Final Report: Chicago Crime

Introduction

Urban crime remains a critical topic of study for policymakers, law enforcement, and researchers alike, and Chicago provides a compelling case for exploration. As one of the largest cities in the United States, Chicago's diverse neighborhoods and socio-economic disparities create a complex crime landscape. With its history of gang activity, economic inequality, and persistent challenges with violent and property crimes, the city offers a unique opportunity to study crime patterns and trends.

This project focuses on uncovering insights into crime trends in Chicago, analyzing data from 2019 to 2024. The selected timeframe includes critical societal shifts, such as the COVID-19 pandemic and its aftermath, providing a lens into how external events influence criminal activity. By examining these patterns, we aim to contribute to the broader understanding of urban crime and provide actionable insights for stakeholders working to create safer communities.

Motivation

The primary motivation for this project stems from the need to understand the factors driving crime in urban settings like Chicago. By analyzing crime data at granular temporal and spatial levels, we can uncover patterns that reveal the dynamics of public safety. The goal is to identify key trends that can inform resource allocation, policy decisions, and community engagement efforts.

The insights derived from this analysis are not just theoretical. They can be applied directly to improving public safety strategies, allocating law enforcement resources more effectively, and identifying socio-economic factors that influence crime rates. Ultimately, this project seeks to aid in creating equitable, data-driven solutions that address the root causes of crime while enhancing the safety and well-being of Chicago's residents.

Dataset

The Chicago Crime dataset offers a comprehensive record of reported crimes in the city, covering incidents from 2001 to the present, minus the most recent seven days. The data, extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system, includes key details about each crime while ensuring victim privacy by limiting address specificity to the block level. Each record is enriched with temporal, spatial, and categorical information, providing a robust foundation for analyzing crime trends and patterns in Chicago.

Notable columns in the dataset include “Primary Type”, which categorizes crimes into general types like "HOMICIDE," "BATTERY," and "BURGLARY," and “Description”, which offers more granular details about the incident. Spatial data is provided through “Block”, as well as precise latitude and longitude coordinates, enabling geospatial analysis and mapping of crime hotspots. Temporal information, such as the date and time of the crime, allows for time-series analysis to explore patterns over days, months, or years. Additionally, policing details, including beat, district, and whether an arrest was made, offer insights into law enforcement responses and resource allocation.

To make the dataset more manageable and focused, we narrowed it down to records from 2019 to 2024. This time frame provides a recent and relevant window for analysis while reducing the dataset size for easier processing. The selected range captures the most current trends and ensures insights align with present-day conditions, such as shifts in crime patterns during and after the COVID-19 pandemic. By focusing on this five-year span, we aim to uncover actionable insights and identify key patterns that can inform policy decisions, law enforcement strategies, and community interventions.

Data Cleaning and Missingness

Handling missing data was an essential step in preparing this dataset for analysis, given the significant proportions of missing values in several key columns. The most notable columns with missing data were “Location Description”, “Latitude”, and “Longitude”, which provide critical spatial context for crime incidents. Addressing these gaps thoughtfully ensured that the integrity of the dataset was preserved, while enabling meaningful insights into Chicago crime patterns. A snapshot of the original dataset is shown below.

ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	...	Ward	Community Area	FBI Code	X Coordinate	Y Coordinate	Year	Updated On	Latitude	Longitude	Location
25953	JE240540	05/24/2021 03:06:00 PM	020XX N LARAMIE AVE	0110	HOMICIDE	FIRST DEGREE MURDER	STREET	True	False	...	36.0	19.0	01A	1141387.0	1913179.0	2021	11/18/2023 03:39:49 PM	41.917838	-87.755969	(41.917838056, -87.755968972)
26038	JE279849	06/26/2021 09:24:00 AM	062XX N MC CORMICK RD	0110	HOMICIDE	FIRST DEGREE MURDER	PARKING LOT	True	False	...	50.0	13.0	01A	1152781.0	1941458.0	2021	11/18/2023 03:39:49 PM	41.995219	-87.713355	(41.995219444, -87.713354912)
13279676	JG507211	11/09/2023 07:30:00 AM	019XX W BYRON ST	0620	BURGLARY	UNLAWFUL ENTRY	APARTMENT	False	False	...	47.0	5.0	05	1162518.0	1925906.0	2023	11/18/2023 03:39:49 PM	41.952345	-87.677975	(41.952345086, -87.677975059)
13274752	JG501049	11/12/2023 07:59:00 AM	086XX S COTTAGE GROVE AVE	0454	BATTERY	AGGRAVATED P.O. - HANDS, FISTS, FEET, NO / MIN...	SMALL RETAIL STORE	True	False	...	6.0	44.0	08B	1183071.0	1847869.0	2023	12/09/2023 03:41:24 PM	41.737751	-87.604856	(41.737750767, -87.604855911)
13203321	JG415333	09/06/2023 05:00:00 PM	002XX N Wells st	1320	CRIMINAL DAMAGE	TO VEHICLE	PARKING LOT / GARAGE (NON RESIDENTIAL)	False	False	...	42.0	32.0	14	1174694.0	1901831.0	2023	11/04/2023 03:40:18 PM	41.886018	-87.633938	(41.886018055, -87.633937881)

Table 1. Original raw data

The “Location Description” column had 7,558 missing values, with the majority linked to crimes classified as Deceptive Practice (7,031 cases). Deceptive Practice includes offenses like fraud, forgery, and identity theft, which are often not tied to a specific physical location. These crimes frequently occur in intangible environments, such as online platforms or via telephone, where providing a location description is inherently difficult. The remaining missing location descriptions were distributed across crime types like Battery (129 cases), Other Offense (117 cases), and Burglary (101 cases). These gaps likely stem from challenges in documenting precise locations during certain incidents, such as altercations or thefts in unmonitored areas. Given the nature of this data, we chose to retain the missing values as NaN, reflecting the reality of unknown or ambiguous locations. Imputation in this context would have risked introducing inaccuracies or misrepresentations, as there is no clear or consistent alternative for filling these gaps.

The “Latitude” and “Longitude” columns had an even larger proportion of missing data, with 18,905 missing values each, posing significant challenges for geospatial analysis. Geospatial data is crucial for visualizing and analyzing crime patterns across Chicago, but these gaps necessitated creative solutions. To address the missing coordinates, we leveraged the “Block” column, which provides block-level location information, and used a geocoding API to approximate latitude and longitude values. Through this approach, we successfully filled geospatial data for 100 cases, prioritizing feasibility within the constraints of free API usage. This approach demonstrated the potential for further enhancements but was limited to the subset of cases we could process without incurring additional costs.

Under different circumstances, we would have extended this process to fill the remaining 18,000+ missing values, ensuring a comprehensive dataset for spatial analysis. However, this limitation was acceptable for the scope of this project, as we prioritized transparency and practical constraints. The remaining missing geospatial values were left as NaN, indicating that they could not be determined based on available data.

Our analysis of missingness highlights important patterns in the data. Deceptive Practice crimes not only lacked Location Description values but were also the most frequent category with missing geospatial data, aligning with the nature of these offenses. Other crimes with notable missingness in latitude and longitude, such as Theft, Narcotics, and Other Offense, reflect potential inconsistencies or omissions in reporting. These patterns emphasize the challenges of documenting and geocoding crime data accurately in a complex urban environment like Chicago.

Beyond addressing missing values, we performed other essential data cleaning steps to prepare the dataset for analysis. This included dropping irrelevant columns and standardizing column names for consistency and machine readability. The dropped columns included Location, X Coordinate, Y Coordinate, IUCR, and FBI Code. These columns were deemed unnecessary for our analysis, as they either duplicated information available elsewhere in the dataset or did not provide meaningful insights for our goals. For instance, the Location column was redundant given the inclusion of Latitude and Longitude, while the X Coordinate and Y Coordinate columns represented an alternative geospatial format not needed alongside latitude/longitude data. Similarly, the IUCR (Illinois Uniform Crime Reporting) and FBI Code columns were specific to classification schemes that were not relevant to our analysis objectives.

To ensure the dataset was easy to work with, we standardized column names by converting them to lowercase and replacing spaces with underscores. This step made the dataset more machine-readable and consistent with common data science practices, streamlining operations like querying and merging. These changes simplified working with the data while retaining all the essential information needed for our analysis. Together, these cleaning steps ensured that the dataset was well-structured and focused, setting a strong foundation for meaningful exploration and insights into Chicago crime trends. Below is a snapshot of the cleaned dataframe.

id	case_number	date	block	primary_type	description	location_description	arrest	domestic	beat	district	ward	community_area	year	updated_on	latitude	longitude
25953	JE240540	05/24/2021 03:06:00 PM	020XX N LARAMIE AVE	HOMICIDE	FIRST DEGREE MURDER	STREET	True	False	2515	25.0	36.0	19.0	2021	11/18/2023 03:39:49 PM	41.917838	-87.755969
26038	JE279849	06/26/2021 09:24:00 AM	062XX N MC CORMICK RD	HOMICIDE	FIRST DEGREE MURDER	PARKING LOT	True	False	1711	17.0	50.0	13.0	2021	11/18/2023 03:39:49 PM	41.995219	-87.713355
13279676	JG507211	11/09/2023 07:30:00 AM	019XX W BYRON ST	BURGLARY	UNLAWFUL ENTRY	APARTMENT	False	False	1922	19.0	47.0	5.0	2023	11/18/2023 03:39:49 PM	41.952345	-87.677975
13274752	JG501049	11/12/2023 07:59:00 AM	086XX S COTTAGE GROVE AVE	BATTERY	AGGRAVATED P.O. - HANDS, FISTS, FEET, NO / MIN...	SMALL RETAIL STORE	True	False	632	6.0	6.0	44.0	2023	12/09/2023 03:41:24 PM	41.737751	-87.604856
13203321	JG415333	09/06/2023 05:00:00 PM	002XX N Wells st	CRIMINAL DAMAGE	TO VEHICLE	PARKING LOT / GARAGE (NON RESIDENTIAL)	False	False	122	1.0	42.0	32.0	2023	11/04/2023 03:40:18 PM	41.886018	-87.633938

Table 2. Cleaned dataframe

By addressing missing values with thoughtful methods and documenting limitations, we ensured that the dataset remains robust for analysis. Our approach balanced the need for completeness with the preservation of data integrity, enabling meaningful insights while acknowledging areas where further resources could enhance the dataset. This transparency sets the foundation for informed, actionable conclusions in understanding Chicago's crime dynamics.

Spatial Analysis Supplementary Data

For this analysis, we aimed to explore spatial patterns of crime in Chicago's community areas by integrating various datasets and creating visualizations that overlay crime, population, and socioeconomic disadvantage on geographical maps. The ultimate goal was to identify correlations between crime density, population distribution, and economically disadvantaged areas to inform data-driven policymaking and resource allocation.

The supplementary datasets used in this project were obtained from a combination of government and publicly accessible sources. Community area boundaries and socioeconomic data were retrieved from the City of Chicago's geographic and economic development datasets, including detailed geojson files and CSVs containing spatial boundary information. Population data, essential for normalizing crime statistics, was sourced from census data provided by Act for Children. The final supplementary data for our spatial analysis included: community areas geopandas geometry boundaries, total population counts per community area for 2024, Chicago's border as a geopandas geometry, and geopandas boundaries for areas deemed socioeconomically disadvantaged by the city of Chicago as of 2024. These were classified via measures of median home income, poverty rates, and unemployment rates.

The cleaning process began by the precleaned crime data, filtering by the top ten crimes, then summing by crime counts per community area, while filtering out rows with missing geocoordinates. Duplicates and null entries were removed to ensure accuracy. The cleaned crime dataset was aggregated by community areas, with unique identifiers facilitating integration with geographic boundary data.

Spatial data required basic preprocessing. The community area geometries were loaded using GeoPandas, ensuring compatibility with other geospatial datasets. Socioeconomic data and population information were merged into the main GeoDataFrame after standardizing key aggregation values. Population data underwent simple cleaning, including removal of commas and type conversion for numerical analysis.

To enable meaningful comparisons, crime frequencies were normalized by total population within each community area. Our normalization method was simply total crime count divided by population per community area. Additional metrics, such as normalized assault and theft rates, were calculated in the same manner to reveal disparities in crime rates relative to population density.

General Trends and Overview of Chicago Crimes

This section provides a general overview of the landscape of crime in Chicago, offering insights into both temporal trends and the prevalence of various crime categories. By summarizing the results of our data exploration process, this section aims to set the stage for deeper analysis by highlighting key patterns and findings that emerged during the initial stages of the project. These general trends give context to the broader dataset, allowing us to better understand the scale and scope of crime in Chicago from 2019 to 2024.

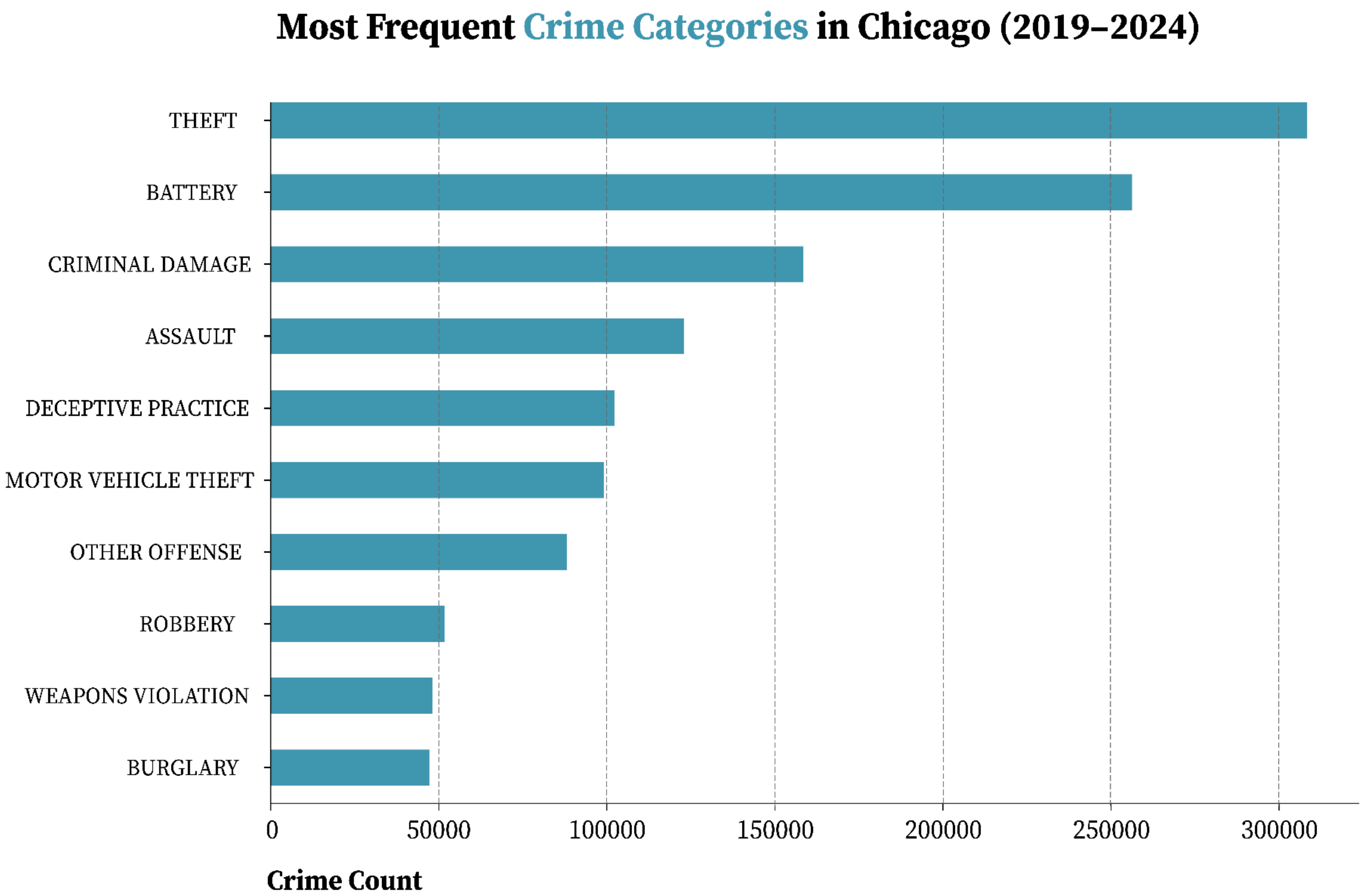

Figure 2. Most frequent crime categories in Chicago (2019 - 2024)

This visualization, "Most Frequent Crime Categories in Chicago 2019–2024," highlights the distribution of crime types over the time period. The x-axis represents the count of reported incidents for each crime category, while the y-axis lists the corresponding crime categories. Each bar in the visualization represents the total number of reported incidents for a specific crime category from 2019 to 2024. Theft stands out as the most frequently reported crime, far exceeding other categories. This finding underscores the persistent challenge of property crimes in urban environments, particularly in densely populated and commercially active areas. Battery and criminal damage also feature prominently, reflecting ongoing issues with interpersonal violence and property-related offenses. Together, these categories account for the majority of reported incidents, painting a picture of the city's primary public safety concerns. Other crime types, such as assault, motor vehicle theft, and deceptive practices, further illustrate the diverse challenges faced by law enforcement and policymakers.

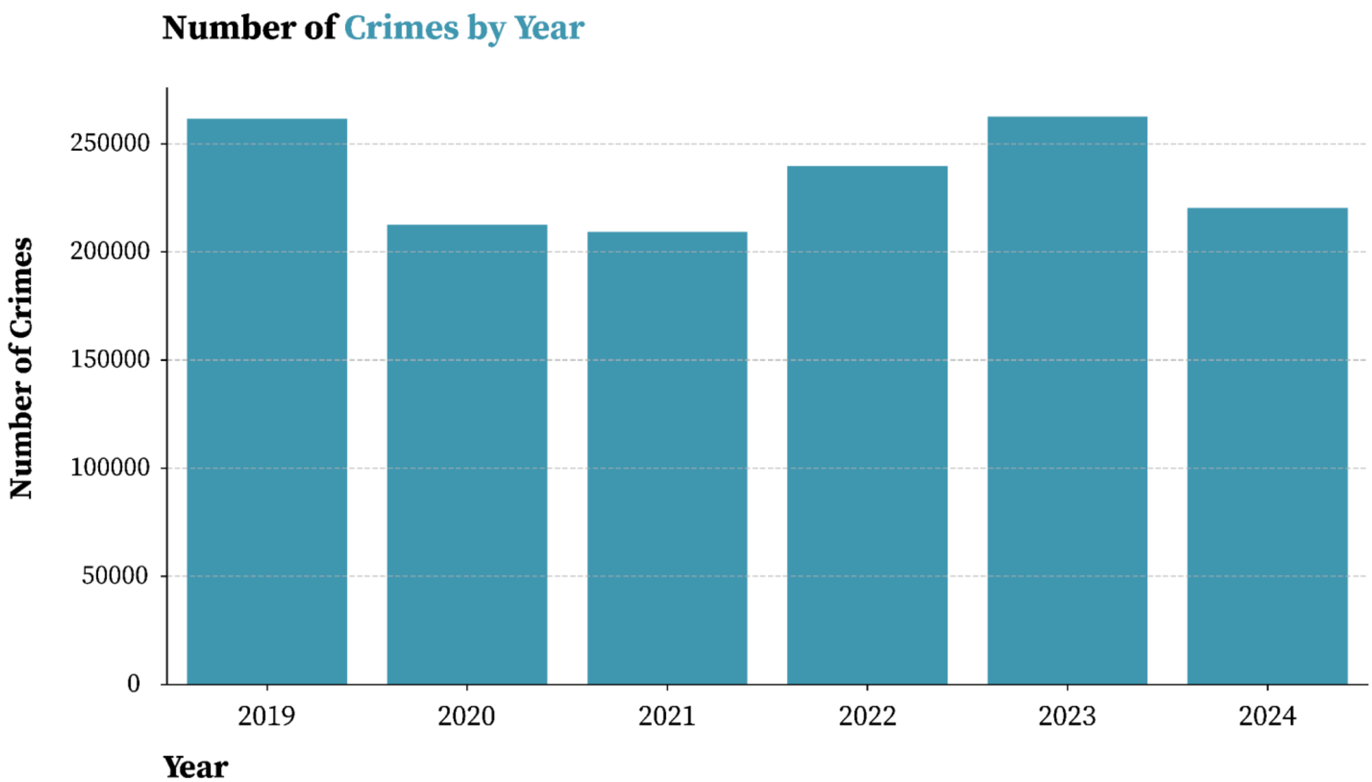

Figure 1. Number of Crimes by Year

The bar chart above, "Number of Crimes by Year," provides historical perspective, capturing fluctuations in reported crimes over time. The x-axis represents the years, while the y-axis shows the total number of crimes reported in each respective year. Each bar corresponds to the total volume of crimes for that specific year, providing a clear view of how crime levels have fluctuated over time. Notably, a decline in reported crimes can be observed during the early years of this period, coinciding with the COVID-19 pandemic. This decline aligns with national trends, as lockdowns and reduced public activity led to fewer opportunities for certain types of crimes. However, as public spaces reopened, crime rates began to rebound, reflecting a return to pre-pandemic dynamics. Additionally, the data for 2024 only extends through November, which likely explains why the total number of reported crimes appears lower compared to prior years.

While the aggregated data provides an overall perspective, breaking it down into individual crime categories can reveal unique trends and patterns that may not be apparent in the broader view. Each crime type might present a different narrative, reflecting varying influences and dynamics over time. This detailed exploration in figure 3 will help us better understand the nuances behind the fluctuations in reported crimes.

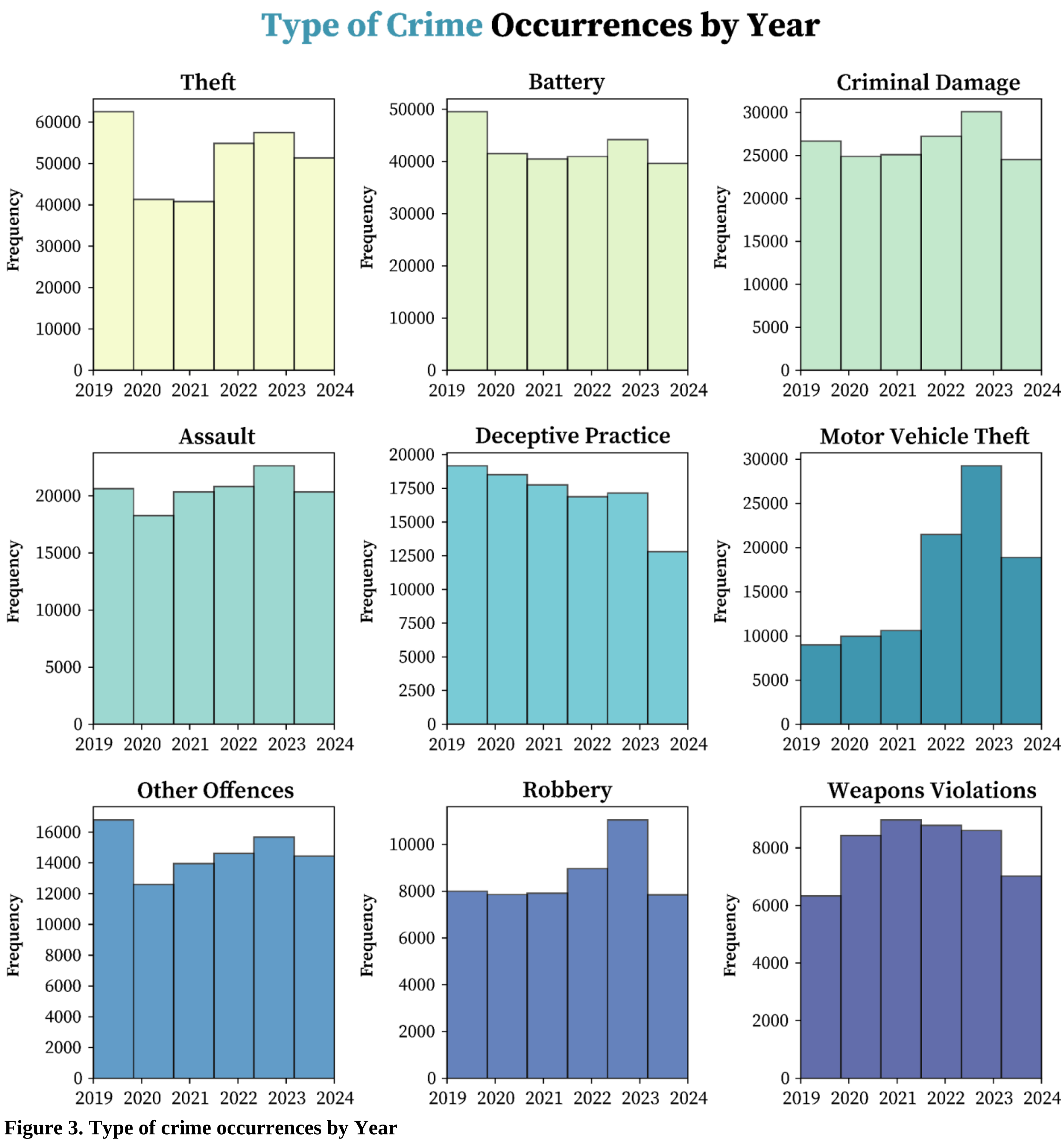

Figure 3. Type of crime occurrences by Year

This visualization provides insights into the top nine most commonly reported crimes in Chicago from 2019 to 2024, broken down by year. The x-axis represents the years in the selected time frame, while the y-axis indicates the total number of incidents for each crime type. Each bar corresponds to the yearly count of a specific crime, with separate colors used to distinguish between the different crime categories. This bar chart allows for a side-by-side comparison of how the frequency of each crime type has changed over time, offering a clear perspective on temporal trends.

The data reveals some interesting patterns across these top crime categories. Theft remains the most frequently reported crime, but it exhibits more variation than categories like assault or other offenses. Specifically, theft saw notable dips during the COVID-19 pandemic, likely due to reduced public activity, closures of businesses, and fewer opportunities for property-related crimes. Battery and criminal damage, while also prominent, show less dramatic fluctuations, suggesting that these crimes are less influenced by pandemic-related factors and may be more stable aspects of Chicago's crime landscape.

One of the most striking patterns is the sharp rise in motor vehicle theft following the pandemic. Starting in 2021, motor vehicle thefts almost doubled or tripled compared to pre-pandemic levels. While the exact reasons for this spike are unclear, several factors may have contributed. The pandemic disrupted global supply chains, leading to shortages in new cars and a surge in used car values. This economic dynamic could have made motor vehicles more attractive targets for theft. Additionally, changes in policing priorities during and after COVID, coupled with potential increases in organized crime targeting vehicles, might have exacerbated this trend.

Robbery also shows notable spikes in 2022 and 2023 following a decline during the pandemic years. These increases likely reflect the return of public activity and economic pressures, creating more opportunities for confrontational theft. These trends highlight the need to explore socio-economic factors driving crime in urban settings.

Deceptive practices, on the other hand, appear to be in a general decline throughout the observed period, with a slight uptick in 2023 compared to 2022. This trend suggests a potential decrease in the reporting or prevalence of these crimes over time. The consistent decline might reflect increased public awareness or improved detection and prevention methods. Alternatively, it could signal a shift in criminal tactics away from this category toward other methods that may be harder to track or classify under this type.

The spikes in certain crimes after the pandemic underscore the complex and multi-faceted nature of crime trends. Economic pressures, shifts in law enforcement priorities, and evolving societal dynamics all likely played a role. These insights are crucial for understanding how external factors influence crime patterns and identifying where targeted interventions might be most effective. By exploring these trends, the visualization highlights the interplay between stability and variability in Chicago’s crime dynamics, offering a foundation for deeper analysis.

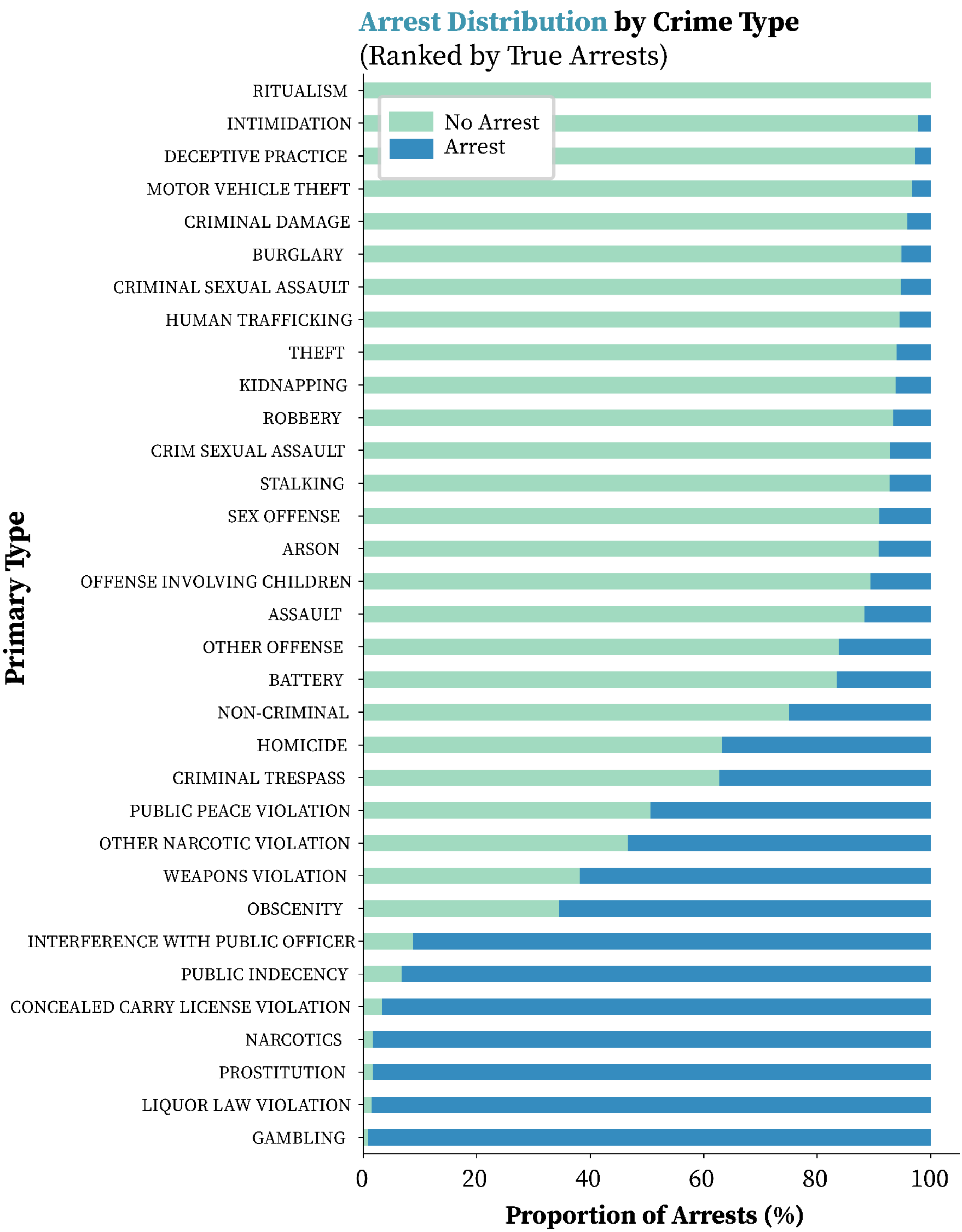

Figure 4. Arrest Distribution by Crime Type

This visualization provides insights into the distribution of arrests for various crime types in Chicago, ranked by the proportion of true arrests. The x-axis represents the proportion of arrests as a percentage, while the y-axis lists the primary crime types. Each bar is divided into two sections: one representing cases that resulted in arrests and the other representing cases without arrests. The distinct colors highlight the contrast between these outcomes, enabling a clear comparison of arrest rates across different crime categories. This bar chart offers a detailed perspective on how law enforcement handles specific crimes, shedding light on variations in arrest success rates and the potential challenges associated with addressing certain offenses.

For instance, crimes such as gambling, narcotics, and liquor law violations have a notably high proportion of arrests, suggesting targeted enforcement strategies and perhaps more straightforward detection mechanisms. On the other hand, crimes like ritualism, intimidation, and deceptive practice exhibit much lower arrest proportions, indicating challenges in apprehension or enforcement for these offenses. These differences emphasize the need to understand the nature of each crime and the obstacles law enforcement faces in addressing them.

The ranking of crimes by true arrests reveals critical insights into the priorities and successes of law enforcement agencies. The high arrest proportions for certain crimes reflect where resources and attention are concentrated. Conversely, moderate or low arrest proportions for crimes like offenses involving children and assault could highlight complexities in these cases, such as the need for substantial evidence or the sensitivity of the investigations. Furthermore, crimes like criminal sexual assault and stalking have relatively low arrest rates, which may point to systemic issues such as underreporting, difficulty gathering evidence, or broader societal barriers that hinder effective enforcement.

These disparities in arrest proportions have broader implications for public policy and resource allocation. High arrest rates for crimes like narcotics might reflect the success of specialized task forces and a prioritization of certain offenses. However, the low arrest rates for critical crimes such as human trafficking underscore the need for further investment in resources, training, and policy interventions. Addressing these gaps is essential to improving outcomes in areas where enforcement currently lags.

Seasonality and Time-of-Day Effects in Crime

In this section, we analyze the temporal patterns of crimes in Chicago by distinguishing between daytime and nighttime occurrences, taking into account the city's geographical location and time zone. Using accurate sunrise and sunset timings for different seasons, we classified crimes based on the time of day they were reported, ensuring that the seasonal variations in daylight hours were properly accounted for. This analysis provides a deeper understanding of how the occurrence of crimes varies throughout the day and across seasons. By examining statistics for each crime type, as well as their distribution across months, we uncover nuanced insights into how temporal factors influence criminal activity in Chicago.

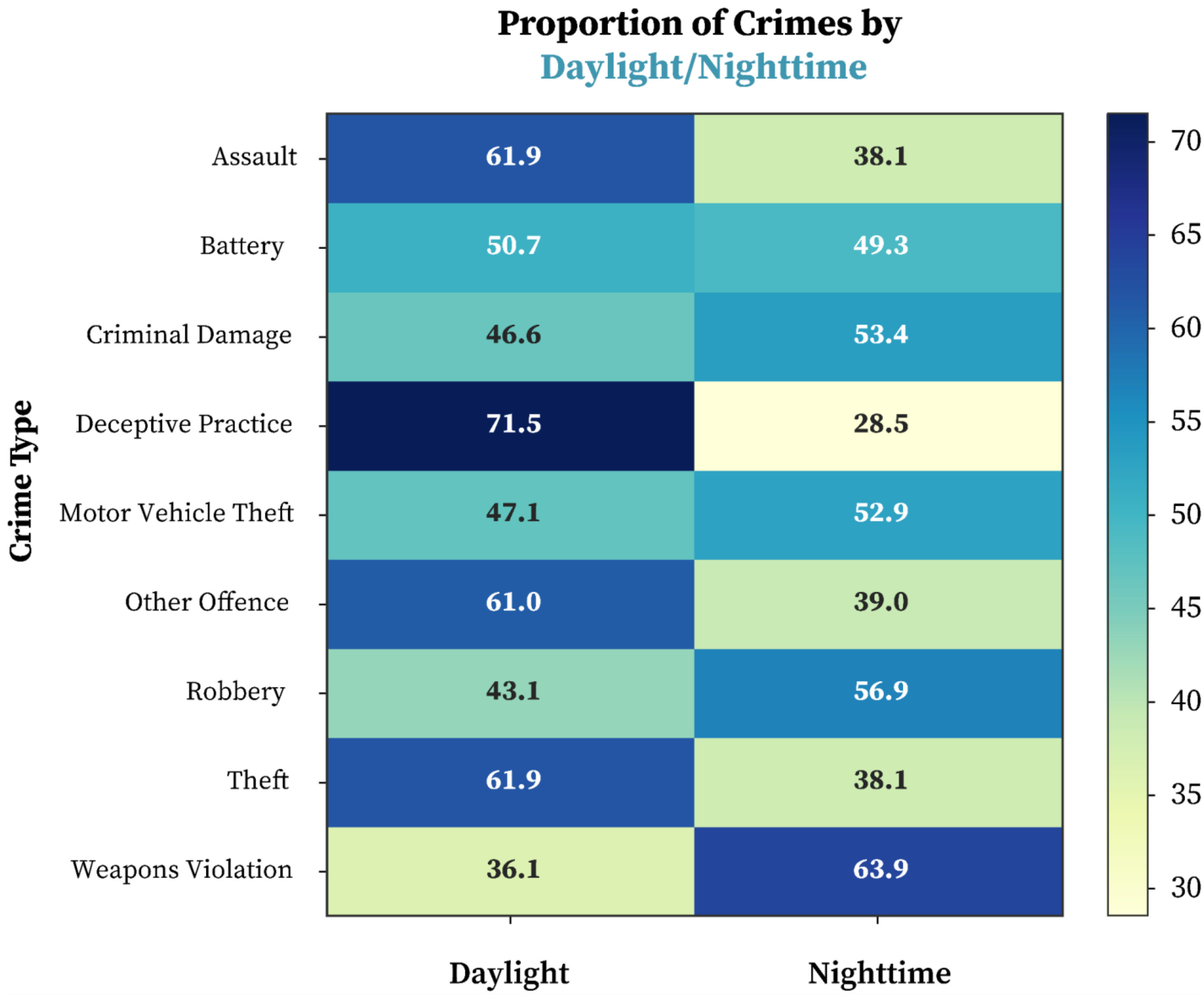

Figure 5. Proportion of Crimes by Daylight/Nighttame

Figure 5 above illustrates the distribution of the top nine crimes in Chicago, categorized by whether they occurred during the day or at night. The x-axis represents the percentage of crimes occurring in either daylight or nighttime, while the y-axis lists the crime types. Each horizontal bar is divided into two segments, with the left side representing crimes committed during daylight hours and the right side representing those that occurred at night. The colors emphasize the relative proportions for each category, making it easy to compare how these crimes vary by time of day. The primary takeaway is that some crimes are more likely to occur during the day, while others predominantly take place at night, reflecting distinct behavioral or situational patterns.

Diving deeper, we observe that crimes like deceptive practice (71.5% daylight) and theft (61.9% daylight) are significantly more likely to occur during the day, likely due to higher activity levels in commercial and public spaces when people are out shopping, commuting, or conducting daily activities. In contrast, crimes such as weapons violations (63.9% nighttime) and robbery (56.9% nighttime) are more frequent at night, possibly due to reduced visibility and fewer witnesses, creating conditions that embolden offenders.

Interestingly, some crimes like battery and motor vehicle theft show a more balanced distribution, indicating that they occur almost equally during the day and night, suggesting factors other than time of day may drive these crimes.

This analysis highlights the importance of time-specific strategies for law enforcement and crime prevention. For instance, resources might be concentrated in commercial areas during the day to address theft and deceptive practices, while nighttime patrols could focus on areas prone to violent crimes or weapons violations. Understanding these temporal patterns is essential for optimizing resource allocation and tailoring interventions to the nature of each crime type.

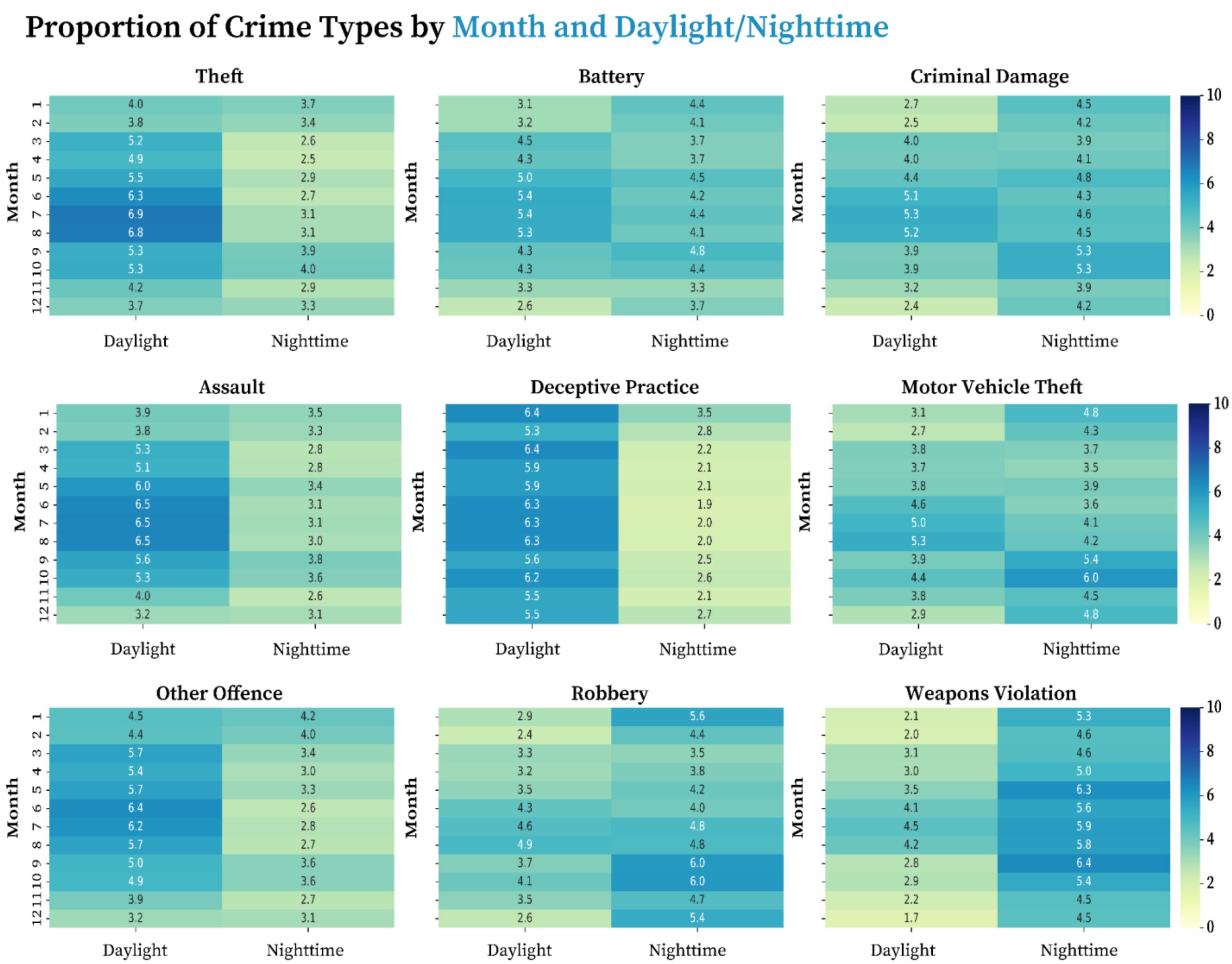

Figure 6. Proportion of Crime Types by Month and Daylight/Nighttime

Compared to the previous figure, the visualization in figure 6 introduces an additional layer of complexity by incorporating seasonality and dynamics throughout the year. It examines how the top nine crime categories vary across months and time of day in Chicago. The x-axis represents whether crimes occurred during daylight or nighttime, while the y-axis denotes the months of the year, ranging from January (1) to December (12). Each subplot corresponds to a specific crime type, with the color intensity representing the proportion of crimes. Darker shades indicate a higher proportion, allowing for a clear comparison of the seasonal and diurnal patterns of crimes. The primary takeaway is that crimes exhibit distinct seasonal and temporal trends, influenced by environmental factors such as daylight hours and social behaviors throughout the year.

Certain crimes, such as theft and deceptive practice, show a higher proportion during daylight hours across most months, reflecting the influence of increased human activity during the day. For example, deceptive practice shows consistently higher proportions during daylight across all months, aligning with business hours and public interactions. In contrast, crimes such as robbery and weapons violations tend to have stronger nighttime proportions, particularly during warmer months, likely due to increased outdoor activity and lower visibility during evening hours.

Seasonal variations also emerge across crime categories, providing insights into how environmental factors influence criminal behavior. For instance, battery and assault display a notable rise in both daylight and nighttime proportions during the summer months, potentially linked to increased outdoor gatherings and social interactions during warmer weather. On the other hand, crimes like motor vehicle theft show higher nighttime proportions in colder months, suggesting opportunistic behavior when vehicles are left unattended for extended periods. These patterns underscore the impact of seasonal dynamics, such as longer daylight hours in summer or darker, colder conditions in winter, on the prevalence of specific crimes.

This analysis reveals how seasonality interacts with time of day to shape crime patterns. Crimes like criminal damage and other offenses exhibit relatively balanced proportions across months, suggesting they may be less influenced by seasonality or time of day and more driven by consistent socio-environmental factors. By identifying these nuanced trends, this visualization provides valuable insights for law enforcement and policy-making. For example, resource allocation can be tailored to specific times of the year and day to address crimes more effectively. Additionally, understanding the dynamics of seasonal crime fluctuations can guide community awareness campaigns and targeted prevention strategies to mitigate risks during peak times.

Spatial Relationships

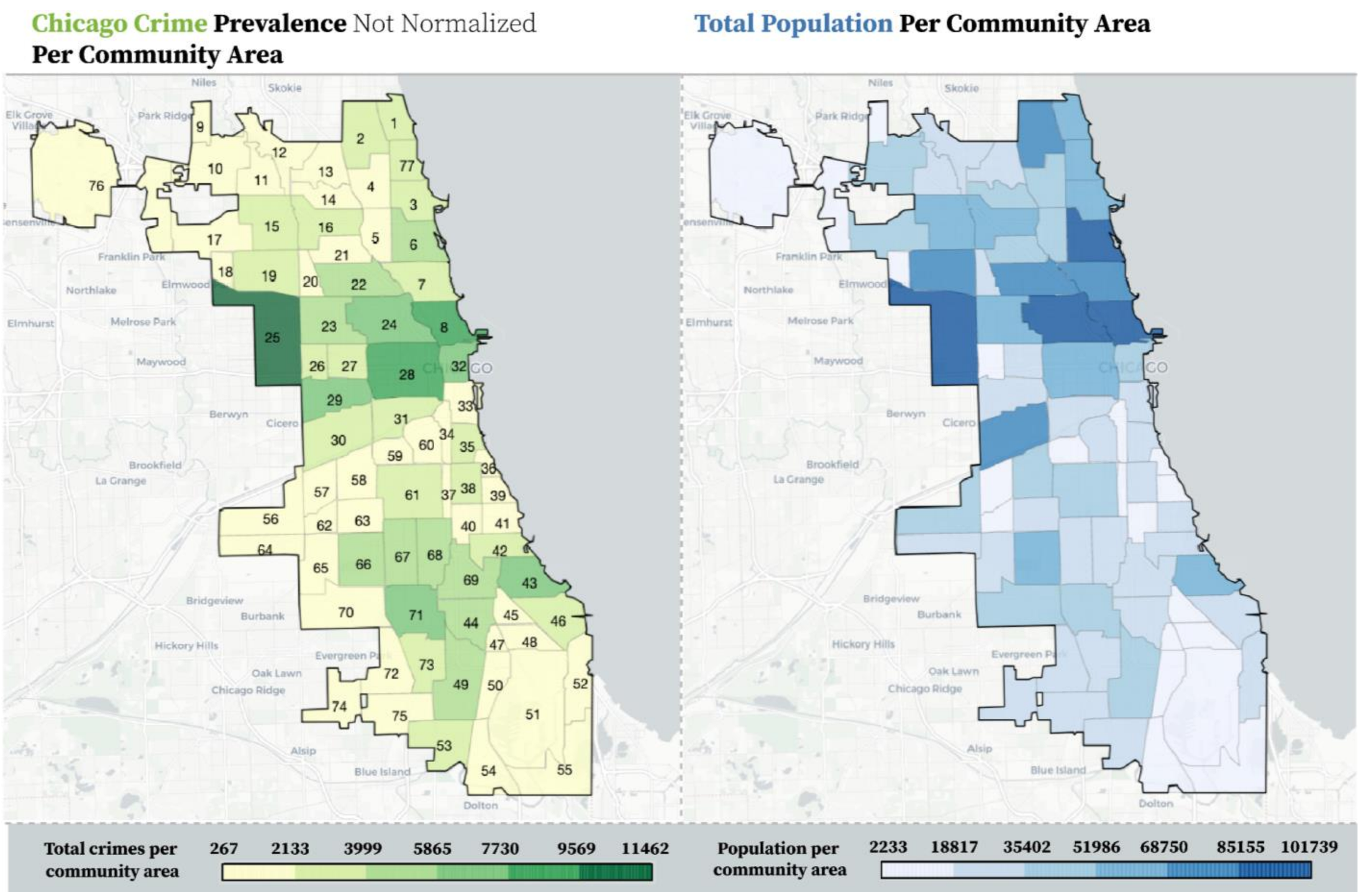

Figure 7. Chicago Crime Prevalence Per Community Area and Total Population Per Community Area

This figure provides a comparative spatial analysis of crime prevalence (not normalized for population) and population density across Chicago's community areas. The heatmap on the left visualizes the raw sum of the top ten crime types in Chicago in 2024 per community area. The shade of green scales darker with greater numbers of crimes in a specific community area. The top ten crimes are, theft, battery, criminal damage, assault, deceptive practice, motor vehicle theft, other offenses, robbery, weapons violation and burglary, respectively. Additionally, each number contained in each community area corresponds to the full name of the community area as shown in appendix item #1. The map on the right illustrates the total population within those same areas in 2024. The corresponding legends at the bottom clarify the scale of crime frequency and population density, with darker shades indicating higher values. We understand that measures like crime frequency and population in real life are not confined by community area borders as represented here, however, summarizing these measures by community areas is the most effective way of comparing across metrics.

A key finding evident from these maps is the strong correlation between crime prevalence and population density. Areas with higher population counts, such as central and northern community areas, also exhibit elevated crime frequencies. This relationship suggests that higher population density may contribute to an increased number of reported crimes, likely due to greater urban activity, higher social interactions, and an increased likelihood of criminal activity being witnessed and reported.

In the southern and peripheral regions of Chicago, where population density is lower, crime frequencies are also observed to be significantly reduced. This analysis does not align with expectations, as southern Chicago is notoriously perceived to have higher crime respective to other areas in Chicago.

Overall, the side-by-side comparison highlights the need to contextualize raw crime data with demographic information to avoid misleading conclusions. While densely populated areas show more absolute crime, further investigation, such as normalization by population, is necessary to discern whether certain areas experience disproportionately higher or lower crime rates relative to their population sizes.

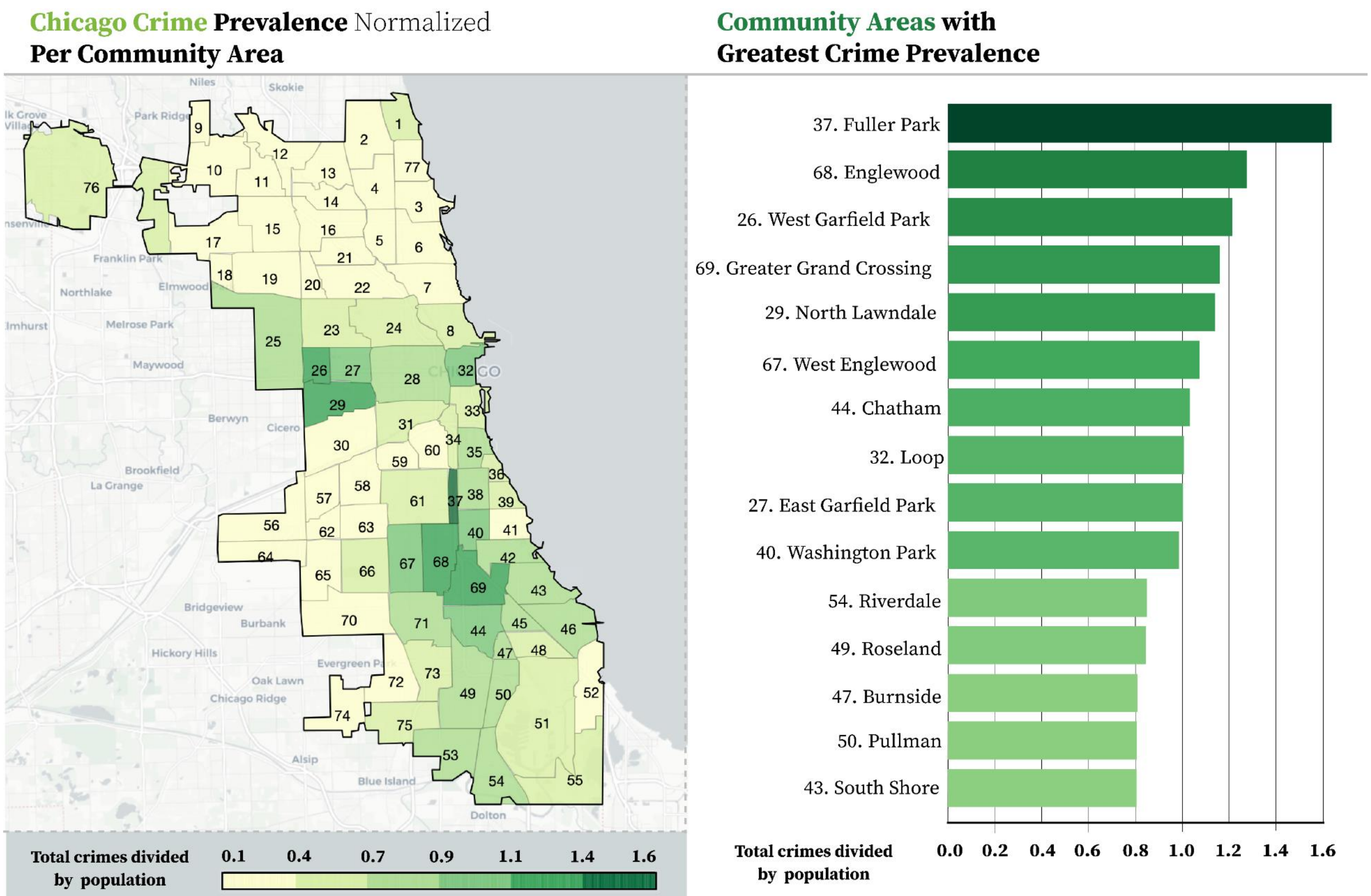

Figure 8. Chicago Crime Prevalence Per Community Area (Normalized) and Community Areas with Greatest Crime Prevalence:

This figure refines the spatial analysis of crime prevalence by normalizing crime counts per community area based on population size. On the left, the map illustrates normalized crime prevalence, where darker shades indicate higher crime rates per capita. On the right, a bar chart highlights the community areas with the highest normalized crime rates, ranked from Fuller Park at the top to South Shore at the bottom of the top tier.

Unlike the previous visualization of raw crime frequencies, this normalization reveals a more nuanced relationship between population density and crime rates. Specifically, community areas with high population density, which previously exhibited high absolute crime counts, now show lower crime rates per capita. This suggests that while densely populated areas experience more crimes in total, they are not necessarily the most dangerous when considering population size.

Conversely, several less populated areas emerge as having disproportionately high crime rates per capita. For example, Fuller Park, Englewood, and West Garfield Park rank at the top of the normalized chart, reflecting their higher crime rates relative to their smaller populations. These findings highlight areas of concentrated criminal activity that may not be evident from raw crime counts alone.

Overall, this figure underscores the importance of normalizing crime data by population to avoid misinterpreting raw crime frequencies. While high-population areas may appear as crime hotspots in absolute terms, the per capita analysis shifts focus to smaller, underserved communities where crime disproportionately affects residents. This insight is critical for targeted interventions and resource allocation to address community-specific safety concerns.

In our next two analyses, we explore alternative explanations for increased crime per capita in Chicago.

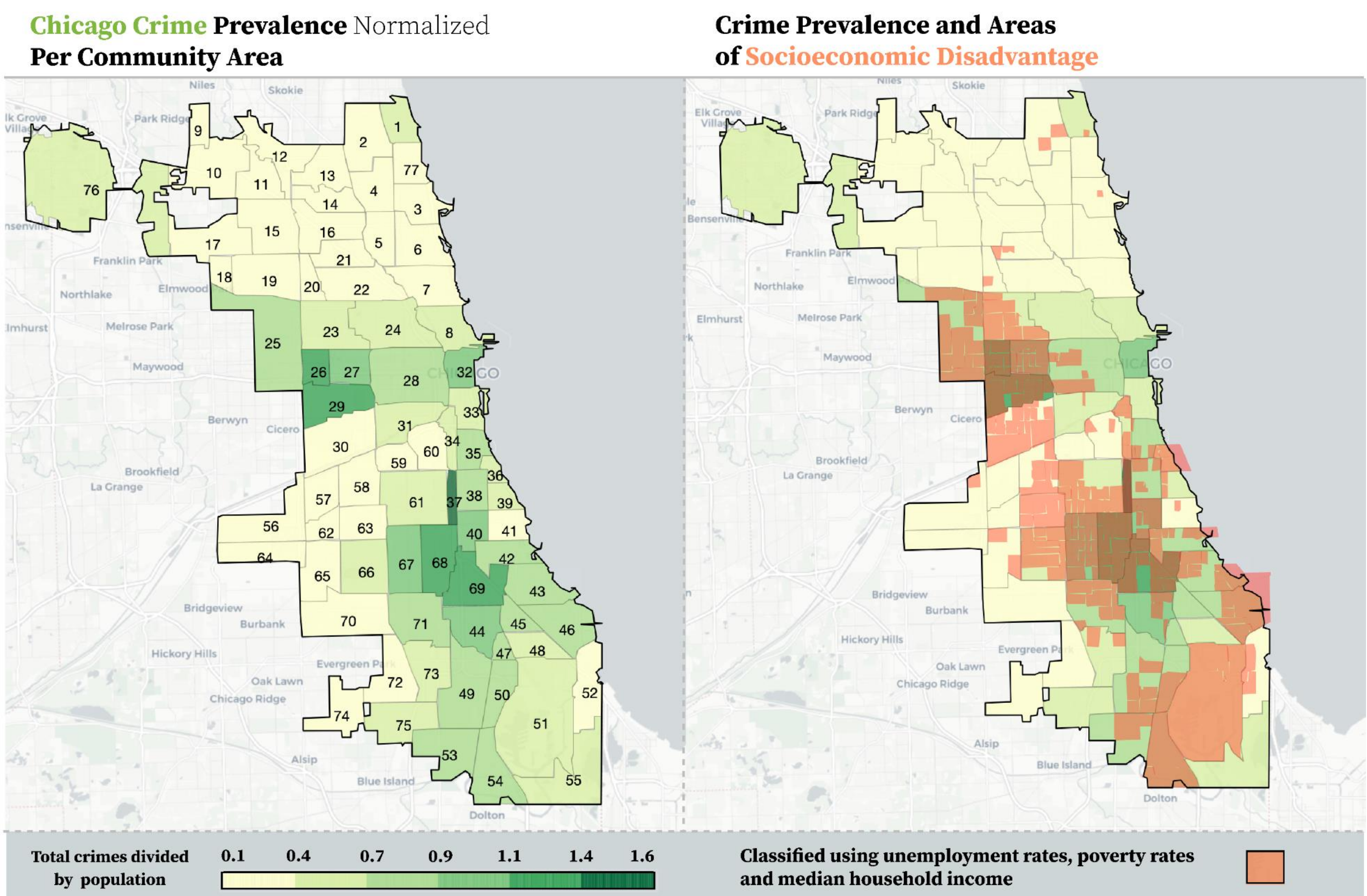

Figure 9. Chicago Crime Prevalence Per Community Area (Normalized) and Areas of Socioeconomic Disadvantage

This figure compares normalized crime prevalence across Chicago's community areas (left map) with areas classified by the City of Chicago as socioeconomically disadvantaged (right map). The left map displays crime rates adjusted for population size, with darker shades indicating higher per capita crime prevalence. The right map overlays areas of socioeconomic disadvantage in orange, defined using indicators such as household income, poverty rates, and unemployment rates.

A clear spatial correlation emerges when comparing the two maps. Community areas with high normalized crime rates—such as Fuller Park, Englewood, West Garfield Park, and Greater Grand Crossing—largely overlap with regions classified as socioeconomically disadvantaged. These findings suggest a strong relationship between socioeconomic hardship and elevated crime rates per capita, indicating that systemic socioeconomic challenges may play a significant role in shaping crime prevalence.

Conversely, more affluent or economically stable areas of Chicago, particularly in the northern and central regions, exhibit both lower levels of socioeconomic disadvantage and lower normalized crime rates. This further highlights the spatial inequality across the city, where communities facing higher poverty, unemployment, and lower household incomes experience disproportionate levels of crime relative to their population.

This analysis underscores the importance of considering socioeconomic factors when analyzing crime data and planning interventions. The alignment between crime rates and socioeconomic disadvantage highlights the need for holistic policy approaches that address the underlying social and economic challenges driving crime, rather than focusing solely on crime prevention. Such findings are crucial for guiding equitable resource allocation and addressing systemic inequalities in urban communities.

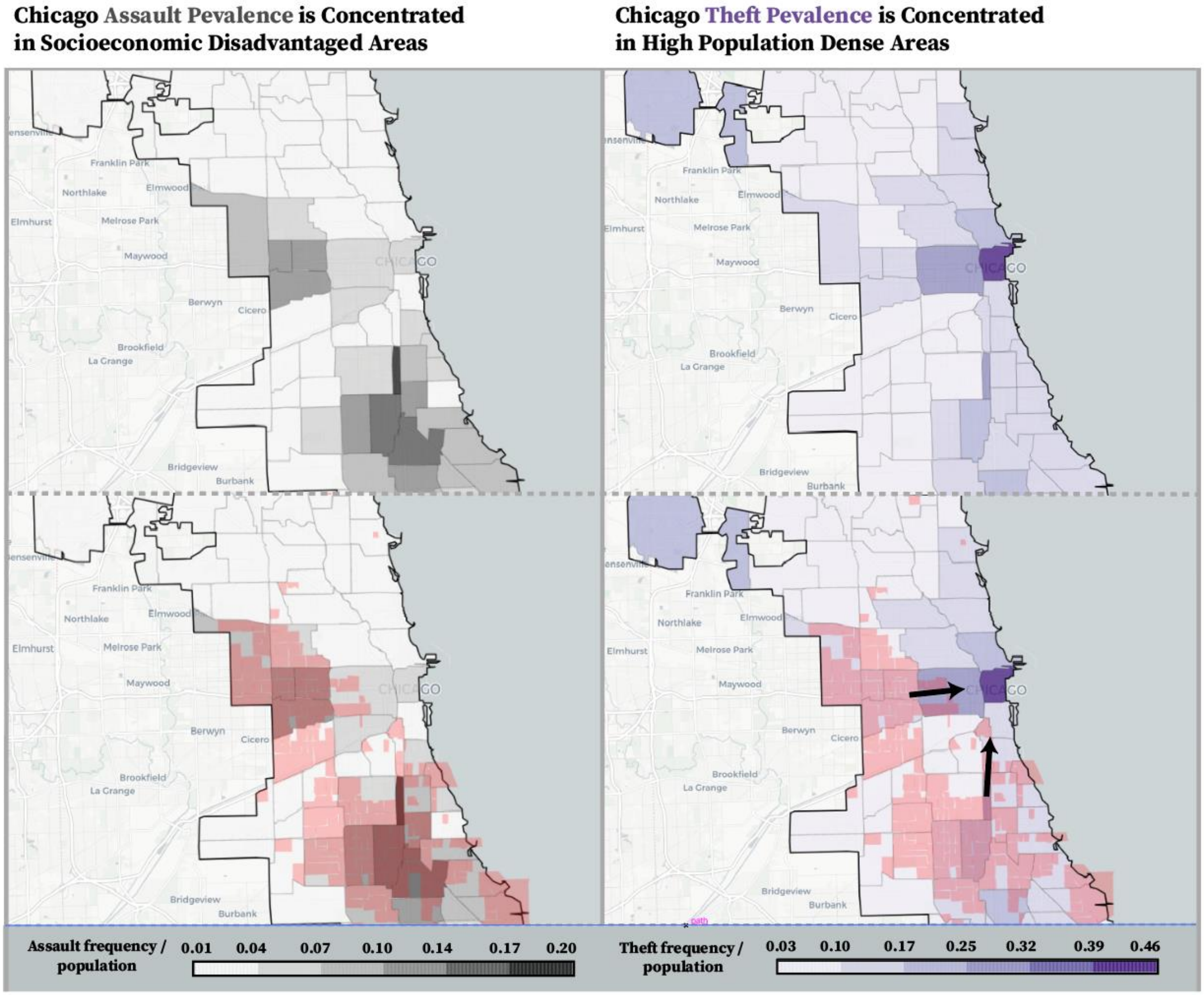

Figure 10. Chicago Assault Prevalence is Concentrated in Socioeconomic Disadvantaged Areas & Chicago Theft Prevalence is Concentrated in High Population Dense Areas

Different crimes also exhibit different spatial behaviors in Chicago, as seen here in figure 10. The left panels show the prevalence of assault across Chicago. The top-left heatmap highlights areas with the highest assault rates per capita, which are concentrated in the southern and western parts of the city. When overlaid with socioeconomic data in the bottom-left panel, a clear overlap emerges: the regions with the highest assault rates coincide strongly with areas of socioeconomic disadvantage. This suggests that socioeconomic factors such as poverty, unemployment, and lower household incomes may play a significant role in influencing the prevalence of assault.

The right panels focus on theft. The top-right map illustrates that theft prevalence per capita is primarily concentrated in areas of high population density, notably the Loop and its surrounding neighborhoods. These findings indicate that theft, unlike assault, is more closely associated with high-density commercial and residential areas where opportunities for property crimes are greater. In the bottom-right panel, the overlay of theft prevalence with socioeconomic disadvantage reveals a less pronounced relationship. While some overlap exists, the high rates of theft in the Loop and other densely populated areas suggest that theft may be driven more by urban activity and less directly by socioeconomic hardship.

Overall, the figure highlights a divergent pattern between assault and theft. Assault appears more strongly correlated with areas of socioeconomic disadvantage, reflecting broader structural inequalities. In contrast, theft prevalence aligns more closely with population density and urban activity, as seen in the concentration of theft around commercial hubs like the Loop.

Further analysis could reveal more nuanced information about crime in Chicago and other cities. For example, a clustering analysis could segment urban areas that behave similarly, and reveal key differences in crime patterns, socioeconomic conditions, or demographic factors. This approach could identify neighborhoods that share similar profiles of crime types or rates, even if they are geographically distant, allowing policymakers to implement targeted strategies that address the specific needs of these clusters. Additionally, integrating temporal data could provide insights into how crime patterns evolve over time and identify periods of increased vulnerability, such as seasonal trends or spikes during economic downturns. By combining spatial and temporal clustering with socioeconomic indicators, future analyses could offer a more comprehensive understanding of the complex factors influencing crime, ultimately leading to more effective and equitable interventions.

Conclusion

This analysis of Chicago crime data from 2019 to 2024 offers critical insights into the city's crime dynamics, emphasizing temporal, spatial, and socio-economic trends. Key findings highlight the variability of crime types, with theft being the most prevalent and motor vehicle theft witnessing a post-pandemic surge. Temporal patterns reveal distinct daytime and nighttime trends, while seasonality underscores shifts in crime activity influenced by environmental factors.

Spatial analysis uncovers a strong correlation between high crime prevalence and socio-economic disadvantage, particularly in underserved areas like Fuller Park and Englewood. The normalization of crime rates by population further identifies communities disproportionately affected, shifting the focus from densely populated areas to smaller, vulnerable regions. The divergence between assault and theft patterns demonstrates the influence of socio-economic factors versus urban density on different crime types.

These insights underscore the necessity for data-driven interventions that address underlying social and economic inequities alongside targeted crime prevention strategies. By integrating these findings into policy-making, stakeholders can foster safer, more equitable urban communities.

References

City of Chicago. (2024). *Chicago Geopandas boundary*. Retrieved from https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-City/ewy2-6yfk

City of Chicago. (2024). *Chicago socioeconomic disadvantaged areas Geopandas boundaries*. Retrieved from https://data.cityofchicago.org/Community-Economic-Dev

Illinois Action for Children. (2024). *Census data by Chicago community area*. Retrieved from https://www.actforchildren.org/wp-content/uploads/2024/03/Census-Data-by-Chicago-Community-Area-2024.pdf

Kaggle. (n.d.). *Chicago crime dataset*. Retrieved from https://www.kaggle.com/datasets/chicago/chicago-crime

Appendix:

1. Community Area and Community Area IDs

Community_Area_ID	Community_Area_Name	Community_Area_ID	Community_Area_Name	Community_Area_ID	Community_Area_Name
1	Rogers Park	21	Avondale	61	New City
2	West Ridge	22	Logan Square	62	West Elsdon
3	Uptown	23	Humboldt Park	63	Gage Park
4	Lincoln Square	24	West Town	64	Clearing
5	North Center	25	Austin	65	West Lawn
6	Lake View	26	West Garfield Park	66	Chicago Lawn
7	Lincoln Park	27	East Garfield Park	67	West Englewood
8	Near North Side	28	Near West Side	68	Englewood
9	Edison Park	29	North Lawndale	69	Greater Grand Crossing
10	Norwood Park	30	South Lawndale	70	Ashburn
11	Jefferson Park	31	Lower West Side	71	Auburn Gresham
12	Forest Glen	32	Loop	72	Beverly
13	North Park	33	Near South Side	73	Washington Heights
14	Albany Park	34	Armour Square	74	Mount Greenwood
15	Portage Park	35	Douglas	75	Morgan Park
16	Irving Park	36	Oakland	76	O'Hare
17	Dunning	37	Fuller Park	77	Edgewater
18	Montclare	38	Grand Boulevard		
19	Belmont Cragin	39	Kenwood		
20	Hermosa	40	Washington Park		