Crime Analysis After Pandemic in Chicago and Los Angeles

Authors: Serena Tang, Sophathriya Sen, Alicia Martinez, Dima Kim
Department of Information Systems, California State University Los Angeles
CIS 5200-01System Analysis and Design
e-mail : xtang13@calstatela.edu, ssen5@calstatela.edu, amarti761@calstatela.edu, dkim171@calstatela.edu

Abstract: This report presents a temporal-spatial and visual analysis of violent crime in Cook County, Chicago, and Los Angeles using the crime dataset from 2020 to the present. The combination of additional datasets of independent factors such as education attainment, unemployment, household income, and population density allow meaningful insights into these influential factors that may or may not impact violent crimes between the two cities. The analysis was conducted using Hive Hadoop, Python, Power BI, Excel, and ArcGIS. Leveraging these tools, we aim to explore and investigate these key factors across both cities as it has one of the highest crime rates in the country. Our analysis will involve building a map story and visual analytics of violent crimes and their relations to the listed independent factors.

1. Introduction

This project uses Hadoop and Hive as an exclusive space between authors to have access to and process datasets. There are various datasets included in this project, such as crime, education, unemployment, income, household, and population, which are two datasets. Therefore, there are a total of ten datasets that need to be cleaned, uploaded and joined. These datasets will be joined by a standard column called "tract ID," representing a small geographic area within a county. We have chosen these datasets, specifically for Chicago's Cook County and Los Angeles, because crime rates are significantly high in the United States.

2. Related Work

These datasets are collected from the United States Census Bureau. Annually, the Census Bureau publishes reports estimation along with demographics components such as births, deaths, and migration. These factors can be categorized by different characteristics such as age, gender, race, as well as by geographic levels such as national, state, and county.[1] Although these datasets are relatively cleaned; however, more cleaning are required due to many fields included that needed to be removed as some of the fields are not relevant for visual analysis. Spatial aggregation will be conducted on the crime dataset. After cleaning both the crime dataset and the related topic datasets, we will use the tract ID as a key to integrate the crime data with the selected topic datasets.

3. Specifications

The total ten datasets comprise data containing numbers and types of crimes, our focus, and independent datasets of education, unemployment, income, and population density across both cities. The dataset is a size of 2.9 GB and exhibits data over the past four years after the pandemic, specifically related to crimes. Although our primary dataset is not a relatively big size, we included the other independent factors and assumed the methodology and processes could be applied to a larger dataset. Table 1 shows the files and size of the files from the dataset.

Table 1. Data Specification

Data Set	Size (Total 2.9GB)
Data Coverage (in year)	2020 - 2024

The table below show the platform specifics utilizing Hadoop platform for our project.

Table 2. Platform Specification

Cluster Version	Oracle Big Data Service
Number of Nodes	5
CPU Speed	2.5 GHz
Memory Size	30 GB

4. Implementation Flowchart

The dataset collected from the Census Bureau contains many details of different crime types, its description, and the total number of crimes that happened in the span of the last four years (2020 to 2024). The entire data manipulation process is visually shown in the flowchart below (Figure 1). As mentioned before, data cleaning was done through Python before it was uploaded to Hadoop due to many unnecessary columns that would overcomplicate the process of uploading to Hadoop. Hence, this approach was concluded and proceeded with cleaned data uploaded to Hadoop. There are two approaches to visually analyzing these datasets. First, the sole crime dataset of both cities was cleaned by Python, uploaded, and visualized using the ArcGIS Online tool. Second, the four related topic datasets were cleaned by Python, uploaded, joined with crime datasets, and visualized using the Power BI tool.

Figure 1. Data Manipulation Flowchart

5. Data Cleaning

Raw files were downloaded from the United States Census Bureau and cleaned using Python. Columns such as tract ID, year, crime counts, crime types, and potential columns were selected for future visual analysis. A thorough cleaning process on Python will be shown in the tutorial for readers to follow easily.

First, we downloaded the crime dataset and the Chicago City and Los Angeles City shapefiles from the United States Census Bureau. Then, we use Python to aggregate them to get the dataset in which crime counts belong to the tract ID. We download the education, unemployment, income,

household, and population raw datasets and use Python to clean them, ensuring the topics dataset only keeps the tract ID and the necessary columns for visualization.

Second, we uploaded the cleaned crime counts and four topics datasets, all within tract ID in HDFS, and created an external table for both cities in Hadoop. Then, we join the crime table with each topic for each city in four years. After joining tables in Hadoop, we export the tables we need to use in visualization.

6. Analysis and Visualization

After data cleaning in Python and joining processes in Hadoop, ten CSV files were downloaded to a local computer—from Hadoop—in preparation for visual analytics.

6.1 Crime Analysis of Chicago

Utilizing an online tool, ArcGIS [1], on the crime map of Chicago between 2020 and 2024 (Figure 2), five key communities in the City of Chicago consistently exhibited high crime levels, with crime counts exceeding 1,000. These areas include the Loop, Austin, Humboldt Park, North Lawndale, and the Near South Side. In the southern part of Chicago, crime incidents are more dispersed, indicating a diverse distribution of criminal activity.

The highest crime counts have risen significantly, from approximately 2,000 in 2020 to over 3,000 in 2024, reflecting a worsening trend in crime severity following the pandemic. A comparison of maps from 2020 to 2024 reveals a clear expansion of high-crime areas, underscoring the growing geographic spread of criminal activity during this period.

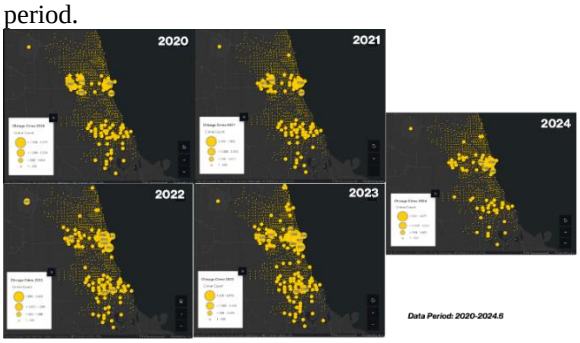

Figure 2. Chicago Crime Map 2020-2024

Focusing on the main communities around downtown Chicago (Figure 3), particularly the Loop, we observe a notable increase in both the number of crimes and the geographic area impacted between 2020 and 2024. The Loop community stands out as a critical hotspot, reflecting the most significant growth in crime activity.

Although the data for 2024 only represents a partial year, it already indicates a sharp rise in incidents compared to prior years. In 2023, the Loop recorded the highest annual crime count at 3,478. Over the observed period, this single community has reported more than 8,500 crime calls, underscoring its status as a focal point for law enforcement and community safety concerns.

This highlights the urgent need for targeted intervention strategies in high-density urban areas like the Loop to address the escalating crime trends and their implications for community well-being.

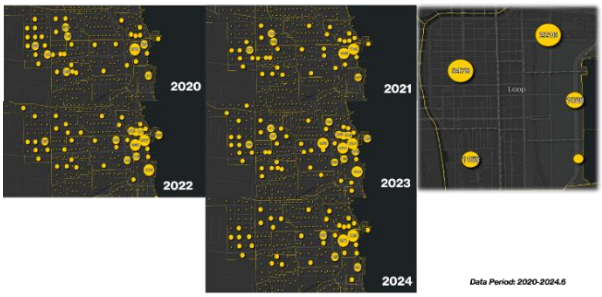

Figure 3. Chicago Downtown Crime Map 2020-2024

6.2 Crime Analysis of Los Angeles

From 2020 to 2024 (Figure 4), crime frequency in Los Angeles not only increased but also expanded to cover a broader area. Three key hotspots emerged: Central City, Los Angeles International Airport (LAX), and South Los Angeles. A particularly notable trend since 2022 is the visible expansion of crime from Downtown Los Angeles into Southeast Los Angeles, reflecting a growing geographic spread.

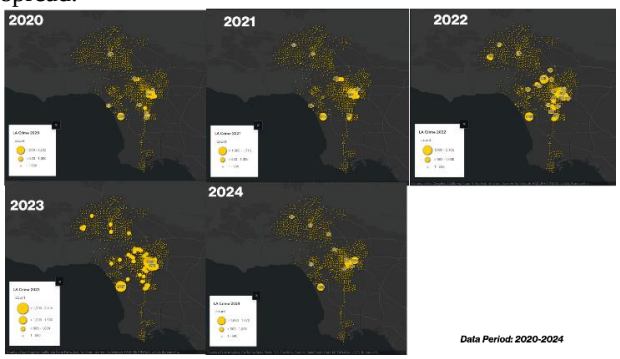

Figure 4. Los Angeles Crime Map 2020-2024

The highest crime counts have surged significantly, rising from 1,282 incidents in the LAX area 2020 to 2,416 in Central City 2023, marking an alarming escalation.

What's particularly interesting is the Hollywood community's crime pattern: in 2022, crime incidents spiked sharply to nearly 1,000 and have remained consistently high over the subsequent two years, hovering around the same level.

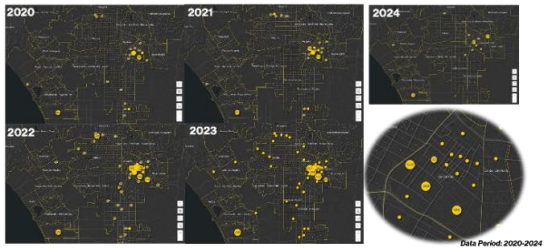

Figure 5. Los Angeles Downtown Crime Map 2020-2024

The Central City community in Downtown Los Angeles has experienced a dramatic surge in crime frequency over the past four years (Figure 5). The highest annual crime count doubled from approximately 1,000 in 2020 to nearly 2,500 in 2023. Even with incomplete data for 2024, the community has already recorded 1,672 incidents, showing a continuing upward trend.

[1] ArcGIS Online: https://www.arcgis.com/index.html

In 2020, the community logged 6,186 calls, but by 2023, this figure had more than doubled to over 14,000. This stark increase underscores the escalating seriousness of the safety challenges faced by residents and businesses in the downtown area.

The crime trends around the LAX area tell a similar story. In 2020, the community reported 1,282 incidents, rising to 2,167 by 2023—a significant increase over three years. Early data for 2024 suggests this upward trend is persisting, pointing to the area's growing vulnerability despite its critical role as a transportation hub.

6.4 Crime and Education Level Analysis

Utilizing Power BI, visualizations for Chicago and Los Angeles's crime and the community's education attainment level are categorized by age group from 2020 to 2022. We examine these trends to uncover how educational attainment may influence public safety.

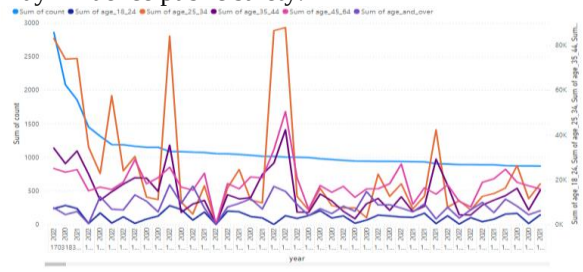

Figure 6. Chicago's Crime and Education Level vary by Age Group

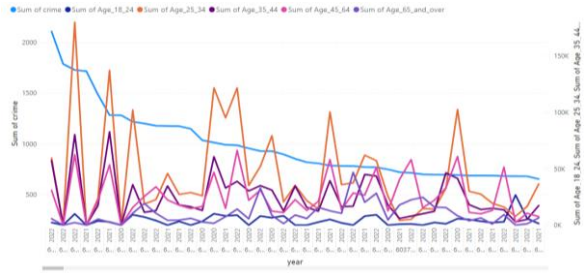

Figure 7. Los Angeles's Crime and Education Level vary by Age Group

Observation has been made that crime rates are higher in the early time frame but steadily decline over the period. Meanwhile, graduation rates for most age groups remain relatively stable, slightly fluctuating. Considering the variability, possible reflection changes are made due to the influences of socioeconomic conditions or local policies exhibited in both charts—Figure 6 and Figure 7.

Overall, these findings reinforce the critical role of education in shaping safer communities. While both cities underscore the link between education and crime rates, each city demonstrates unique characteristics to consider when designing interventions to shape safer communities.

6.5 Crime and Unemployment Analysis

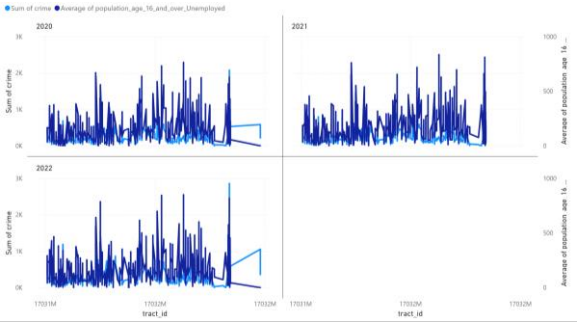

Figure 8. Chicago's Crime and Unemployment Population of Age 16 and Over

Figure 8 illustrates the trends in unemployment and crime with the main focus on the year 2020 to 2022. The crime rates appeared to have fluctuated significantly, which aligns with the spike in unemployment.

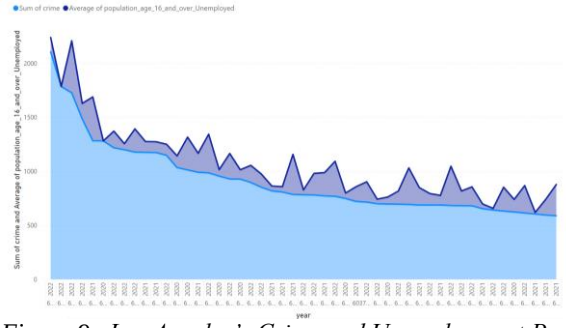

Figure 9. Los Angeles's Crime and Unemployment Population of Age 16 and Over

The observation of Figure 9 demonstrates that the variations in unemployment rates appear less dramatic than in Chicago. Although the crime rates seem to stabilize slightly despite continued unemployment, the possibility indicates differing socio-economic dynamics. This observation underscores how localized factors may have influenced the relationship between crime rate and unemployment.

Therefore, both Figure 8 and Figure 8 highlight the interplay between economic conditions and public safety; the trends display that unemployment and crime are interconnected, suggesting a relationship between the two factors. Although both charts have shown a slight fluctuation, the number of unemployment will always stay below the number of crimes.

6.6 Crime and Income Los Angeles

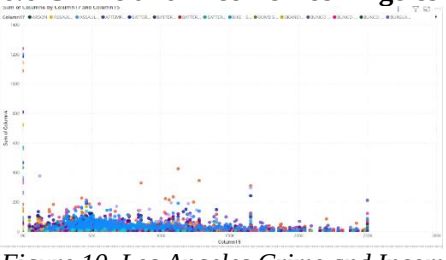

Figure 10. Los Angeles Crime and Income Ranging from 0-250k

It was concluded that the following crimes were widespread amongst this particular income range: Battery Simple Assault, Theft Plain Petty (valued under $950), Burglary from Vehicle, Vehicle Theft, Assault with a

Deadly Weapon, Grand Theft (valued over $950), Vandalism (including church-related vandalism), and Domestic Abuse Assault.

The next observation concluded that these are the popular crimes for the income range of 100k—250k: burglary, Identity Theft, Violation of a Temporary Restraining Order, Trespassing, Lewd Letters/Telephone Calls, Domestic Abuse, Grand Theft, and Document Forgery.

6.7 Crime and Income Chicago

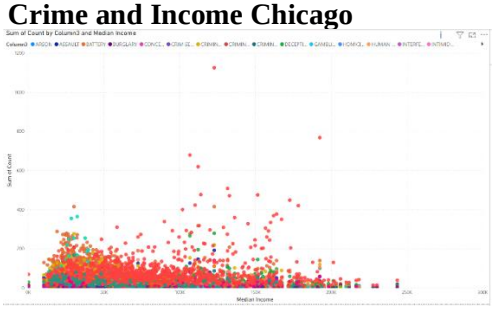

Figure 11 Chicago Crime and Income Ranging from 0-250k

Observation of Figure 10 reveals the types of crimes committed across the income spectrum ranging from $0 to $250,000. The primary emphasis was placed on crimes committed within the $50,000 and below range to identify any consistent trends and to draw comparisons with crimes observed in higher income brackets. It was determined that the following crimes were particularly prevalent within the $50,000 and below range: Battery, Theft, Assault, Weapons Violations, and Motor Vehicle Theft.

Figure 11 reveals the types of crimes committed across the income spectrum ranging from $0 to $250,000. To identify any consistent trends and draw comparisons with crimes observed in higher income brackets, the primary emphasis was placed on crimes committed within the $50,000 and below range. It was determined that the following crimes were particularly prevalent within the $50,000 and below range: Battery, Theft, Assault, Weapons Violations, and Motor Vehicle Theft.

6.8 Crime and Population within Los Angeles

Figure 13 Heatmap of Los Angeles Population

The first ethnic group analyzed is African Americans. The data indicates a higher prevalence of the following crimes within this community: assault with a deadly weapon and aggravated assault, simple battery, burglary, burglary from a vehicle, criminal threats (with no weapon reported), identity theft, and vehicle theft.

The second group examined is the Asian population. Among this group, commonly reported crimes include vehicle theft, simple battery, burglary, assault with a deadly weapon, and vandalism.

Finally, the analysis turns to the Mexican population and the crimes most frequently reported within this community. These include intimate-partner simple assault, theft from motor vehicles, burglary, and identity theft.

6.8 Crime and Population within Los Angeles

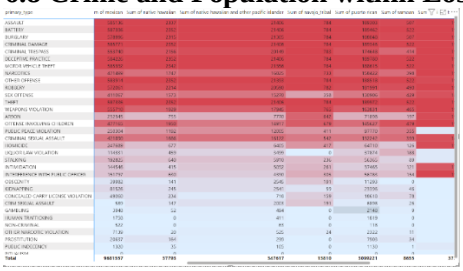

Figure 14 Heat Map of Chicago Population

The first ethnic group analyzed is African Americans. The following highlights a set of prevalent crimes specific to this community. These include assault, battery, criminal damages, deceptive practices, and narcotics.

The second group American Indians under examination reveals a distinct set of commonly reported offenses, such as assault, burglary, criminal trespassing, robbery, and weapons violation

Lastly, the analysis delves into the third ethnic group Cuban population, focusing on the types of crimes most frequently associated with this community. These offenses are battery, theft, criminal damage, assault, and robbery.

7. Conclusion

Ultimately, this study reveals critical post-pandemic crime trends in Chicago and Los Angeles, linking rising crime rates to factors like education, unemployment, income, and population dynamics. Key hotspots, including the Loop and Central City, saw significant escalations, highlighting the need for targeted interventions. Future work with larger datasets and broader scopes could refine these findings and support more effective policy strategies. Visit the project's GitHub[2] for detailed resources.

References

[1] Bureau, U. C. (2021, December 16). *U.S. Census Bureau at a glance*. Census.gov.

[2] K. Hwang, *Computer Arithmetic*, John Wiley, 1997.

[2] GitHub Link: **https://github.com/xuewentang/5200group2/tree/main**