

Grant Savage

Data 512

11/9/2021

A5 Extension Plan

## **Motivation**

I plan on studying the impacts of COVID 19 and domestic violence cases in Cobb County and the surrounding Atlanta area. I have anecdotally heard that domestic violence is up, and I've seen some statistics reported on for large cities, however I have not seen data to back this claim for Atlanta or Cobb County. Surfacing this data provides social service agencies and non-profits with hard statistics and visualizations they can use to fight for funding for the services targeted to solve this issue. Surfacing this data is also useful because it will provide insight into what problems we will face in future pandemic-like lockdowns.

## **Research Questions**

How was domestic violence influenced by the pandemic?

Was there a month where violence peaked?

Have domestic violence cases dropped since the vaccine was available?

Hypothesis:

Domestic violence significantly increased during the pandemic

Domestic violence cases post vaccine are lower than domestic violence cases pre-vaccine.

## **Data**

I plan on using the Atlanta Police Department's raw crime data download (<https://www.atlantapd.org/i-want-to/crime-data-downloads>). The available data lists every crime reported by the APD and includes information such as location, date, and type of crime. I will also use the data provided in A4 to conduct this analysis. This data is extremely granular, and I will be able to quickly summarize daily domestic violence cases for Cobb County. There are possible ethical concerns around using data provided by the APD. The largest risk is that domestic violence is not equally reported. There may be underrepresented communities of people experiencing domestic violence who do not report these crimes to the police. It is also possible there may be contamination of data, police must use their own discretion when filing police reports.

## **Unknowns and Dependencies**

An unknown is the ratio of reported domestic violence cases to total domestic violence cases. For my analysis to be valid I am relying on this ratio to remain the same during the period I am analyzing. I am dependent on the APD's encoding of domestic violence cases to remain the same during the period I am analyzing.

## Methodology

First, there will be a decent amount of data cleaning and merging that will be required. The critical steps are as follows:

1. Flag the APD data for domestic violence cases
2. Limit the APD data to domestic violence only cases
3. Filter the APD data to Cobb County
4. Combine the APD data and the Covid19 data into a single data frame (merge on date)

Next, I can begin to answer the research questions.

"How was domestic violence influenced by the pandemic?" can be answered by simply charting the reported domestic violence cases per capita over time. I plan on using a connected scatter plot and highlighting key dates of the pandemic such as first case reported in Cobb, US mask recommendation, and vaccine availability. I will also answer this question by running a hypothesis test to see if domestic violence has meaningfully increased during COVID-19. To do this I will perform a t-test on Q3 2019 data and Q3 2020 data. The visual will be a useful tool to quickly show relative rates of domestic violence and any seasonality, while the t-test will more answer the question more concretely.

"Was there a month where violence peaked?" is an interesting question. I will need to group the data by month, then I plan on answering this in two ways. I will first answer it with a simple connected scatter plot, and I'll be able to visually identify the month with highest rates of domestic violence. Next, I plan on removing seasonality from the data and finding the month where violence peaked with seasonality factored in. I'll likely use the `seasonal_decompose` function provided in python's [statsmodels](#) package.

"Have domestic violence cases dropped since the vaccine was available?" can be answered by a t-test for reported cases per-capita before and after the vaccine was available. I will target the same months on differing years to account for seasonality. A t-test is an appropriate method to use to test if there is a meaningful statistical difference between the mean of two sets of data. In this case the two sets are pre-vaccine rates of domestic violence and post-vaccine rates.

## Timeline

My project proposal is broken up into 3 primary research questions:

1. How was domestic violence influenced by the pandemic?

2. Was there a month where violence peaked?
3. Have domestic violence cases dropped since the vaccine was available?

November 14<sup>th</sup>

- Have data cleaned and research question 1 answered and visuals are produced

November 21<sup>st</sup>

- Research question 2 is answered and visuals are produced

November 28<sup>th</sup>

- Research question 3 is answered and visuals are produced
- A6 is drafted
- Skeleton of final report is established

December 5<sup>th</sup>

- A6 presentation is complete
- Draft of final report is complete

December 12<sup>th</sup>

- Final report is complete