Crimes in Chicago Data Analysis Project

Sherry Au, Kevin Holt, Daniel Shock, Andrea Kang, Vincent Benoit and Trevor Breen

Executive Summary

- Crime in Chicago is a significant issue
- The “What” - types of crime:
    - The most common crimes reported in the city between 2001 and 2017 are: Theft, Battery, Criminal Damage, Narcotics, and Other Offences
        - The majority of these crimes (87%) were classified as “Non-Domestic”
    - Only 30% of crime resulted in an arrest
    - Troublingly, since 2009, there have been more homicides resulting in non-arrest than arrest
- The “When” - crime trends over time
    - Levels of crime between 2001 and 2016 have exhibited a bimodal frequency distribution (using time as the x-axis)
        - Between 2001 and 2005, crime peaked in 2001 and reached a nadir by 2004
        - Between 2006 and 2010, reported crime was back up to an all time high (peaking in 2008)
        - Between 2010 and 2011, there was a drastic reduction in crime, followed by a modest and steady decrease up until 2016
        - The frequency of most crimes appear to be decreasing, however human trafficking and concealed carry license violations seem to be increasing in frequency
    - Crime seasonality
        - On average, the frequency of reported crimes reach a low in February, peaking over the months of July and August.
    - Crime and time of day
        - Reports of crime peak between the hours of 8 pm and midnight, reaching daily lows between 4 and 6 am.
- The “Where” - where does crime occur in Chicago
    - The majority of crimes physically occur in the street, in the residence, and on the sidewalk in Chicago
    - The majority of crimes reported in the city were in the Westside, Central, and Far Southeast side

Introduction

Crime is a pervasive issue in most major urban centers of developed countries. Our understanding of the way crime develops and proliferates is underpinned by the collection of reliable data. With this data, a fuller understanding of the distribution and drivers of crime may emerge. Statistical mathematics provides a robust lens through which the patterns of crime may be studied (1). For instance, some studies have found that types of crimes cluster geographically to varying degrees (2). Others have found that the levels or rates of crime can peak during certain times of the year (including holidays, festivals, and school breaks) (3). Being able to fully describe these patterns enables cities and law enforcement agencies to fully understand the magnitude of crime in their cities, allocate resources temporally and geospatially, and, most importantly, devote resources and social support into programs aimed at neighborhood stabilization.

Objectives

For this project, data sets outlining the crimes in Chicago between the years of 2001 and 2017 (obtained as CSV files from the Kaggle website [4]) will be used for the analysis. Chicago has one of the highest crime rates in the United States of America, therefore a deeper understanding of the " what, when, where, and how" is the foundation for the design and implementation of workable solutions to the problem. Our objectives are four-fold:

1. To analyze the **categories/types of crimes** that occur in Chicago.
    - Specifically, the datasets provide crime classifications in multiple forms, including Illinois Crime Reporting codes. This crime type is captured in two variables, namely "Primary type" and "Description"
    - Further, the crimes can be identified as either domestic or non-domestic as defined under the Illinois Domestic Violence Act. This data can be used to summarize the number of crimes committed under "Primary type" and their domestic vs non-domestic status to determine the relative frequency of occurrence
    - The Chicago homicide rate is also a much discussed topic. By looking at the homicide category under "Primary type," we can investigate the frequency of homicides and where they are most likely to occur.

2. The **time of crime occurrence** will be fully explored.
    - Specifically, the dataset provides the date upon which the crimes occurred. The analysis will focus on extracting the date, month, year, and time for each crime.
    - These created variables will be used this to describe and evaluate temporal trends in the data over the various type classification of crimes.

3. The **geospatial occurrence of crime** will be fully explored.
    - The data provides several attributes which will be used to analyze the location of the crimes.
    - We will use the location description attribute to conduct analysis on where crimes typically occur such as:
        i. In residence,
        ii. On the street
        iii. Neighborhood blocks (approximate street address),
        iv. Beats which provides a small geographical area categorized by police,

v. district which provides the district categorized by police,
vi. ward which provides are categorized by city council,
vii. community area which provides areas categorized by the City of Chicago,and
viii. Latitudinal and longitudinal coordinates.

- Using these attributes, the areas with the highest crime rate will be identified. Also, the types of crime in each area, as well as whether trends over time will be evaluated.

4. The **outcomes of the crime (arrest vs non-arrest)** will be analyzed.
    - This analysis will focus on the relationship between various predictors of interest (e.g. type of crime, the location of crime, time of crime) and whether or not an arrest was made.

Hypotheses

Based information gleaned from various sources (e.g. news outlets, peer reviewed literature, Donald Trump's Twitter account), it is hypothesized that the number of crimes in Chicago has increased over the years with the type of crimes changing as well due to shifts in economic, social, ethnic, and political parameters over time. Specifically, we expect that the rates of crime and homicide have increased, and that there is a temporal relationship, with most crimes occurring at night and over the summer. There is likely a seasonal effect on where crime occurs as well - winter months likely see more indoor crimes committed. There is also likely geospatial clustering of crime in the city - most violent crime likely occurs in poorer, disadvantaged neighborhoods where social and economic support services are severely lacking.

Analyzing the "what, when, where, how" and studying the risk factors for arrests made will provide vital insight into the nature and trend of crimes committed in Chicago.

Data Preparation

General data preparation

For the analysis, datasets created from the Chicago Police Department's "Citizen Law Enforcement Analysis and Reporting" system were downloaded as CSV files from Kaggle (4). The datasets (a total of 4 - consisting of years 2001-2004, 2005-2007, 2008-2011, and 2012-2017) were downloaded from the Kaggle website and concatenated to produce a single dataframe upon which to work. The pandas functions info(), head() and isna() were used to better understand the data. The columns 'Updated on' and 'Case Number' were removed, as they added no value to satisfy project objectives (i.e. the "what, when, and where" questions). The 'IUCR' column was also removed as this identifier was used to identify 'Primary Type' and 'Description' attributes but the identifier itself does not provide additional information for our analysis.

Preparation of location data

Data was explored by neighborhood, ward, community zone, and latitude/longitude. Python was used to describe the distribution of crime across the city of Chicago. Due to the high number of null values present in the data for the variables ward, community zone, and address, the latitude and

longitude variables were used to visualize the geographic distribution of crime across the city using using Tableau Public.

Preparation of time data

For the analysis of timing of crimes, the data provide the month, day, year and time in the 'Date' field. String split methods (str.split()) were used to separate month, day, and year using "/" as our delimiter and using " " as the delimiter to generate a time variable. New attributes for day, month, year and time were created to facilitate further time analysis.

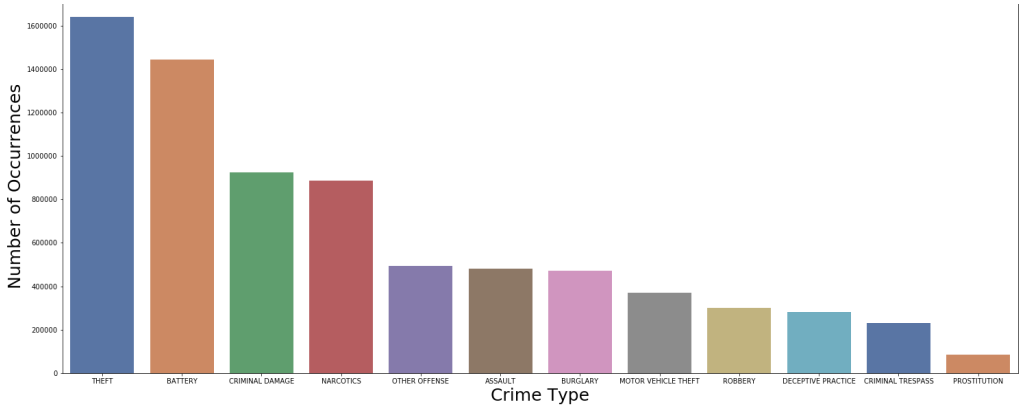
Comments on data quality

The quality of the data was generally quite good. The majority of rows for several variables contained non-null values, especially with respect to **primary crime type and description** (convenient, as these were outcomes of concern used in the analysis). The variables outlining geographic information presented a challenge for the analysis. The specific address of the crime was partially redacted to protect the identity of those involved - this variable was of limited value. Another concern pertained to the variables for ward and community area - both variables had null values in over half of their observations. Many location variables, such as beat, ward, community area and district are assigned identification codes by the Chicago Police department and the City of Chicago. This required additional research to understand what these variable values/attributes represented.

Data Analysis/Results

Crime Categories/Arrests

Figure 1. Top 12 Primary Types of Crime, by Frequency



There were over 20 different primary types of crimes, with the top 12 primary types of crimes (by frequency) committed in the city of Chicago between the years 2001 and 2017. Overall, the top 5 crimes were: Theft, Battery, Criminal Damage, Narcotics, and Other Offences (Figure 1). Homicides was not one of the top 12 crime categories (in terms of frequency) - it was actually the 21st most frequent type of crime, which may be surprising given the media narrative on the high homicide rate in Chicago.

Table 1. Summary of Type Crime Descriptions Under Primary Type "Other Offense"

	Primary Type	Description	count
290	OTHER OFFENSE	TELEPHONE THREAT	164930
272	OTHER OFFENSE	HARASSMENT BY TELEPHONE	121077
281	OTHER OFFENSE	OTHER VEHICLE OFFENSE	41486
294	OTHER OFFENSE	VIOLATE ORDER OF PROTECTION	38193
266	OTHER OFFENSE	FALSE/STOLEN/ALTERED TRP	25678
271	OTHER OFFENSE	HARASSMENT BY ELECTRONIC MEANS	23084
280	OTHER OFFENSE	OTHER CRIME INVOLVING PROPERTY	13749
282	OTHER OFFENSE	OTHER WEAPONS VIOLATION	9771
292	OTHER OFFENSE	VEHICLE TITLE/REG OFFENSE	8142
279	OTHER OFFENSE	OTHER CRIME AGAINST PERSON	7541

The Primary Type "Other Offense" provides little information on the type of crime committed. Indeed, it was used as a "catch-all" categorization that described miscellaneous crimes not falling under other categories. Table 1 outlines the ten most frequent description of crimes falling under the "Other Offence" umbrella.

Figure 2: Proportion of Non-Domestic vs. Domestic Crime

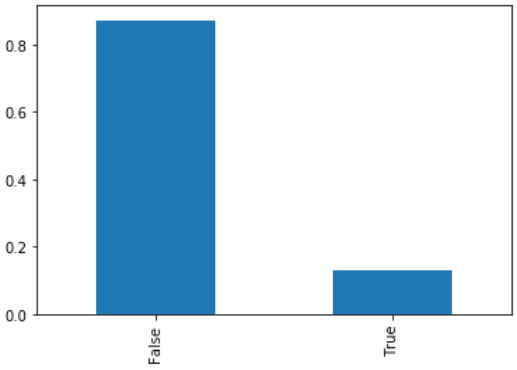
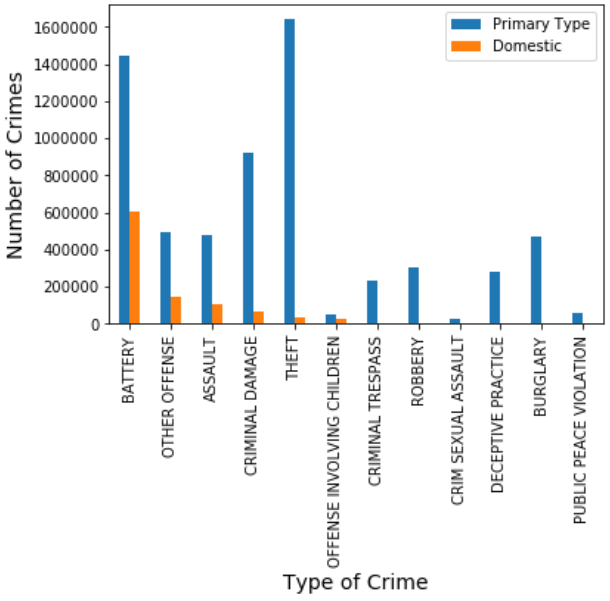


Figure 3: Top 12 Crimes by Total Instances of Domestic Crime

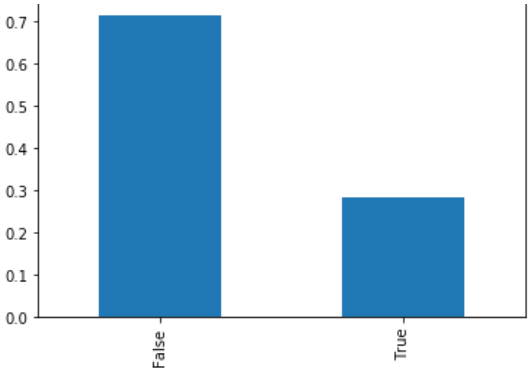


w

The dataset further provides insight on whether a committed crime was "Domestic" or in nature. Only 13% of all crimes were categorized as domestic (with the remaining 87% classified as

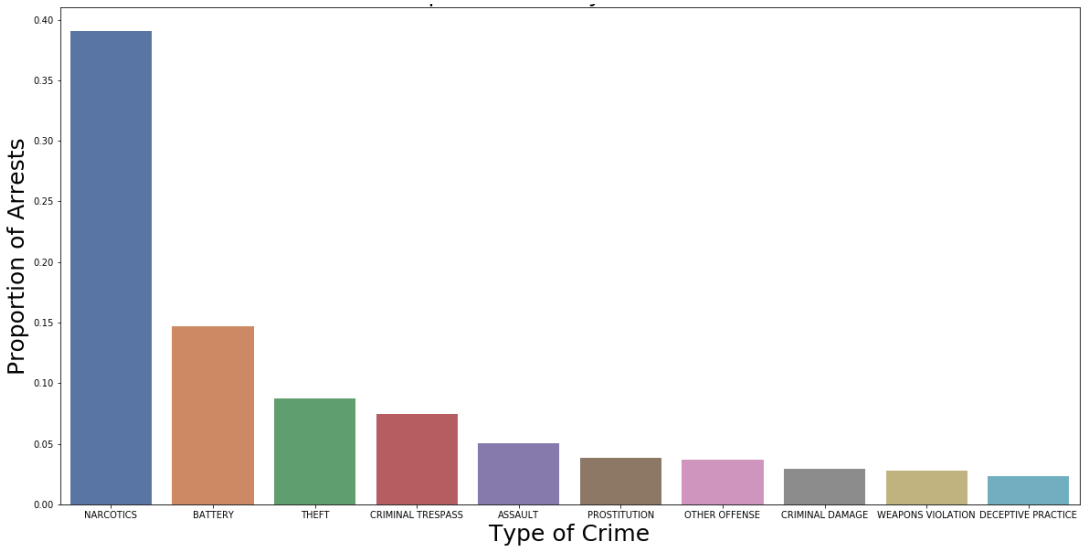
Non-Domestic). Taking a deeper look, the primary categories of domestic crime include battery, “Other Offence”, assault, and criminal damage (Figure 3). This provides additional context to the nature and motives of crimes committed, as we can see only a small proportion of crime is considered “Domestic”.

Figure 4: Proportion of Crimes Resulting in Arrest



Another objective of the analysis was to evaluate whether or not the reported crime resulted in an arrest. In many cases, the crime was simply not serious enough to be deemed an arrestable offence. In other cases, the perpetrator was likely simply not arrested. Figure 4 outlines the relative proportion of crimes that result in an arrest (30%).

Figure 5: Top 10 Crimes by Arrest Rate



When evaluating only those reported crimes that resulted in an arrest, the “top 12” list differs from Figure 1. Specifically, Narcotics, Battery, and theft are the top 3 crimes when analyzing only those observations that resulted in an arrest (Figure 3). It’s interesting to note that Narcotics has a significantly higher arrest rate of nearly 40%, while more violent crimes like battery and assault are 15% and below. Further, homicide doesn’t even make the top 10, due to its extremely low arrest rate.

Figure 7: Chicago Annual Homicides by Arrest vs. Non-Arrest from 2001 to 2017

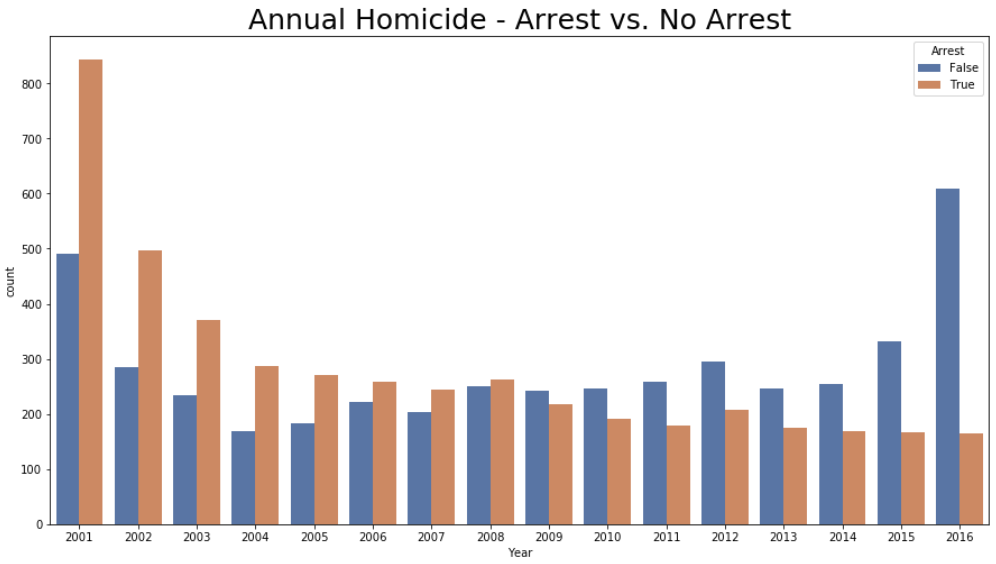
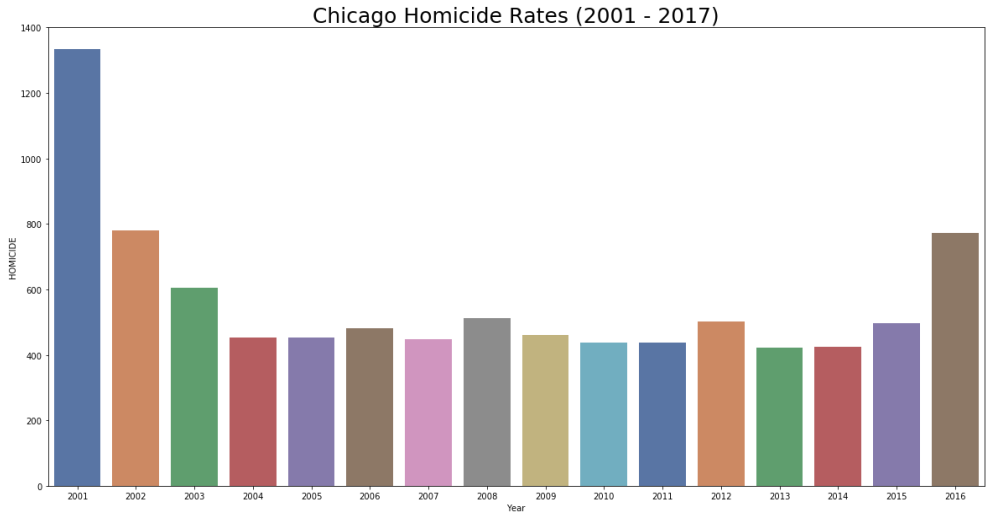


Figure 6: Chicago Homicide Rates from 2001 to 2017



Chicago has one of the highest total number of murders per year in the United States of America (https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/tables/table-8/table-8.xls/view). As such, a deeper analysis of homicide rates was conducted. Figure 6 outlines the annual murders recorded in Chicago between 2001 and 2017. Of note, the dataset is incomplete for 2017, hence the low number of homicides recorded. The figure indicates there was a significant spike of over 1,300

homicides in 2001, with the rate dropping and leveling to 500 murders/year starting by 2004 and continuing to 2015. The year 2016 saw an increase in the number of homicides to nearly 800.

For over 60% of homicides, a perpetrator was arrested. This trend was particularly evident between the years of 2001 and 2008. However, in 2009 the proportion of homicides with an arrest began to shift - relatively fewer homicides committed were associated with arrests (Figure 7). In 2009, the number of non-arrests first eclipsed the number of arrests, and the gap has widened since. In 2016, over 70% of homicides occurred without an arrest. Some of this shift may be driven by the lag in time between the reporting of a crime and when the crime is committed.

Time Series

Another objective of the analysis was to explore the timeframe of crimes in Chicago, including analyzing yearly, seasonal, and time of day trends across the years, stratified by different types of crimes.

Figure 8. Number of Crimes in Chicago between 2001 and 2016.

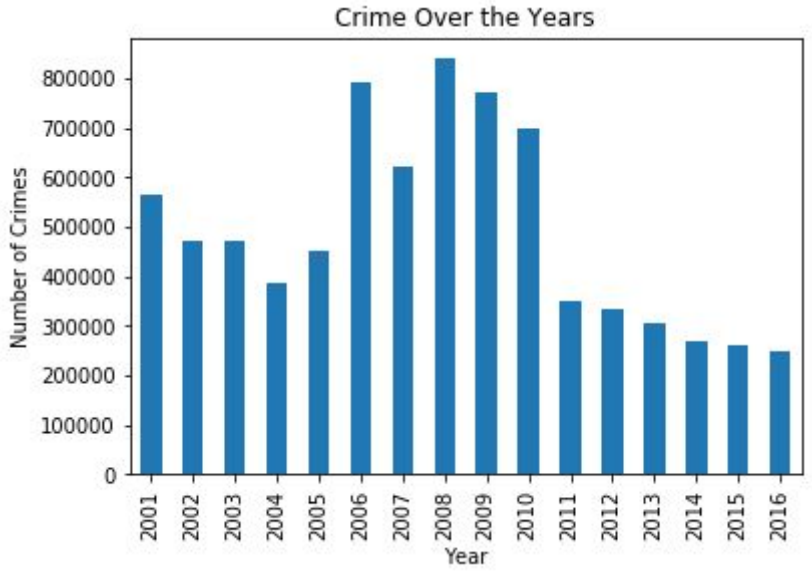
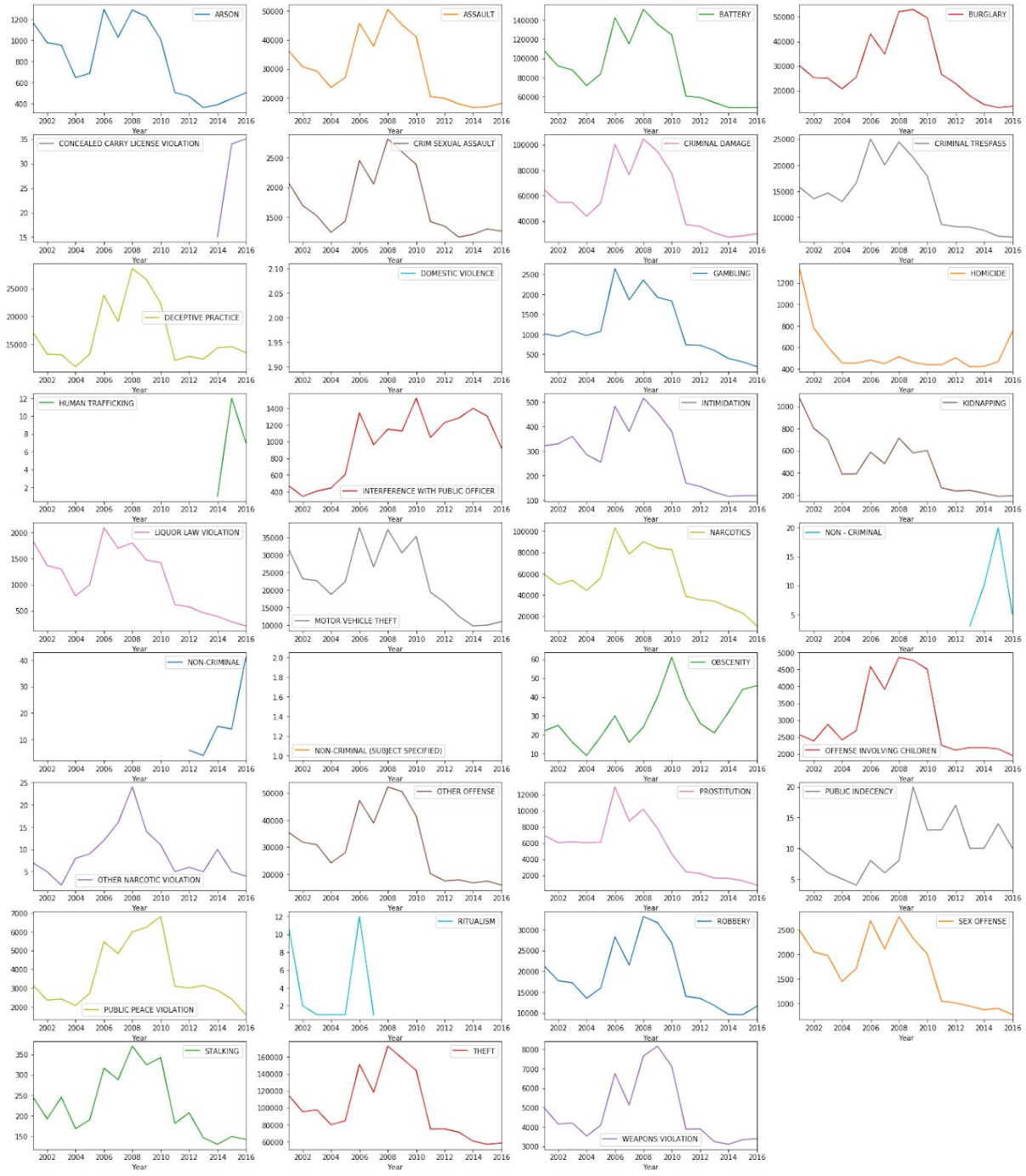


Figure 8 outlines the trend of crimes between 2001 and 2016. Our analysis discards the last year of our sample, i.e. 2017, given only data on homicides were available. Three distinct phases stand out over the period studied:

1. A downward trend took place in the early 2000s (2001-2004);
2. Crimes precipitously increased in 2006, stabilizing around 800,000 crimes per year for the following 5 years; and
3. A downward trend after 2008after 2008

The yearly occurrence of crime by type is outlined in Figure 9. Most types of crime have seen falling crime rates (similar trend to Figure 8), but some exceptions exist. For instance, the homicide rate was at peak levels in 2001 and declined quickly until stabilizing at lower levels in 2004. There appears to be an increase in the number of yearly homicides in recent years (2015 and 2016). As well, interference with public officer, public indecency, and obscenity all appear to be rising. One other interesting observation is that it appears as though there are a few crimes (human trafficking and concealed carry license violations) for which the Chicago have just begun collecting data for (or classifying them as such).

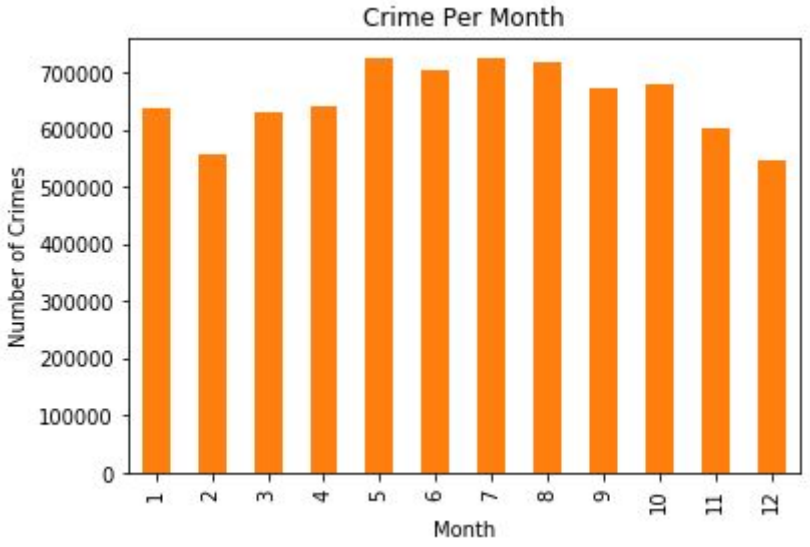
Figure 9. Yearly number of crimes committed in Chicago between 2001 and 2016, stratified by crime type.



The next step was to investigate whether there were seasonal effects in crimes committed. Based on the monthly crime rates (Figure 10), there appear to be seasonality in crime rates: crime rates reach their nadir in February and then spike from July – August (during summer). Some have argued that higher temperatures and longer daylight hours can lengthen the amount of time people spend

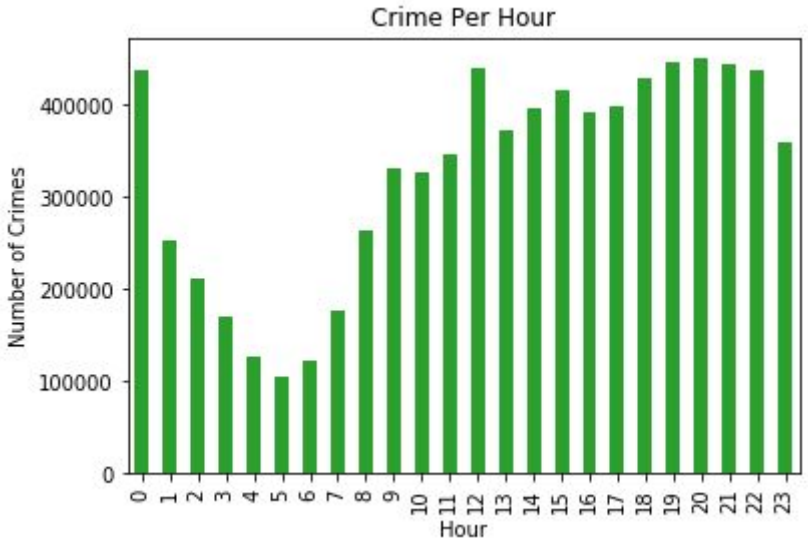
away from their homes. This in turn could lead to more people spending more time in public (and away from their homes). More empty homes could lead to more thefts and burglaries (3). The heat and humidity may also contribute to irritability, stress, and aggressiveness, priming those predisposed to violence (5).

Figure 10. Monthly crime rates in Chicago by month in which the crime was reported (2001-2016)



Time of day also affects the amount of crime committed. Times between 8 pm to midnight are the most dangerous times in Chicago; whereas between 4 am and 6 am were the safest hours (Figure 10). As expected, crimes in Chicago mainly happen late in the day and decline in the early morning. This trend fits with previous research (5). One interesting future analysis might attempt to add a denominator to the data - Look at the crime rates controlling for the number of people that are actually awake during those hours. Perhaps, after controlling for the fact that there are likely fewer people awake in the morning, this rate might be relatively higher than it is currently portrayed as being. Such data, though useful, may prove difficult to collect, but may uncover interesting conclusions.

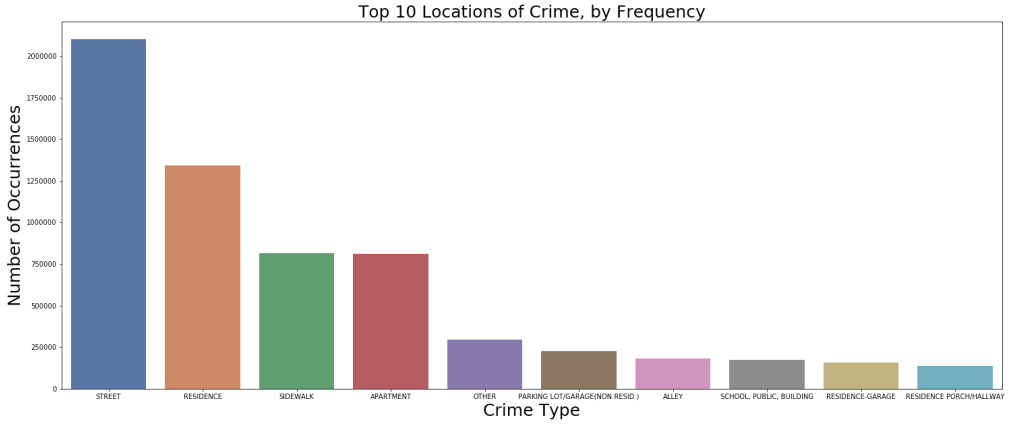
Figure 11. Crimes committed in Chicago by hour of the day (24 hour clock), years 2001-2016



Geographical/Location

One major objective of the analysis was to describe where crimes were most likely to take place. Generally speaking, the "Location Description" field indicated the category of location where the incident occurred (e.g. on the street, in a residence, on the sidewalk etc...). Figure 12outlines the top 10 locations for crime with "Street" (over 60%), "Residence" and "Sidewalk" rounding out the top 3 locations of crime on the list.

Figure 12. The top 10 physical locations where crime occurred in Chicago between 2001 to 2017.



Geographically, the variable "Community Area" aids in understanding the "where" of crimes in Chicago. Chicago has a total of 77 community areas - geographic divisions that cover multiple neighborhoods. Table 2 outlines the ten areas with the highest incidence of crime between 2001 and 2017. Austin, the South Shore, and the Near North Side are the 3 community areas with the highest incidence of crime in Chicago. Recall, however, that the variable Community Area had more than half of all observations with null values. No definitive conclusions can be made regarding location based solely on 50% of the data.

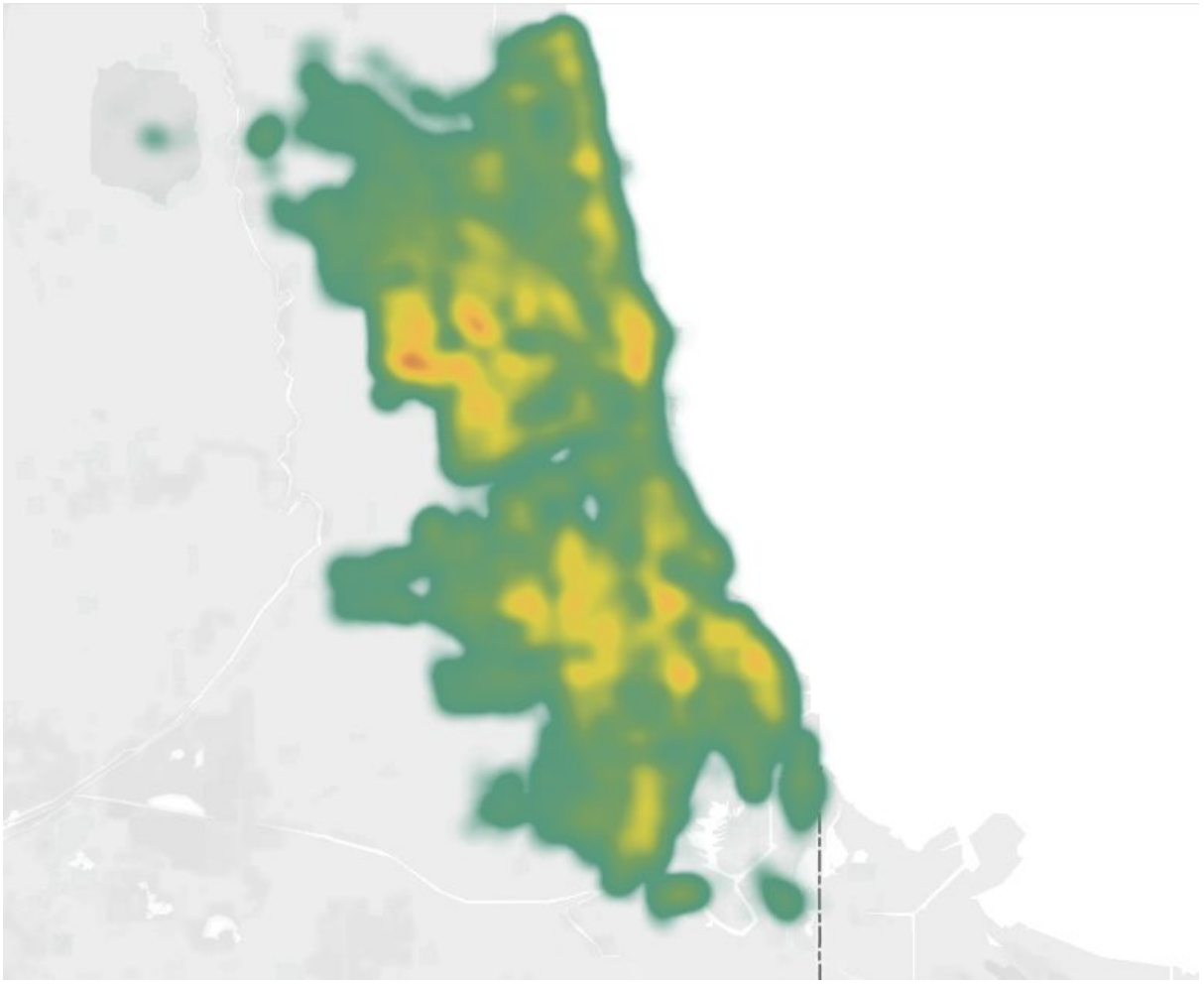
Table 2. The 10 community areas of Chicago with the highest incidence of crime, 2001-2017.

Community Area	Community Area Name	Incidences of Crime
25.0	Austin	467069
43.0	South Shore	237361
8.0	Near North Side	235378
23.0	Humboldt Park	229000
67.0	West Englewood	216979
24.0	West Town	213776
71.0	Auburn Gresham	207155
28.0	Near West Side	205913
29.0	North Lawndale	202866
68.0	Englewood	197639

With this in mind, the analysis focused on latitude and longitude when evaluating the relative geospatial concentration of crime in Chicago. Figure 13 outlines a heatmap representation of the relative frequency of crime across the Community Zones of Chicago (Figure 14). When comparing the community areas with the heatmap, there is significant agreement. The community zones (Figure 14) Austin, Humboldt Park, and West Town geographically correspond with the areas of high crime frequency featured in Figure 13.

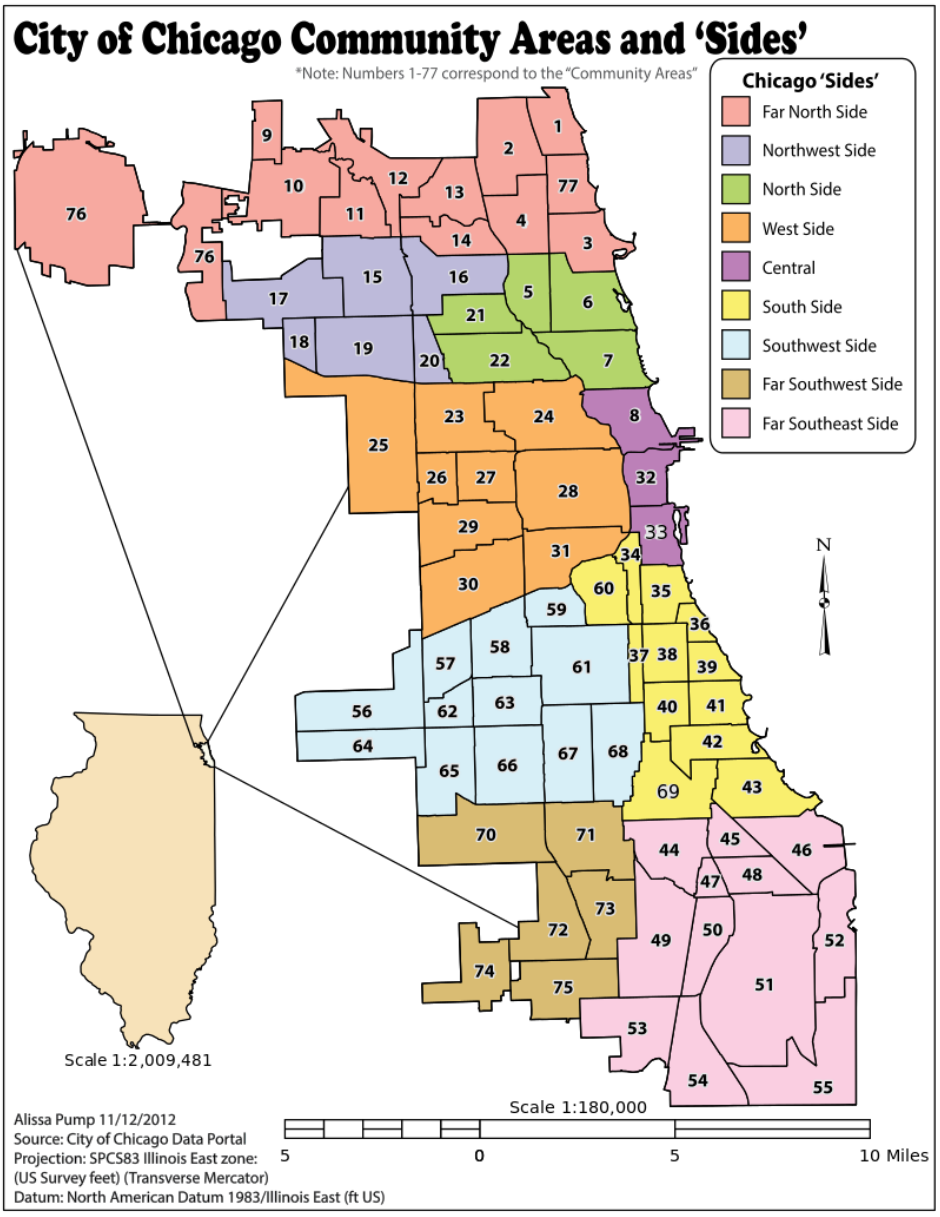
Figure 13. Heat map of crime frequency in Chicago (2001-2017)

Group 17: Crimes in Chicago Data Analysis



Heatmap can be found at:
https://public.tableau.com/profile/andrea8193#!/vizhome/Book1_15643448908320/Sheet1

Figure 14. Map of Chicago Community Areas



Attribution: By Alissapump - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=22655083

Conclusion

Crime in Chicago is a significant source of concern for public safety and wellbeing. In an age where the media can highlight, magnify, and sensationalize issues, it is of paramount importance for government and law enforcement officials to focus on evidence-based approaches to understanding issues. Crime is no exception to this rule. To make meaningful decisions regarding the allocation of resources and money, cities and law enforcement agencies must understand the "who, what, when, and where" of crime in their jurisdiction. This was the objective of the current study.

First, the study sought to understand what types of crimes were most common in Chicago between 2001 and 2017. The top 5 classes of crimes reported in Chicago were: theft, battery, criminal damage, narcotics, and other offences. The majority of crimes were classified as "non-domestic"

(87%). Interestingly, only about one third of reported crimes resulted in an arrest. Perhaps more troubling was the fact that greater than 50% of all homicides committed in the city in recent years DID NOT result in an arrest. This analysis begs the question as to why there was such a low arrest rate, especially concerning homicides in Chicago. Future research should focus on more fully understanding why this was so.

Second, the analysis sought to understand when crimes occur over time. Over the years of the study, there were three "phases" of crime level in Chicago. Between 2001 and 2005, crime levels significantly declined, from a high in 2001 to a nadir in 2004. The second phase was between 2006 and 2010, seeing crime levels reach an all-time high in 2008. Then, between 2010 and 2011, there was a drastic reduction in crime levels. Thereafter, the frequency of crime reports have been constantly decreasing, year-over-year. There was also a seasonal trend noted across the years of analysis. Reported crime levels were highest in the summer months (July-August) and lowest in February. Some have postulated that, as people are more likely to be out of their house and in public areas during the summer, this sets up the perfect situation (empty houses) for burglary and theft. Additionally, hot weather may promote feelings of irritability and aggression in those predisposed to violence. More research is necessary.

Looking at the time of day that crimes are typically committed, levels peak between 8 pm and midnight. The cloak of darkness likely fosters criminal behavior and enables crime. Whereas, the lowest levels of crime were between the hours of 4 and 6 am. This is likely driven by the majority of the population being asleep during these hours.

Finally, the study focused on describing where crime was most likely to occur in Chicago. The majority of crimes occurred in the street, in the residence, and on sidewalks in Chicago. When evaluating the heat map for determining the geospatial distribution of crime, the more complete results using latitude and longitude matched with the analysis describing community areas/zones, despite more than 50% of the observations for community area being "null". The areas of highest crime reporting in Chicago were in the Westside, Central, and Far Southeast side of the city.

Overall, the study provides preliminary information into the "what, when, and where" of crime in Chicago. Armed with this information, legislators and enforcement officials can better allocate resources to prevent and manage crime.

References

1. D'Orsogna MR, Perc M. Statistical physics of crime: A review. Phys Life Rev [Internet]. 2015;12:1–21. Available from: http://dx.doi.org/10.1016/j.plrev.2014.11.001
2. Oliveira M, Bastos-Filho C, Menezes R. The scaling of crime concentration in cities. PLoS One. 2017;12(8):1–13.
3. Towers S, Chen S, Malik A, Ebert D. Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective. PLoS One. 2018;13(10):1–27.
4. City of Chicago. Crimes in Chicago: An extensive dataset of crimes in Chicago (2001-2017). Kaggle [Internet].Available from: https://www.kaggle.com/currie32/crimes-in-chicago
5. Carbone-Lopez K, Lauritsen J. Seasonal variation in violent victimization: Opportunity and the annual rhythm of the school calendar. Journal of quantitative criminology. 2013; 29(3):399–422. https://doi.org/ 10.1007/s10940-012-9184-8
6. OJJDP Statistical Briefing Book. Online. Available: https://www.ojjdp.gov/ojstatbb/offenders/qa03401.asp?qaDate=2016. Released on October 22, 2018