Eli Corpron
Data 512
11/7/22

Project Part 2: Extension Plan

Motivation/problem statement:

Covid has had a huge and varied impact on our society. It has caused societal strains from the huge number of people who got sick from and died, from the temporary contraction of our economy and the current explosive activity that's causing inflation, and the numerous psychological impacts of forced separation from quarantine and masking policies. I am also continually hearing about how there is currently a huge increase in crime.

To me there seems to be sort of a pulling interaction caused by the pandemic. On one hand people were staying at home a lot more, not going out, and the government covid relief meant for a while many people didn't have to immediately worry about losing where they live or running out of money. Pulling in the other direction we had extended periods of isolation from quarantine rules. Also, while immediate worries about money may have been paused for many, they didn't go away. On top of that, as we moved through the pandemic we had increasing inflation to worry about. I am curious if these impacts from the pandemic caused an increase or decrease in the amount of crime people caused. For this question I will focus on El Paso County in Colorado.

Research questions and/or hypotheses:

What I am looking to discover is if the stresses of the pandemic lead to an increase in crime in El Paso county in Colorado. The data I have access to is the amount of people who died from Covid in El Paso, if masking mandates were in effect for a specific date, and the number of crimes committed in El Paso county. The number

of deaths and masking mandates will be a stand in for how bad Covid was at the time. My question then is did the death rate and masking policies have an impact on the number of crimes committed in El Paso, Colorado? My hypothesis is that they did; that there is significant evidence that the number of crimes increased when death rates and masking were in effect.

The conclusion will be relatively specific to El Paso county. If I had time I would like to extend my research by looking at other counties and how their crime rates reacted to similar stressors and then compare them to El Paso. Unfortunately I do not have the time or the necessary for that extension.

Data to be used:

I will use three data sets to test this hypothesis. The first two are from Part 1 of the project: the masking mandate data from the CDC and the Covid death data from John Hopkins University. The crime data I will be using is from Colorado Crime Statistics (https://coloradocrimestats.state.co.us/tops).

Colorado Crime Statistics is funded by the Colorado Automobile Theft Prevention Authority, which is a part of the Colorado State Patrol. Since the source of the data is the Colorado state government, I trust that it is accurate data. The website gives access to the raw data, along with some filtering functionality. I can get crime data for specifically El Paso county, broken down by year, month, and crime type from 2020-2021, which matches the date range for the Covid and masking mandate data. The screen shot of the data shows the crimes per 1000 people per month in 2020, with crimes broken down by overall type.

Incident Date	2020											
Incident Month	January	February	March	April	May	June	July	August	September	October	November	December
Offense Type												
All Offense Types	5.56	4.70	4.64	4.72	4.90	5.01	5.28	5.64	5.34	5.09	5.18	6.14
Crimes Against Person	1.03	0.90	0.86	0.83	1.03	1.03	1.03	1.10	1.02	0.87	0.81	0.85
Crimes Against Property	3.95	3.33	3.39	3.37	3.36	3.58	3.72	4.03	3.78	3.71	3.94	4.96
Crimes Against Society	0.57	0.48	0.39	0.52	0.51	0.40	0.53	0.51	0.55	0.51	0.43	0.33

Unknowns and dependencies:

Since all of my data is from and about El Paso county I feel pretty confident about conclusions made about El Paso. It will be difficult to justify extending the results to other places in America or even just for general statements about the impact of Covid on crime.

I also have some general worries about how the data is collected and how that can impact the results. From what I can tell the crime statistics are collected when a crime is charged or reported to the police. However I am unsure of what happens when incidents are investigated. If later on a crime is decided to not have happened or is given a different classification I am unsure how that impacts the crime count. It is also possible for a single crime to be given multiple offense types, e.g. assaulting someone when committing a burglary. These could lead to an inflation in crime counts, though without specific knowledge about the collection methodology or the rates at which incidents are assigned multiple offense types it is difficult to tell what impacts these potentially inflated numbers have on the results.

Methodology:

The hypothesis I am looking to explore is if there is significant evidence that the number of crimes increased when covid death rates increased and when masking mandates were in effect. Masking mandates are a categorical variable, covid death rates are a discrete variable along with crime rates. These variable types lead me to think of linear regression, with a null hypothesis that masking mandates and covid death rates have no impact on the crime rate in El Paso county.

Linear regression has a number of assumptions that need to be satisfied in order to be valid. The first is that there is in fact a linear relationship between the crime rate and masking mandates and covid death rates. The second is that the variables are independent. The third is that there is constant variance in the residuals. The last is that the residuals are normally distributed. All of these assumptions will require at least some exploration of the data before I can say if they are satisfied, but the assumption of independence between masking mandates and covid death is at best dubious. From part 1 it looked like masking mandates had little effect on the Covid death rate, but that is certainly an assumption I will need to monitor and could have an impact on the reliability of the results.

I have already gathered the data. The masking mandate and Covid death rate data can easily be used from Part 1, and I have access to the El Paso county crime rate already as well. I have two linear regression models I want to compare. The first uses the crime rate per 1000 people and the death rate as variables, the other uses the number of crimes and the cumulative death amount. The reason for this is that I don't think I have a Covid death statistic that is equivalent to a population normalized crime rate. The closest would be the death rate, but that isn't quite the same as a per capita Covid death rate.

My plan for presentation is to present the model against the data. I would essentially give a visual confirmation on if the model makes sense. I would also present some of the alternative models I tried (if I had tried multiple models) along with their accuracy and how they performed on the model assumptions. Lastly I would present on possible meanings that could be extracted from the model. Another way to present the

data would be through a time series. Since my intended model is going to be a linear regression model it won't be the only way I present the data, but it could be helpful in understanding both the data and the results of the model.

Timeline to completion:

I have already collected the data, but I still have a number of milestones to complete. The first milestone is to complete all necessary data cleaning for the crime data. The second milestone is to test the assumptions for a linear regression model. The third is to compare the model using per capita crime rates against the number of crimes committed. The third milestone would be to attempt to improve, tweak, or find other ways to create a more accurate model. The fourth milestone would then be to visualize the results, and the fifth and last milestone to write up the report.

We are presenting on December 8th, with slides due December 5th. That leaves roughly 4 weeks to complete the first four milestones. My goal is to have a milestone finished each week, but ideally I would have at least a rough start to the visualization the week before December 5th, or in three weeks. That would give about a full week to create my slides, and about another week to write up my report.