INFO 250

8/28/18

Analyzing the Crime in Chicago from 2012-2017 (Working Title)

James Hemmen, Raj Patel, Hong Son

**Abstract:**

Year over year, Chicago has been reported as one of the most dangerous cities in the United States. With a notorious volume of gang and general crime violence, Chicago has seen patterns and possible foreseeable trends. Numerous variables can affect the frequencies observed. The research presented explores crime in Chicago in terms of location, time, the act of crime, and the reasons for changes and stagnation in crime.

**Introduction**

Chicago has a representation of being a violent state where the crime rate is one of the highest, higher than the US average. The Chicago Polent’s Bureau of Records pushlishes records of incidents beginning in the 20th century that are released to inform the public of the safety in Chicago. The report shows starting in this period (2012-2017), the number of crimes have surged up in the mid 2010s from prior. For the dataset, we wanted to find out more about the crimes, such as types of crimes or where and when the crimes happen, and possible consequences such as arrest rates from the crimes in Chicago.

The dataset was obtained from Kaggle that was obtained from the Chicago Police Department’s CLEAR (Citizen Law Enforcement Law Enforcement Analysis and Reporting) [1]. The dataset reports incidents with variables such as crime identifier, location information, time information, type of crimes, and prosecution information. The dataset excludes murder but contains all other crimes.

Python was used to transform the date format, MM/DD/YYYY HH:MM:SS AM/PM, into different date format in order to create different types of visualizations and obtain more insights from the data. Tableau was used to created the visualizations in order to see and understand more about the data.

**Data Munging**

To prepare the data, Python was utilized for format conversions and exploring the data. Initially, the Pandas library, used primarily for data manipulation, is loaded. Then the data file is read in and loaded into a dataframe. There are 1,456,714 rows of data (instances) and there are 23 variables (features). Afterwards, two new columns, ‘NewDate’ and ‘MonthNo’, are engineered. With the help of Lamda functions, the timestamps are broken apart to just the dates. Then, the dataframe is converted back to a CSV file and saved.

**Visualization I**

|  |  |
| --- | --- |
|  |  |

Figure 1: Arrest to Report Ratio;Year by Year (2012-2017)

**1.1 Type of Visualization:**

The first visualization shows whether the number of people that were arrested and not arrest from the number of records pertaining to the years 2012 to 2017. We used a line chart to graph the instances because we believe a line charts best shows the changes in between the years in regards to arrest rates over time. We wanted to identify through the years if there was a huge or slight change in the arrest rates. The x axis concerns the date in terms of years. The y-axis is the number of records. The colors red indicate the suspect(s) was not arrested and green indicates the suspect(s) was arrested.

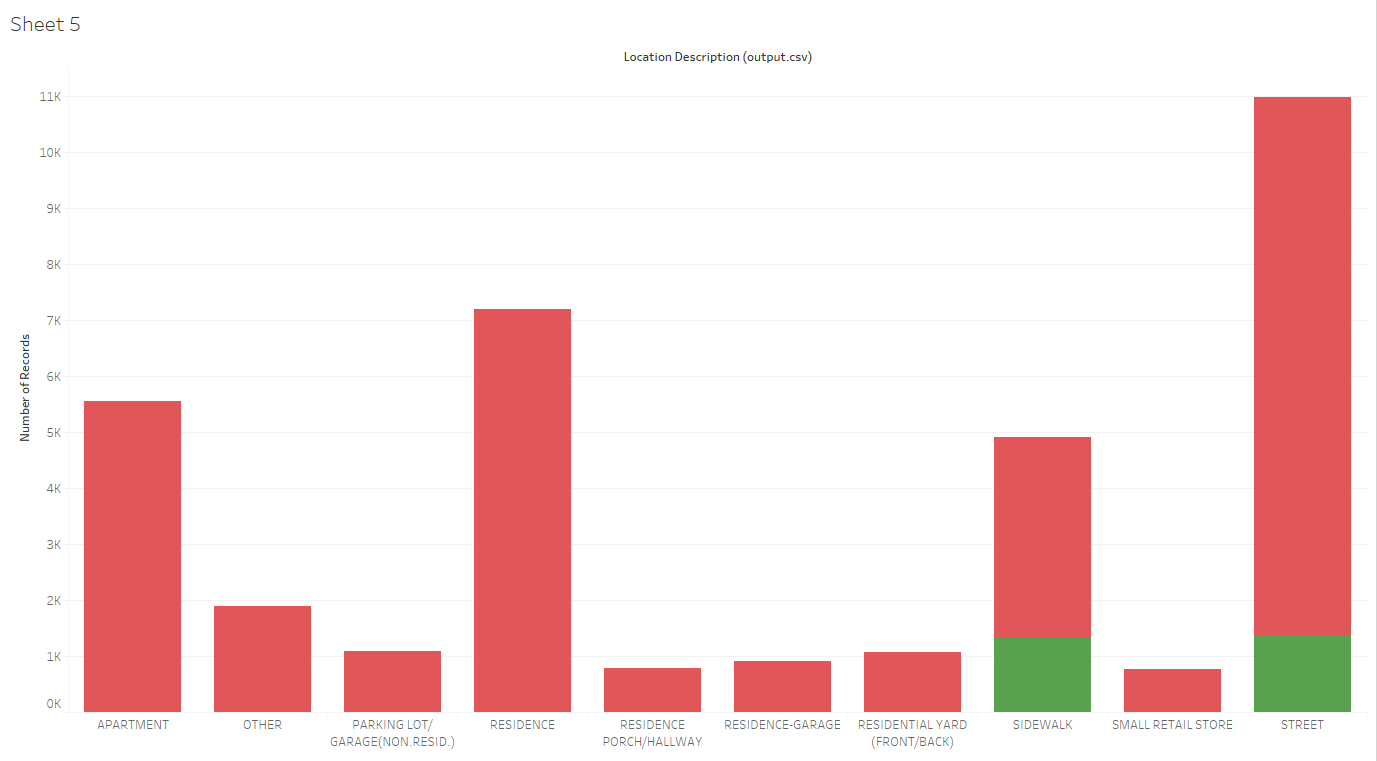
**1.2 Results:**

2012 had the lowest number of records of crimes. There is a small increase in crimes until 2014. 2015 had the highest amounts of crimes, an increase of around 25000 in number of crimes. 2016 cool down a bit from 2015 but remained higher than prior years. The number of arrests increased when there were more crimes.

**1.3 Interpretation:**

From the visualization, there is a peak from the year 2015. 2015 was the year in which there was major attention on police brutality, when police were viewed as not following conduct. Because citizens believed that the police were abusing their power, they decided to go against the police by increasing the participation in crimes [2].

**Visualization II**



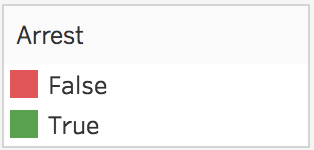


Figure 2: Breakdown of Arrests vs. Reports by Location

**2.1 Type of Visualization:**

Our second visualization is a bar chart showing the total number of crimes reported vs the total number of arrests made, split by location description. The graph was produced with a filter used to show the top 10 locations descriptions. This allows us to both see what areas are most prone to crime as well as which areas usually yield the highest arrest rates. The x axis concerns the location description. The y-axis is the number of records. The colors red indicate the suspect(s) was not arrested and green indicates the suspect(s) was arrested.

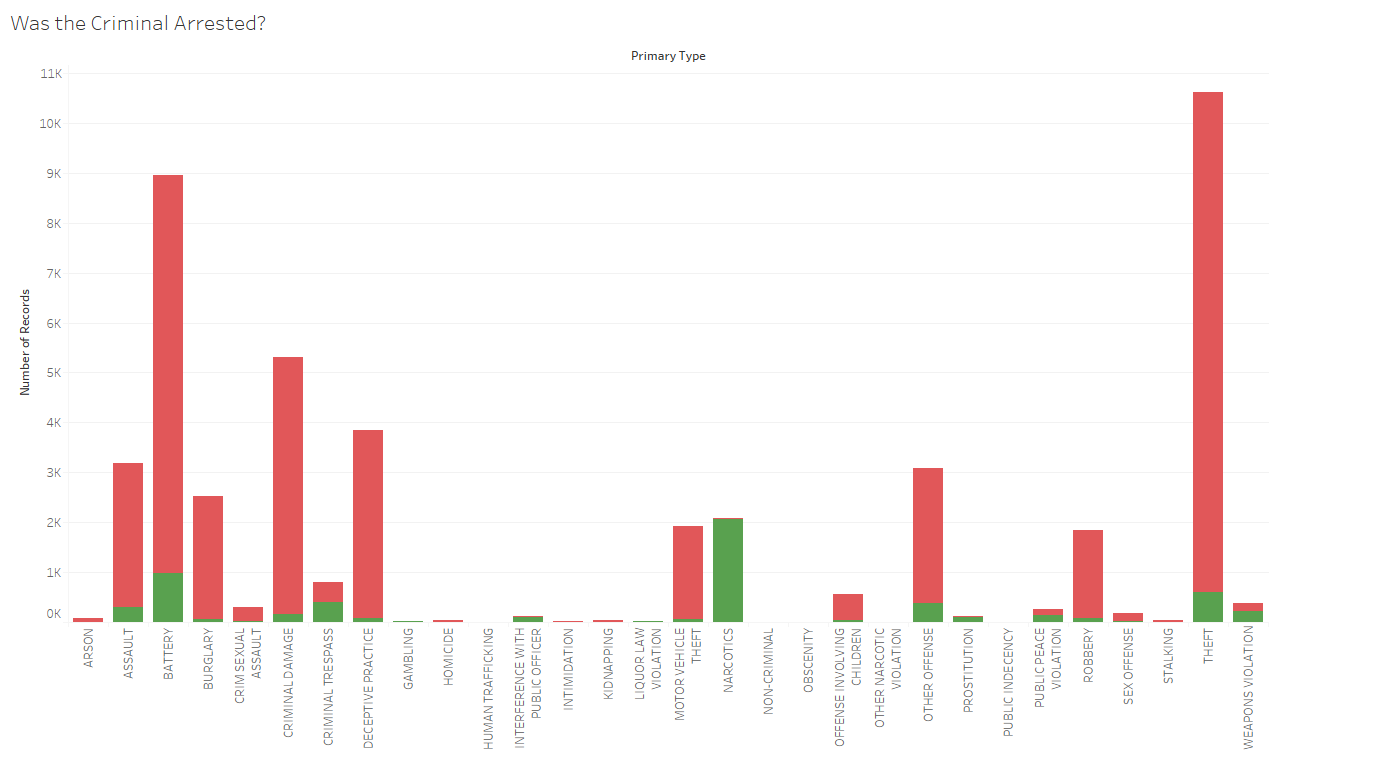
**2.2 Results:**

Streets had the highest number of crimes committed for location description. Residence had second highest, followed by apartment. Sidewalk and street were the only 2 that had arrest indicated by the green bars while all the other location descriptions did not have any arrest.

**2.3 Interpretation:**

Interestingly, the only two locations with arrests high enough to even be visible to us are the sidewalk and the street. There are many reasons why this may be the case. Firstly, these are locations where police patrol regularly. When a crime occurs on the street, an officer can immediately respond and react to the situation at hand. When a crime occurs somewhere else less public (like a residence), somebody has to first call the police and then give a report once they show up. A full investigation has to occur before anybody can be prosecuted for the crime. At every step along the way, fewer people will be willing to cooperate. The culture of the area is to not trust the police and to never “snitch”, so oftentimes crimes that aren’t directly witnessed by a police officer go unreported and perpetrators go uncharged.

**Visualization III**



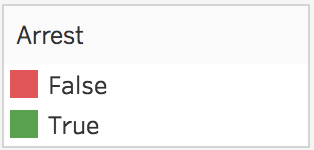


Figure 3: Breakdown of Arrest vs. Primary Type

**3.1 Type of Visualization:**

This visualization is a bar chart showing the total number of crimes reported vs the total number of arrests made, split by primary type. This allows us to both see what type of crimes are committed and which crimes have the highest arrest rates. The x axis concerns the primary type, with the y-axis is the number of records. Green indicates reports which led to an arrest, while red are unsolved/inconclusive cases.

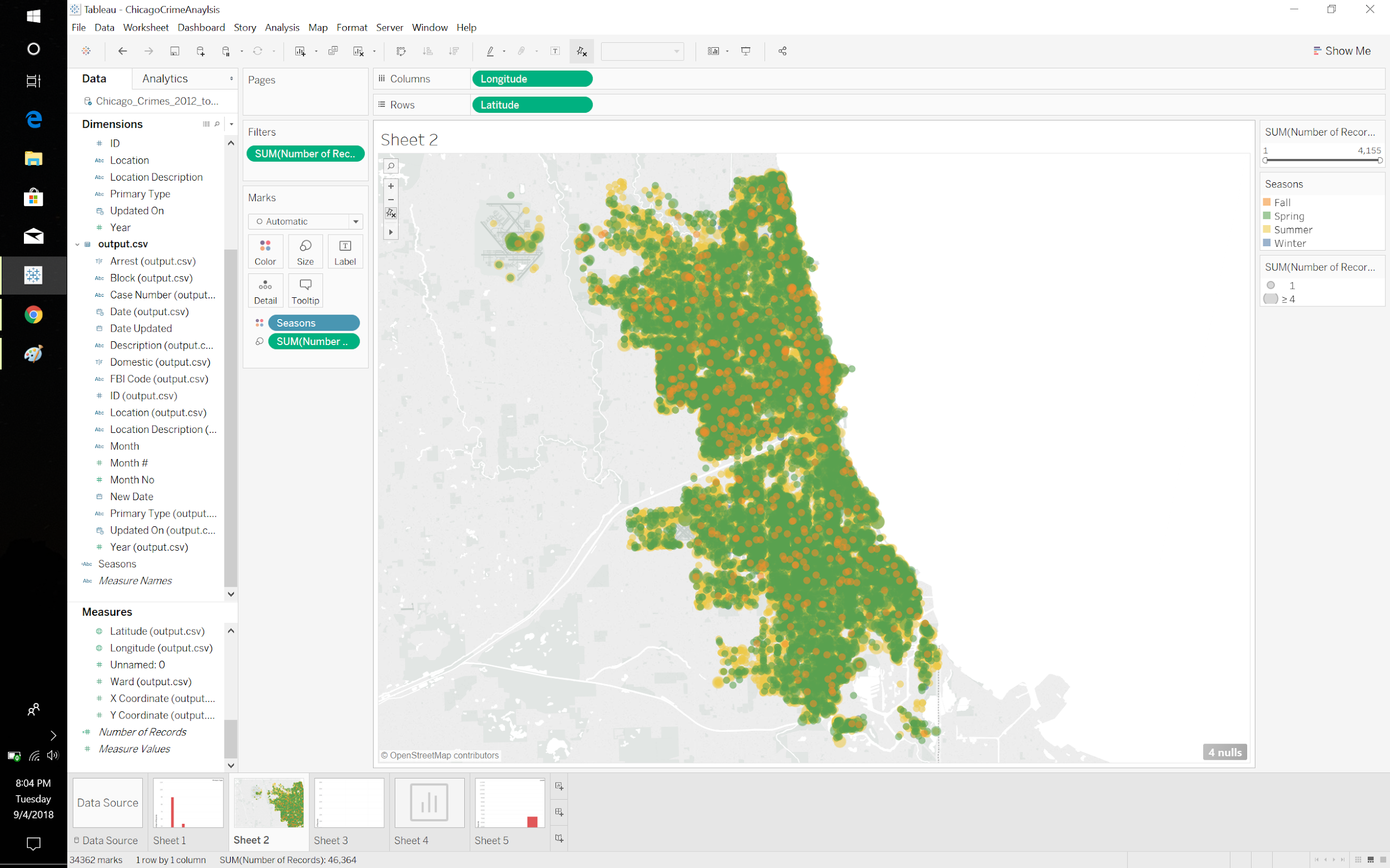
**3.2 Results:**

Theft had the highest number of records for primary type. Narcotics had ~100% of reports yield arrests.

**3.3 Interpretation:**

Two interesting observations from this graph were the high number of robbery reports and the essentially 100% arrest rate for narcotics charges. Robbery can be assumed to be so high for two reasons: it is a common crime, and people who have their items stolen are more likely to go to the police. Many times people who have items worth stealing will not believe in the “don’t go to the police” idea, so they are more willing to make a report. As for the narcotics arrests, that can be explained by the fact that it is nearly impossible to be caught in America with narcotics and not get arrested. These are extremely highly controlled substances, and anybody caught with them is guaranteed to see jail time.

**Visualization IV**



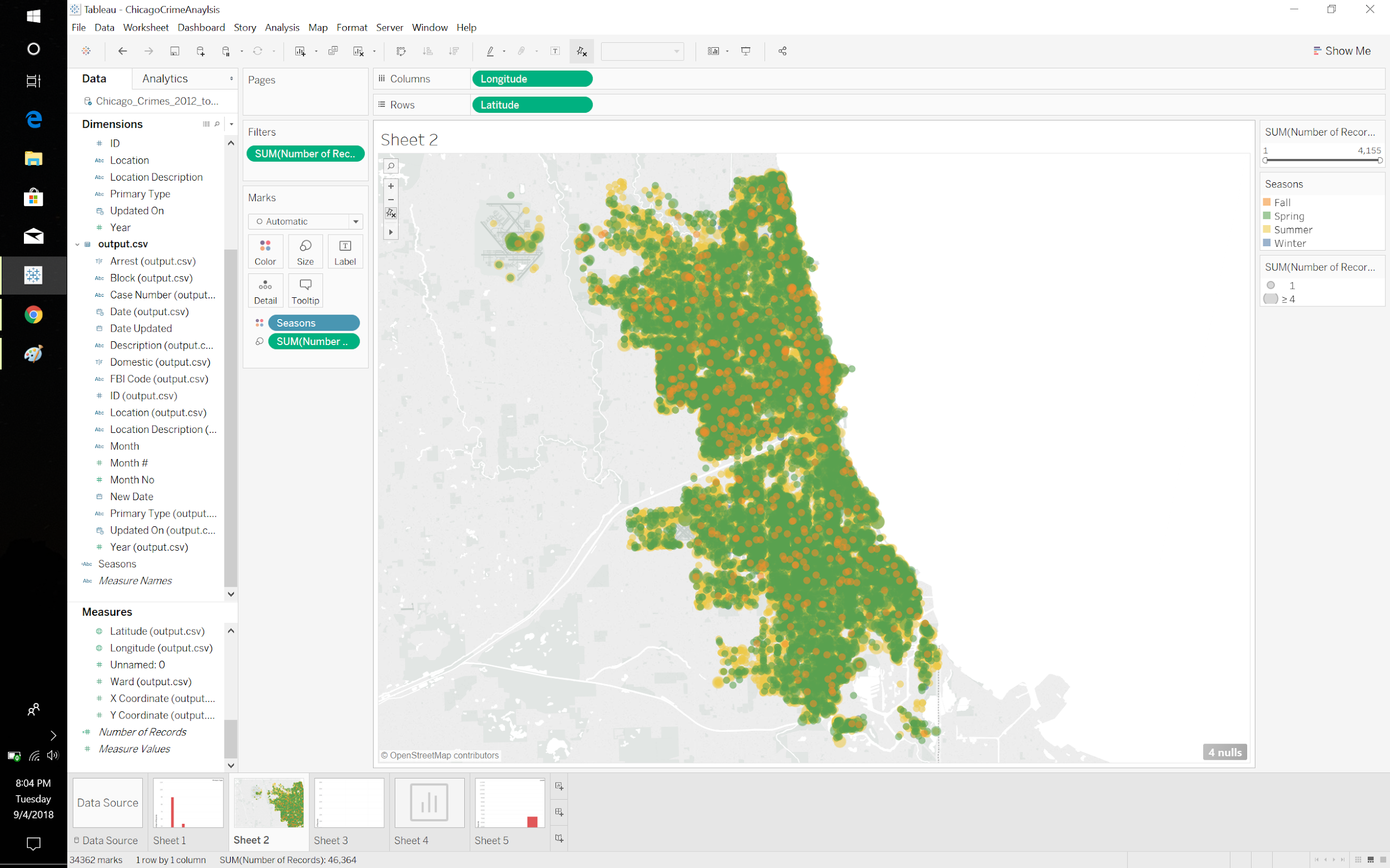


Figure 4: Crimes and Seasons

**4.1 Type of Visualization:**

This visualization is a symbol graph that shows the areas that are prone to all types of crimes and the seasons that crimes have occured from 2012-17 to get an idea if weather has an possible implications on crimes. Latitude and Longitude was used to map the locations of the incidents. The colors categorizes the seasons that the crime occurred using the months to create the calculated field. Orange represents Fall, green is Spring, yellow is Summer, and blue is winter. Number of records was used for size to see which area has the most crimes.

**4.2 Results:**

The graph is mainly populated by green dots indicating spring. The graph is also populated by yellow and orange dots. Size is useful in this graph because it can tell us the severity of the area where the cluster areas tend to have larger number of crimes compared to the smaller areas if we want to look at locations.

**4.3: Interpretations:**

Weather might be a possible cause for when crimes happen. There are very few occurrences in the winter. According to the visualization, spring is the time where crimes are mostly likely to happen followed by summer. There are a few occurrences of fall. There have also been occurrences in Chicago International airport on the left of the map. So when the temperatures are low, there tends to be very little crime. The mid temperatures have the highest chances of being in crimes. When the temperature is hot, there are many crimes but not as much as the mid temperatures.

**Similar Study**

Northwestern University conducted a study about violent crime in Chicago in late 2017 [3]. The primary goal of this study was to identify trends in gun violence within the city. One interesting discovery that was made was the effect of the time of year on the number of violent crimes committed.

In the graph below, provided by Northwestern University, it is clear that murders drastically increase during the warm months. While this was somewhat reflected in our data (by a graph not shown in this paper), our dataset involved all types of crime rather than just murder. Interestingly, murders committed without a gun follow what is essentially an identical pattern as gun murders. This information can help with properly allocating resources making sure that there will be as many police on the street as possible during the summer when the crime rate spikes so dramatically.

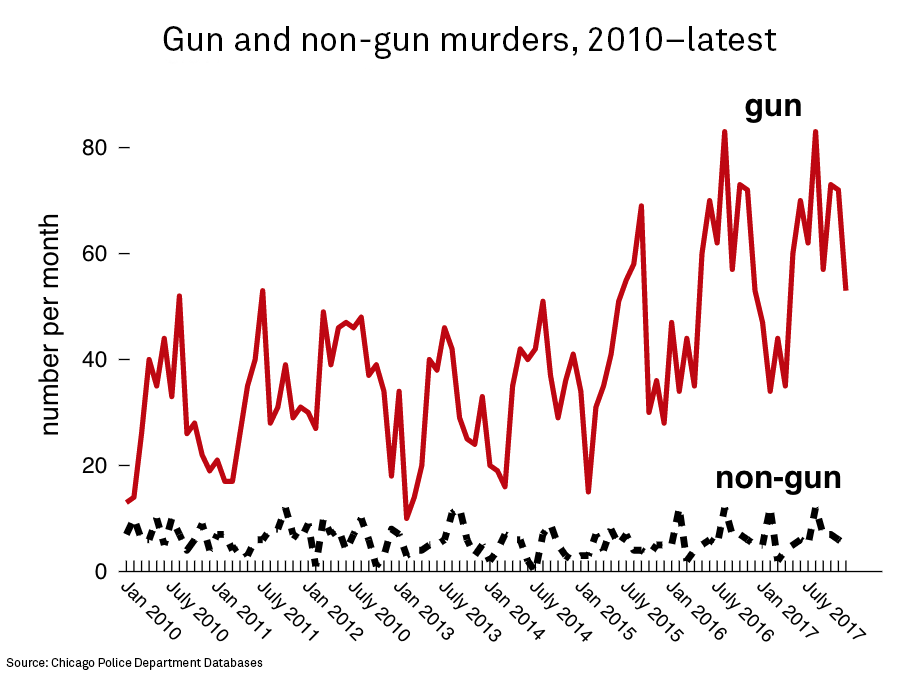


Figure 5: Gun and Non-Gun Murders, 2010 - July 2017

**Conclusion**

As shown figure four, Chicago is a city riddled with crime with little to no pockets of low crime. Because of this, it has been the focus of politicians and researchers alike. Figure 2 shows how the vast majority of reports are coming from street crimes rather than in home problems, hinting at the idea that crimes in the home may not always be reported, and figure 3 shows just how many crimes go without arrest, showing the need for more officers on the job. Without ever visiting or even reading about the city, creating these visualizations has shown us the dangers of Chicago, as well as the times when it it is safest and the types of danger a visitor may encounter.

**References**

[1] Currie32. “Crimes in Chicago | Kaggle.” *Countries of the World | Kaggle*, 28 Jan. 2017, 12 www.kaggle.com/currie32/crimes-in-chicago.

[2] Gorner, Jeremy. “Chicago Violence, Homicides and Shootings up in 2015.” 123*Chicagotribune.com*, 4 Jan. 2016, 333www.chicagotribune.com/news/local/breaking/ct-chicago-police-violence-2015-met1-20160133301-story.html.

[3] “News.” *The Long-Term Effects of Lead: Flint and Beyond: Institute for Policy Research - 123Northwestern University*, 123www.ipr.northwestern.edu/about/news/2018/crime-in-chicago-research.html.