Introduction

The main issue we aim to address in this project is the evolving landscape of crime in Chicago, with a specific focus on how crime trends have changed over time, which communities are most affected, which areas experience the highest crime rates, and how we can use these insights to identify safer neighborhoods within the city. This topic is particularly significant because Chicago's high crime rates remain a concern. Our primary goal is to analyze crime patterns to help individuals, better understand where safety risks are most prevalent and which areas offer greater security. A key aspect of our investigation involves examining how crime affects different demographic groups. By analyzing data based on age, gender, and race, we can identify which groups are most frequently targeted or involved in criminal incidents. This will help us understand the broader social and economic factors contributing to specific trends. Such information is essential for informing residents about areas where they are more likely to face risks and where they can feel safer. By analyzing crime data over the years, we examine how crime disproportionately impacts certain demographic groups and assess crime distribution across different areas, providing insights into the varying levels of crime in each region and highlighting disparities in victimization rates.

Datasets

For our project, we used three datasets. Two obtained from The Home Of The U.S. Government's Open Data (Data.gov) website and one from The Chicago Data Portal. All datasets were downloaded as CSV files for further analysis. The first dataset, "Violence Reduction - Victim Demographics - Aggregated," includes aggregate data on index crimes, such as homicides, battery, and robbery. It also includes the date, time, type location (zip code, block) of the crime, reported to the Chicago Police Department from 1991 to the present (it is updated daily). The data is broken down by quarter and demographic factors such as age, sex, and race. It also contains the type of location where the crime occurred (i.e. park, house, school, etc), and the day of the week it occurred (Monday, Tuesday…). The second dataset, "Crimes - One Year prior to Present," contains more recent crime incidents (over the past year), extracted from the CPD's system, offering detailed crime occurrences at the block level. The second dataset provides us with the date and time of the occurrences, location (block), as well as primary and secondary descriptions. Both datasets also contain the latitude and longitude of the location of the crime. The second dataset, "Crimes - One Year prior to Present," contains more recent crime incidents (over the past year), extracted from the CPD's system, offering detailed crime occurrences at the block level. The second dataset provides us with the date and time of the occurrences, location (block), as well as primary and secondary descriptions. Both datasets also contain the latitude and longitude of the location of the crime. The third dataset, "Boundaries - Neighborhoods", is from the Chicago Data Portal and provides geographic information about neighborhood boundaries in Chicago. It has 98 rows and five columns, giving the primary and secondary name, area, length, and multipolygon of each unofficial neighborhood. A multipolygon is a collection of multiple sets of ordered x, y coordinates "that may be overlapping, disjoint, or nested," (Polygons vs multipolygons in GIS - March 17, 2025). It may be used to represent a "collection of land parcels, where each parcel together makes up a larger area, such as a city," (Polygons vs multipolygons in GIS - March 17, 2025).

These datasets relate to our topic by providing victim demographics, crime types, crime locations, and Chicago neighborhood locations. The Violence Reduction dataset is 61080 rows x 38 columns. The Crimes dataset is 254083 rows x 17 columns. While the first dataset looks at violent crimes from 1991 to present and the second looks at a wider range of crimes over the past year, the two are related by their ability to categorize, describe, and provide the location for the crime and the demographics of the victim. We intend to relate the two using victim demographics and crime types over different periods of time. Descriptions can be categorized using key words since every description is unique. We can also use the date and time to relate the times of day or year crimes occur more frequently. We can also look at what days of the week crime occurs more frequently by using the day of the week column in the "Violence Reduction - Victim Demographics - Aggregated" data set and a function that tells us the day of the week from the date in the "Crimes - One Year prior to Present" dataset. We can analyze each dataset (Violence Reduction and Crimes) separately to examine violent crimes or past year trends. Finally, for location in particular we can relate all three datasets by using numpy and scipy in python to get the centroid of each neighborhood and then the nearest neighborhood for each crime instance (after merging the Violence Reduction and Crime datasets).

Q1: Are specific demographic groups disproportionately impacted by violent crimes in Chicago?

For Question 1, we aim to identify potential patterns between different types of violent crimes and the demographic backgrounds of the victims, specifically focusing on three key demographics: race, age, and sex. To solve this problem, our team extracted information from the "Violence Reduction - Victims of Homicides and Non-Fatal Shootings" dataset, which provided demographic data on victims for crimes violent committed over the years. We were unable to utilize the "Crimes - One Year Prior to Present" dataset, as it did not provide the demographics of victims. Since our CSV file contains other details beyond our focus, such as location information, we need to filter our dataset to include only the relevant variables. Specifically, we will only include the columns: age, sex, race, victimization type (victimization_primary). By narrowing our dataset to these key attributes, we can focus solely on understanding the patterns of victimization based on race, age, and sex.

For the first part of our question, we want to focus on victimization by crime type and sex, where we focus on three specific types of violent crime (HOMICIDE, BATTERY, or ROBBERY), and analyze their distribution by sex (FEMALE OR MALE). To do this, we extract only "sex" and "victimization_primary" columns from the filtered dataset, ensuring that we only focus on relevant variables. Next, we include only records where the crime type is HOMICIDE, BATTERY, or ROBBERY, while also excluding cases where the sex of the victim is labeled as "UNKNOWN" to maintain accuracy in the analysis. Afterwards, we group the remaining data by both crime type and sex, then aggregate the dataset by counting the occurrences of each combination. Finally, we first calculate the total number of victims for each crime type. Next, we compute the percentage of victims for each sex per crime type by dividing the individual counts for female and male by the total victims for that crime type, multiplying by 100 to express it as a percentage. This ultimately shows us how victimization varies with sex.

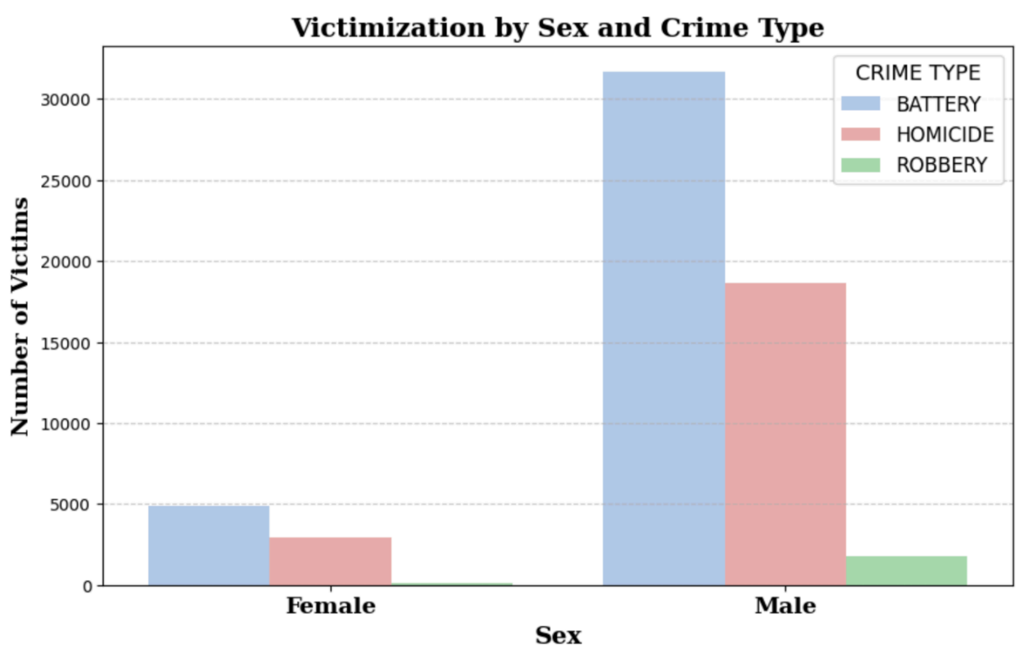

Figure 1.1: *Victimization by Sex (Female and Male) and Crime Type (Battery, Homicide, Robbery)*

From Figure 1.1, we can note that males are significantly more likely to be victims across all three crime types. For example for battery, 86.66% of victims are male, 13.34% are female. Second, for homicides, 86.39% of victims are male, 13.61% are female. Finally, for robbery, 94.36% of victims are male, and 5.64% being female. Initially, this data was surprising, as we assumed women would be more likely to be victims of violent crimes. However, a Professor Richardson from University of Maryland stated that the rate of violent victimization is substantially higher among men than women. One reason for this disparity is that men are often more exposed to violent environments. They are more likely to stay in high-crime areas at night and possibly be joining violent organizations. Additionally, poverty can play a significant role, as economic hardship can force individuals into dangerous situations, increasing exposure to crime. Furthermore, cultural expectations of masculinity encourage assertiveness and competition, which can lead to lower impulse control and higher risk-taking behaviors. In moments of conflict, aggression may become the default response, escalating situations into dangerous and potentially violent encounters. While men experience higher rates of violent crime, certain crimes against women are significantly underreported, particularly domestic violence and sexual assault. For example, rape is considered the most underreported crime, with 63% of sexual assaults not reported to police (NSVRC) due to stigma and societal pressures. Finally, law enforcement may also tend to document crimes that have male victims more thoroughly than other crimes such as rape. A consideration to keep in mind is that we do not know the motivation behind these crimes in the dataset, whether or not geographic proximity played a role in the crime or whether there was a personal conflict beforehand, the insights we have gathered is from external research to help interpret our data. Additionally, our dataset is severely lacking data points. For instance, online sources indicate approximately 7,000 battery cases, 11,000 robberies, and 600 homicides occur annually. However, when multiplied over 33 years, the expected case counts are significantly higher than what our dataset contains, suggesting that robberies and battery cases are notably underrepresented.

For the second part of our question, we want to focus on victimization by crime type and age groups, where we focus on three specific types of violent crime (HOMICIDE, BATTERY, or ROBBERY), and analyze their distribution by age groups (0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+). To do this, we extract only "age" and "victimization_primary" columns from the filtered dataset, ensuring that we only focus on relevant variables. Next, we include only records where the crime type is HOMICIDE, BATTERY, or ROBBERY, while also excluding cases where the age of the victim is labeled as "UNKNOWN" to maintain accuracy in the analysis. Afterwards, we group the remaining data by both crime type and age, then aggregate the dataset by counting the occurrences of each combination. Finally, we first calculate the total number of victims for each crime type. Next, we compute the percentage of victims for each age group per crime type by dividing the individual counts for each age group by the total victims for that crime type, multiplying by 100 to express it as a percentage. This ultimately shows us how victimization varies with age distribution.

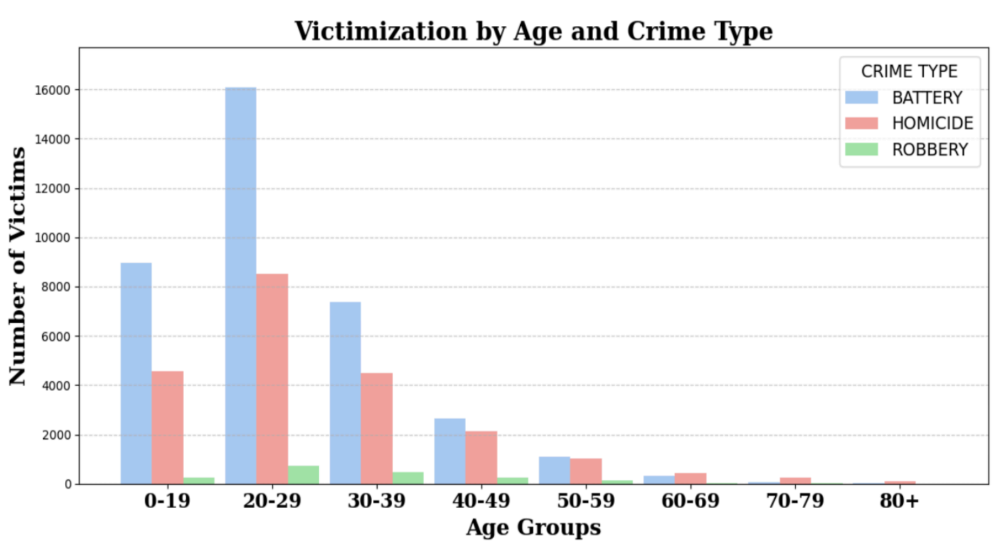

Figure 1.2: *Victimization by Age (0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+) and Crime Type (Battery, Homicide, Robbery)*

After conducting analysis, we have noted that the highest percentage of battery victims are aged 20-29 (44%), followed by 0-19 (24.5%) and 30-39 (20.2%). In Figure 1.2, we can note that the highest percentage of homicide victims are aged 20-29 (39.5%), followed by 30-39 (20.9%) and 0-19 (21.1%). Third, the highest proportion of robbery victims are aged 20-29 (38.2%), followed by 30-39 (24.3%). We also noted that victimization decreases steeply after 40-49 across all three crime types. Our team was surprised to find that individuals aged 20-29, along with 0-19 and 30-39, are the most frequent victims across all three crime types: battery, homicide, and robbery. This pattern could possibly cause a significant prevalence of violence among youth, which is supported by global research. According to the World Health Organization (WHO), youth violence is a major public health issue affecting individuals aged 10–29. This violence takes form in many ways including bullying, physical fights, assault, and gang-related activity, and are often caused by early exposure to violence within their family, low commitment to education, and early involvement. With over 193,000 homicides occurring annually among individuals aged 15–29 worldwide, this really underscores the scope of the issue, and recognizes young people as both perpetrators and victims in crime trends.

For the third part of our question, we want to focus on victimization by crime type and race, where we focus on three specific types of violent crime (HOMICIDE, BATTERY, or ROBBERY), and analyze their distribution by race (Asian/Pacific Islander, Black, Indigenous, Black Hispanic, White, White Hispanic). To do this, we extract only "race" and "victimization_primary" columns from the filtered dataset, ensuring that we only focus on relevant variables. Next, we include only records where the crime type is HOMICIDE, BATTERY, or ROBBERY, while also excluding cases where the race of the victim is labeled as "UNKNOWN" to maintain accuracy in the analysis. Afterwards, we group the remaining data by both crime type and race, then aggregate the dataset by counting the occurrences of each combination. Finally, we first calculate the total number of victims for each crime type. Next, we compute the percentage of victims for each race per crime type by dividing the individual counts for each race by the total victims for that crime type, multiplying by 100 to express it as a percentage. This ultimately shows us how victimization varies with race.

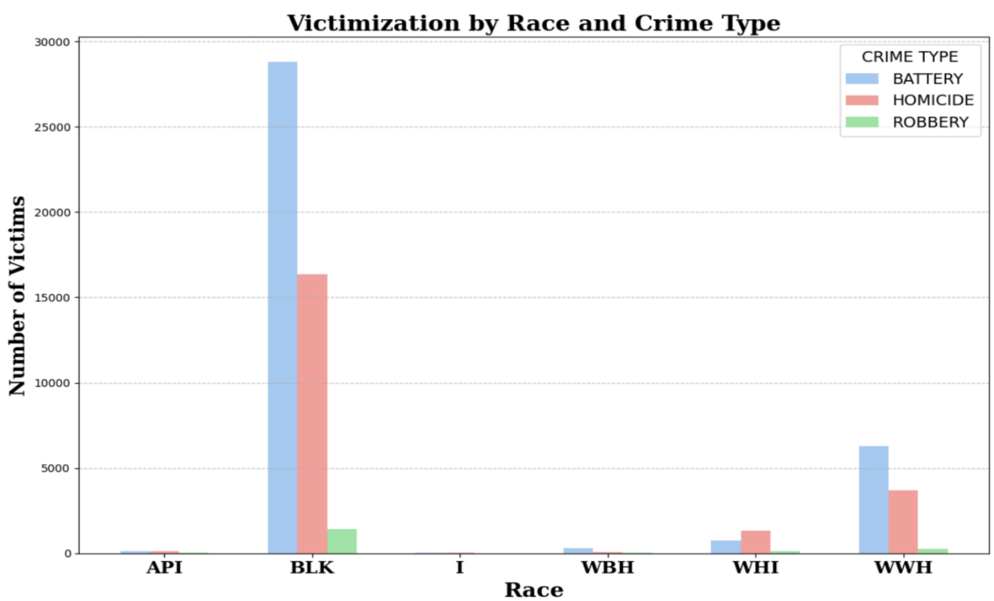

Figure 1.3: *Victimization by Race (Asian/Pacific Islander, Black, Indigenous, Black Hispanic, White, and White Hisspanic) and Crime Types (Battery, Homicide, Robbery)*

After conducting analysis, we have noted that black individuals have the highest victimization rates across all crime categories. In Figure 1.3, it can be noted that 79.49% of battery case victims, 75.88% of homicide cases victims, and 76.28% of robbery case victims are Black. White Hispanic victims are the second most affected group. For example, 17.32% battery case victims, 17.09% homicides case victims, 14.39% robbery case victims are White Hispanics. We also noted that Asian/Pacific Islander, Indigenous, and Black Hispanic victims have significantly lower percentages across all crime types. The data we got could possibly highlight a significant racial disparity in crime victimization, which is likely influenced by socioeconomic factors, systemic inequalities, and disparities in law enforcement practices. Professors from Georgia State University said, “Striking racial gaps, rooted in a legacy of structural racism, have left generations of Black people with disproportionately less wealth and education, lower access to health care, less stable housing and differential exposure to environmental harms (Barry)”. From the Bureau of Justice Statistics, individuals living in households with significantly lower incomes experience violent crime at a rate more than twice as high as those in households earning $75,000 or more. This reflects the broader socio economic challenges faced by many Black communities, which may contribute to higher crime rates in some areas. At the same time, studies suggest that crimes committed in predominantly White communities would be less likely to result in arrests, possibly wanting to maintain a certain public image. Additionally, law enforcement tends to put their focus predominantly on non-White communities. When certain communities experience excessive policing, it can undermine trust in law enforcement, discouraging cooperation with authorities and perpetuating systemic disadvantages that continue to affect these communities today.

Q2: How have violent crime rates in Chicago evolved over the years, and have there been shifts in the demographic backgrounds of victims?

We used the "Violence Reduction - Victims of Homicides and Non-Fatal Shootings" dataset for question 2, as it provided demographic data on victims over the years and showed changes in crime types. To answer the first part of the question—how violent crime rates in Chicago have evolved over the years—we focused on Homicides, Battery, and Robbery, as these crimes had the most data, enabling us to identify more accurate patterns over time. After that, we grouped the data by year and month and counted the number of incidents for each combination of year and sum per each crime type. Then, to capture a smooth trend without the noise of month-to-month fluctuations we used a moving average, which essentially smooths the data by averaging the count of incidents over a window of 10 months, which allows us to see long-term trends more clearly. Then, we plotted a line graph for each crime type in a single figure to observe how each one has changed over the years.

To explore whether there have been shifts in the demographic backgrounds of crime victims over the years, we used a similar approach. We analyzed crime victims by race, sex, and age over time. For each of these categories, we created a new table with 'date' as the first column and the corresponding demographic factor (sex, age, or race) as the second column. For each table we created, we excluded any rows where the corresponding demographic factor was listed as unknown. After that, we started grouping the data and noticed that certain categories lacked sufficient information, so we decided to create new categories. For example, when analyzing the race data, we combined 'Asian Pacific Islander (API)', 'Indigenous (I)', and 'WBH' into a single 'Other' category. Similarly, when analyzing the age data, we grouped the age ranges '40-49', '50-59', '60-69', '70-79', and '80+' into a single '40+' category due to insufficient data for each individual range. Once we had the updated data tables for each demographic factor, we plotted smoothed crime trends over time (grouped by year and month) for each category, creating three separate figures.

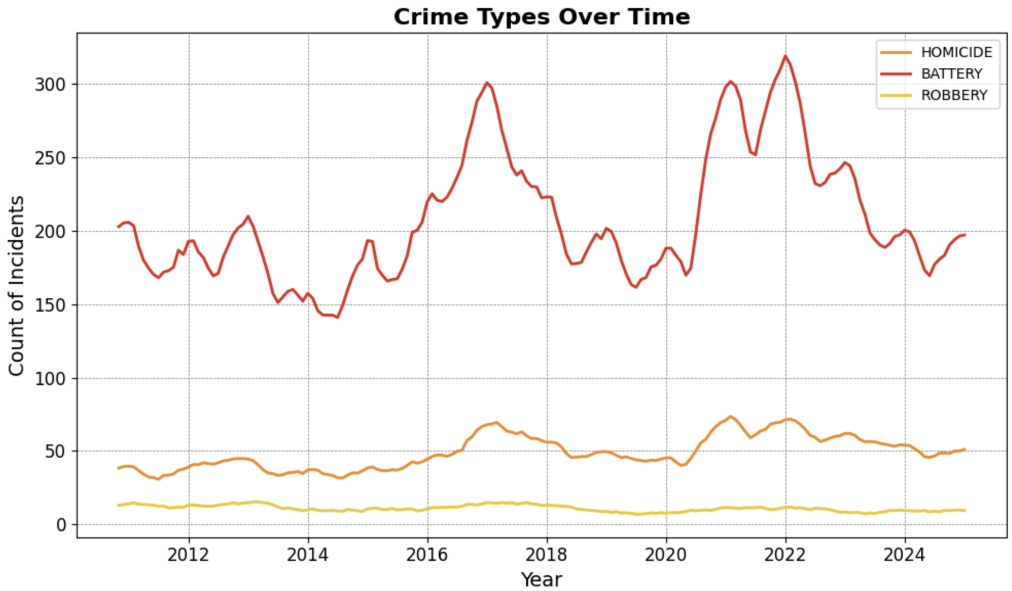

Figure 2.1: *Plot showing the smoothed trends of crime incidents for three types of victimization (Homicide, Battery, and Robbery) over time (1991-2025)*

After conducting our analysis, several key patterns and trends emerged. To start with, we decided to analyze crime types over time. Figure 2.1 clearly shows how, over time, battery is much more common than homicide, and homicide is more common than robbery. We decided to dive deeper and understand why this was the case and concluded that battery includes various forms of physical assault and is less severe than homicide, which could be one of the main reasons why it occurs more frequently. Additionally, many battery incidents occur due to spontaneous disputes between individuals, making them more common than homicides or robberies, as these types of crimes often require more planning or premeditation (Ramos, 2023). Given that Homicide is a more serious and irreversible crime that involves death it statistically happens less often than non-lethal violence. However, our findings show that Homicide is still more common than robbery and this may be because robberies, while frequent, often involve non-lethal violence or theft and may be reported less frequently or under different classifications. Homicide is a more serious crime that requires much more attention from law enforcement, making it a more visible and perhaps more frequently recorded event, even though it may be less common. One reason why battery may be more common than robbery is because in the case of multiple offenses, the incident is classified based on the most serious offense (Chicago Police Department, 2017).

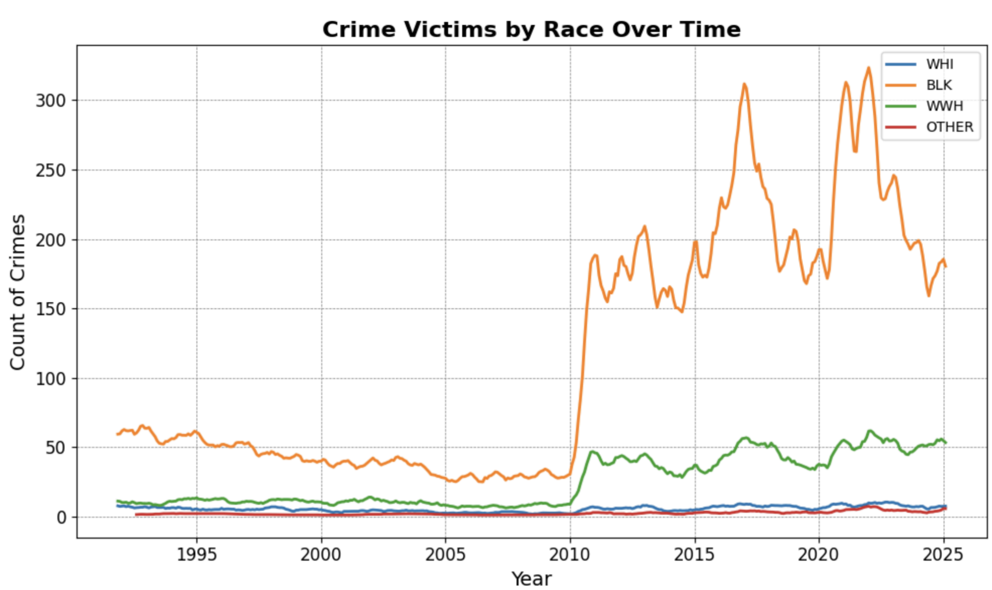

Figure 2.2: *Plot showing smoothed trends of crime over time (1991-2025) by different races (White, Black, White & Hispanic, and Other)*

To answer the second part of question 2, we began by analyzing how crime has evolved over time for each racial group. Figure 2.2 shows a sudden increase in the number of crimes around 2010. We concluded that this may have occurred due to improvements in the Chicago Police Department's (CPD) crime reporting and data transparency. Around the early 2011, the CPD implemented CompStat, a system that likely helped improve how crimes were tracked and reported. Additionally, with advancements in technology, the CPD may have enhanced its database systems for recording incidents and improved digital reporting platforms, making it easier for officers to log and track incidents (Chicago Police Department, 2017). By looking at Figure 2.2, we were also interested in understanding why Black victims (BLK) were the most common, followed by White Hispanics (WWH). First, we concluded that this may be due to the demographic distribution in Chicago. We discovered that, after White people, who account for 32.7% of the population, Black people make up 28.4% of Chicago's population (DataUSA, 2022), which could explain why Black individuals are commonly targeted as victims. However, this seemed counterintuitive, as White people, the largest demographic, had one of the lowest crime counts. We then considered other factors contributing to higher victimization rates among Black individuals. The first factor was historical segregation and economic disinvestment, which have concentrated Black communities in under-resourced neighborhoods, leading to higher crime rates. Additionally, this economic segregation increases inequality and potentially unemployment rates, which may also raise the likelihood of crime in these areas (Sweeney & Fry, 2021). The black poverty rate in Chicago (27.7%) is above the city's average (16.9%) (Coffey, 2024). Lastly, discrimination and media portrayals of violence in Black neighborhoods, often reinforcing negative stereotypes, may influence public policy and further marginalize these communities, ultimately making black people more vulnerable to crime.

White Hispanics (WWH), who make up around 10% of Chicago's population (DataUSA, 2022), were the second most common group of crime victims. Similar to Black individuals, Hispanics face higher poverty rates (16.1% of Hispanics in poverty) compared to White Chicagoans (8.5%) (Coffey, 2024), largely due to living in neighborhoods with limited resources. This results in high unemployment and economic disparity, making them more vulnerable and exposed to being targeted by crime. One reason Hispanics may have a lower crime count than Black individuals is not only demographic distribution but also immigration and language barriers. Many Hispanics in Chicago are undocumented, which may discourage them from reporting crimes to the police due to fear of deportation (Theodore, 2013). Finally, regarding the "Other" racial categories, they likely have the lowest crime count due to the small proportion of the Chicago population they represent.

Figure 2.3: *Plot showing smoothed trends of crime over time (1991-2025) by sex (Female and Male)*

We were curious whether men or women were more targeted by crime, and Figure 2.3 clearly revealed that men have consistently been more victimized over time. To understand if this was due to demographics, we found that women make up 51.5% of Chicago's population, while men account for 48.5% (Neilsberg Research, 2025). This clearly shows that demographic distribution was not the main cause of the situation. Therefore, as mentioned earlier, we concluded that one reason may be that men are statistically more likely to engage in risky behaviors, such as criminal activities or involvement in street life, which ultimately makes them more vulnerable to becoming victims of crime (Harris & Jenkins, 2023). Another reason could be that certain neighborhoods in Chicago are known for gang violence and criminal activity. As a result, these areas may experience higher rates of victimization among young men, particularly those who are seen as potential recruits or rivals in gang-related conflicts. Another potential reason for underreporting of crimes by women is the fear of retaliation. According to data from the Bureau of Justice Statistics, a significant number of simple assaults (often classified as battery) and homicides are committed by someone the victim knows. In the United States, approximately 1 in 4 women will experience violence at the hands of a partner, yet around 44% of domestic violence cases go unreported (Connections for Abused Women and their Children, 2023). Many women fear retaliation when reporting crimes, particularly sexual assault or domestic violence. Additionally, cultural norms and a lack of trust in the justice system may discourage them from seeking help, which could explain why women appear to be less victimized by crime than men.

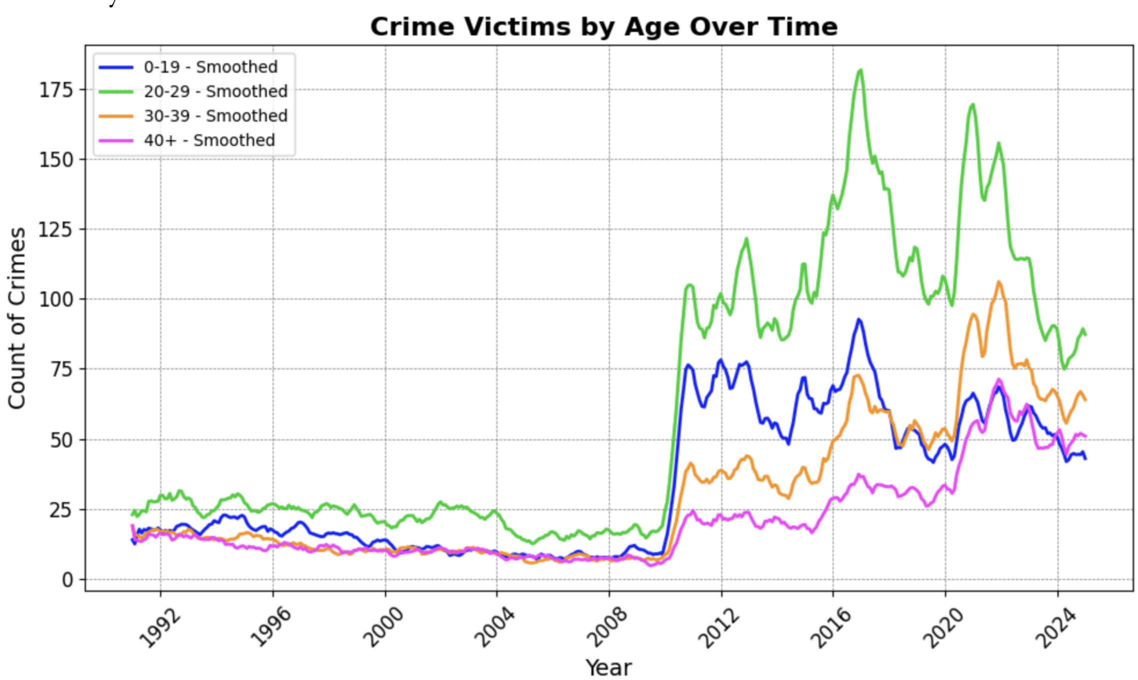

Figure 2.4: *Plot showing smoothed trends of crime over time (1991-2025) by age groups*

We were also very curious to understand if there was a potential relationship between crime victims and age groups. Figure 2.4 clearly shows that over time, the age group of 20-29 years old has been the most targeted by crime. To determine whether this was due to the population distribution in Chicago, we decided to investigate further. We discovered that the age group 0-19 accounted for 22.18% of the population, 20-29 accounted for 17.14%, 30-39 accounted for 17.38%, and those aged 40+ accounted for 42.68% (Neilsberg Research, 2025b). This shows that the 20-29 age group accounts for a relatively large portion of the population. According to the

2019-2023 American Community Survey, the largest age group in Chicago, IL, was those aged 25 to 29. Therefore this may be one of the reasons why this age group is the most targeted by crime. Another main reason may be because people in their 20s tend to be more vulnerable to crime because of immaturity, active social lives, and therefore more exposure to risky situations. (Wojciechowski, 2024).

Another particularly interesting trend we notice was that the age group form 0-19 years old historically was the second most targeted age group but now it's the least targeted age group. We were curious to understand this trend. After conducting some research, we discovered that over the years, CPD has implemented various efforts to reduce violence among this age group. For example, they developed strategic policing initiatives to quickly respond to violence and disrupt violent social networks. One key initiative was the expansion of the Gang School Safety Teams (GSST), a collaborative effort between CPD and Chicago Public Schools to monitor gang conflicts within schools and on social media, and proactively work to de-escalate potential violence. This has proven to be an effective approach, expanding into more than 35 schools and helping to reduce youth violence. (Emanuel, 2013). Another successful program is Chicago CRED (Create Real Economic Destiny), which focuses on reducing gun violence by engaging individuals at high risk of involvement in crime. The program provides education and coaching and has led to a 50% reduction in gunshot victimization. By offering support and opportunities for a better life, CRED has proven to be an essential tool in reducing youth violence and improving the future for Chicago's most vulnerable communities (Kulke, 2023). Furthermore, another notable trend we observed is the increase in victims aged 40 and older over recent years. This may be partly explained by the large proportion of individuals over 40 in the Chicago population, as well as other social factors. Recently, offenders have been targeting victims aged 60-80 due to their vulnerability and ease of attack. Particularly in Chicago’s South side the elderly have been victims of a series of armed robberies in Armour Square near Chinatown (Sullivan, 2024).

Finally, another notable trend observed across all the graphs is the peaks in crime counts during 2016-2017 and 2020-2023. The first spike in 2016 may have been linked to the decline in "stop and frisk" practices by the CPD following December 2015. According to the University of Utah, there is a strong correlation between the reduction in stop and frisk and the rise in homicides and shootings, particularly in African-American and Hispanic communities (Law et al., 2018; Park, 2018). While we did not find direct data explaining the peak in crimes from 2020-2023, we believe this increase may have been influenced by the effects of COVID-19 and the rise in unemployment, which could have contributed to more violence and crime.

Q3: How often do certain crimes occur across different locations and times of the day, week, and year in Chicago?

To answer the question of how often do certain crimes occur across different geographic locations, times of the day, days of the week, and months of the year in Chicago, we used data from all three datasets. We started by merging the information from our Violence Reduction and Crimes dataset. We separated violent crimes from nonviolent crimes so that we could visualize our findings easier. Since we wanted to analyze how crime frequencies changed over times of the day, days of the week, and months of the year, we grouped data from both our merged dataset in that manner for each corresponding part. After doing so, we calculated the frequency and percentages of each crime type across the times of the day, days of the week, and months of the year for each category (violent and nonviolent). We used the frequencies and percentages to create stacked bar plots and stacked percentage line plots of the crime types over the specified time periods.

To analyze how often crimes occurred in different geographic locations in Chicago, we used both our merged dataset and our Boundaries dataset. We wanted to see which neighborhood had the most occurrences of crimes overall.. We found the centroid (central latitude and longitude point) of each neighborhood in Chicago, then we assigned each reported crime in the merged dataset to its nearest neighborhood.

To carry out the methods outlined above, we began by loading the Violence Reduction and Crimes datasets as separate dataframes into Google Colab and cleaning them. We renamed columns for simplicity and casted the columns to their proper types (for example, we casted our 'date' columns for both dataframes to be of type datetime. We then used the 'date' column to create three more columns for each dataframe ('year','month','time_of_crime', and 'day_of_week') to conduct our analyses. We then ensured that both data frames had a column describing the crime reported. For our Crimes dataframe, we created one column that concatenated both the primary and secondary crime descriptions. For our Violence Reduction dataframe, we filled missing fbi descriptions with the text 'NON-FATAL' shooting because we discovered from looking at the data that all the missing descriptions were from crimes that involved gunshot injuries, but were labeled as 'NON-FATAL'. Since both data frames contained a unique case number column, we merged the cleaned crime and violence dataframes on the 'case_number' column using a full outer join to retain unmatched cases from both datasets. We also added suffixes to the column names to distinguish the source of each data field (from 'crimes' or 'violence') due to matching column names. After doing so, we began making columns unique to this merged dataframe that we will refer to as 'matching_cases'. The columns are as follows: date, month, day of the week, year, time of crime, latitude, longitude, block, location description. To create these columns, we used our crimes dataset as our primary source and filled missing values with information from our violence dataset (because our crimes dataset was more descriptive). Finally, we dropped all other columns of matching_cases.

We created a new dataframe called 'cases_loc' so that we could work on matching_cases without erasing the original, cleaned data. We dropped all rows with missing crime descriptions from cases_loc so that we could create a new column called 'general_crime_description.' We created a function that used regex patterns to categorize the crime descriptions, and we applied this function to the description column of cases_loc. The regex patterns and subsequent general_descriptions were chosen by generalizing what was often seen in the location descriptions and continuously checking to see what type of crimes were being casted as 'Other' to update our general descriptions. We then used our unique general crime descriptions to create a list for violent crimes and nonviolent crimes based on the general descriptions.

We started by plotting crime frequencies over the days of the week. We first created a dataframe called crime_counts by grouping cases_loc by the day of the week and then the crime description. We filtered for our violent crimes using our list. Then we created a pivot table with the day of the week as our index, our crime descriptions as our columns and the crime counts as the values. This allowed us to create a stacked bar chart showing distribution of violent crimes across the days of the week, highlighting any patterns in crime occurrence throughout the week. We did the same for nonviolent crimes by filtering cases_loc using our nonviolent crimes list. To make our stacked percentage line plots, we converted our counts pivot tables to a percentages pivot table (the percentages for the particular day of the week) and plotted it for both violent and nonviolent crimes. To plot crime frequencies and percentages over the months of the year, we followed the same structure, but we grouped cases_loc by month first instead of by day of the week.

By plotting violent and nonviolent crime distributions across the days of the week, we find that there is not much variation for violent or nonviolent crimes. For nonviolent crimes, the counts stay pretty consistent, around 25,000 per day. For violent crimes, we see a slight increase of roughly 3,200-4,000 extra cases on Saturday and Sunday. Other than that, it stays around 18,100 cases throughout the weekdays. The violent crimes are overwhelmingly categorized by battery and violence, with assault and homicide in a relatively far second. Nonviolent crimes are led by theft, with property damage and vandalism in far second. The rest in each category are small enough to consider negligible proportions wise.

Plotting crime distributions across the months of the year however revealed some profound trends. In Figure 3.1, we see that crime occurs the least in the winter months of November through February. Crime then increases until it peaks in the summer months of June and July, then decreases until November again. We think one explanation for this may be that people are committing less crimes during the winter because they are staying inside more since it is too cold. The average temperature during Chicago's cold season is below 43 degrees fahrenheit, while during the warmer months the average high is 73 degrees fahrenheit (Weatherspark.com). It is also worthwhile to note that battery and violence is consistently the most frequent violent crime type.

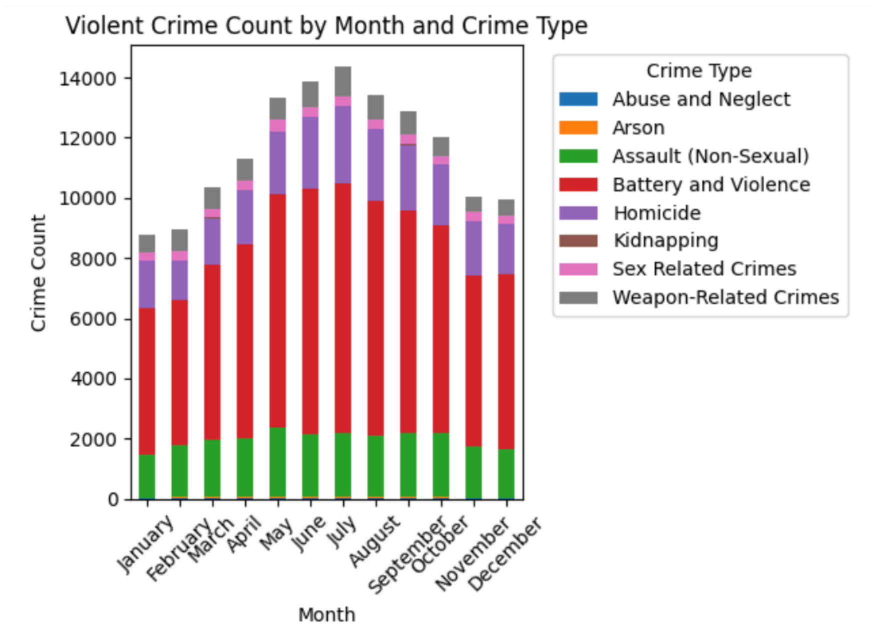

Figure 3.1: *Stacked Bar graph showing the number of crime incidents for violent crimes over months of the year (1991-2025)*

For nonviolent crimes, there is a lot less variation in crime counts across the months. Generally the data stays around 15,000 cases. But, it follows the trend of the lower months being the colder months. Like violent crimes, the month with the least cases is January, the coldest month of the year in Chicago (Weatherspark.com). The cases still peak in July as well, the warmest month in Chicago (*Weatherspark.com*).

To plot crime frequencies and percentages over different times throughout the day, we started by group the day into four six-hour sections: overnight (12:00 AM - 5:59 AM), morning (6:00 AM - 11:59 AM), afternoon (12:00 PM - 5:59 PM), and evening (6:00 PM - 11:59 PM). We extracted the hour of occurrence from our crimes_loc dataframe and made 'hour' a new column in cases_loc. We then created a function to apply these categories to the 'hour' column and create a new 'section_of_day' column. Finally, we grouped cases_loc by the section of day and general crime description and created a pivot table based on those counts. We used these counts to create a stacked bar chart for both violent and nonviolent crimes. To create our stacked percentage line plots, we converted the pivot table of counts into a pivot table of percentages for both violent and nonviolent crimes by summing across the table values and dividing each value by the total sum of that row. This gave us the proportion of each crime type relative to the total number of crimes in each time section.

By plotting crime distributions over sections of the day, we discover noticable differences as well. As seen below in Figure 3.2, violent crimes are most frequent in the evening and least in the morning. With afternoon and differing sparingly. We continue the pattern of battery and violence being the most frequent crime category, with assault overtaking battery and violence as second place in the afternoon and morning.

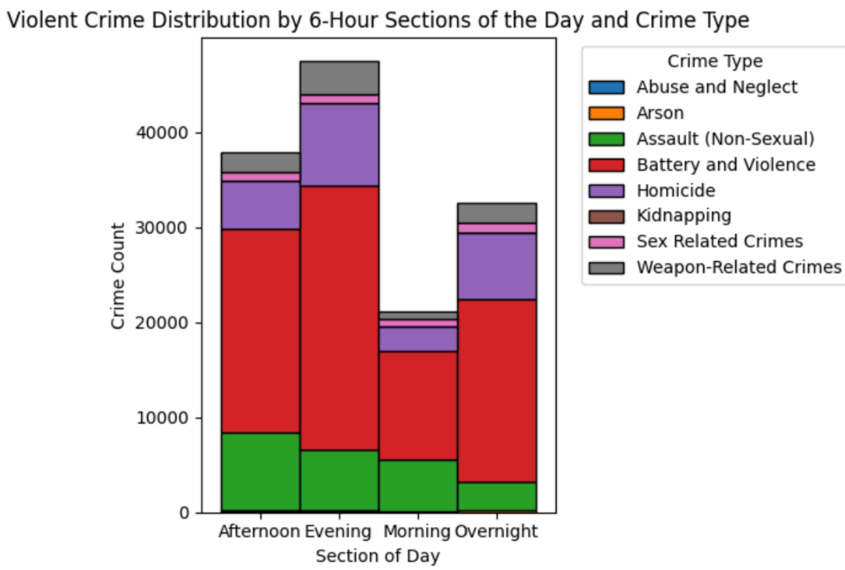

Figure 3.2: *Stacked bar graph showing the number of crime incidents for violent crimes over sections of the day (1991-2025)*

In Figure 3.3, we see that nonviolent crimes are at its height in the afternoon, with the bar graph showing a decrease as we move across evening , morning, and overnight. There is a negligible difference between morning and overnight. But, the least amount of crime occurs overnight. This makes sense that there would be less activity, since people are sleeping. From the data, we see a noticeable increase in the percentage of property damage and vandalism overnight. This makes sense as well since it is easier to go unnoticed and commit these crimes under the cover of night.

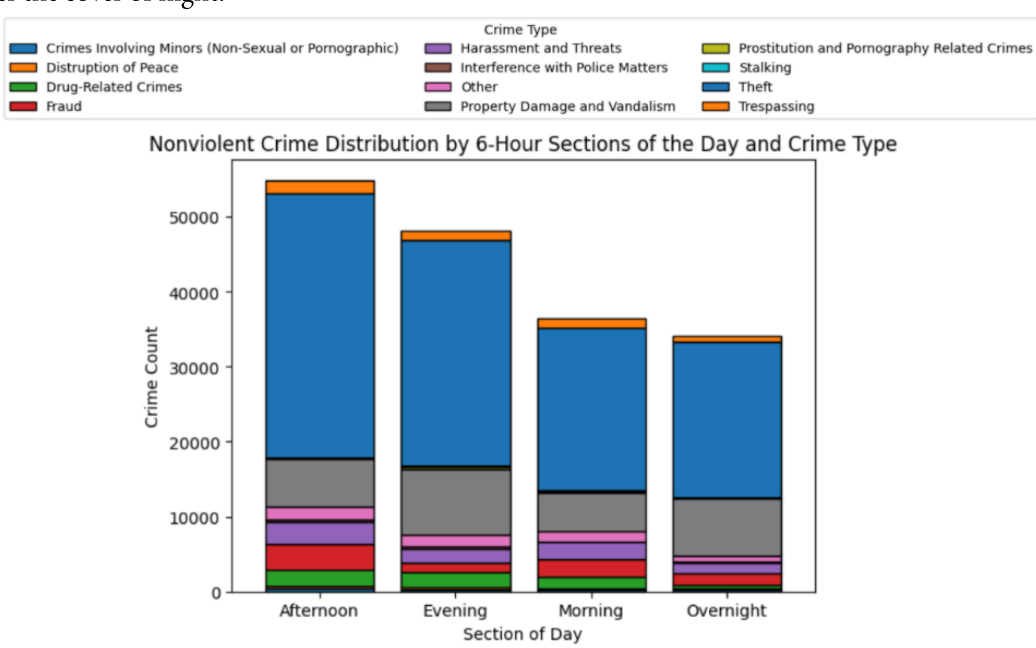

Figure 3.3: *Stacked bar graph showing the number of crime incidents for nonviolent crimes over sections of the day (1991-2025)*

For our final section of analysis, we looked at the frequency of crimes in different geographic locations in Chicago. We loaded our Boundaries dataset and renamed the columns. After dropping unneeded columns we were left with two columns: 'coordinates' and 'primary_name'. The primary name refers to the name of the neighborhood; it is simply text. However, 'coordinates' is really the geometric object known as a 'MultiPolygon'. We converted our coordinate column from a string to a proper MultiPolygon using shapely. Then we used the centroid function from shapely and applied it to our coordinates column to create a new centroid column for every neighborhood in Chicago. We then created columns for both the x and y coordinates of the centroid. Before proceeding, we redefined cases_loc to be our original matching_cases data and deleted entries with missing longitude and latitude coordinates since we would not be able to locate them geographically. This deleted 983 rows, or 0.31% of our rows. We then created a function to find the closest neighborhood for every reported crime and applied the function to our longitude and latitude values in cases_loc. We placed this in a new column called 'nearest_neighborhood'. To find the frequency of overall crimes in each neighborhood, we grouped by the nearest neighborhood to get the counts using size. We placed these counts in a dictionary with the neighborhood name as the key and the counts as the value. Then, we applied this dictionary to our Boundaries dataframe to create a new 'crime_count" column. Using plotly, we created an interactive map by plotting these centroids over a map of Chicago , with the color and size of each point corresponding to the frequency of overall crimes.

By plotting the counts for every neighborhood, we were able to visualize the counts by proportion in Figure 3.4 . We found that the top ten neighborhoods in Chicago with the highest crime counts are as follows: Austin with 15,677 crimes, Englewood with 12,070 crimes, Garfield Park with 10,441 crimes, North Lawndale with 10,071 crimes, Auburn Gresham with 9,722 crimes, Humboldt Park with 8,746 crimes, South Shore with 8,265 crimes, Grand Crossing with 7,798 crimes, Chicago Lawn with 6,696 crimes, and Chatham with 6,339 crimes. According to an article published in January 2024, the top ten neighborhoods with the highest crime are as follows: Washington Park, West Garfield Park, Englewood, North Lawndale, East Garfield Park, West Englewood, Austin, Grand Crossing, Riverdale, South Shore (FOX 32 Chicago). The order of our top ten list as well as the inclusion of Auburn Gresham, Humboldt Park, Chicago Lawn, and Chatham. It is important to recognize that since we found the closest neighborhood (via centroid) to each crime, it is possible that crimes occurring in one neighborhood were closer to the centroid of a different neighborhood. Additionally, while there are over 200 neighborhoods in Chicago, our Boundaries dataset only named 98, meaning some were grouped together or omitted (Chicago's neighborhoods).

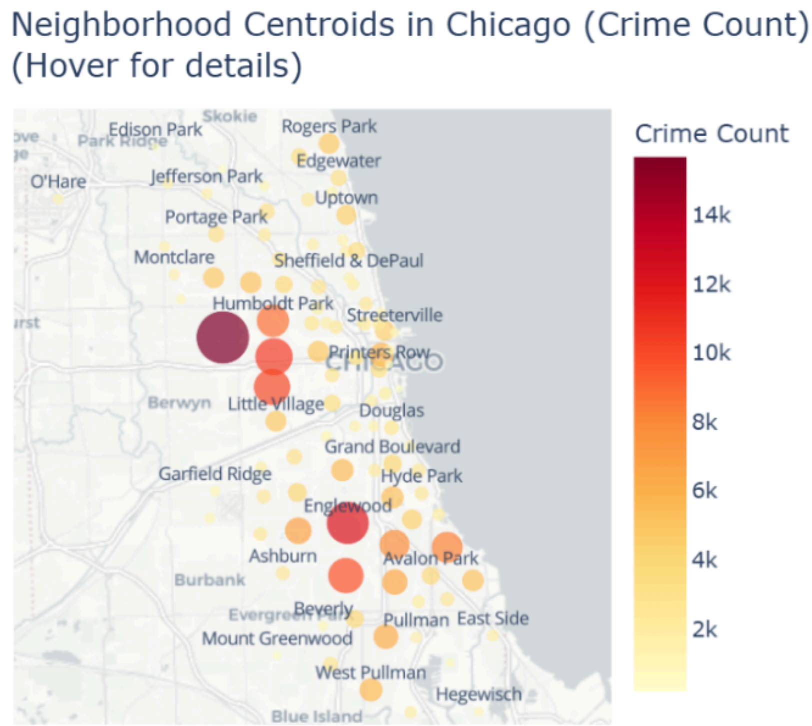

Figure 3.4: *Plotly map showing the number of crime incidents for data from 1991-2025 across the neighborhoods of Chicago (size and color relative to crime frequency).*

Distribution

For this project, since we were three individuals, we decided to divide the work by question. Elisa Duan was responsible for the coding, graphs, and analysis for Question 1. Mariana Prado handled Question 2, while Chelsea Vital took charge of Question 3. Afterward, we met to work on the introduction, dataset information, summary, and compiled the final report.

Summary

Our analysis highlights the evolving landscape of violent crime in Chicago. For Question 1, we observed the disproportionate impact of violent crime on men, as well as Black and Hispanic individuals, particularly those in the 20-29 age group being the most affected. This is likely due to their heightened social activity and exposure to risk.

For Question 2, our analysis revealed that battery was the most common crime, followed by homicide and robbery. Demographic trends showed higher victimization rates among Black and Hispanic individuals. Over time, crime trends also indicated increased victimization of men and those in the 20-29 age group, with a noticeable rise in crime against older adults and a decrease in violent crime among younger individuals. We also identified significant crime spikes in 2016-2017 and 2020-2023, which were likely linked to changes in police practices and the socioeconomic impacts of the COVID-19 pandemic.

For Question 3, with the merged datasets, we found that violent crime instances were slightly higher on weekends and in the evenings, while nonviolent crimes remained steady throughout the week. Crime trends also showed a seasonal peak in the summer, with a dip in the winter. Notably, two neighborhoods—Austin and Englewood—had the highest crime rates, with battery and violence being the most common violent crimes, and theft being the leading nonviolent offense overall.

In the future, we aim to strengthen our analysis by exploring correlations between various demographic combinations. For example, we can explore the correlation of specific races and age groups, sex and age, and race and sexes being targeted by crime. By examining these nuanced intersections, we hope to uncover more detailed insights into how different demographic groups are impacted by violent crime. Additionally, we think it could be beneficial to do predictive modeling, anticipating future crime patterns based on race, sex, and age. This predictive approach could provide valuable insights and help inform proactive strategies for crime prevention. Furthermore, regarding our analysis of crime locations in question three, we aim to improve our methodology by moving beyond centroid-based approximations of neighborhood crime. Instead, we hope to explore location-specific data within neighborhoods to better understand crime distribution patterns and pinpoint high-risk areas with greater accuracy. We can provide a more precise representation of crime hotspots and enhance our understanding of the spatial dynamics of crime in Chicago.

References

1. Barry, Celeste. 2023. 'One in Five: Disparities in Crime and Policing – the Sentencing Project', *The Sentencing Project* <https://www.sentencingproject.org/reports/one-in-five-disparities-in-crime-and-policing/> Accessed 10 Mar. 2025.
2. "Boundaries - Neighborhoods: City of Chicago: Data Portal." Chicago Data Portal, data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9. Accessed 17 Mar. 2025.
3. Chicago Police Department. "Statistical Reports – Chicago Police Department." *Chicagopolice.org*, 2017, www.chicagopolice.org/statistics-data/statistical-reports/. Accessed 10 Mar. 2025.
4. "Chicago's Neighborhoods." Chicago Studies | The University of Chicago, chicagostudies.uchicago.edu/neighborhoods. Accessed 17 Mar. 2025.
5. City of Chicago. "Violence Reduction - Victim Demographics - Aggregated | City of Chicago | Data Portal." *Data.cityofchicago.org*, City of Chicago, 21 May 2021, data.cityofchicago.org/Public-Safety/Violence-Reduction-Victim-Demographics-Aggregated/gj7a-742p/about_data. Accessed 4 Feb. 2025.
6. Coffey, Chris. "Poverty Hits Black, Hispanic Chicagoans Hardest, but Education, Jobs Fix That." *Illinois Policy*, 11 June 2024, www.illinoispolicy.org/poverty-hits-black-hispanic-chicagoans-hardest-but-education-jobs-fix-that/. Accessed 10 Mar. 2025.
7. Connections for Abused Women and their Children. "Stigma and Domestic Violence." *CAWC*, 27 Feb. 2023, www.cawc.org/news/stigma-and-domestic-violence/. Accessed 10 Mar. 2025.
8. DataUSA. "Chicago, IL | Data USA." *Datausa.io*, 2022, datausa.io/profile/geo/chicago-il. Accessed 13 Mar. 2025.
9. Emanuel, Mayor. *CHICAGO'S KEY INITIATIVES to REDUCE YOUTH VIOLENCE a Review of Investments and Results*. 2013. Accessed 10 Mar. 2025.
10. FOX 32 Chicago. "These Are Chicago's 10 Most Dangerous Neighborhoods, According to Propertyclub." FOX 32 Chicago, 3 Jan. 2024, www.fox32chicago.com/news/chicago-most-dangerous-neighborhoods. Accessed 17 Mar. 2025.
11. Harris, Christine R., and Michael Jenkins. "Gender Differences in Risk Assessment: Why Do Women Take Fewer Risks than Men?" *Judgment and Decision Making*, vol. 1, no. 1, 1 Jan. 2023, pp. 48–63, www.cambridge.org/core/journals/judgment-and-decision-making/article/gender-differences-in-risk-assessment-why-do-women-take-fewer-risks-than-men/3386EA020D940A2805EA3785662E7832, https://doi.org/10.1017/S1930297500000346. Accessed 10 Mar. 2025.
12. Kulke, Stephanie. "Chicago Community Violence Intervention Program Shown to Reduce Gun Violence." *News.northwestern.edu*, 6 Nov. 2023, news.northwestern.edu/stories/2023/11/chicago-community-violence-intervention-program-shown-to-reduce-gun-violence/. Accessed 10 Mar. 2025.
13. Law, Utah, et al. *What Caused the 2016 Chicago Homicide Spike? An Empirical Examination of the ' ACLU Effect' and the Role of Stop and Frisks in Preventing Gun Violence Recommended Citation*. 2018. Accessed 17 Mar. 2025.
14. Neilsberg Research. "Chicago, IL Population by Age - 2023 Chicago, IL Age Demographics | Neilsberg." *Www.neilsberg.com*, 22 Feb. 2025, www.neilsberg.com/insights/chicago-il-population-by-age/. Accessed 13 Mar. 2025.

15. Neilsberg Research. “Chicago, IL Population by Gender - 2025 Update | Neilsberg.” Neilsberg.com, 2025, www.neilsberg.com/insights/chicago-il-population-by-gender/. Accessed 9 Mar. 2025.
16. Park, Madison . “Chicago Police Count Fewer Murders in 2017, but 650 People Were Killed.” CNN, 1 Jan. 2018, www.cnn.com/2018/01/01/us/chicago-murders-2017-statistics/index.html. Accessed 17 Mar. 2025.
17. Polygons vs Multipolygons in GIS - March 17, 2025.” Mapscaping.Com, 23 Jan. 2023, mapscaping.com/polygons-vs-multipolygons-in-gis/. Accessed 17 Mar. 2025.
18. Ramos, Elliott. “Here’s What’s Happening with Crime in Chicago in 2023 - CBS Chicago.” Www.cbsnews.com, 29 Dec. 2023, www.cbsnews.com/chicago/news/heres-whats-happening-with-crime-in-chicago-in-2023/. Accessed 17 Mar. 2025.
19. Richardson, Joseph, et al. *Male Victims of Violence*. 2017. Accessed 11 Mar. 2025.
20. Statistics about Sexual Violence, www.nsvrc.org/sites/default/files/publications_nsvrc_factsheet_media-packet_statistics-about-sexual-violence_0.pdf. Accessed 11 Mar. 2025.
21. Sullivan, Richard M. “Chicago Seniors Targeted in Recent Wave of Robberies; Police Urge Vigilance in Deering District.” *Hoodline*, 22 Nov. 2024, hoodline.com/2024/11/chicago-seniors-targeted-in-recent-wave-of-robberies-police-urge-vigilance-in-9th-district/. Accessed 9 Mar. 2025.
22. Sweeney, Annie, and Paige Fry. “In 2021, Violence Spread into Chicago Neighborhoods That Normally Enjoy a Sense of Security, Rattling Nerves and Putting City Leaders on the Defensive.” *Chicago Tribune*, 23 Dec. 2021, www.chicagotribune.com/2021/12/23/in-2021-violence-spread-into-chicago-neighborhoods-that-normally-enjoy-a-sense-of-security-rattling-nerves-and-putting-city-leaders-on-the-defensive/. Accessed 11 Mar. 2025.
23. Theodore, Nik. *Insecure Communities: Latino Perceptions of Police Involvement in Immigration Enforcement*. May 2013. Accessed 11 Mar. 2025.
24. Thompson, Alexandra, and Susannah Tapp. 2023. *Criminal Victimization, 2022*, *Bureau of Justice Statistics* <https://bjs.ojp.gov/document/cv22.pdf>. Accessed 13 Mar. 2025.
25. USAFacts. “Data Says Domestic Violence Incidents Are Down, but Half of All Victims Don’t Report to Police.” USAFacts, 6 Oct. 2021, usafacts.org/articles/data-says-domestic-violence-incidents-are-down-but-half-of-all-victims-dont-report-to-police/. Accessed 10 Mar. 2025.
26. “Weatherspark.Com.” Chicago Climate, Weather By Month, Average Temperature (Illinois, United States) - Weather Spark, weatherspark.com/y/14091/Average-Weather-in-Chicago-Illinois-United-States-Year-Round#google_vignette. Accessed 17 Mar. 2025.
27. Wojciechowski, Charlie. “Violence in Chicago Has Changed. An Outreach Group Explains Why and What Needs to Be Done.” *NBC Chicago*, NBC Chicago, 5 July 2024, www.nbcchicago.com/news/local/chicago-violence-changes-gangs-shootings-police/3482134/. Accessed 11 Mar. 2025.
28. World. “Youth Violence.” *Who.int*, World Health Organization: WHO, 31 Oct. 2024, www.who.int/news-room/fact-sheets/detail/youth-violence. Accessed 10 Mar. 2025.