Ariel Shurygin

Data 512 Project Write up

Introduction

Understanding the sources of crime is a body of work spanning many fields of study, economics, psychology, and sociology to name a few. Looking at crime from any of these perspectives always has one similarity, crime is a cause and effect relationship. Certain actions taken by the federal reserve, or an individual, or society at large influences the crime rates in neighborhoods across the USA (figure 1). During the COVID-19 epidemic the United states saw a surge in crime rates not seen in many years. This study makes some assumptions about pivotal moments during the pandemic and analyzes their impacts on certain kinds of crime in Denver county Colorado.

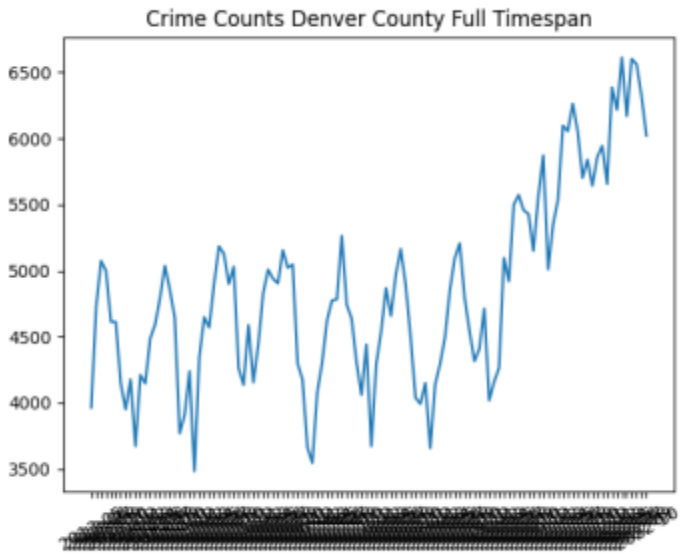

Figure 1: Crime Counts Denver County 2013-2022

Background/Human Centered Problem

Many conversations have been had in popular media about crime during the pandemic, mostly focusing on fear mongering around prisoner releases (1) or lawlessness in the streets (2). Many media outlets had ulterior motives to promote some sort of political action in one way or the other as is often the case in American politics. However there is a depth of research around the push and pull factors around crime (3). In particular many studies point out the positive correlation between unemployment rates and crime. The clash between popular reporting about Covid induced crime and existing studies was the primary motivation behind this project and one of the main human centered forces. Crime, and the ways we talk about crime is a uniquely human issue. The prevalence of crime depends heavily on both the ways we talk

about it, the ways we treat it, and the data we use to measure it. For example the types of crimes that are reported versus not can impact the amount of money that goes towards treating the issue, and thus can cause it to grow or shrink. The self referential nature of crime makes it very difficult to measure from a purely data driven perspective. There is value in looking at crime as a cause and effect relationship, to see how it responds to events in the real world rather than just predicting it with some machine learning algorithm.

Hypothesis

Lets take three major moments from the pandemic: the introduction of the mask mandate, lockdowns, and the first round of stimulus checks and see how they impacted crime rates in Denver County Colorado. The three hypothesis are as follows:

1) The introduction of mask mandates increased public disorder crime by 15%, as those who actively broke masking mandates put strain on police forces
2) Lock down increased domestic abuse by 20% as victims no longer could avoid their abusive partners and were more likely to be stuck in the home with them.
3) Stimulus checks reduced robbery and financial crimes by up to 10%, as people could finally put food on their tables, even for a week or two

Denver County Covid-19

Denver county is the county surrounding Denver, the largest city in Colorado, home to around 700,000 people in 2020 and with a land area of 154 miles squared. Over the course of the pandemic the county followed the same trend as most other major metropolitan areas in the United States, enacting masking policies in response to large rises in COVID cases and removing them once those cases reduced. Here is a chart of the weekly case count, the total case count, and whether or not there was a masking order in the county.

Figure 2: Weekly/Total Covid cases (red/blue) and masking mandates (orange)

Methodology

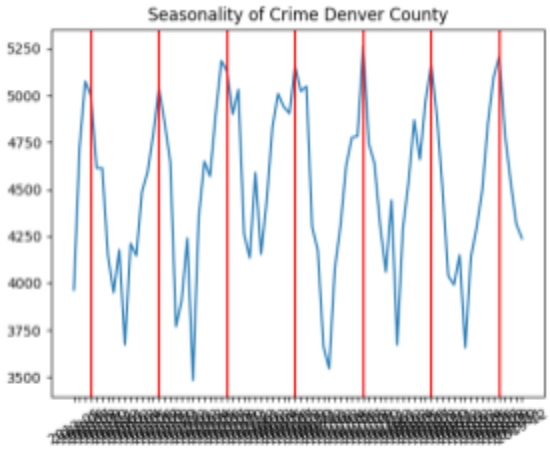

Figure 3: Crime rates Denver County with each August highlighted

Before deciding how to measure changes in crime rates I looked at crime before the pandemic, specifically 2013 - 2019. There is a fairly predictable trend with peaks in crime rates in August and troughs in January. Using the seasonality of crime I could create a distribution of crime for a particular month. For example the month of August has higher than average crime so it would be modeled by a slightly different distribution than January crime. Assuming negligible shifts in Denver County politics (which is a big assumption but out of scope of this project) we can use a normal distribution from year to year with mean and standard deviation to model the likelihood of a certain amount of crime in that month. With the generated crime distribution cumulative density functions can test whether or not the pandemic had a statistically significant impact on crime.

Making accurate predictions about crime is a difficult human-centered issue as it attempts to turn every victim and perpetrator into a number. Without ethnographic research as to why an individual commits crimes the personal stories are lost in a fog of data. To compensate for these ethical issues I attempted to choose statistical analysis methods that made as few assumptions as possible. In short I made three major assumptions. Firstly, crime for a particular month follows a normal distribution, meaning the crime rates of August last year do not impact this August and they have an average value they tend to stay around. Secondly, I assumed the crimes themselves were independent, meaning if domestic abuse rose this was not necessarily correlated with a rise in robbery crime. I believe for the crimes studied in each of the three hypotheses the independence of crime rates does apply, while this assumption may not hold for all crime combinations. Lastly, I assumed that beyond covid related policy, there was no major shift in policing tactics in Denver between 2019 and 2020, aka funding set to expire or some preplanned shifts.

Findings

After digging into the data I found that my findings did not line up with research in the field, nor did it follow my own expectations. I was unable to disprove the null hypothesis for my first two predictions, and actually showed the opposite trend in my third hypothesis. Let it be known before figures are shown that a lack of a full picture in this analysis may have contributed to a false narrative being painted. Statistics in combination with incomplete information can give incorrect or misleading results.

Figure 4: Statistics without the full story may be misleading

Hypothesis 1: The introduction of mask mandates increased public disorder crime by 15%, as those who actively broke masking mandates put strain on police forces.

In short this hypothesis could not be shown in the data, in this scenario the null hypothesis was that the introduction of mask mandates had no statistically significant impact on public disturbance crime when compared to pre pandemic years. The mask mandate was first introduced in August of 2020, meaning that if we compare the number of public disturbance calls to police in August 2020 vs all other pre pandemic years, there should be some statistically significant difference, here are the results:

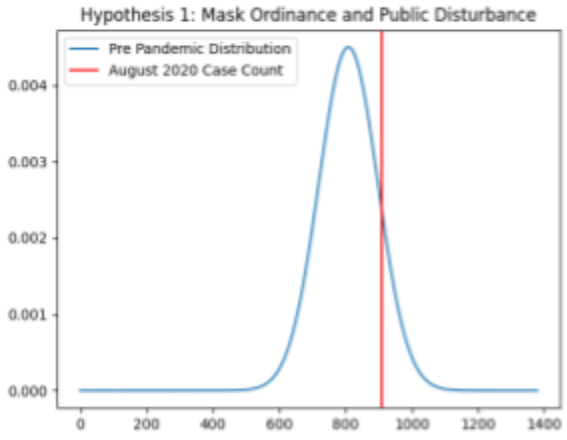

Figure 5: Case Distribution August pre Pandemic Compared to Observed Value

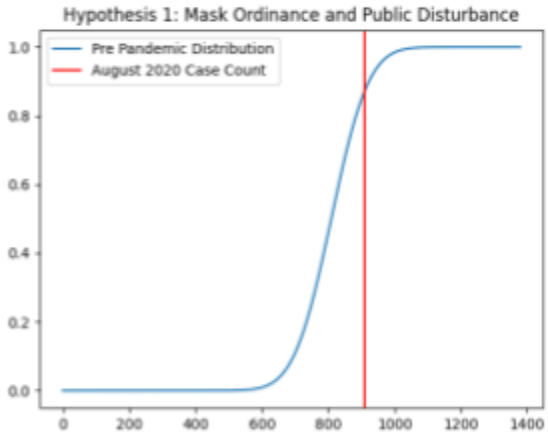

Figure 6: Public Disturbance CDF

In short we could not disprove the null hypothesis because there was a 13% chance that the month of August had a public disturbance case count at or above the observed value in 2020, which was 909 cases of public disturbance. This is the importance of taking into account standard deviation. While the observed crime rate in 2020 (909 cases) was 12% higher than the average rate for the years before it, the standard deviation of the pre pandemic distribution accounted for this shift.

Hypothesis 2: Lock down increased domestic abuse by 20% as victims no longer could avoid their abusive partners and were more likely to be stuck in the home with them.

Yet again this hypothesis could not be shown with statistical significance. The null hypothesis in this case being no significant change between domestic abuse rates during the lockdown as opposed to the same period of time before. In this case the lockdown occurred From March 26th to April 26th 2020 (4).

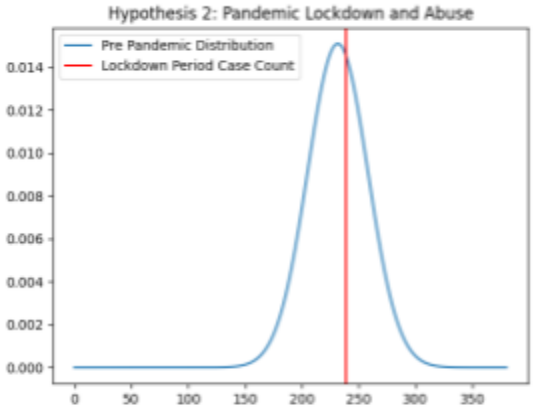

Figure 7: Case Distribution March-April pre Pandemic Compared to Observed Value

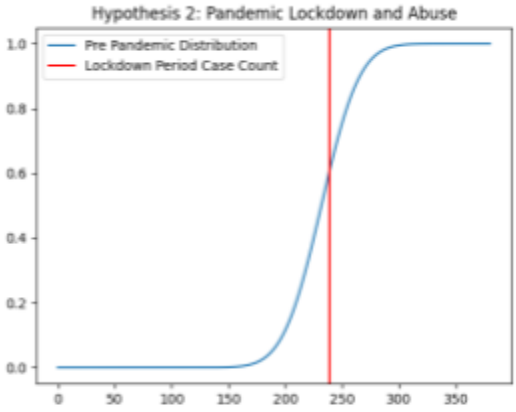

Figure 8: Domestic Abuse CDF

Again we could not disprove the null hypothesis because there was a 38% likelihood that the observed domestic abuse case count (239 cases) could appear in the pre pandemic distribution. Again while the observed value was 3% higher than the average for before the pandemic, this was covered by the standard deviation between the years of 2013-2019

Hypothesis 3: Stimulus checks reduced robbery and financial crimes by up to 10%, as people could finally put food on their tables, even for a week or two

To test this hypothesis the month directly after the stimulus checks were released was measured. When compared to the distribution before the pandemic it was clear that robbery was much higher than before, with a P value of 0.00164 this result is considered statistically significant, however this is not the full picture.

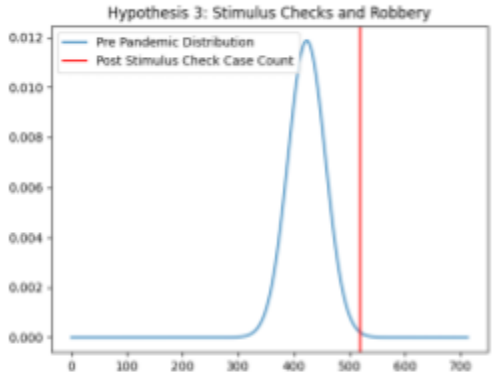

Figure 9: Case Distribution April pre Pandemic Compared to Observed Value

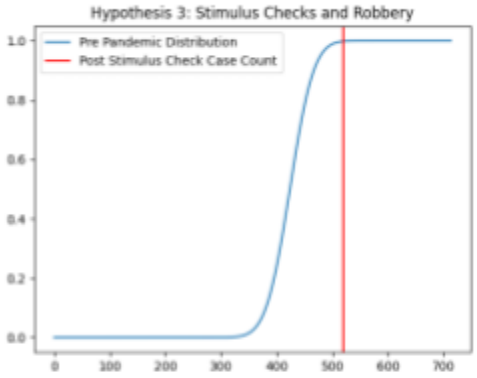

Figure 10: Robbery CDF

What this image is stating is that compared to <u>before</u> the pandemic the stimulus checks had no impact on burglary crime, but did it have an impact compared to crime rates in the weeks before? To do this we can simply plot a chart of daily crime counts before and after the stimulus checks and visibly inspect. Unfortunately this method is not fruitful either. While the stimulus checks did lift many people out of poverty temporarily, it did not meaningfully impact crime.

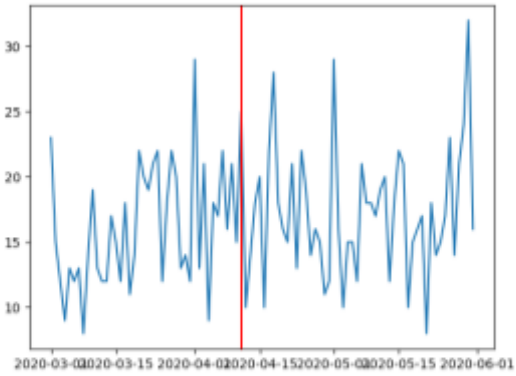

Figure 11: Daily Burglary counts in the weeks before/after the stimulus checks

Discussion/Implications

These findings are interesting because they show the ways that one may arrive at incorrect or partial conclusions when looking only at crime counts, without accounting for deviation. For example, technically the first hypothesis involving public disturbance and mask mandates was accurate, there was a visible 15% increase in calls during the month in which mandates were introduced. However when taking into account the standard deviation of calls year to year before the pandemic, this could have occurred by chance, we can not reject that null hypothesis. Future research could correct these findings by introducing thick data, or more full datasets to paint a more exhaustive picture of Denver county during those years.

During the course of this project I was thinking about these 6 human centered points highlighted in our week 7 reading about bias in data:1) Inconclusive evidence leading to unjustified actions 2) Inscrutable evidence leading to opacity, 3) Misguided evidence leading to unwanted bias, 4) Unfair outcomes leading to discrimination, 5) Transformative effects leading to challenges for autonomy and informational privacy, and 6) Traceability leading to moral responsibility. In specific points 1, 3, and 4 proved to be most visible in this study. As mentioned above an incomplete picture can often move lawmakers in the incorrect decision. For example, looking at the conclusions of hypothesis 3 one might think that giving more money to those in need does nothing to robbery rates in the county, and thus continue to underfund communities, leading to more crime and even less investment. That entire chain of unfair outcomes leads to even more misguided evidence and unwanted bias. While this study was meant to reinforce statements made by professionals and researchers, its use of only a single dataset made it difficult to do so, instead becoming an example of how bias can spring up in a research project.

Limitations

1) As mentioned in the methodology section of the report one of the most blatant limitations of this project was its insistence of looking at crime as a number, instead of a complex relationship between humans and systems. Some of the assumptions made by representing crime as a normal distribution, specifically independence in crime and time, contributed to incorrect predictions.
2) The NIBRS dataset is not perfect. The NIBRS dataset is collected with the aid of the FBI and excludes certain types of crimes for privacy's sake. For example any crime involving minors, as well as cases of sexual assault are all excluded from the NIBRS dataset. This exclusion may have had consequences on the analysis, specifically hypothesis two, which dealt with abuse as a result of the lockdown. I predict that if that data was anonymized and included (with safety measures such as date-time fuzzyness), the reported values would more closely match the hypothesis.
3) How far back in NIBRS do we measure? For the sake of this analysis the cutoff point chosen was the beginning of 2013. Technically NIBRS data did exist in 2012 and before, however it was more partial and was thus excluded. When comparing 2012 to 2013 data there is a systemic increase in the mean crime rates which is not explained by some seismic shift in Denver area politics. Leaving me as the researcher to assume that the increase in crime was actually caused by a change in reporting, and not actual lawlessness.

Conclusion

Attempting to model the cause and effect nature of crime is a difficult task with just data. No matter how complete or accurate a single dataset is, there is a need for thick data and ethnographic studies on crime in order to properly model it. This study set out with three hypothesis:

1) The introduction of mask mandates increased public disorder crime by 15%, as those who actively broke masking mandates put strain on police forces
2) Lock down increased domestic abuse by 20% as victims no longer could avoid their abusive partners and were more likely to be stuck in the home with them.
3) Stimulus checks reduced robbery and financial crimes by up to 10%, as people could finally put food on their tables, even for a week or two

However, due to a combination of factors it was unable to convincingly show any of the three points. Despite this, no results open the door for a conversation about how they were generated, and the flaws in expectations that lead to them. When dealing with human centered issues it is important to be thorough and cognizant of bias, never taking data as gospel and asking questions about its veracity and methods.

References

1) https://www.foxnews.com/politics/kamala-harris-wants-low-risk-inmates-released-amid-fears-of-covid-19-spreading-in-prisons
2) https://www.foxnews.com/media/what-happens-land-becomes-lawless-dr-oz
3) https://www.ojp.gov/ncjrs/virtual-library/abstracts/economic-factors-crime-and-delinquency-critical-review-empirical
4) https://covid19.colorado.gov/public-health-orders-and-executive-orders

Data Sources

- COVID Cases Daily: https://www.kaggle.com/datasets/antgoldbloom/covid19-data-from-john-hopkins-university
- Masking Mandates by County: https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i

- Denver Crime Data: https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime
- COVID timeline for “beginning” of the pandemic: https://www.cdc.gov/museum/timeline/covid19.html#:~:text=January%2010%2C%202020,%2DnCoV)%20on%20its%20website.
- Lockdown order dates were sourced via: https://covid19.colorado.gov/public-health-orders-and-executive-orders
- Stimulus check rollout dates were sourced via: https://www.fool.com/the-ascent/personal-finance/stimulus-checks/#:~:text=March%2027%3A%20The%20CARES%20Act,around%205%20million%20per%20week.