

# The Effect of Mayoral Changes and Locality on Chicago Crime Rates

By: Mark Belsis, Alex Funches, Ivan Murillo  
Group 21

## ***Introduction***

For the project we are using two datasets that contain observations representative of crime incidents in the City of Chicago in the years of 2010 and 2018. These datasets were extracted by the city of Chicago from the Chicago Police Department's CLEAR system or the Citizen Law Enforcement Analysis and Reporting system of Chicago and uploaded to the Chicago Data Portal.

Chicago Crime Data 2010: <https://data.cityofchicago.org/Public-Safety/Crimes-2010/q4de-h6yq>

Chicago Crime Data 2018: <https://data.cityofchicago.org/Public-Safety/Crimes-2018/3i3m-jwuy>

The aim of this project is to investigate how a mayor's policies may affect crime rate and the setting it occurred in. Desiring to see this visually in a dataset the final year of each of the previous two mayors' terms were selected which would allow for quantitative comparison and help us capture any effect their terms and enacted policies may have had on the crime rate. This research interest guided our questions of interest which includes finding the wards which have the highest number of reported crimes, their type, and if crimes are more likely to occur in a particular location.

## ***Methods***

Both files have a similar structure and most of the same variables. The Chicago crime data from 2010 has 370,375 observations across 18 variables while the Chicago crime data from 2018 has 267,985 observations across 22 variables. The datasets share 18 variables with descriptive information on the crime type, date of occurrence, and location both in the general area with a ward provided and exact coordinates such as the block and longitude and latitude geographic coordinates. Additionally, the crimes are further categorized by description and

categorical variables are used for if the crime was domestic or resulted in an arrest. For identification the variable Case Number is used in both datasets with the addition of ID for the 2018 dataset.

Both data files were merged relatively easily by sorting by Date and then simply merging the data files in a new DATA step. 13 of the 22 variables contain categorical, character variables that are used for identification and to describe the location and date of occurrence for the crimes committed. The remaining 9 variables are numeric representations of the police geographical area, community, district, and geographic coordinates used for identification and since none of these variables are numeric counts, we did not examine the means or range of these variables. Instead we started our validation and cleaning process by looking for missing values and assessing whether they can be inferred or corrected. Using an SQL query, we find that there are a couple of thousand missing values in the XCoordinate and YCoordinate variables but since these variables won't be utilized, we can ignore them. We do have 1,016 missing values for the LocationType variable, and it seems that they are usually accompanied by a 'Deceptive Practice' Offense with a 'Financial Identity Theft' description so it may be that these crimes don't exactly have a physical location which means we can't really fix them. There are also 22 missing values for Ward which is the City Council District the crime took place in, but we believe that these 22 values are missing because the crimes occurred in a location outside of the City Council districts. There is also a single missing value for CaseNumber but given that we don't know if the CaseNumber follows any specific assignment procedure we will just ignore this observation.

Overall these missing values shouldn't pose too much of a problem in our analysis as the merged dataset contains 638,360 observations and the missing values only amount to around 1% of the observations.

Another aspect that was examined were the individual levels of both CaseNumber and Ward as we have some idea of how many levels there should be. For CaseNumber we expect each one to be unique, but we find that using a FREQ procedure shows us that there are only 638,304 non-missing values. With 1 missing we find that 55 Case Numbers are not unique however this can be explained as in the description of dataset it is indicated that in cases of murder each victim's information is recorded and if we check the Case Numbers we do indeed see that the crimes were murders for these duplicate case numbers. For Ward we see that there are 50 non-missing levels which is what we want as there are 50 wards represented.

The variables that we will be using for our analysis are highlighted in the 'Variable Levels' table in the results section. Offense and LocationType are categorical variables while Ward and Year are numeric identifiers.

## Results

### Missing Values in the merged dataset

Total number of the observations	Number of the missing values for Case Number	Number of the missing values for Date	Number of the missing values for Block	Number of the missing values for IUCR	Number of the missing values for Offense	Number of the missing values for Description	Number of the missing values for LocationType	Number of the missing values for Arrest	Number of the missing values for Domestic	Number of the missing values for Beat
638360	1	0	0	0	0	0	1016	0	0	0

Number of the missing values for Ward	Number of the missing values for Code	Number of the missing values for XCoordinate	Number of the missing values for YCoordinate	Number of the missing values for Year	Number of the missing values for Latitude	Number of the missing values for Longitude	Number of the missing values for Location
22	0	4925	4925	0	4925	4925	4925

Variable Levels

Number of Variable Levels				
Variable	Label	Levels	Missing Levels	Nonmissing Levels
<b>CaseNumber</b>	Unique ID of case incident	638305	1	638304
<b>Date</b>	Date of crime	252678	0	252678
<b>Block</b>	Block of incident	31278	0	31278
<b>Offense</b>	Classification of crime	33	0	33
<b>Description</b>	More in depth description of the crime	358	0	358
<b>LocationType</b>	Description of location such as Street, Apartment, Residence, etc.	149	1	148
<b>Location</b>		243826	1	243825
<b>IUCR</b>	Illinois Uniform Crime Reporting code to classify criminal incidents	356	0	356
<b>Arrest</b>	TRUE or FALSE variable on whether an arrest was made	2	0	2
<b>Domestic</b>	TRUE or FALSE variable on whether the crime was domestic	2	0	2
<b>Beat</b>	Indicates beat (police geographical area) in which crime occurred.	303	0	303
<b>Ward</b>	An numerical ID for a local authority area	51	1	50
<b>Code</b>	FBI Code of Offense	26	0	26
<b>XCoordinate</b>		61608	1	61607
<b>YCoordinate</b>		97252	1	97251
<b>Year</b>	Year of crime	2	0	2
<b>Latitude</b>		243738	1	243737
<b>Longitude</b>		242425	1	242424
<b>ID</b>	ID of observation	267986	1	267985
<b>Update</b>	Date crime record was updated	1389	1	1388
<b>District</b>	District Number crime occurred in	24	1	23
<b>Community</b>	Community Crime occurred in	79	1	78

As we see in the variable's levels table, CaseNumber has its one missing value and its 55 repeated instances which we don't want to change. LocationType has 149 different levels with one being missing which again given that the crimes associated with these instances don't have a specific physical location it is understandable for them to be missing.

The first question we aimed to answer was whether crime rate changed across mayors and this was accomplished using a simple FREQ procedure. Below we see the overall crimes committed in each year and we find that from 2010 to 2018 there is indeed a significant drop in crime, about a 27.6% decrease to be exact.

*Number of Crimes Committed by Year*

Year of crime		
Year	Frequency	Percent
2010	370375	58.02
2018	267985	41.98

The second question we aimed to answer was which Wards have the highest crime rate, what types of crimes are committed the most in these high risk Wards, and are these results consistent across years. On the next page you'll see a breakdown of the top 5 Wards with the highest crime rates in each year. The first thing to notice in the Top 5 Ward tables is that the top 5 highest crime rate Wards were the same in 2018 as they were in 2010 just in a different order. Its also important to note that each Ward experienced an overall decrease in crimes committed except Ward 42 which went from 3<sup>rd</sup> to 1<sup>st</sup> from 2010 to 2018 and had an increase of 4,711 crimes committed. It's also interesting to note that while counts dropped, except for Ward 42, the percent share in each Ward of the overall crimes committed in all Wards rose across the board.

Top 5 Wards: 2010

Ward	COUNT	PERCENT
28	16900	4.56316
24	14210	3.83684
42	13443	3.62974
2	13388	3.61489
27	12991	3.50770

Top 5 Wards: 2018

Ward	COUNT	PERCENT
42	18154	6.77436
24	12652	4.72123
28	11935	4.45367
27	11241	4.19470
2	10080	3.76146

Below and on the next page you'll see the top 5 offenses committed in each of the top 5 Wards. For the most part, the type of offense committed in each Ward was relatively consistent and this is especially true for Wards 24 and 27 which had the same top 5 offenses in in the same order across 2010 and 2018 just with different counts. Ward 28 was fairly consistent as it just had a change of order for its top 5 offenses with Narcotics and Battery switching places from 2010 to 2018. Wards 2 and 42 were less consistent as they saw a change in some of the offense listed in the top 5 from 2010 and 2018. Ward 2 had Narcotics and Other Offense drop out of the top 5 and in their place was Deceptive Practice and Assault offenses. For Ward 42 there was only a shift away from Criminal Trespass offenses toward Assault offenses.

Top 5 Offenses (Ward 2): 2010

Offense	Ward	COUNT
THEFT	2	4442
BATTERY	2	1902
NARCOTICS	2	1322
CRIMINAL DAMAGE	2	1075
OTHER OFFENSE	2	767

Top 5 Offenses (Ward 2): 2018

Offense	Ward	COUNT
THEFT	2	3835
BATTERY	2	1426
DECEPTIVE PRACTICE	2	1077
CRIMINAL DAMAGE	2	880
ASSAULT	2	699

Top 5 Offenses (Ward 24): 2010

Offense	Ward	COUNT
NARCOTICS	24	3214
BATTERY	24	3007
THEFT	24	1786
CRIMINAL DAMAGE	24	1301
ASSAULT	24	878

Top 5 Offenses (Ward 24): 2018

Offense	Ward	COUNT
NARCOTICS	24	2567
BATTERY	24	2544
THEFT	24	1359
CRIMINAL DAMAGE	24	1134
ASSAULT	24	890

Top 5 Offenses (Ward 27): 2010

Offense	Ward	COUNT
THEFT	27	2983
NARCOTICS	27	2527
BATTERY	27	2188
CRIMINAL DAMAGE	27	1242
ASSAULT	27	677

Top 5 Offenses (Ward 27): 2018

Offense	Ward	COUNT
THEFT	27	3098
BATTERY	27	1755
NARCOTICS	27	1625
CRIMINAL DAMAGE	27	879
ASSAULT	27	676

Top 5 Offenses (Ward 28): 2010

Offense	Ward	COUNT
NARCOTICS	28	4686
BATTERY	28	3290
THEFT	28	1931
CRIMINAL DAMAGE	28	1225
ASSAULT	28	963

Top 5 Offenses (Ward 28): 2018

Offense	Ward	COUNT
BATTERY	28	2637
NARCOTICS	28	1969
THEFT	28	1531
CRIMINAL DAMAGE	28	1128
ASSAULT	28	954



Top 5 Offenses (Ward 42): 2010

Offense	Ward	COUNT
THEFT	42	7033
DECEPTIVE PRACTICE	42	1351
BATTERY	42	1235
CRIMINAL DAMAGE	42	587
CRIMINAL TRESPASS	42	571

Top 5 Offenses (Ward 42): 2018

Offense	Ward	COUNT
THEFT	42	9348
DECEPTIVE PRACTICE	42	3017
BATTERY	42	1821
CRIMINAL DAMAGE	42	722
ASSAULT	42	696

For the final question, we sought out to answer what locations were the most likely for a crime to occur. Below we see a top 10 list of the highest occurring locations for a crime to take place. With nearly a quarter of the total crimes, the street is the most likely place for a crime to occur followed by residences and apartments. These three locations alone account for over 50% of the total crimes that occurred. More distinct designations could be helpful in examining the location of crimes since there is some ambiguity in the difference between a street and a sidewalk as both are listed in the top 10 list.

Top 10 Locations for a Crime to Occur

LocationType	COUNT	PERCENT
STREET	152102	23.8650
RESIDENCE	106068	16.6422
APARTMENT	77357	12.1374
SIDEWALK	65095	10.2135
OTHER	23319	3.6588
PARKING LOT/GARAGE(NON.RESID.)	17943	2.8153
ALLEY	13515	2.1205
SMALL RETAIL STORE	13428	2.1069
RESTAURANT	12506	1.9622
VEHICLE NON-COMMERCIAL	11620	1.8232

