A5 – Extension Plan

Corina Geier

11/10/2021

I have two potential areas I hope to perform an analysis of. First, I would like to look at how Covid-19 cases affected the unemployment rate in Baltimore County. I would like to perform this analysis to understand the effect Covid had on the labor industry in Baltimore county, as employment can have a big influence on poverty and the economy, as well as social factors such as social unrest or conflict. I hope to learn how well Baltimore County managed the employment rate while dealing with a pandemic. 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](https://www.pnas.org/content/118/35/e2105482118). Understanding how the pandemic effected people’s jobs is very important to ensuring people have more job security in the future. This is what makes the problem human centered, as it explores how Covid shaped people’s employment and lives.

I also plan to look at crime statistics in Baltimore County and see if there is any relationship between Covid cases and the amount of crime in a given month. This will tie in with my first analysis as high unemployment rates could also have an effect on crime, making it difficult to separate the two analyses. One crime area I will pay specific attention to is assault, as [research from UC Davis](https://www.ucdavis.edu/news/covid-19-isolation-linked-increased-domestic-violence-researchers-suggest) suggests that the pandemic was linked to an increase in domestic violence assault cases. I hope to learn if there is a strong correlation between covid and crime rates and if specific crimes increased or decreased during the pandemic. This is human centered as a correlation between the pandemic and crime has very real consequences for people and if the problem is properly understood measures can be taken to prevent similar consequences from happening in the future.

My first research question is “how was the economy in Baltimore County influenced by the number of covid cases in the area?”. My hypothesis is that there is a positive correlation between the Covid infection rate and the unemployment rate in Baltimore County.

My second research question is “Is the crime rate correlated with the number of Covid cases in Baltimore County?”. My hypothesis is that higher cumulative covid cases correlate to higher crime rates, based on other factors that covid cases have an influence on. This hypothesis is partially reliant on my first hypothesis, as I have a secondary hypothesis that higher unemployment rates correlate to higher crime rates.

I plan to use additional data from the [Bureau of Labor statistics for unemployment data](https://data.bls.gov/pdq/SurveyOutputServlet) and data from the Baltimore County Government for information on crime rates which comes in two separate files, one for [2020 data](https://app.powerbigov.us/view?r=eyJrIjoiOTg4M2E3ZGQtMzIyNC00ZDk0LWEzNGItZWM3MWM4NGQ0YWM4IiwidCI6Ijk0NGZhOWJhLTg0NTQtNDEzZC1iOWU2LWJmNDBhZjFkNmE5YiJ9) and one for [2021 data](https://app.powerbigov.us/view?r=eyJrIjoiZmEyNTg0ODEtYmM0Ni00YzMyLWI3N2QtMWI5NzNmMDE3MWE0IiwidCI6Ijk0NGZhOWJhLTg0NTQtNDEzZC1iOWU2LWJmNDBhZjFkNmE5YiJ9). 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. This will help me look at unemployment rate vs infection rate for a given month. The data is in the public domain except for previously copyrighted photographs and illustrations. I’m looking at this particular dataset 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](https://en.wikipedia.org/wiki/Bureau_of_Labor_Statistics). A possible ethical consideration when using this dataset is that it is coming from a government agency, which means it might not be accounting for people not documented by the government and may have a selection bias as the data is collected using a computerized questionnaire. This data set will help me answer my question by providing unemployment data from throughout the pandemic which I will be able to use in a time series format to figure out what, if any, correlation exists between infection rate and unemployment. This expands on the Common Analysis as I will utilize the calculations I created for infection rate over time for use in this analysis.

The crime rate dataset contains monthly data with the number of crimes reported in a given month. The crime counts are separated by crime type. Since the 2020 and 2021 dataset have slightly different crime types I will focus my research primarily on the crime categories of Arson, Assault, Burglary, Homicide, Motor Vehicle Theft, Sexual Assault, and Robbery which are all shared categories between the two datasets. The data is all publicly available for use. I’m using this specific dataset since it’s coming from the Baltimore County government, so I can trust that they have recorded all the reported crimes from their own internal databases, as police departments are government agencies. A possible ethical consideration for using this dataset might be that as it is coming from the Baltimore County government and police department, it only reports 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, I will make sure to highlight that this only accounts for reported crimes and not all crimes. This data will help answer my question of if crime rate is correlated with Covid cases by allowing me to look at how crime rates have changed over time and how they’ve changed for different crime categories. Similar to my first research question, this will expand on the Common Analysis by allowing me to use my prior calculations while looking at a new data set to draw insights.

One factor that I can think of that’s outside my control is that I do not have consistent data from before the pandemic for crime rates. This might make it hard to access the impact of the pandemic on crime rates, as I’ll only be able look at how crime rates have changed throughout the pandemic as cases fluctuate.

I plan to use cross-correlation for both of my research questions as it’s a reliable way to compute to what degree two time series are dependent on one another, as noted in [towards data science](https://towardsdatascience.com/computing-cross-correlation-between-geophysical-time-series-488642be7bf0). I will probably shift the dependent vectors, crime and unemployment, by one month to account for a time lag in correlation. Cross correlation can help indicate whether the values of one time series are predictable based on the values of another lagged time series. There is a package in python called statsmodel.api that allows for an easy implementation of cross correlation. A correlation closer to 1 would tell me that the input and response vectors are heavily positively correlated. For my presentation I plan to discuss significant correlations that I find using my new datasets.

My timeline to completion is listed below:

* Data collection – Nov 13
* Data cleaning – Nov 16
* Cross correlation implemented – Nov 18
* Testing this methodology – Nov 24
* Visualizing results – Nov 28
* Process documentation – Dec 2
* Rough draft of slides – Dec 2
* Slides finalized – Dec 7
* Final report – Dec 14