**Environmental Data Science Final Project**

Jonathan Bernard Widjajakusuma (jbw7)

April 30th 2024

**Background and Introduction**

The topic I will be researching for my final project is a important one, because understanding the main factors contributing to high crime rate is crucial for developing system and strategies to reduce crime, and hence creating a much safer community for all. My research question for this project is **"What economic factors contribute most to crime rates in Chicago?"**. I’m mainly focusing in Chicago, as it has been known to have a relatively high crime rate compared to other cities across the United States. By identifying which economic variables have the strongest contribution on crime in Chicago can help bring valuable insights to reduce crime rate in the City. By focusing on a highly populated urban center like Chicago may lead to finding solutions that are applicable to other rapidly growing cities such as New York, Jakarta, Los Angeles, and Tokyo.

My hypothesis to my research question is that unemployment rate will be the most significant factor as I think people who are unemployed may lack the ability to financially support their daily needs. Hence, this could drive them to engage in criminal activities to obtain basic needs and money. As this was proven during the Covid-19 pandemic when millions of people suddenly became unemployed and as a study conducted by Northeastern university reported “30 percent increase in the number of homicides from 2019 to 2020”(Northeastern University, 2023). This shows that during economic hardship, even individuals without previous crime history may resort to criminal activities out of desperation.

Existing research have also explored the relationship between economic factors and crime rates. Based on the FBI's 2017 Crime in the United States report, Chicago ranked 3rd in the number of violent crimes. Furthermore 1st in the number of murders nationwide (FBI, 2017). Furthermore, Forbes reported that as of March 2024, the state of Illinois had an unemployment rate of 4.8% which was the 4th highest in the United States (Forbes, 2024). These reports clearly shows the importance of investigating the potential relationship between economic factors in crime in Chicago.

**Data Used**

For this project, I used two datasets along with Chicago's Census tract data. The first dataset, titled "Crimes - Map," was downloaded from the Chicago Data Portal (https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6). This dataset is the only version available and contains data ranging from 2023 to 2024 in which I used the entire range of the dataset in my analysis. I chose this dataset because it contains essential parameters that I require to conduct my analysis, such as the date of occurrence and specific coordinates of each crime. Although the dataset contains dozens of other parameters, I focused solely on the Date of Occurrence and Location variables, as the goal of this project is to analyze the spatial distribution of crime rather than focusing into specifics like crime types.

The second dataset, "Chicago Health Atlas Data," was downloaded from the Chicago Health Atlas website (https://chicagohealthatlas.org/indicators/UMP?topic=unemployment-rate) and I specifically selected the of social and economic indicators. This dataset is also the only version available and ranges from 2018 to 2022. However, to keep temporal proximity as close as possible to the crime dataset, I only conducted analysis from 2021 to 2022. The dataset includes essential economic factors such as Unemployment Rate, Economic Diversity, and Median Household Income, making it an ideal choice for this project.

Lastly, I obtained Chicago's Census Tracts data from the "Boundaries-Census Tract—2010" dataset (https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik), which is also downloaded from the Chicago Data Portal. This dataset represents the only version available and comes from 2010.

**Methods**

**Data Wrangling**

For the crime dataset, I first removed unnecessary columns and kept only the relevant columns, which is LOCATION. I then aggregated the number of crimes based on LOCATION to calculate the crime count for each unique location using the aggregate() function. I then proceed to removed duplicate rows based on the LOCATION column to ensure unique locations and removed rows with missing coordinate values using the filter() function from the dplyr package. I did the same for the economic dataset, I also removed unnecessary columns and kept only relevant variables. I removed rows with missing coordinate values using the filter() function from the dplyr package. I converted both datasets into sf objects using CRS 4326 and performed a spatial join between the census tract shapefile and the crime and economic data. I then aggregated the crime count by census tract using group\_by() and summarize() to calculate the sum of crime counts for each census tract.

I chose the crime count per census tract as our spatial variable because it allows me to analyze the spatial distribution of crime across different areas of the city. By examining how crime counts vary across census tracts and comparing them with economic factors such as unemployment rates and median household incomes, I can gain valuable insights into the potential relationships between the economic environment crime rates. This spatial variable aligns with my research question and hypothesis, enabling me to investigate whether areas with higher levels of economic hardship often experience higher crime rates.

**Data Visualization**

For this I used the tmap package to create interactive heatmaps for each crime count, unemployment rate, median household income, per capita income, and poverty rate. This allowed me to visually explore (and for readers to better understand) the relationship between the different factors and crime rate. However, as I could not see any clear connection, therefore I decided to perform correlation analysis and spatial autocorrelation analysis.

**Correlation Analysis**

I merged the aggregated crime data with the economic data using a spatial join based on the census tracts. I then converted the merged sf object into a data frame to select the relevant columns for correlation analysis. I then printed the correlation matrix for analysis.

**Spatial Autocorrelation Analysis using LISA**

To perform spatial autocorrelation analysis using Local Indicators of Spatial Association (LISA), I first converted the merged crime and economic data from an sf object to a SpatialPolygonsDataFrame. I then created a spatial weights matrix based on queen contiguity using the poly2nb() function and converted the weights to a listw object using nb2listw(). After ensuring the crime count variable was all numerical values, I calculated Local Moran's I using the localmoran() function and added the results to the census tract data. Finally, I converted the data back to an sf object, and generated a LISA map using tmap to visualize the spatial autocorrelation of crime counts.

**Spatial Autocorrelation Analysis using Global Moran’s I**

The steps I took for this analysis is very similar to the analysis using the LISA method, but after making sure that the crime count variable only contained numerical values, I used moran.test() to the crime count variable and the spatial weights matrix to calculate the Moran’s I statistic, measuring the degree of spatial autocorrelation in the data.

**A diagram of data analysis

Description automatically generated**

**Results and Discussions**

Crime Map

**A map of a city

Description automatically generated**

Unemployment Map Median Household Income Map

**A map of a city with numbers and a graph

Description automatically generatedA map of a city with numbers and a screen

Description automatically generated**

Per Capita Income Map Poverty Rate Map

**A map with numbers and a screen

Description automatically generatedA map of a city with numbers and a table

Description automatically generated**

The heatmaps produced using the map package provide a visual representation of the spatial distribution of crime counts, unemployment rates, median household incomes, per capita incomes, and poverty rates across the city's census tracts. There are some patterns by visually analyzing the maps, specifically, areas with higher crime counts tend to be concentrated in the southern and western parts of the city, which are regions that also exhibit higher unemployment rates and lower median household incomes. These observations aligns with my hypothesis that unemployment and economic disadvantage may contribute to higher crime rates. However, there are some inconsistencies in the spatial patterns. Some census tracts with relatively high median household incomes and low poverty rates still experience slightly high crime counts. While the heatmaps allow for a general visual overview, they do not clearly reveal any strong correlations between the variables.

**Correlation Analysis**

**A number of numbers on a white background

Description automatically generated**

The correlation matrix shows the correlation coefficients between the different variables. The results indicate that crime count has a weak positive correlation with the unemployment rate (0.1297) and poverty rate (0.1342) and a weak negative correlation with median household income (-0.0868) and per capita income (0.0007). These correlations slightly indicates that higher unemployment rates and poverty rates are slightly associated with higher crime counts, while higher median household incomes are slightly associated with lower crime counts. However, the correlations are relatively weak and do not provide strong evidence for a direct causal relationship.

**Spatial Autocorrelation Analysis using LISA**

A map of a city

Description automatically generated

The LISA map shows the spatial clustering of crime counts across the city. The map shows areas of high-high as blue and low-low as grey spatial clustering, indicating census tracts with high crime counts surrounded by other high-crime tracts and low-crime tracts surrounded by other low-crime tracts. The presence of spatial clustering suggests that crime rates are not randomly distributed across the city but have spatial patterns. The spatial clustering of crime rates revealed by the LISA analysis further shows that crime is not randomly distributed across the city but has distinct spatial patterns. The presence of high-high and low-low clusters suggests that areas with similar crime rates tend to be located near each other, forming hotspots and coldspots of criminal activity.

**Spatial Autocorrelation Analysis using Global Moran’s I**

A white paper with black text

Description automatically generated

The results shows a highly significant positive spatial autocorrelation in the crime rates across the census tracts of Chicago. The Moran’s I statistic is calculated as 0.3044, with a p-value that is less than 2.2e-16, hence further indicating that spatial distribution of crime rates is not random but there are strong clustering patterns. Further showing that criminal activities tend to be located near each other.

Based on the results, the initial hypothesis that the unemployment rate would be the most significant contributor to crime rates in Chicago is not strongly supported. While the unemployment rate shows a weak positive correlation with crime count, it is not the strongest correlation among the variables analyzed. The LISA analysis reveals that crime rates exhibit spatial clustering, indicating the presence of other factors outside the economic variables considered in this project. From the Global Moran’s I results, its clear that crime is not random, so its still crucial to research what other factors plays a role in these clustering of crime activities.

**Limitations**

* Limited variable selection: The analysis focused on a limited set of economic variables. Other relevant variables, such as education levels, housing conditions, or social factors, may provide additional insights into the factors contributing to crime rates.
* Temporal mismatch: The crime data and economic data used in the analysis cover different time periods. The crime data is from 2023 to 2024, while the economic data is from 2018 to 2022. This temporal mismatch may introduce some limitations in interpreting the results, as the economic conditions may have changed over time.
* Analysis level: The analysis is conducted at the census tract level; therefore, the results at the individual level may lead to inaccurate results, as the relationships observed at the aggregate level may not be true for individuals within each census tract.

**Conclusions**

In this final project, I investigated the relationships between crime rates and various economic factors in Chicago. From the analysis I conducted in this project revealed weak correlations between crime rates and the selected economic factors. The findings does not support my initial hypothesis that the unemployment rate would be the most significant contributor to crime rates in Chicago. The spatial autocorrelation analysis indicated the presence of spatial clustering of crime rates, suggesting that other factors beyond the economic variables considered may contribute to the spatial distribution of crime in the city.

It is important to acknowledge the limitations of this study, such as the limited set of variables, the temporal mismatch between data sources, and the aggregate level of analysis. Future research could reduce these limitations by using a broader range of variables, aligning data temporally, and exploring different levels of analysis. Despite these limitations, this project still contributes to the topic on the factors influencing crime rates in urban areas and clearly shows the need for a more comprehensive approach that considers various social, economic, and environmental factors when addressing crime in Chicago and other urban areas.

**References**

“Boundaries - Census Tracts - 2010: City of Chicago: Data Portal.” *Chicago*, data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik. Accessed 3 May 2024.

Campisi, Natalie. “Unemployment Rates by State: See Your State Rank.” *Forbes*, Forbes Magazine, 25 Apr. 2024, www.forbes.com/advisor/personal-finance/unemployment-rates-by-state-04-25-24/.

“Crimes - Map: City of Chicago: Data Portal.” *Chicago*, data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6. Accessed 3 May 2024.

Dureva, Desislava. “How Crime Rate Changed during the Pandemic - Neu MSCJ.” *Public Affairs Master’s Degree Programs at Northeastern*, 22 Feb. 2024, publicaffairs.northeastern.edu/articles/us-crime-rate-during-pandemic/.

“Unemployment Rate - Chicago Health Atlas.” *Chicagohealthatlas.Org*, chicagohealthatlas.org/indicators/UMP?topic=unemployment-rate. Accessed 3 May 2024.

Vigderman, Aliza. “Which U.S. Cities Are the Most Dangerous?” *Security.Org*, 2 Feb. 2024, www.security.org/resources/most-dangerous-cities/.

“Violent Crime.” *FBI*, FBI, 10 Sept. 2018, ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topic-pages/violent-crime.