that unemployment rate and crime reports are associated. Additionally, while my cross correlation did yield a statistically significant strong positive correlation, it wasn't meaningful in a qualitative sense, and it seemed that the state stay-at-home order had the biggest impact on unemployment rate. As a human-centered data scientist it's important to not only look at the statistics but also to consider the problem from a qualitative perspective and think about how the results will affect people. Due to this, I believe more granular data is needed to fully understand the relationship between the three variables as my crime report data came in at a monthly level and had changes in reporting from one year to the next.

My time series graph on masking mandates and Covid-19 infection rate per 1000 people-days did show that masking seemed to lead to a reduction in Covid-19 infection rate after implementation of a masking policy. This could indicate that taking measures such as masking could help in avoiding future government lockdowns to reduce spikes in unemployment rate.

Future research could build on this study by first considering government shutdowns and gathering more information on this to examine it as a potential confounding factor for rises in unemployment rates. Additionally, it would be interesting to explore hospitalization rates, to see if this had a larger effect on government shutdowns than infection rate alone. Future work concerning the analysis around crime reports could be to investigate more granular data on crime reports as well as to continue the analysis in future years as the reporting has become much more accurate as of 2021. Additionally, outside factors such as 2020 protests, stay-at-home orders and seasonality could be explored as confounding factors for rises in crime reports.

Limitations

The most significant limitation I faced in my analysis was the lack of publicly available granular data for both crime reports and unemployment rate. While I had daily data for both Covid-19 cases and whether a masking mandate was in effect, I could only find monthly data for crime reports and unemployment rate in Baltimore County. Due to this lack of data, I was unable to find significant relationships and further analysis would be needed to draw more accurate conclusions. Even when using the weekly data from the state of Maryland, it only included unemployment rate of those on unemployment benefits, a limitation as it didn't include all those unemployed.

I was also unable to make comparisons across 2020 and 2021 for crime reports due to changes in reporting. This made it hard to measure the effect of Covid-19 infection rate on crime report count as the pandemic went on, as I could only look into one year at a time. Additionally, there was no credible information on crime report counts from before 2020 that I could access to attempt to understand the baseline for crime report counts before the pandemic started.

Another limitation is the assumption I made in each analysis of there being no confounding factors. For instance, when I investigated mask mandate vs Covid-19 infection rate I assumed mask mandates were the only factors that could be affecting infection rate, while in reality many other factors are at play including vaccinations and social distancing. While there were no other factors considered, I believe some meaningful conclusions can still be drawn just from looking at these two factors in my analysis.

Conclusion

I initially wanted to understand the relationship between masking mandates and Covid-19 infection rate, as well as between crime reports, unemployment rate and Covid-19 infection rate. My first research question was "How did masking policies change the progression of confirmed Covid-19 cases from February 1, 2020 to October 15, 2021 in Baltimore County?". By creating a visualization, I was able to see that lower infection rate seemed to follow implementation of a masking mandate, with the rate starting to rise again after the mandate was lifted.

My second research question was if Covid-19 infection rate and unemployment rate in Baltimore County were correlated with each other. My hypothesis was that they would be positively correlated, while the null hypothesis stated no correlation between the two variables. I ended up failing to reject this null hypothesis as I found no statistically significant correlation between these two variables. Even looking at a more granular level of weekly insured unemployment rate in the state of Maryland yielded no significant correlation. Additionally, the cross correlation for this data yielded no meaningful relationship either, as the optimal lag time was 7 months. It seemed that the state stay-at-home order might have been a better indicator for the rise in unemployment rate, especially as unemployment rate began declining as the initial order ended, even with the fluctuations in infection rate.

Finally, I asked if Covid-19 infection rate and crime report count in Baltimore County were correlated. My hypothesis was that they would be positively correlated, with a null hypothesis that there was no correlation. I also ended up failing to reject this null hypothesis with no statistically significant correlation in the year 2020 or 2021.

This analysis should inform an understanding of human centered data science as one must consider the meaning of results outside of what is statistically significant. Other studies have shown relationships between unemployment and crime, Covid-19 and unemployment, and masking mandates and decrease in Covid-19 cases and the reader should always ensure they are considering multiple credible sources. This study has been made reproducible if the reader wished to reproduce it on their own and see the limitations first-hand. Questions such as those asked in this study are important to understand so that the real effects on people can be understood and measures can be taken to improve upon things such as unemployment or crime rate. The first step in solving these problems is to ask and understand these human centered questions.

References

1. Boundless. "Boundless Economics." *Lumen*, Boundless Economics, https://courses.lumenlearning.com/boundless-economics/chapter/introduction-to-unemployment/.
2. Marani, Marco, et al. "Intensity and Frequency of Extreme Novel Epidemics." *PNAS*, National Academy of Sciences, 31 Aug. 2021, https://www.pnas.org/content/118/35/e2105482118.
3. Raphael, Steven, and Rudolf Winter-Ebmer. *Identifying the Effect of Unemployment on Crime* | the ...* University of California Berkeley, https://www.journals.uchicago.edu/doi/full/10.1086/320275.
4. Guy, Gery. "Association of State-Issued Mask Mandates and Allowing on-Premises Restaurant Dining with County-Level Covid-19 Case and Death Growth Rates - United States, March 1–December 31, 2020." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 27 May 2021, https://www.cdc.gov/mmwr/volumes/70/wr/mm7010e3.htm.
5. "Tracking the COVID-19 Economy's Effects on Food, Housing, and Employment Hardships." *Center on Budget and Policy Priorities*, https://www.cbpp.org/research/poverty-and-inequality/tracking-the-covid-19-economys-effects-on-food-housing-and.
6. "The Office of Governor Larry Hogan." *Governor of Maryland*, https://governor.maryland.gov/coronavirus/.

Data Sources

Baltimore County Government 2020 monthly crime reports

Baltimore County Government 2021 monthly crime reports

Bureau of Labor statistics for unemployment data

CDC data on masking mandates by County

Kaggle repository of John Hopkins University COVID-19 Data

Maryland State Weekly Insured Unemployment