Max de Saint-Exupéry, Charitha Vennapusa, Maya Ghosal
Professor Trimble
Introduction to Data Science II
May 13th, 2022

How Crime Rates in Chicago differ over time, place and type

Brief description of the dataset:

For this project, we selected the Crimes dataset (2001 - Present) from the Chicago Data Portal as our foundation of study. This dataset was extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) map, and includes 22 fields for every crime recorded by Chicago PD since 2001, including the type of crime; a more specific description of each crime; location of crime by coordinates, street address, block and civilian district; time the crime was recorded to have occured (both in terms of date and time of day); whether the crime occured within someone's home and whether or not the perpetrator(s) was arrested. Given that the complete dataset, which is updated on a daily basis, covers criminal activity over the last 21 years, we chose to focus our study more by filtering the dataset based on the date and the year to only include all records from March 1st, 2019 to March 31st, 2022, which covered a total of 688,825 crimes recorded over the last three years. This gave us the opportunity to analyze how crime rates have changed with COVID since the dataset includes both pre-Covid and post-Covid years, as well as examining subjects of interest such as when and where crime rates are at their highest and lowest, the frequency of certain crimes throughout the day, and the likelihood of arrest depending on the location of the crime.

Why might we care about this dataset? What is its significance?

Chicago is a city renowned for its immensely high crime levels, with its March 2022 crime rate of 3,926 per 100,000 people being over 67% higher than the national average (Get Jerry). Given the city's reputation, analyzing the CLEAR dataset is important in understanding the causes of crime in the city, which can lead to formulating effective solutions. Finding correlations between crime and time or location can help police departments allocate resources better to combat crime. For example some parts of Chicago have crime levels similar to that of relatively safer cities such as New York, while the seven most dangerous police districts have homicide rates around 100 per 100,000 residents, falling just below those of the most dangerous cities on Earth (US 99). Thus, the data can help determine which communities need more police reinforcement or more welfare programs implemented. Another reason why analyzing crime datasets is so important is because police departments can see whether their previously implemented programs have led to a decrease in crime or not.

How crime rates have changed over the last 3 years (graph A):

In regards to graph A, we can observe an overall trend of crime increasing in the spring and summer and decreasing in the fall and winter, with said fluctuations likely being linked to changes in temperature throughout the year. This is because warmer weather not only encourages people to move around but also aggravates humans, thus increasing the chance of irrational criminal activity. This is supported by many studies that have shown that increasing temperatures have led to rises in crime rates. For example, one study states that "the relationship between temperature and crime became linear…[crime] increased steadily with ambient temperature up into the mid-90s" (British Journal of Criminology). A more recent study conducted by the IOP claimed that "we expect an increase in violent crime of 1.8%, 3%, 5%, and 7% for a 1.5 °C, 2 °C, 3 °C, or 4 °C warmer world with greater warming also leading to greater projection uncertainty" (IOP Science).

From the graph, we can also observe that crime rates drastically dropped in April 2020 and February 2021. Further research indicates that the drop in April 2020 was largely driven by the state-wide stay-at-home order enforced at the time in the early months of the COVID-19 pandemic,with NPR reporting that crime decreased overall during the pandemic compared to previous years because people had stopped working, traveling, and doing other sorts of economic activities (NPR). Similarly, NBC Chicago reported that public crime was down, stating that "the 30% drop in overall crime [came] largely from double-digit reductions in criminal sexual assaults, robberies, aggravated batteries, burglaries, thefts and carjackings", which make up the majority of crimes

reported by CLEAR (NBC). Meanwhile, the drop in the crime rate in February 2021, to a level near that of April 2020, was likely driven by the drastic fall in temperature at the time, which incentivized people to stay at home and thus reducing the number of criminals on the street. Studies indicate that the numerical drop in crime rates stemmed from a significant decline in carjacking incidents during this period, though slightly offset by a rise in gun-related homicides (Chicago Sun Times). In addition to the sharp decline in crime rates when lockdowns started, our chart indicates that post COVID-19 crime rates are lower compared to the prior year before and also tend to oscillate. To confirm this, we would need crime data from prior years as 2019 could have just been an outlier.

How crimes are distributed geographically across the city (graph B):

The scatter plot illustrates the longitude and latitude coordinates of every crime recorded by the Chicago Police Department between March 1, 2019 and March 31, 2022. It should be noted that a few data points were removed from the graph as the location of a few crimes recorded by Chicago PD occurred outside of city jurisdictions, and we wanted to visualize criminal activity within the city. Upon observing the graph, one may observe how there exist several line-like areas with not a single crime reported. Upon comparing the scatter plot above and the map of Chicago, we observe that the empty, white-line patches correlate with rivers and highways, both of which are locations with little to no reported crime as these are logically inplausible places to commit crimes. Another key trait of the scatterplot is that the density of recorded crimes, illustrated by the density of points around certain coordinates, indicate which locations serve as crime hotspots. Upon further cross-referencing with a Chicago map, we find that, in terms of crimes counted and recorded by Chicago Police, the most crime-heavy burrows of the city were the west side and central Chicago, with our data indicating that the four out of the five most crime-heavy community areas, which account for over 20.04% of all crimes recorded in Chicago between March 2019 and March 2022, were located in these districts. This correlates with various other reports on crimes in Chicago by location, such as a 2014 CNN article stating that the Austin community area, the most common location of reported crime based on the database, experienced homicide rates at least 10 times higher than those in other parts of the city (CNN). Further research indicates that the high crime levels in certain districts stem from rampant numbers of west and central Chicago households living below the poverty line ($31,000 per year at present), subpar education funding and lack of workforce opportunities in these neighborhoods, in turn encouraging criminal activity (Chicago Sun Times). By contrast, neighborhoods that we found to have the lowest amount of criminal activity such as Edison Park largely resided in the far north side, where median household income was reported at or above $100,000 per year in 2020 (Money Inc). Thus, the disparity between low crime levels in wealthier neighborhoods and immensely high crime levels in poor neighborhoods demonstrates the impact of wealth gaps throughout the city on the distribution of criminal activity.

Hourly crime rates of most popular crimes (graph C):

The stacked bar chart depicts the number of crimes segregated by the hour of the day when they occured, focusing on the top 5 most frequently occurring crime types. There are two very interesting patterns that this graph produces that are worth analyzing: (1) how the overall crime rate changes according to time of day (in hours) and (2) how ratio of various different types of crimes changes according to time of day (in hours). Corresponding to the first aspect, we can deduce that the crime rate generally peaks in the afternoon and early evening hours (from about 12pm-6pm) and also around midnight, and generally declines in the late evening, early night, and early morning hours. The decline in crime rates, which is especially steep during early morning hours can trivially be explained by the fact that people are asleep in their homes, so the likelihood of a crime committed by an external party is highly unlikely. The sharp increase in crime rate at midday could also trivially be explained by using exactly the opposite reasoning: at midday, people are usually out in public and moving around, which increases the likelihood of crimes being committed. However, what seems non-trivial is the spike in crime rate at midnight. At midnight, the streets tend to be emptier, which gives way to certain types of crimes to occur successfully. This could further be understood by analyzing factor 2. Comparing the bars at noon and midnight, it can be deduced that the ratio of the types of crimes committed undergo small changes. This change in ratio is particularly noticeable in theft and battery crimes. At midnight, there are about as many battery crimes (if not more) as there are of theft, but at noon, the ratio is skewed in favor of theft. This phenomenon is very well

explained by a study conducted by The Sleep Judge in USNews, which explains why many thefts tend to happen during daylight whereas more serious crimes are usually carried out in the dark around midnight. According to the study, it is easier to carry out battery crimes discreetly at night, when streets are both quieter and emptier, which ultimately results in a lower possibility of there being witnesses to the crime (USNews). Meanwhile, thefts, especially simple ones, are easier to carry out discreetly in the presence of a crowd, when the streets are busy, such as around noon. Thus, certain types of crimes are responsible for the spike in midday crime rates while others cause a spike in midnight crime rates.

Modeling: what is the correlation between the number of thefts in a neighborhood and the arrest rate for said crime? (graph D):

In our modeling, we wanted to determine whether the ranking of a neighborhood in terms of theft-related crime count would impact the arrest rate in that neighborhood for said crime. To accomplish this, we first ranked the Community Areas in ascending order based on the number of thefts that occured in said neighborhood, then calculated the proportion of theft crimes where an arrest occurred. The positive slope of the line of best fit created from this model showed that areas with high reports of crime resulted in arrests more often compared to areas with low reporting of theft compared to the number of thefts recorded. A correlation coefficient of 0.434 was found, showing that there was a moderate linear relationship between ranking and proportion of arrests. Through modeling, we were attempting to predict the chances of someone getting arrested given that they committed a theft (y) and got the linear equation $y = 0.000516x + 0.0327$ with x being the rank of the community area they committed the theft in. In this case, the slope of 0.000516 tells us that with every unit increase in the ranking of a Community Area (the neighborhood had a higher number of crimes than the one before it), the proportion of arrests increased by about 0.0516% on average. The intercept of 0.0327 tells us that if a theoretical neighborhood were to have a rank of 0, thereby having the lowest theft crime count, then there would be a 3.27% chance of the perpetrator being arrested for said crime. Another interesting revelation stemming from this modeling is that it shows that there is bias when determining whether someone is guilty of a theft. In a fair world, we should have a correlation and slope of 0 because the area the crime is committed should not have an affect on arrest rates. However, given the positive correlation here, we can conclude that officers serving in areas with more thefts arrest more criminals even in proportion to total thefts. In fact, according to the line of best fit, we see that crimes committed in the Community Area with the highest number of crimes was nearly twice as likely to result in an arrest than crimes committed in the Community Area with the lowest number of crimes.

What others have done with this data:

In addition to our own research and modeling, we also explored how other researchers used the same Chicago PD database to visualize and predict criminal activity in Chicago. For example, a 2018 report published by Christopher Tan used the same database to extract recurring phrases in the crimes entered in each row in order to determine the most common locations and methods of criminal activity. In doing so, they found that assault and robbery with handguns were the most highly reported crimes in Chicago, and often led to the perpetrator being arrested more often compared to other types of crime. Furthermore, they also used the same phrase-extraction technique to extract the street address from each crime's block location, thereby determining that crimes more frequently occured on streets named after Martin Luther King Jr or on larger streets such as Michigan Avenue. Similarly to us, they also used the data to visualize the frequency of crimes throughout the year, finding that, on average, criminal activity was at a low during February and peaked during the summer months due to high temperatures sending people out of their homes, as well as stimulating aggression.

On the subject of trends between temperature and crime rates, a report published by Paul Reeping and David Hemenway in BMC explored the association between weather and daily shootings in Chicago between 2012 and 2016. To accomplish this, the researchers collected data from the Chicago Tribune Breaking News Desk, which collects data through news reports on cases of interpersonal injury, including gun-related homicides, as well as data from Weather Underground. Furthermore, the researchers also considered how variables and outliers such as the day of week, holidays, humidity and precipitation could affect their results. Overall, the researchers found that, on both weekdays and weekends, the city temperature for a certain day was strongly positively correlated with the number of shootings per day.

Conclusion/what did we learn?

Overall, we had some key findings with this dataset. First, we learned that in Chicago, crime rates are seasonal such that warmer weather incentivizes higher crime rates. Additionally, COVID-19 did impact crime rates in that there was an overall drop in crime whilst the aforementioned crime pattern continued to occur. Second, we learned that areas with high median incomes had low occurrences of crimes whereas areas where the majority of residents lived under the poverty line had high crime rates, such as the west/central boroughs of Chicago. Third, we learned that crime is more likely to happen during the daytime when most people are awake and out. Lastly, we learned that thefts committed in areas with high reports of crime were more likely to result in arrests, showing that there is some sort of bias in how police treat perpetrators. All of these findings are important in determining what type of solutions to implement. For example, more resources, police officers, and programs should be enforced in areas with high poverty rates and especially during the summer and afternoon.

Contribution of statement:

Overall, the group was fairly evenly divided. We each worked on one plot (Charitha on crime rates over the past 3 years, Max on crimes by coordinates and Maya by hourly crime rates), including the creation of our own graphs and evidence-driven analysis of said graphs. Furthermore, Charitha worked on the conclusion and modeling portion, Max worked on researching how others had used the same dataset as us to form their own conclusions and Maya worked on the introduction and brief description of the dataset. However, we all helped each other on our individual responsibilities, so it can be argued that everyone contributed to every part of the project to some degree.

Bibliography:

- **Main data:** "Crimes - 2001 to Present: City of Chicago: Data Portal." *Chicago Data Portal*, Chicago Police Department, 13 May 2022, revised 13 May 2022, https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2.
- Christensen, Jen. "Tackling Chicago's 'Crime Gap'." *CNN*, Cable News Network, 14 Mar. 2014, https://www.cnn.com/2014/03/13/us/chicago-crime-gap/.
- Cohn, Ellen G. "Weather and Crime." *The British Journal of Criminology, Vol 30, No. 1*, University of British Columbia, 1990, https://ibis.geog.ubc.ca/courses/geob370/students/class07/crime_weather/misc/weather_and_crime.pdf.
- Corley, Cheryl. "Crime Has Declined Overall during the Pandemic, but Shootings and Killings Are Up." *NPR*, NPR, 20 July 2020, https://www.npr.org/2020/07/20/892418244/crime-has-declined-overall-during-the-pandemic-but-shootings-and-killings-are-up.
- Favre, Lauren. "Study Finds Crime in Big Cities Is More Likely during the Day." *U.S. News & World Report*, U.S. News & World Report, 12 June 2019, https://www.usnews.com/news/cities/articles/2019-06-12/study-finds-crime-in-big-cities-is-more-likely-during-the-day.
- Fulton, R.E. "What You Should Know about Chicago Crime Rates." *GetJerry.com*, Jerry, Inc., 4 Mar. 2022, https://getjerry.com/car-insurance/chicago-crime-rates.
- Harp, Ryan D., and Kristopher B. Karnauskas. "Global Warming to Increase Violent Crime in the United States." *Environmental Research Letters*, IOP Publishing, 3 Mar. 2020, https://iopscience.iop.org/article/10.1088/1748-9326/ab6b37.
- Hart, Rob. "As Violent Crime Rises, 'Safety Gap' between Chicago Neighborhoods Widens." *US 99*, US 99, 29 Oct. 2021, https://www.audacy.com/us99/news/local/as-violence-rises-safety-gap-between-neighborhoods-widens.
- "Overall Chicago Crime down 30% in April, Police Say." *NBC Chicago*, NBC, 1 May 2020, https://www.nbcchicago.com/news/local/overall-chicago-crime-down-30-in-april-police-say/2264894/.
- Reeping, Paul M., and David Hemenway. "The Association between Weather and the Number of Daily Shootings in Chicago (2012–2016) - Injury Epidemiology." *BioMedical Central*, BioMedical Central, 22 June 2020, https://injepijournal.biomedcentral.com/articles/10.1186/s40621-020-00260-3.
- Sun-Times Wire. "Carjackings Drop in February, but Shootings and Murders Remain High: Police." *Chicago Sun-Times*, Chicago Sun-Times, 1 Mar. 2021, https://chicago.suntimes.com/crime/2021/3/1/22306821/chicago-crime-stats-february-2021-gun-violence-homicide-police.
- Tan, Christopher. "Chicago Crime Data Analysis (Python Project)." *CJANGO*, 15 Oct. 2018, https://cjango.wordpress.com/portfolio/chicago-crime-data-analysis-python-project/.
- Vix, Bill. "The 10 Richest Suburbs in Chicago." *Money Inc*, 31 July 2020, https://moneyinc.com/richest-suburbs-chicago/.

Chart A:

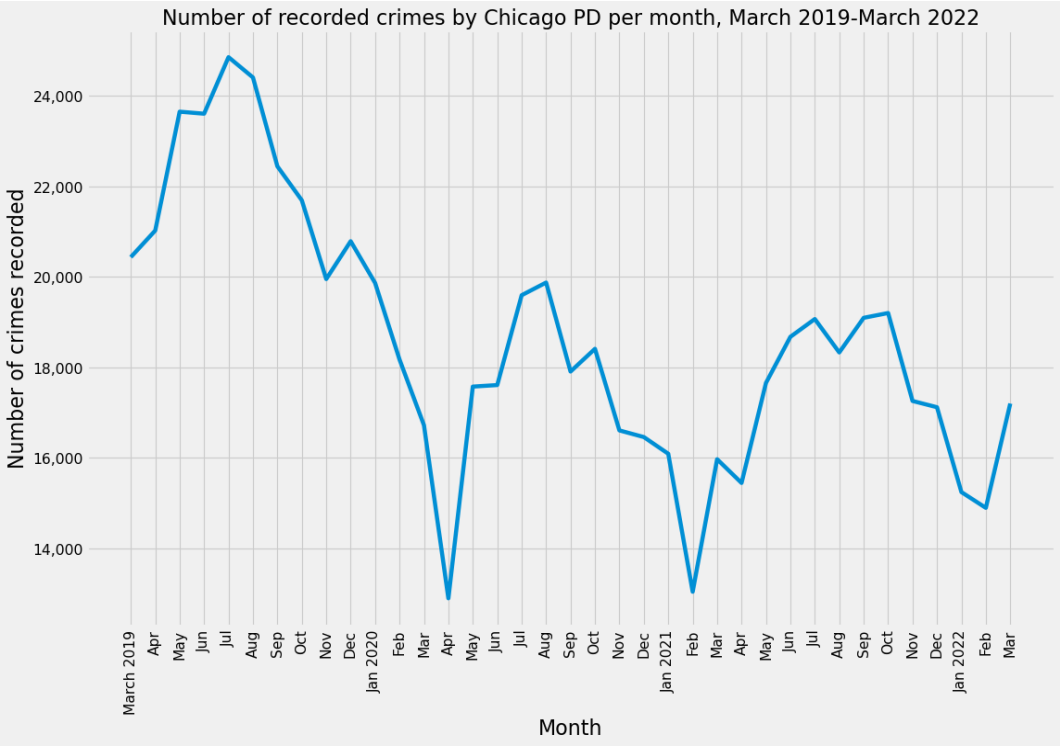
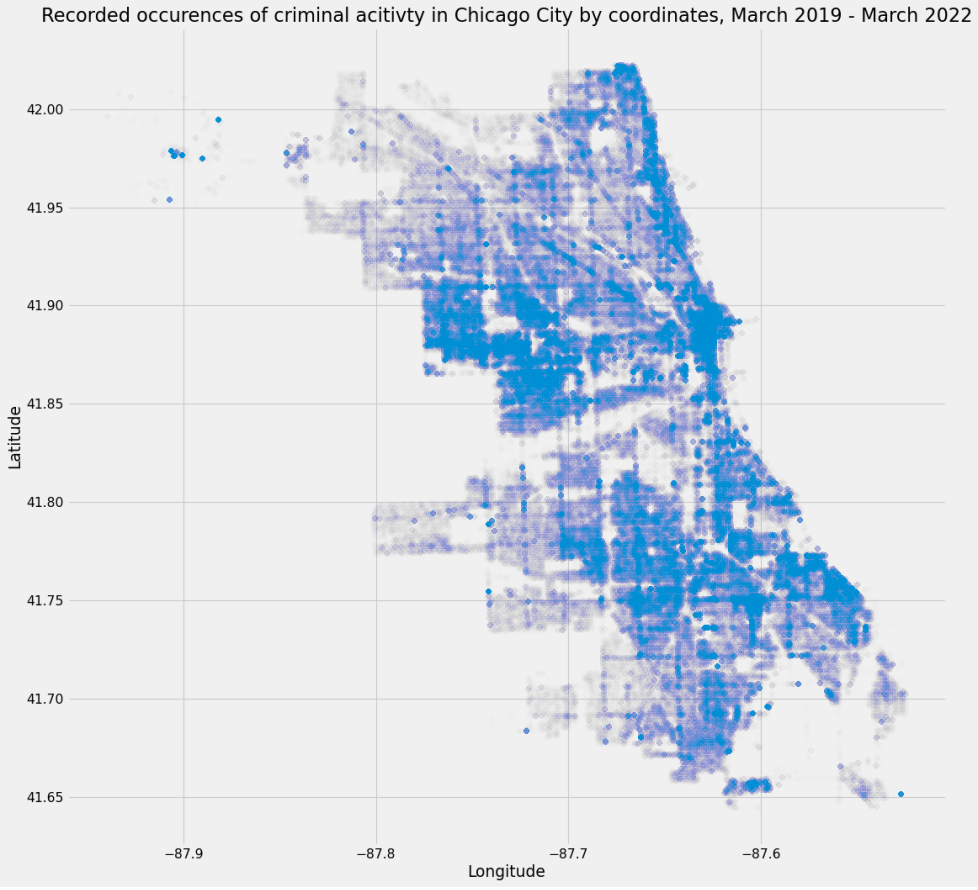
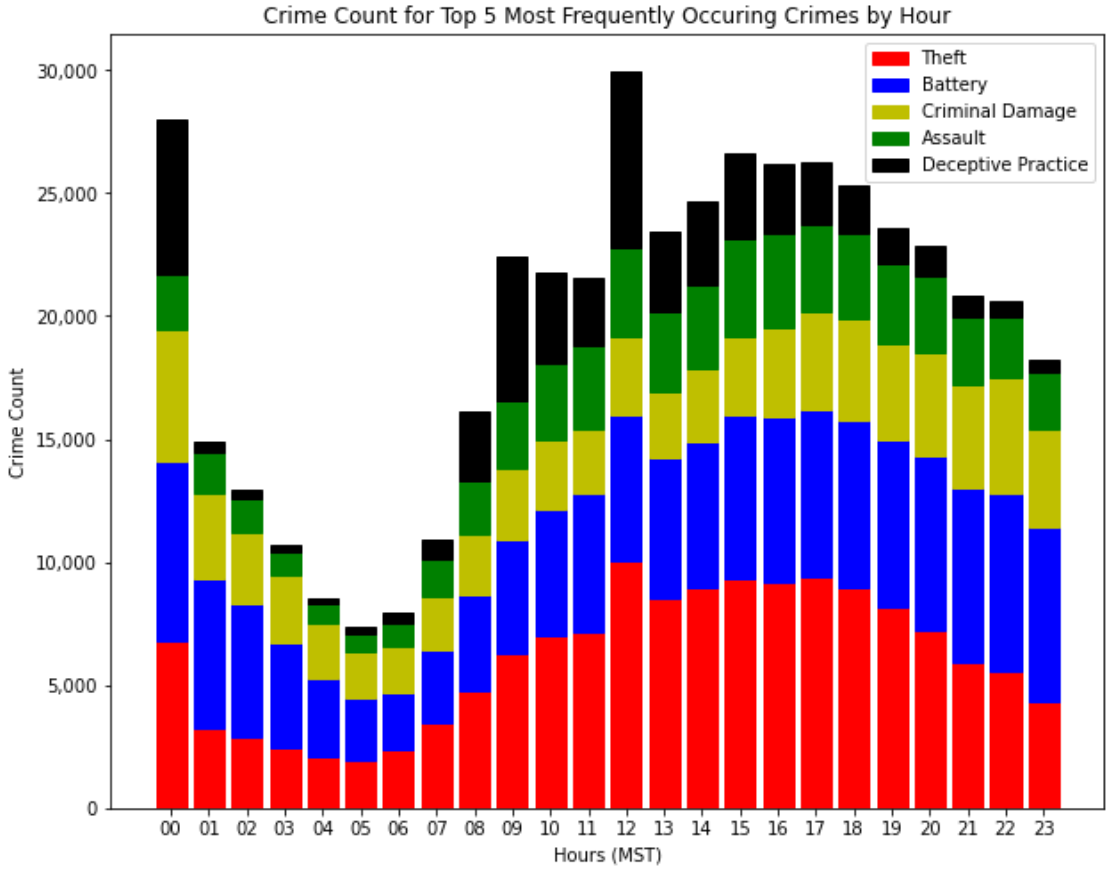


Chart B:



Graph C:



Graph D: