**An Analysis on the Relationship between Weather and Crime in Chicago, IL**

Kevin McManus

Colorado State University Global

MIS 581: Capstone – Business Intelligence and Data Analytics

Dr. Kimberly Ford

July 4, 2021

Table of Contents

[Abstract 3](#_Toc75962700)

[Introduction 4](#_Toc75962701)

[Objectives 4](#_Toc75962702)

[Overview of Study 5](#_Toc75962703)

[Research Hypothesis 5](#_Toc75962704)

[Literature Review 6](#_Toc75962705)

[Research Design 7](#_Toc75962706)

[Methodology 8](#_Toc75962707)

[Methods 9](#_Toc75962708)

[Limitations 10](#_Toc75962709)

[Ethical Considerations 10](#_Toc75962710)

[Findings 12](#_Toc75962711)

[Data Discussion 12](#_Toc75962712)

[Findings Discussion 14](#_Toc75962713)

[Conclusion 21](#_Toc75962714)

[Recommendations 21](#_Toc75962715)

[References 22](#_Toc75962716)

# **Abstract**

This study was completed to analyze the relationship between the types of crimes committed by individuals and different weather factors in Chicago, IL. The Chicago Police Department crime dataset, which is available to the public, was combined with a weather dataset collected at the Chicago Midway Airport to create the dataset used for this study. Relationships of many weather factors and the types of crime committed were investigated and revealed that average daily wind speed, precipitation, maximum temperature, minimum temperature, and weather where thunder was present all impacted the type of crime committed. A closer look was taken at the relationship between maximum temperature and the types of crime committed and found that as the maximum temperature increased, the types of crimes committed became more violent. An additional step was taken to see how temperatures considered “hot” and “cold” impact the location of crimes committed. The study found that crime locations shift as the weather shifts between “hot” and “cold” as crime frequencies change. By identifying these existing relationships between weather factors and the types of crimes committed, members of the Chicago Police Department can prepare for any potential rise in crimes or ensuring that their department is properly staffed if an influx of crimes is suggested in future months based on the weather forecast.

# **Introduction**

Policing in the United States has evolved over time from a small and loosely structured department to a one that is large and strict with structure that may even resemble those of the armed forces. As the structure of police departments have changed over the years, so has the criticism of those departments. What was once a sector of the city that was well respected and loosely watched over has turned into an area of the city that is put under a microscope each moment of the day with news stories that portray the good and bad actions of those first responders. The Chicago Police Department (CPD) is no exception to this rule and has faced their own fair share of scrutiny and scandals while doing their best to enforce the law and keep the citizens of Chicago safe. The CPD is a great organization to analyze how crime plays a part in the daily lives of its citizens and how external factors, such as weather, can impact the types of crime committed on a given day from 2014 through 2020 to create a positive public perception.

# **Objectives**

By completing this analysis between crime and weather, the City of Chicago Police Department will be able to identify if there is a relationship between different weather factors and types of crimes committed, such as violent, non-violent and property crimes. The study will also determine if there is a relationship between the types of crimes committed and the maximum temperature for the day. Lastly, this analysis will determine what weather factors can help predict the type of crimes that will be committed by offenders. A look will also be taken to determine if temperature influences the location of where the crimes are committed. By properly being able to predict the types of crimes committed and locations, CPD can properly staff different precincts and beats to ensure that their citizens can rely on the support of the department.

# **Overview of Study**

The two datasets, *Crimes – 2001 – Present* and *Chicago Weather,* will be joined together to provide additional insights into the already established study of how weather affects the types of crimes to assist in better evidence-based decision-making (O’Leary, 2021). The crime data for each incident, in addition to the climate data points that occurred on the day of the incident, can be combined to allow for an analysis to be completed on how weather impacts the types of crimes that are committed in the city, along with any trends that may be found that correspond with the city’s climate trends. By being able to identify how weather may impact crimes in a particular city, crime cycles may be uncovered which could allow police departments to conduct additional trainings on crimes that may begin to increase due to seasonality.

# **Research** **Hypothesis**

RQ 1: Is there a relationship between all external weather factors and the types of crimes committed?

H0 1: There is not a relationship between all external weather factors and the types of crimes committed.

HA 1: There is a relationship between all external weather factors and the types of crimes committed.

RQ 2: Is there a relationship between the maximum temperature for the day and the types of crimes committed?

H0 2: There is not a relationship between the maximum temperature for the day and the types of crimes committed.

HA 2: There is a relationship between the maximum temperature for the day and the types of crimes committed.

RQ 3: Will one or more weather variables be able to predict the type of crime committed?

H0 3: One or more weather variables will not be able to predict the type of crime committed.

HA 3: One or more weather variables will be able to predict the type of crime committed.

# **Literature Review**

Weather is a very volatile and unpredictable natural event that occurs throughout the country on any given day. Given how random weather patterns can be, curiosity can enter one’s mind on how criminal activity can be affected by those weather conditions. There have been various studies completed by researchers to try to understand how weather plays a role in violent criminal activities. These studies found that violent crimes increase with temperature rises (Michel et al., 2016; Butke & Sheridan, 2010; Ranson, 2014). The studies completed by Michael et al. (2016), Butke and Sheridan (2010), and Ranson (2014) each took place in difference cities throughout the United States outside of Chicago, IL that provide a base for this analysis. Given the research locations that have taken place, with the plethora of information collected by the CPD, Chicago presented itself as an ideal candidate for analysis.

This analysis takes the findings from previous studies a step further as it raises additional curiosities. In the prior studies listed above, the analysis is completed utilizing temperature and violent crimes only. This analysis suspects that temperature will play a significant impact on the types of crimes committed, such as violent crimes, but will also examine if the types of crimes committed are statistically impacted by other types of weather conditions, such as snow or rain.

The location of the weather station used for the weather data plays an important part to the study. Hofer et al. (2019) discusses the difficulty that the location can play in a study as the exact weather attributes is not fully representative of the area. Careful selection should be utilized when selecting a weather station location, which Chicago-Midway Airport has 100% coverage of the city for their attributes. However, because the station is on the southern end of the city, areas on the northern side could not be fully accurate. This is due to temperature variances that could be a result of the heavily populated areas such as downtown in comparison to more suburban areas of the city. Areas of potential improvement for the weather will be discussed in a later section of the study.

Throughout the research for this study, while Hofer et al. (2019) discussed the location of weather data and its impact on crime, there seemed to be a lack of information on how weather attributes could impact the location of crimes. The lack of this information could be due to a variety of reasons, such as lack of crime location or insufficient weather data locations. Attempting to analyze this type of information could prove to be difficult using latitude and longitude coordinates for each incident, so completing a visual analysis of the crimes committed under certain weather conditions could prove to be useful. It would be interesting to see how crime density areas could potentially shift based on weather patterns and conditions. Any potential shifts in density could be a great benefit to CPD by allowing to properly staff certain beats or community areas should trends be identified.

# **Research Design**

The research design for the project goes over the steps taken to determine the statistical significance between weather and crime in the City of Chicago. A closer look will be taken into the quantitative methodology and research methods utilized in this project.

## **Methodology**

The City of Chicago has many datasets that are available to the public for their own downloading and analysis. The *Crimes -2001 – Present* dataset is a very informational dataset to the city and their criminal analysts. This dataset has been chosen as it provides information about each criminal report that is made by the CPD and allows for the analysis of data over any desired time that is within the dataset’s beginning collection period of January 1, 2001, to seven days prior to current day. This dataset is updated daily to provide the most up to date data to the public. This dataset does undergo cleaning and manipulation prior to completing any analysis.

For the crime dataset, the *Domestic* variable needed to be converted to a binary variable, 1 representing true and 0 representing false, as it is originally a character when the dataset was exported. For this research project, the *IUCR,* *FBI.Code*, *Description, Location.Description, Year,* and *Updated.On* were not needed as only *Primary.Type* will be analyzed. The *Primary.Type* variable was grouped into violent, non-violent, and property crime groupings, which are assigned values of 3, 2, and 1 respectively, for the combined dataset to see how the weather has an impact on each category. These classifications are based on CPD’s classification of violent and property crimes, with those not listed falling in the non-violent category. Lastly, the crime data will be filtered to only include those crimes that occurred from 2014 to 2020. Once the filtered crime data has been completed, along with the removal of unnecessary variables in both datasets, they will be merged into a single dataset that will be utilized for the statistical testing and analysis.

The National Oceanic and Atmospheric Administration (NOAA) is an organization that makes climate data available to the public that is collected across different stations throughout the United States. Data from this public dataset will be used for the Chicago-Midway station that covers 100% of the city in their data points for each day of recorded data, according to the NOAA. This *Chicago Weather* dataset will provide insights into the temperature, precipitation, wind speed, and weather conditions in the city on any day.

In the weather dataset, the cleaning includes removing the *SNOW* and *TAVG* variables. These were removed as they do not have any values other than 0, which would not have a statistical impact on analysis due to these values remaining constant. *Station* and *Name* were also removed as these variables are unnecessary since all crimes reported are from the City of Chicago with the weather data all be collected from the Chicago-Midway Airport.

## **Methods**

Correlation testing will be completed to identify relationships between crime and the different weather variables/conditions. Butke and Sheridan (2010) emphasize the point that a temperature should be set as the defining point between what is considered “hot” and what would be considered “cold” to allow for the proper analysis of how colder and hotter temperatures play a part in the crime rates throughout the City of Chicago. For this project, 70 degrees Fahrenheit will be the dividing temperature. SAS will be utilized by completing a correlation test to identify if the relationship between *TMAX* and *Primary.Type* to see if they are positively or negatively correlated and if that relationship is statistically significant. SAS will also be used to run a second correlation test between *Primary.Type* and all the weather variables to see the relationships between each variable and *Primary.Type* to study their impact on the type of crime committed and their significance.

Regression analysis will be completed to identify all variables in the merged dataset that affect crime. Ranson (2014) completed a regression analysis to see how weather has impacted crime in historical data and identified that increasing temperatures resulted in more violent criminal activity throughout the United States. Nominal logistic regression tests allow for users to evaluate relationships between predictor variables and a response variable that has 3 or more groupings. A nominal logistic regression test in SAS will be utilized to determine which weather variables could assist in accurately predicting a crime as violent, non-violent, or property.

Visualizations will be created in Tableau showing the density of crime locations for the entire merged dataset to see where the hotspots are located within the city. After creating this visual, 2 additional visuals will be created to display the hotspots of crime when *TMAX* is above and below 70 degrees Fahrenheit to see potential variations while the temperature was considered “hot” and “cold” outside. All three of the created visuals will be compared to one another to see how temperature impacts the crime density throughout the city.

## **Limitations**

It is crucial to note that in the study completed by Hofer et al. (2020), it is important to take into consideration the distance of crimes in relation to the location of the weather station where the weather data was collected. While the station may account for the entire coverage of a city, Hofer at al. (2020) identified that the distance from the weather station can create a skew for data as the distance away from the station can cause for a variation in values of the data points. This piece of information can create a skew in the visuals of the crime densities given their distance from the weather stations as rain or temperatures at the station may not be present on the opposite end of the city.

## **Ethical Considerations**

In the United States, the right to privacy through the collection of data by government agencies, such as CPD, is protected by the Fourth Amendment (Daskal, 2015). For the government agency to obtain the content of data that may be sent through emails, agencies must obtain warrants based on probable cause (Daskal, 2015). This is an important note that pertains to the CPD as they are bound by laws that prevent them from accessing different types of data to help protect the individual rights of a person. Davis (2012) identifies that the distinction between PII and anonymized data are critical for the privacy of data and protecting the identities of those involved. By being able to set a clear distinction in the data policy on what makes pieces of data anonymized and PII, companies or the Chicago Police Department can ensure that the PII information is not shared and removed from the data. As new data is collected, the CPD will be able to properly classify pieces of data that would fall under PII and those that would be classified as anonymized data.

Ethical concerns with the usage of big data include considering if certain pieces of data need to be collected in the first place. If the data point does not have a specific purpose in coming to an appropriate solution, model, or insights, consideration should be taken as to whether the CPD should spend the time and allocate the necessary resources to collect that piece of data. Hand (2018) goes to add that the data collected should be able to benefit the public. Data should not be collected simply because it can be collected, but the collection needs to provide a sense of business value to the organization and public. Another ethical concern revolves around bounded ethicality, which prevents individuals from making an objective ethical deliberation (Feldman & Kaplan, 2020). Bounded ethicality is a large concern when completing an analysis as this impacts the researcher’s ability to not let their bias influence their study and interpretation of the results.

With the combination of the two datasets, statistical analysis can be completed to test the previously mentioned hypotheses. When utilizing hypothesis testing, the determined p-value provides statistical significance to the researcher. When the p-value is less than 0.05, there is strong evidence against the null hypothesis, to where it can be rejected (Hazra & Gogtay, 2016). Hazra and Gogtay (2016) go on to point out that a p-value that is less than 0.01 indicates stronger evidence against the null hypothesis. P-values that are greater than 0.05 provide the researcher with weak evidence against the null hypothesis, so the null hypothesis could not be rejected nor accepted; while a p-value near 0.05 provides marginal evidence against the null hypothesis (Hazra & Gogtay, 2016). These p-values will be utilized when determining the statistical significance for the hypotheses testing for this analysis for the potential benefit of the Chicago Police Department.

# **Findings**

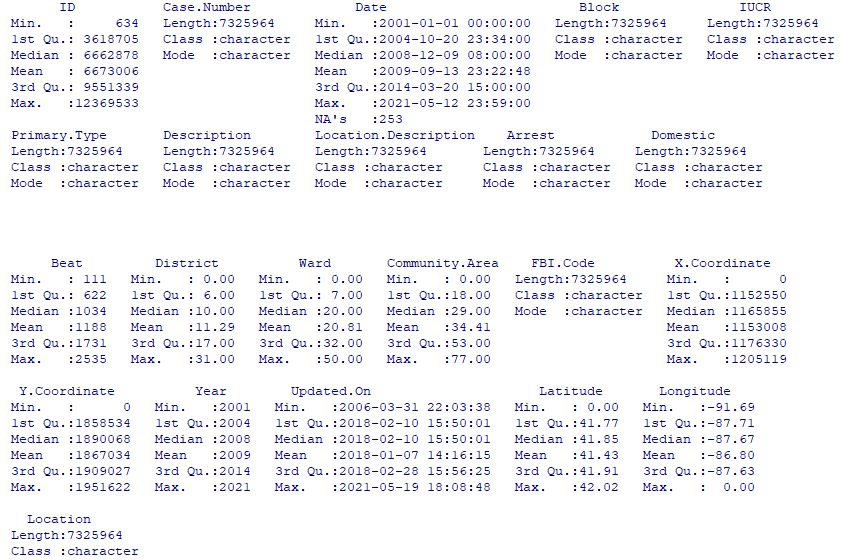
This section of the paper is broken up into two sections. The first section discusses the data that will be utilized for the study, including descriptive statistics. In the second section, the findings from the analysis will be discussed and how these prove or disprove the hypotheses previously listed for the study.

## **Data Discussion**

The variables in the *Crime – 2001 – Present* dataset is filled with many integers and character data types for the appropriate variables. There is a numeric data type for the *Latitude* and *Longitude* columns as these present the location of the incidents to allow for proper plotting of the data points. There are two columns, *Date* and *Updated On*, that display both the date and time of the incident and for the most recent update. The *Arrest* and *Domestic* columns are both TRUE/FALSE values that utilizes the words as opposed to a numeric identification.

**Figure 1**

*Crimes – 2001 – Present Summary Statistics*

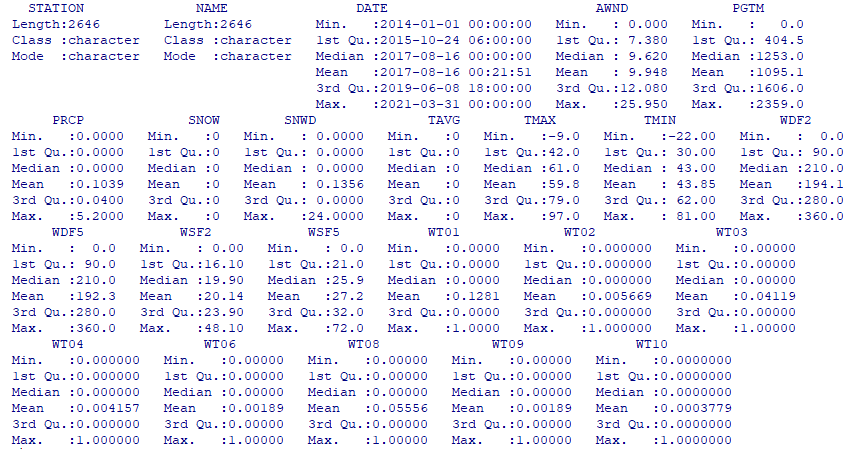


The dataset consists of over 7.3M rows of data ranging from January 1, 2001, to May 12, 2021, which can be seen as the total number of rows in *Case.Number*. There are 253 missing values in the *Date* variable, but these will be excluded from the analysis once a defined time is chosen for the project. There was a significant time before the first update was made to the dataset from when the collection period began. There are 0 values in the *District, Ward, Community.Area, X. Coordinate, Y. Coordinate, Latitude and Longitude* variables, which indicate that there are missing values in those fields as those coordinates do not include the City of Chicago and there are not any district, beats, or community areas in Chicago that include 0 for their values.

The *Chicago Weather* dataset utilizes character, integer, and numeric data types throughout. The column utilized to display if certain weather conditions were present, *WT##,* utilizes an integer value of 1 to signify if the weather condition was present and is left blank/null if the condition was not present.

**Figure 2**

*Chicago Weather Summary Statistics*

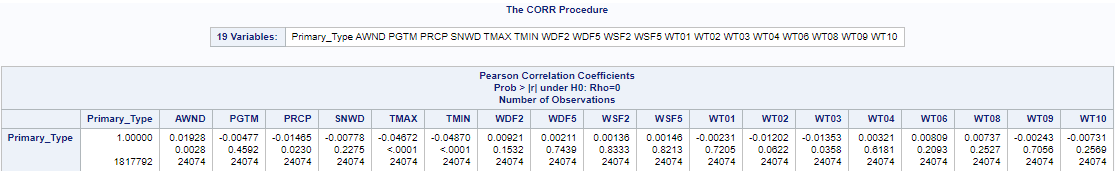


The weather dataset has a total of 2,646 records that span from January 1, 2014, through March 31, 2021. The *SNOW* and *TAVG* variables do not have any values that were included in the dataset other than 0. As all of these values are recorded at the same station, the *Station* and *Name* variables will likely not be needed.

## **Findings Discussion**

Figure 3 shows the partial output of the correlation test between the type of crimes, *Primary\_Type,* and all the weather variables collected. The whole chart is not shown as the study in RQ 1 is only concerned with the relationships between *Primary\_Type* and the weather variables.

**Figure 3**

*Partial Correlation Chart between Primary\_Type and All-Weather Variables*

Based on the correlation chart, there is a statistically significant relationship between the type of crime and average daily wind speed (*AWND),* precipitation (*PRCP),* maximum temperature (*TMAX),* minimum temperature (*TMIN),* and weather where thunder was present (*WT03)*. This is determined as the p-value is less than 0.05. Between these relationships, there is a direct relationship between the type of crime and average daily wind speeddue to a positive correlation coefficient. The remaining statistically significant weather categories have an indirect relationship with the type of crime due to the negative correlation coefficient. The strength of the relationship can be determined by how close the correlation coefficient is to 1 for a positive relationship and -1 for a negative relationship. The values that are close to 0 mean that the relationship is not strong, but rather weak.

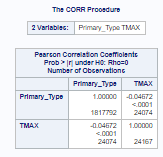
RQ 1: Is there a relationship between all external weather factors and the types of crimes committed?

As not all of the weather variables were found to be statistically significant in having a relationship with the type of crime, the null hypothesis is accepted. This information can be of use to the CPD as they could only focus on the variables that have proved to be statistically significant. By doing this, the CPD would not have to store unnecessary data in their database as it would not have any effect on any models that they could potentially build to predict types of crimes committed.

Figure 4 below shows the correlation chart between the type of crime and the maximum temperature.

**Figure 4**

*Correlation Chart between Primary\_Type and TMAX*

**

The above chart shows that there is a statistically significant negative relationship between the type of crime and the maximum temperature. As the temperature increases, the type of crime goes from property all the way to violent showing that more violent crimes occur in higher temperatures. This negative relationship is what would be expected based on the studies completed by Michel et al. (2016), Butke and Sheridan (2010), and Ranson (2014) which all found that violent crimes increase as temperatures increase.

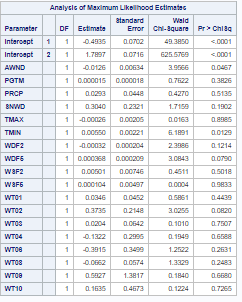
RQ 2: Is there a relationship between the maximum temperature for the day and the types of crimes committed?

Due to the statistically significant relationship that was found, the null hypothesis is rejected. The significance of this relationship can be greatly beneficial to CPD as they can prepare their staff for more violent call for the officers as the city may experience a heat wave. The relationship would also allow CPD to change any potential staffing concerns to beats with a heavier violent crime percentage or potentially offer overtime to ensure all beats are properly staffed.

Figure 5 below shows the logistic regression analysis between the type of crime and all the weather variables to identify if any of the weather variables can predict the type of crime committed.

**Figure 5**

*Logistic Regression Analysis of Primary\_Type and Weather Variables*

**

Based on the chart above, average daily wind speed (*AWND)* andminimum temperature (*TMIN)* are statistically significant in predicting the type of crime committed. This is determined as the p-value is less than 0.05 to identify the significance and the larger the parameter estimate, the stronger impact it has on the type of crime. It is important to note that the average daily wind speed has a negative impact on the type of crime due to the negative parameter estimate value.

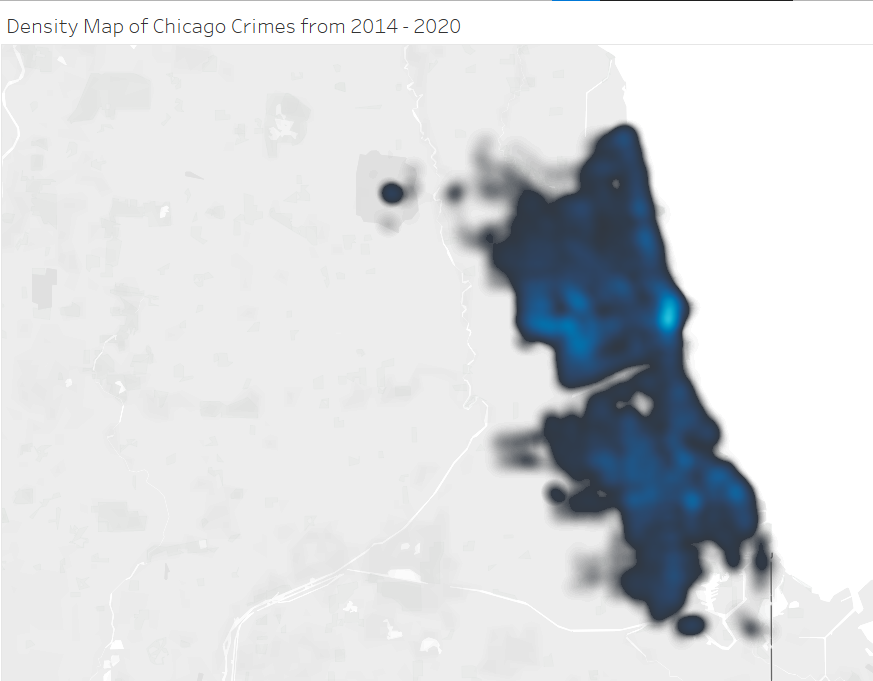
RQ 3: Will one or more weather variables be able to predict the type of crime committed?

Since the average daily wind speed and minimum temperature are statistically significant in predicting the type of crime committed, the null hypothesis is rejected. These results are useful for the city knowing that they can utilize weather data to predict the type of crime committed. However, due to the parameter estimates for each of the two variables, CPD should discuss if the data is useful to them and would be worth the effort and storage space to collect the data.

Figures 6 – 8 all show density maps of the crimes that occurred within Chicago from 2014 through 2020. The density color range changes from black and lightening in color all the way to white as the density increases within the city.

**Figure 6**

*Density Map of Chicago Crimes from 2014-2020*

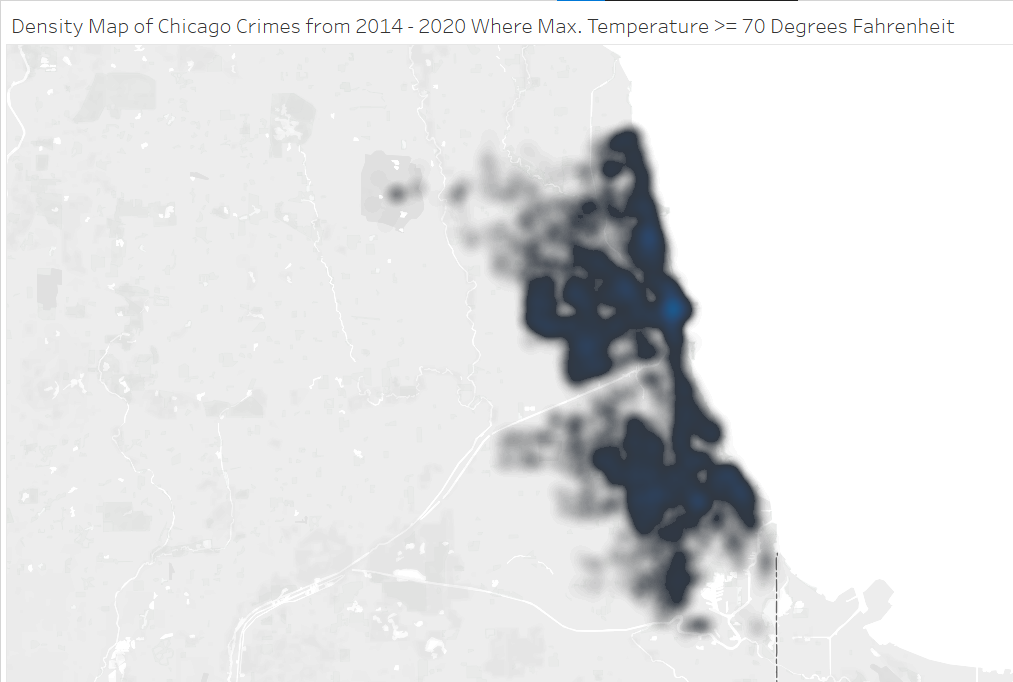
**

In looking at the above visual for the crime density throughout the entire City of Chicago, the largest density of crimes occurs towards the downtown area of Chicago, which is signified by the noticeably light blue/white color. The outskirts of the city are mostly black in color, which would signify that crime densities increase as you move away from the city limits.

Figure 7 below shows the crime densities in Chicago where the maximum temperatures are 70° Fahrenheit or higher. This gives an idea of where the crimes have occurred when the weather is considered “hot”.

**Figure 7**

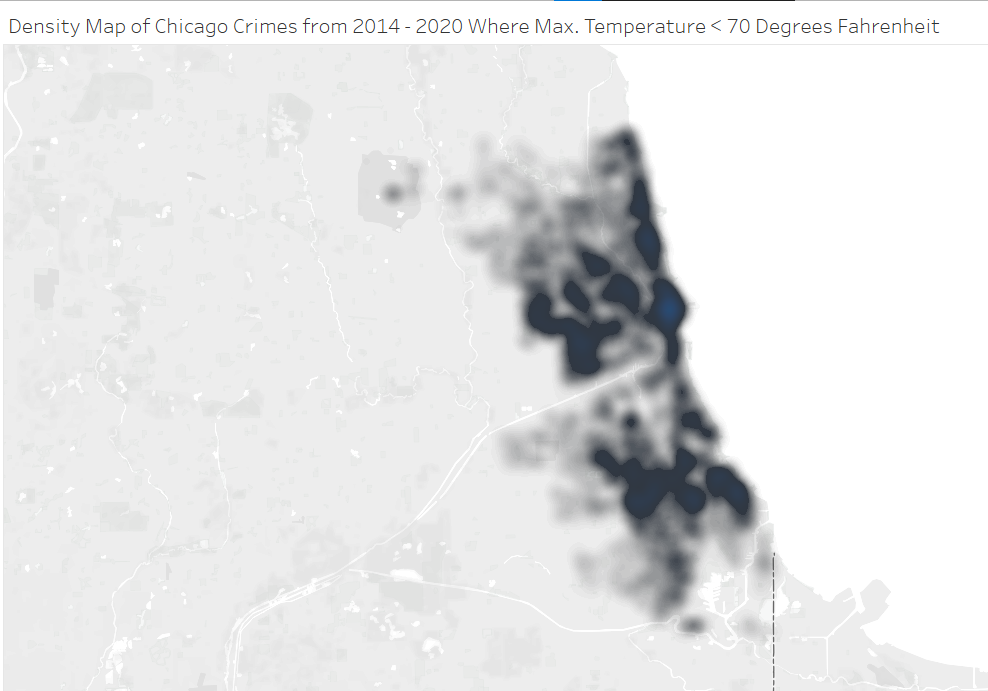
*Density Map of Chicago Crimes from 2014-2020 Where Maximum Temperature ≥ 70°F*

******

When compared with the total crime density of the city, crimes when the temperature is “hot” tend to occur even less frequently around the city limits and concentrate in the downtown area, east of downtown, and the south side of Chicago.

**Figure 8**

*Density Map of Chicago Crimes from 2014-2020 Where Maximum Temperature < 70°F*

******

Crimes when the weather is “cold” outside still occur in the downtown area, but less frequently on the south side and east of downtown, in comparison to the weather when it is “hot” outside. Visually, it would appear that less crimes occur in frequented areas when it is “cold” outside as the strength in color variations and boldness of the dark colors are not as dominant when compared to Figure 7. When comparing the “hot” and “cold” density maps, it is apparent that as the temperatures reach 70°F or higher, the crime densities shift as crimes became more frequent and in repetitive areas of the city.

# **Conclusion**

Throughout the entire country, Chicago spends a large sum of money into their police department to keep their citizens and property safe, while enforcing the law. Crimes occur within the city regardless of the time of day or weather; however, additional research has been completed to determine if any external weather factors play a part in crime itself. The rise in temperature has a statistically significant effect in an increase in violent crimes, while average daily wind speed, minimum and maximum temperature, precipitation, and the presence of thunder all have a statistically significant relationship with the type of crimes committed. By being able to identify weather factors that played a role into crime from 2014 to 2020, the CPD can leverage this data to have an upper hand on preparing their officers for any potential rise in crimes or ensuring that their department is properly staffed if an influx of crimes is suggested in future months based on the weather forecast.

# **Recommendations**

With only two variables of statistical significance and very minimal parameter estimates, whether positive or negative, CPD must ask themselves if using weather data to predict types of crimes committed is worth collecting. However, given the location of where the weather data is collected and its impact on crime as studied by Hofer et al. (2020), it would be beneficial for CPD to collect weather data at each precinct station. By doing this, the crimes committed in each precinct can be run through the same tests above to truly identify if this study is representative of each precinct compared to the whole city as only a single weather station dataset was utilized. Additional attempts to take this research a step further or for duplication can access all research documents, such as copies of this paper, datasets, and codes, at the following GitHub repository, https://github.com/kevinmcmanus16/CapstoneProject.

# **References**

Butke, P., & Sheridan, S. C. (2010). An analysis of the relationship between weather and aggressive crime in Cleveland, Ohio. *Weather, Climate, and Society*, *2*(2), 127–139.

Daskal, J. (2015). Law enforcement access to data across borders: The evolving security and rights issues. *Journal of National Security Law & Policy*, *8*(3), 473–501.

Davis, K. (2012). *Ethics of big data*. O'Reilly Media, Inc.

Feldman, Y., & Kaplan, Y. (2020). Big data and bounded ethicality. *Cornell Journal of Law & Public Policy*, *29*(1), 39–93.

Hand, D. J. (2018). Aspects of data ethics in a changing world: Where are we now? *Big Data* 6(3), 176-190. http://doi.org/10.1089/big.2018.0083

Hazra, A., & Gogtay, N. (2016). Biostatistics series module 2: Overview of hypothesis testing. *Indian Journal of Dermatology*, *61*(2), 137–145. https://doi.org/10.4103/0019-5154.177775

Hofer, I. M. J., Hart, A. J., Martín-Vega, D., & Hall, M. J. R. (2020). Estimating crime scene temperatures from nearby meteorological station data. *Forensic Science International*, *306*. https://doi.org/10.1016/j.forsciint.2019.110028

Michel, S. J., Wang, H., Selvarajah, S., Canner, J. K., Murrill, M., Chi, A., Efron, D. T., & Schneider, E. B. (2016). Investigating the relationship between weather and violence in Baltimore, Maryland, USA. *Injury*, *47*(1), 272–276. https://doi.org/10.1016/j.injury.2015.07.006

O'Leary, Z. (2021). *The essential guide for doing your research project* (4th ed.). Sage Publishing.

Ranson, M. (2014). Crime, weather, and climate change. *Journal of Environmental Economics and Management*, *67*(3), 274–302. https://doi.org/10.1016/j.jeem.2013.11.008