Economic Impact on Crime in Chicago

CSPB 4502

Harper Chen
Computer Science
University of Colorado Boulder
Boulder Colorado US
haipeng.chen@colorado.edu

Jeff Mason
Computer Science
University of Colorado Boulder
Boulder Colorado US
jema2174@colorado.edu

Brendan McDonnell
Computer Science
University of Colorado Boulder
Boulder Colorado US
brmc7151@colorado.edu

1 Problem Statement

The proposed study aims to look at the effect economic conditions have on the crime rate within the city of Chicago. The study will look to evaluate the relationship of the overall crime rate, as well as various types of crime such as the property crime rate, in Chicago versus the unemployment rate and the GDP of the Chicago metropolitan area.

Identifying potential connections between economic factors and crime would allow for the city of Chicago to create criminal justice and economic policies to try and mitigate identified types of crime with economic reform. Any observed relationships could then be studied throughout the US to see if the same relationships exist and can be used to try and alleviate crime rates within other geographic areas.

2 Literature Survey

2.1 An Analysis of the Suspect: The Impact of Economic Conditions on Crime

Study looked at four macroeconomic variables, GDP per capita, unemployment rate, inflation rate, and interest rate and their effect on various crime rates (violent, property, and total). The study concentrates on data related to US economic and crime data from 1961 to 2019. The overall results show that US GDP did not have an effect on any type of crime rate within the US, and the unemployment rate did not have a statistically significant impact on violent crimes. The aggregate economic variables did show that economic conditions contribute to the crime rate. Property crime was shown to be the crime most influenced by economic conditions.[1]

2.2 The Impact of Economic Activity on Criminal Behavior: Evidence from the Fracking Boom

The study looked at the impact on crime rates in areas of North Dakota undergoing economic change as a result of the fracking industry. During the initial phase of area development resulting from the fracking industry crime rates were shown to decrease, but while economic promise had increased in the area economic conditions had not improved during this phase. During the follow-on phase associated with improving economic conditions, crime rates were shown to increase. The author offers the idea that the increase in job opportunity within the areas from the fracking industry results in an influx of people migrating to the fracking areas, and a large proportion of these individuals are young men who are statistically more likely to conduct crime. [2]

3 Proposed Work

3.1 Data Cleaning

Any data points containing NAN or Null values within required attributes such as date or type will be removed from the dataset prior to any form of analysis.

3.2 Data Integration

Three datasets will be linked together for analysis. The datasets will be linked together utilizing date timestamps within all three datasets.

The unemployment rate data will be linked to Chicago crime data via a month-year format, while the GDP data will be linked strictly via a year format.

3.3 Data Reduction

First, the data will be reduced in size in order to help facilitate management of the dataset and account for outlier time periods. Reducing the dataset timeline from 2001 through 2023 to 2006 through 2019 reduces the data points from over seven million to roughly 4 million data points. Additionally, omitting data starting in 2020 omits the brief periods of intense economic turmoil caused by the Covid-19 pandemic that may skew analysis of the data.

Second, a number of data attributes will be reduced that are either considered not pertinent for analysis of the data, or redundant. For example, police beat, case number, and FBI code will be removed from the dataset due to not providing meaningful information for this analysis. Data attributes such as grid location will be removed because location data has been captured by the latitude/longitude attribute.

3.4 Data Transformation

First, the date timestamp within the crime dataset is formatted dd-mm-yyyy HH:MM. While this level of detail may be beneficial for analysis of crime patterns related to days of the week or time of day; the focus is too narrow for the analysis proposed within this paper. In order to link with the unemployment rate dataset the date will be transformed into mm-yyyy. For linking with the GDP dataset the data will be further aggregated to the yyyy format.

Second, the classification of different crimes will be consolidated into fewer categories to facilitate analysis. For example, theft and robbery will be classified as property crimes to be compared in aggregate against other types such as violent crimes.

3.5 Data Analysis

Evaluate correlation between unemployment rate and GDP with the crime rate within Chicago. Identify patterns related to economic indicators and locations within the city. Generate

The proposal is similar in nature to the referenced studies in the Literature Survey. The study will look at similar economic factors like GDP and unemployment rate to the study referenced in 2.1. Similar to the study referenced in 2.2 this proposal will analyze a smaller geographic area. Limiting the scope of the analysis may allow for a better fitting of the correlation between crime rate and these economic factors by potentially reducing the number of confounding factors.

4 Data Set

4.1 Chicago Crimes: 2001 to Present

Dataset contains over 7 million data points with 22 attributes. Tracks reported in crime in the Chicago metropolitan area. Reports on crime type, location, date, crime code, and various description fields of the crime event. Data Source: City of Chicago

https://data.cityofchicago.org/browse?category=Public%20Safety

4.2 Chicago Unemployment Rate

Dataset contains the monthly unemployment rate for the city of Chicago since 2001. Data Source: Bureau of Labor Statistics

https://data.bls.gov/pdq/SurveyOutputServlet

4.3 GDP of the Chicago Metro Area

Tracks yearly GDP of the Chicago metro area (adjusted to 2012 dollars) since 2001. Data Source: Bureau of Labor Statistics

https://www.statista.com/statistics/183827/gdp-of-the-chicago-metro-area/

5 Evaluation Methods

In our evaluation, we will employ a combination of quantitative and qualitative methods to analyze the relationship between crime rates and economic indicators in Chicago using the Chicago Crime dataset. Our evaluation methods will address the following aspects:

1.Identifying correlation between crime rates and economic indicators: We will calculate correlation coefficients (e.g., Pearson's r or Spearman's rho) between crime rates and economic indicators such as GDP and unemployment rate. This will help us quantify the strength and direction of the relationships and determine if economic indicators influence crime rates.

2.Estimating the cost of crime in Chicago: We will perform an economic analysis using the crime data to estimate the direct and indirect costs of crime to individuals, businesses, and the community. This may include costs related to property loss, medical expenses, lost productivity, and the criminal justice system.

3.Analyzing crime at the neighborhood level: We will use spatial analysis techniques, such as Geographic Information Systems (GIS), to map crime incidents and economic data at the neighborhood level. This will allow us to visualize the distribution of crime and economic indicators across Chicago and identify areas with high crime rates and vulnerable economic conditions.

4.Identifying the economic impact of crimes committed against businesses: We will analyze the dataset to determine the types and frequency of crimes committed against businesses and estimate the associated economic losses. This will provide insights into the specific challenges faced by businesses in Chicago due to crime.

5.Informing policies aimed at reducing crime through economic development: Based on our findings, we will provide recommendations for policy interventions that could potentially reduce crime rates by addressing underlying economic factors. This may include initiatives related to job creation, education, and community development.

6.Identifying additional factors contributing to crime rates in Chicago: We will review the existing literature and conduct further analysis as needed to explore other factors that may influence crime rates in the city, such as social, demographic, and environmental factors.

7.Considering limitations and biases in the data: Throughout our evaluation, we will be mindful of the limitations and biases present in the data, and we will address these in our interpretation of the results. This may involve assessing data quality, data collection methods, and potential confounding factors.

8.Incorporating additional data sources and analysis methods: We will acknowledge the need for a more complete understanding of the relationship between crime rates and economic indicators in Chicago and will explore the possibility of incorporating additional data sources and employing more advanced analytical methods to further our research.

By employing these evaluation methods, we will provide a comprehensive analysis of the relationship between crime rates and economic indicators in Chicago, which will inform policy development, contribute to our understanding of the factors influencing crime in the city, and highlight areas for future research.

6 Tools

6.1 Communication

In order to properly facilitate efficient communication between our group members we've decided to utilize several different methods of communication. The first major tool we've enlisted to assist with our communication is Discord. Discord provides our

group with a way to quickly communicate simple messages and ideas with assurances that everyone in the group will receive a notification of the idea. As everyone in our group had some familiarity with the platform it seemed like a natural fit. The next platform we will incorporate to help us communicate is Zoom. While Discord does offer voice and video calling, having a more reliable platform with less issues like Zoom should help to avoid any problems.

6.2 Project Organization and Document Management

By using one centralized platform for all of the external documents related to our project (proposals, presentations, etc.) our group can ensure that organization is our strong suit. We've chosen to use Google Drive and Google Docs to allow us to all work on these important project documents simultaneously. This should allow all of us to contribute from any location by giving us the most flexibility possible to store our in-work documents.

To organize the code that comes with this project along with our completed documents our group has decided to use github. The versatility of github in accepting most any form of documentation should suit this project well as we will be uploading documents, jupyter notebooks, videos, and more to support our findings. The organization that is provided by the very user-friendly interface of github's version control.

6.4 Data Visualization

Arguably just as important as finding interesting trends in our data and answering questions central to the idea of the project, is clearly communicating our findings to our audience in a way which they can easily understand. Data visualization tools provide that bridge between raw numerical data and the user experience by forming an easily interpretable medium. To assist in our data visualization efforts our group will use Tableau and PowerBI. These platforms will allow us to take our findings and convert them into a visually pleasing representation to assist in the video presentation of our work. While we do plan to use Matplotlib for the plotting of our work during our internal data analysis, Tableau and PowerBI will provide additional options that better allow us to present important information to our audience.

6.4 Programming Language and Tools

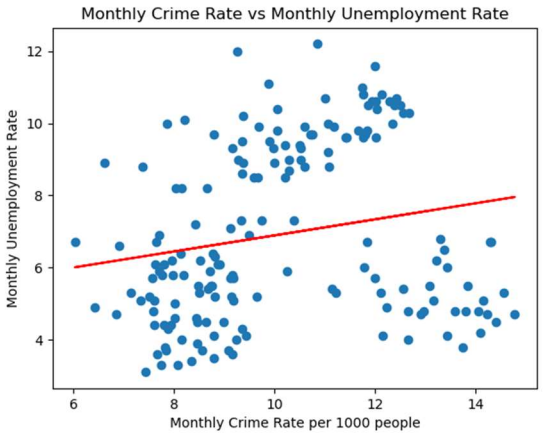
While there are several options available for programming languages including some that are designed for statistical analysis, our group has determined that the best option is a language that we are all comfortable working in, Python. With our group's collective experience in Python we don't anticipate any issues that we won't understand how to resolve. There are several libraries that we anticipate using to help with our efforts. First we anticipate using the Pandas library, this should handle the majority of our data cleaning and general data wrangling. The ability of Pandas to work with CSV files makes it a great choice. Next we will be using the Numpy library. Numpy contains functions for handling basic statistical functions and will be used throughout the project for statistical analysis. In addition to Numpy our group may use the SciPy stats module if we wish to use more complex statistical analysis functions that aren't available in Numpy. The last library that we anticipate using is Matplotlib. Although we will use Tableau and PowerBI for communicating our final results to our audience, Matplotlib provides a great way of viewing our findings internally. Lastly, to organize our code as we work on it, our group will be using Jupyter Notebooks. Jupyter Notebooks will allow our group to separate our ideas into different code blocks and give us great flexibility when it comes to running different sections of code. It will allow us to easily highlight issues that need to be fixed, and save sections that produce the most interesting results.

7 Results

7.1 Overall Crime Rate vs Unemployment Rate

The overall monthly crime rate was defined as the total number of crimes that occurred within a month divided by the current Chicago population multiplied by 1000 to achieve a monthly crime rate per 1000

people. All crime types recorded by the City of Chicago were included within this dataset. The scatter plot below plots the crime rate and unemployment rate linked by the corresponding year and month in which both occurred. The scatter plot displays a weak or insignificant positive relationship between unemployment rate and the total crime rate.



This insignificant relationship is confirmed via the pearson correlation value between the two rates which resulted in a value of 0.195276.

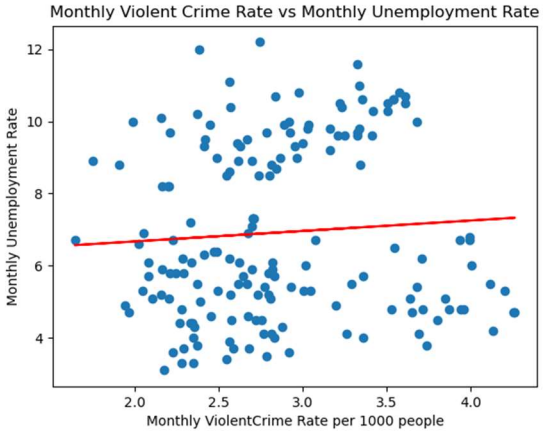
7.2 Property Crime Rate vs Unemployment Rate

Property crimes were defined as any crimes that fell into one of the following five categories of theft, auto theft, robbery, burglary, or deceptive practice. The monthly property crime rate was calculated per 1000 people like the total crime rate. The scatter plot depicts a stronger relationship between property crime than the relationship between total crime and the unemployment rate.

Although the relationship between property crime and unemployment rate is stronger the pearson correlation value is only 0.271569

7.3 Violent Crime Rate vs Unemployment Rate

Violent crimes were defined as any crimes that fell into one of five type categories. The five violent crime categories were assault, battery, homicide, domestic violence, or robbery. The correlation value of the two rates was 0.069969 indicating no relationship between unemployment rate and the violent crime rate within Chicago. This lack of a relationship is evident in the scatter plot with a near flat correlation line and a fairly random dispersal of points.



7.4 Overall Crime vs GDP

8 Milestones

8.1 Milestones Completed

Data Collection and Preparation

- Confirm all rows with null values in the date timestamp attribute are removed
- Confirm all rows with null values for crime type are removed

Datasets Fully Transformed/Integrated

- Crime date timestamps transformed into mm-yyyy and yyyy format
- Crime types are binned into separate categories, property crime, violent crime, other.

Descriptive Statistics and Correlation Analysis

- Conduct correlation analysis to identify potential relationships between crime rates and economic indicators
- Perform regression analysis to model the relationships between crime rates and economic indicators

8.2 Milestones TODO

24 April Progress Report

- Create line graphs and box plots charting the trends of changes in unemployment rate, GDP and the crime rate
- Bayesian Classification analysis of economic factors and crime rates greater or less than Chicago average
- GDP Correlation and Regression analysis
- Submit progress report
- Evaluate milestone deficiencies

1 May

- Continue work on final paper and video presentation
- ReadMe production
- Video Recording
- Finalize code documentation

8 May Final Project

- Video/Paper/Code uploaded to github

REFERENCES

[1] Perez, R. (2022). An Analysis of the Suspect: The Impact of Economic Conditions on Crime. Economics Undergraduate Honors Theses Retrieved from https://scholarworks.uark.edu/econuht/40

[2] Street Brittany (2019). The Impact of Economic Activity on Criminal Behavior: Evidence from the Fracking Boom. The Center for Growth and Opportunity at Utah State University. Retrieved from https://www.thecgo.org/research/the-impact-of-economic-activity-on-criminal-behavior-evidence-from-the-fracking-boom/