Assessment of Daily Correlation Between Crime, Traffic Congestion, and Weather within Chicago

Cade Dombrowski
Marquette University
cade.dombrowski@mu.edu

Reid Holben
Marquette University
reid.holben@mu.edu

David Helminiak
Marquette University
david.helminiak@mu.edu

Liam Fruzyna
Marquette University
liam.fruzyna@mu.edu

Abstract
This paper studies the link between weather, traffic, and crime, three omnipresent features of Chicago. Using an amalgamation of datasets collected for the city using R programming methods, relationships will be established between them. Datasets were selected to encompass years 2013-2015 over the entirety of the geographical area under the Chicago Police Department's jurisdiction. Crimes are linked to the nearest roads, for which traffic congestion measures are calculated. These are then linked with daily weather measurements for subsequent analysis. It is hoped that analyzing the resultant trends will provide insight into how some existing systems might be optimized that heavily rely upon them. For example helping police optimize their patrol and response routes to minimize crime and response times, or simply showing which traffic routes would be better to use during a downpour.

Author Keywords
R; Analytics; Crime; Traffic Congestion; Weather

ACM Classification Keywords
E.m. MISCELLANEOUS

Github
Project code and generated figures may be found at github.com/mail929/MU-COSC3570-Project.

Introduction
Weather, traffic, and crime take place in every city in America daily, shaping much of people's lives. For example, weather determines the type of clothes needed to be worn, type of transportation to be used, and types of activities, or tasks to be accomplished on that day. Additionally, traffic affects noise levels and travel times. Lastly, crime determines police activities, and peoples' choices of location to live, work, and play. It makes sense then that weather, traffic, and crime are factors in a city's quality of life rating, which cities want high in order to attract and retain people. Chicago, according to INRIX, is home to the 6th worst traffic hot spot in the United States [7]. Also, according to the FBI's semiannual UCR report, Chicago's crime rate is one of the highest in the United States [8]. These factors combined with Chicago's incredibly varied weather, which has temperatures ranging from below 0 to over 100 degrees Fahrenheit, heavy snowfalls, heavy rainfalls, and strong winds makes Chicago an ideal candidate for examining the relationship between these phenomena. The data is broad and varied, allowing, after data cleaning, to look for patterns as to

why there are variations, and what effects these have on other datasets. In addition, the traffic and crime datasets contain location information allowing for the examination of spatial patterns. Crime patterns were analyzed regarding both volume of crimes and type of crimes based off weather and traffic. Understanding the relationships between weather, traffic, and crime then makes for the potential to estimate, or predict potential outcomes related to them. For example, perhaps certain locations in Chicago have a high rate of theft during cold weather. Knowing this would allow police to prepare more for these situations during those times and hopefully reduce the overall rate. The correlations found may better enable the City of Chicago to optimize systems that are affected by these three events, as seen in the example above. In a city that has such a high crime rate, being able to predict possible outcomes of crime, and traffic in an area allows people to prepare and be proactive in solving these problems.

Dataset Overview

The dataset for the weather came from NOAA's weather station at Midway Airport and contains daily temperature, as well as precipitation data [11]. Due to Midway's location in the city of Chicago and low variance in the weather across the city, it accurately represents the weather conditions on given dates. The crime data comes from the Chicago Data Portal [4]. In the portal, a massive dataset of all crime reports since 2001 is continuously updated and publicly available for use. Finally, the traffic data was also taken from the Chicago Data Portal, and examined segments of roadways and measured congestion through monitors placed on the public bus system [12]. Unfortunately, the data availability ends in 2015, despite it being clear that the data is still being collected. The overlap period for the data then was in 2013-2015.

There is existing research that covers similar topics with similar datasets. The only previous work looking at domestic violence came from Louisiana State University, where it was found that domestic violence was the only crime that increased with more traffic congestion [10]. In Scotland, it was noted that warmer days during the winter had a demonstrable effect on crime, especially increasing the incidences of robberies. They correlated this to people simply enjoying being outside more in warmer temperatures [16]. In Belgium, rain and wind speed were actually linked to less traffic congestion. It was postulated that because of the adverse driving conditions that less people actually would go on the roadways, therefore making them more open [3]. Researches at Kent State examined the effect of temperature on crime in Cleveland. There was determined to be a linear relationship between rising temperature as well as the incidences of crime, with some areas affected more than others [1]. A related study looked at risky drivers and if they had a higher propensity for crime, and confirmed that a relationship does exist. This ties into the previous study about domestic violence rates [9].

In a very similar set of research, a data scientist looked at the changes between crime and temperature in Chicago, and was able to see a linear relationship [15]. A visualization group showed the effects of weather and other phenomenon on traffic rates in Chicago [5]. Pollution rates and their effects on crime were also examined, and it was found that while violent crime went up with increased pollution, other types of crimes were largely unaffected in Chicago [6]. This strikes against the idea that good weather leads to more criminal activity. Data scientists also created a model with 75.6% accuracy rating in Chicago that was able to use crime and weather data and predict what conditions would be optimal for

crime to occur [14]. Lastly, use of Twitter was tied into crime and weather data, again proving that a verifiable link exists [2]. Weather is not just about crime, its about human activity, and all these studies prove that more activities happen when the weather is better.

Data Cleaning

The datasets, all formatted as comma separated value (.csv) files, needed to be cleaned and combined into one dataset. The datasets were read into an R programming environment in order to begin this process. Inspecting each frame, they all had variables as column titles, and observations as rows. Despite meeting this principle of tidy data, the traffic and crime data had multiple variables stored within a single column, particularly date and time. Also, all datasets included variables that were not applicable to the research being done. The presence of duplicate observations was seen within the traffic data, which had identical timestamps for observations that were made at different times throughout the day. Lastly, there was a large amount of missing data that could not be extrapolated from existing information, such as location of crimes, traffic speeds, and average daily temperatures. Based off these issues an outline for how the data should be cleaned was developed, where data that overlaps in time and location would be used for final analysis.

First, the date and times were split in order to form identical ISO formatting in each dataset Then, each dataset was filtered to include only information from the overlapping period of January 1, 2013 to December 30, 2014. Variable names were adjusted to have the same formatting with no acronyms, all capitalized, and units added in parentheses where applicable. Variables not applicable to the research along with NULL observations, which could not be extrapolated, were removed.

Additionally, for the weather data, fog was originally split into two different kinds, these were combined. Looking for a central tendency of temperature, a daily average was not present within the weather set, therefore the median was calculated instead.

A unique challenge was presented by the traffic data and its duplicate rows. The duplicates needed to be compressed to give a single, daily observation for a segment of road. A hash was formed by multiplying together the integer version of date and the segment id number together. A data frame was formed using the hash as an input to the "aggregate" function along with a custom statistical mode function to determine the most frequent road speed for the segment. The hash was then used to pull back in the associated date and time variables. The road speeds provided within the dataset as miles per hour (MPH) were then grouped to form congestion levels: 0-10 MPH as "Heavy," 10-21 MPH as "Medium," and above 21 MPH as "FreeFlow."

The datasets now needed to be combined into one. The date variables of crime and weather were matched in order to combine those datasets. The traffic, however, needed to be combined to this new combined dataset geographically. To do this, road segment geographical coordinates were pulled from an additional data set and converted into spatial lines [13]. The coordinates given from crime could then be matched to a road segment using the function "snapPointsToLines". The segment ID was thus copied into the combined data frame for use in joining it with the traffic data, forming the final dataset.

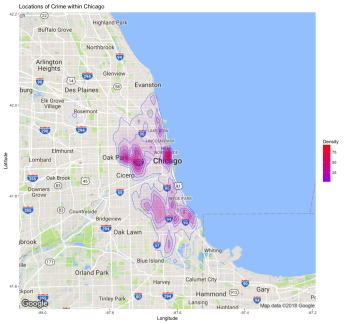

Figure 2: Heatmap of all crimes in the city.

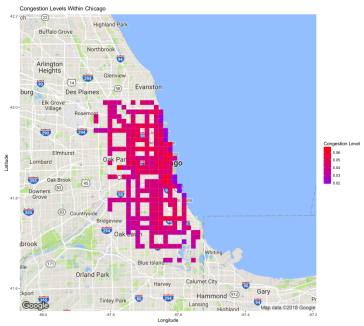

Figure 4: Spatial analysis of congestion levels.

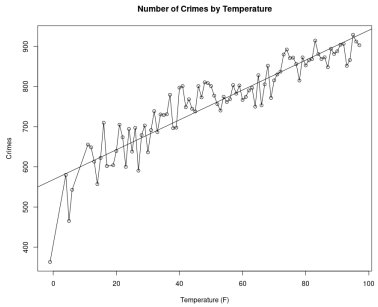

Figure 5: Model of the average number of crimes by daily high.

Data Analysis

Crimes within Chicago range from simple violations, such as for concealed carry licenses and parking, to the extremes of homicide and child abuse. The most common crimes, as seen within Figure 1, are theft and battery. Unsurprisingly, the physical distribution of crime within Chicago (Figure 2) matches very closely with the density of Chicago's population. Predominantly, crime occurs within three regions: the northwest, downtown and southern border with Indiana. However, when examining specific crime types in Figure 3 there are several exceptions to the general trend. Non-criminal, conceal/carry violations and other narcotic violations almost always occur in the northwest region. Ranking downtown Chicago, in combination with the Near North Side, as by far the most dense in terms of total criminal activities sees the northwest section ranked second and the southern third. There also exists an area north of downtown centered on Lake View that has an elevated crime level. While the downtown area crime density is very centrally located relative to it geographically, the northwest is split between two subsections: Oak Park and the 290 interstate. The southern region is the least well defined having no real central location.

Comparing these criminal acts with different traffic congestion levels showed that medium levels are most heavily correlated with crime occurrence, with free flowing conditions second and heavy last. Unfortunately, this does not provide a great deal of insight into how crime might correlate with traffic patterns. Since the data is limited to daily averaged levels, it cannot be determined, with the current available information how a crime may have altered the pattern, or how a pattern could have led to the commitment of a crime. Looking at Figure 4, in comparison with Figure 2, it may be seen that the majority of medium congestion level locations occur in the same geographical locations as reported crimes.

Existing work indicates that there is a strong correlation between rise in temperature and the average amount of crime [15]. As shown in Figure 5 plotting the number of crimes against temperature confirmed this behavior within the datasets. In order to better see this correlation, a linear model was created on the figure to compare the two, resulting in an R^2 values of 0.84 and a P-Value of 2.2e-16. This indicates a high degree of linearity within the dataset, from which it may be concluded that crime tends to increase with temperature. Further, it could be seen that the types of crimes committed, varied with degrees above Chicago's average temperature. Looking at Figure 6 shows that gambling and homicides are far more likely to occur at increased temperatures. Given this very clear relationship, it was suspected that weather type might have similar effects, but examining the data revealed that this was not the case, with the type of weather having little, to no impact on the amount of crime.

Ignoring more general weather types and focusing solely on precipitation there is a variance of crime between 750 to 850 depending on the number of inches of rain, with minimal crime between 1.5 and 2 inches and a peak from 2 to 3 inches (Figure 7). This phenomenon did not meet expectations, as precipitation increases there should be a decrease in temperature and therefore by the previously established relationship, a decrease in crime. This correlation between rain and crime was further examined via a linear model in the same figure. There did appear to be some kind of increase with crime with heavier precipitation, however, the results were not very significant as only a R^2 value of 0.023 and P-Value of

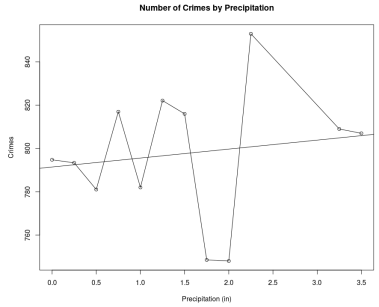

Figure 7: Average crimes per daily rainfall.

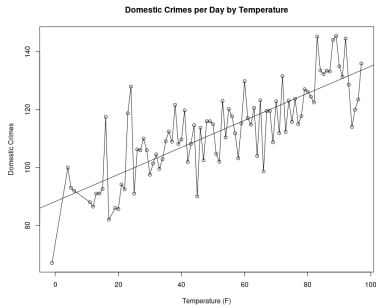

Figure 9: Average number of domestic crimes by temperature.

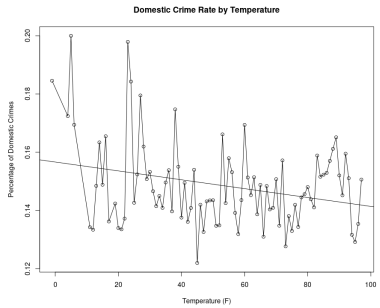

Figure 11: Domestic crime rate by temperature.

0.635 were achieved. The occurrence of crime subtypes by precipitation levels, seen in Figure 8 indicates that arson and prostitution are much less likely when raining, but otherwise crime is generally unaffected by rainfall. One important consideration that has to be made when examining rain data is that it does not rain at regular intervals, so there is not nearly as much data. Further, it rains in very precise amounts for precise amounts of time, meaning it is hard to break into categories of rain amount, or tell if the day itself could be classified as rainy, or if there was a downpour when everyone in Chicago was still asleep. A further variable that should be examined is how the rate of rainfall affects crime, yet there is insufficient data to answer this question.

To further analyze the relative effects of phenomenon on crime, temperature and precipitation were compared to whether crimes were marked as domestic or not, with the intention of visualizing whether crime was simply moving off of the streets and into dwelling areas at higher temperatures and rainfall. To begin, the amount of domestic crimes per day was compared to both the temperature and the precipitation, shown in Figure 9 and Figure 10. The model of temperature against domestic crimes was reasonably linear, resulting in an R^2 of 0.639 and a P-Value of 2.2e-16, however precipitation was not, with only an R^2 of 0.042 and a P-Value of 0.521.

Those models did not tell the full story because they were of domestic crimes per day, and as previously established the overall crimes per day increased with temperature. To investigate without this bias, the domestic crime rate (domestic crimes / total crimes) was compared to both temperature and precipitation. Precipitation again did not provide a clean model, but there was a significant change in the temperature model. As seen in Figure 11 there is now a decrease in the crime rate with temperature, but the model was not very successful and only achieved an R^2 value of 0.077, but with a P-Value of 0.008. These results make sense considering people are more likely to spend time indoors during inclement weather and domestic crimes tend to occur within your family dwelling.

Conclusions

Looking at the analysis of the datasets used, crime increases linearly with temperature and will occur regardless of whether it rains or not. This means that temperature and weather type are not strongly correlated. However, the type of crime does vary with weather type. For example, domestic crime activity was seen to increase when it rains, while arson and prostitution decrease. Geographically, crime is centered on three areas within Chicago: the northwest, downtown and southern border with Indiana. These locations match the originating locations of the greatest number of medium level traffic congestion reports.

The original purpose in studying these datasets was to form a fundamental understanding of how they interrelate and attempt to form a predictive model that might better serve the people of Chicago. Since the datasets were formed on a daily time scale, a great deal of precision was lost, meaning that only relationships that are overwhelming supported by the data can and should be used to form predictive measures. However, that the observations made match with an intuitive understanding of human behavior(s) and existing research shows useful information for how people should interact throughout Chicago, holding onto their wallets downtown, staying indoors during warmer temperatures and finding peace that even though you may be stuck in heavy traffic, the odds of crime are lower in your area.

References

[1] Paul Butke and Scott C. Sheridan. 2010. An Analysis of the Relationship between Weather and Aggressive Crime in Cleveland, Ohio. *Weather, Climate, and Society* 2, 2 (2010), 127–139. DOI: http://dx.doi.org/10.1175/2010WCAS1043.1

[2] X. Chen, Y. Cho, and S. Y. Jang. 2015. Crime prediction using Twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium*. 63–68. DOI: http://dx.doi.org/10.1109/SIEDS.2015.7117012

[3] Mario Cools, Elke Moons, and Geert Wets. 2010. Assessing the Impact of Weather on Traffic Intensity. *Weather, Climate, and Society* 2, 1 (2010), 60–68. DOI:http://dx.doi.org/10.1175/2009WCAS1014.1

[4] Chicago Police Department. 2018. Crimes - 2001 to present. (2018). Retrieved March 6, 2018 from data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2.

[5] R. Grossman et al. 2005. Real Time Change Detection and Alerts from Highway Traffic Data. In *Supercomputing, 2005. Proceedings of the ACM/IEEE SC 2005 Conference*. 69–69. DOI: http://dx.doi.org/10.1109/SC.2005.60

[6] Evan Herrnstadt and Erich Muehlegger. 2015. *Air Pollution and Criminal Activity: Evidence from Chicago Microdata*. NBER Working Papers 21787. National Bureau of Economic Research, Inc. https://EconPapers.repec.org/RePEc:nbr:nberwo:21787

[7] INRIX. 2017. INRIX Identifies the Worst Traffic Hotspots in the 25 Most Congested U.S. Cities. 2017. (2017). Retrieved March 26, 2018 from inrix.com/press-releases/us-hotspots/.

[8] Mason Johnson. 2018. How Chicago's Crime Rates Compare to the Rest of America. (9 3 2018). Retrieved March 26, 2018 from inrix.com/press-releases/us-hotspots/.

[9] Marianne Junger, Robert West, and Reinier Timman. 2001. Crime and Risky Behavior in Traffic. *Journal of Research in Crime and Delinquency* 38, 4 (2001), 439–459. DOI: http://dx.doi.org/10.1177/0022427801038004005

[10] Daniel Brent Louis-Philippe Beland. 2017. Traffic and Crime*. (4 2017). Retrieved March 6, 2018 from lpbeland.com/uploads/7/8/7/5/7875420/traffic_crime_r2.pdf.

[11] Global Historical Climatology Network. 2018. Custom GHCN-Daily CSV. (2018). Retrieved March 6, 2018 from ncdc.noaa.gov/cdo-web/search?datasetid=GHCND.

[12] City of Chicago. 2018a. Congestion Estimates by Segments. (2018). Retrieved March 26, 2018 from data.cityofchicago.org/Transportation/Chicago-Traffic-Tracker-Congestion-Estimates-by-Se/n4j6-wkkf.

[13] City of Chicago. 2018b. Historical Congestion Estimates by Segment. (2018). Retrieved March 6, 2018 from data.cityofchicago.org/Transportation/Chicago-Traffic-Tracker-Historical-Congestion-Esti/77hq-huss.

[14] Alexander Stee and Diego Klabjan. 2014. *Forecasting Crime or Crime Forecasting*. Springer New York, New York, NY, 1682–1682. DOI:http://dx.doi.org/10.1007/978-1-4614-5690-2_100257

[15] Ming Tang. 2017. Weather and crimes in Chicago. (2017). Retrieved March 26, 2018 from rpubs.com/crazyhottommy/weather_crime_chicago.

[16] Lisa A. Tompson and Kate J. Bowers. 2015. Testing time-sensitive influences of weather on street robbery. *Crime Science* 4, 1 (28 Apr 2015), 8. DOI: http://dx.doi.org/10.1186/s40163-015-0022-9

Appendices

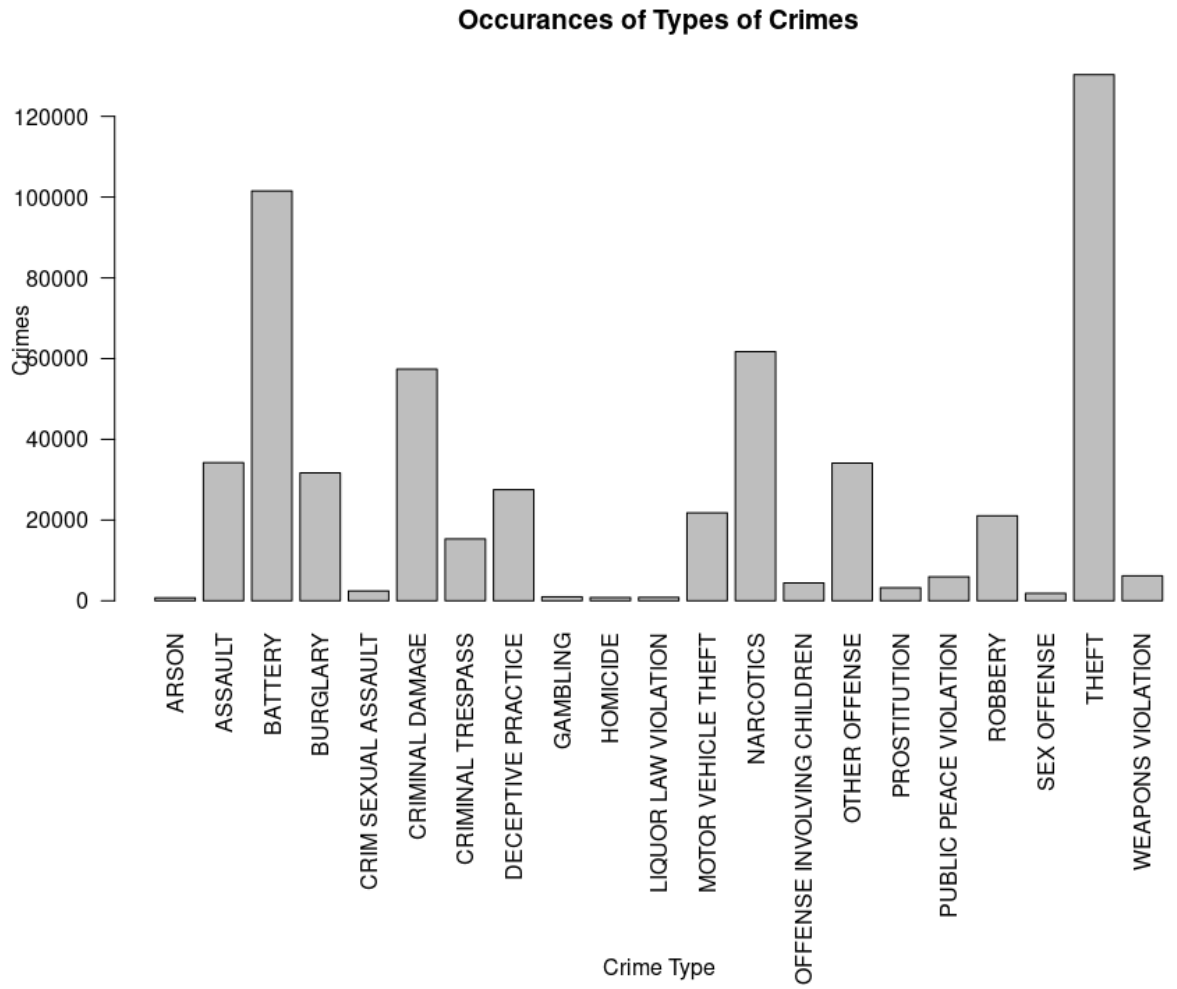

Figure 1: Number of occurrences of each type of crime.

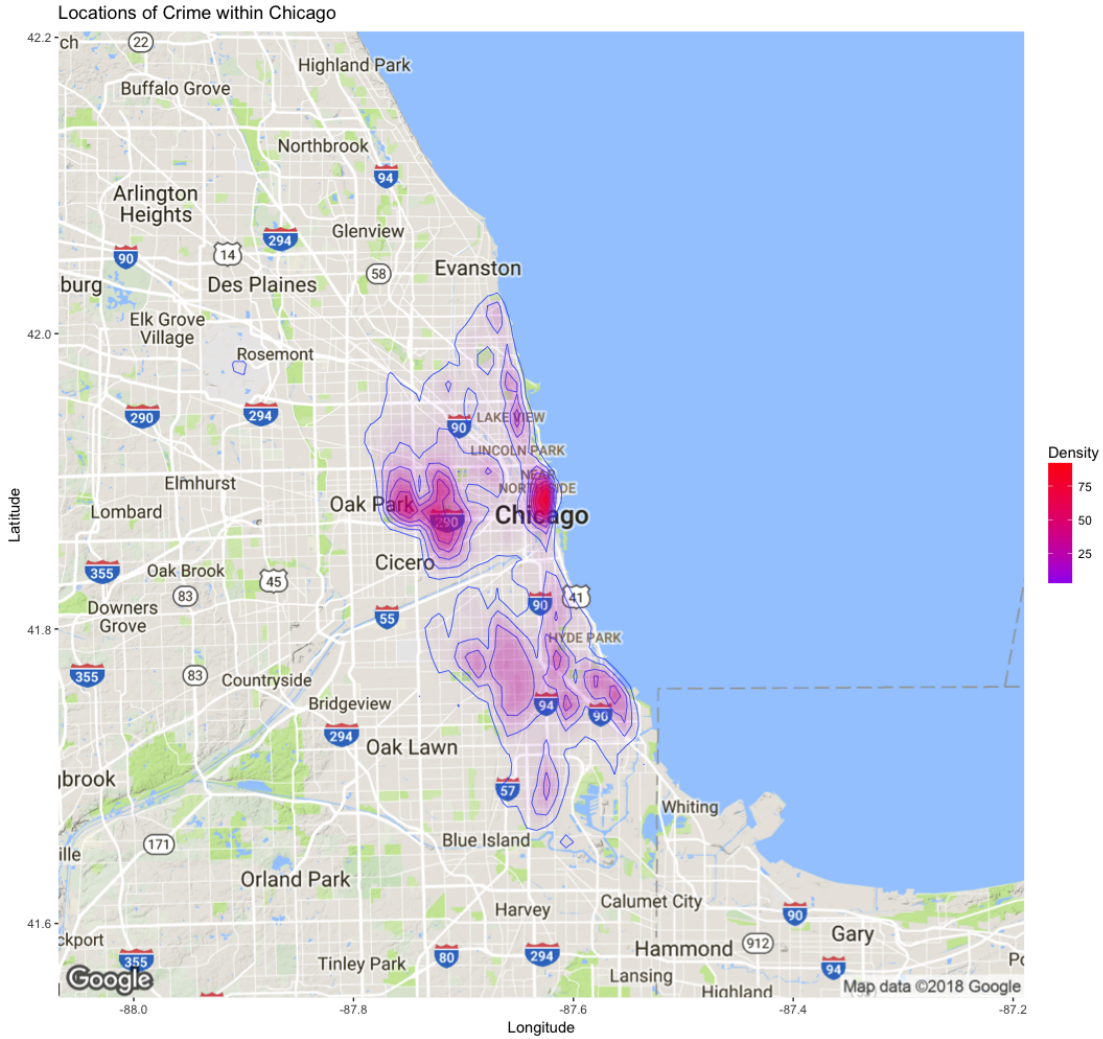

Figure 2: Heatmap of all crimes in the city.

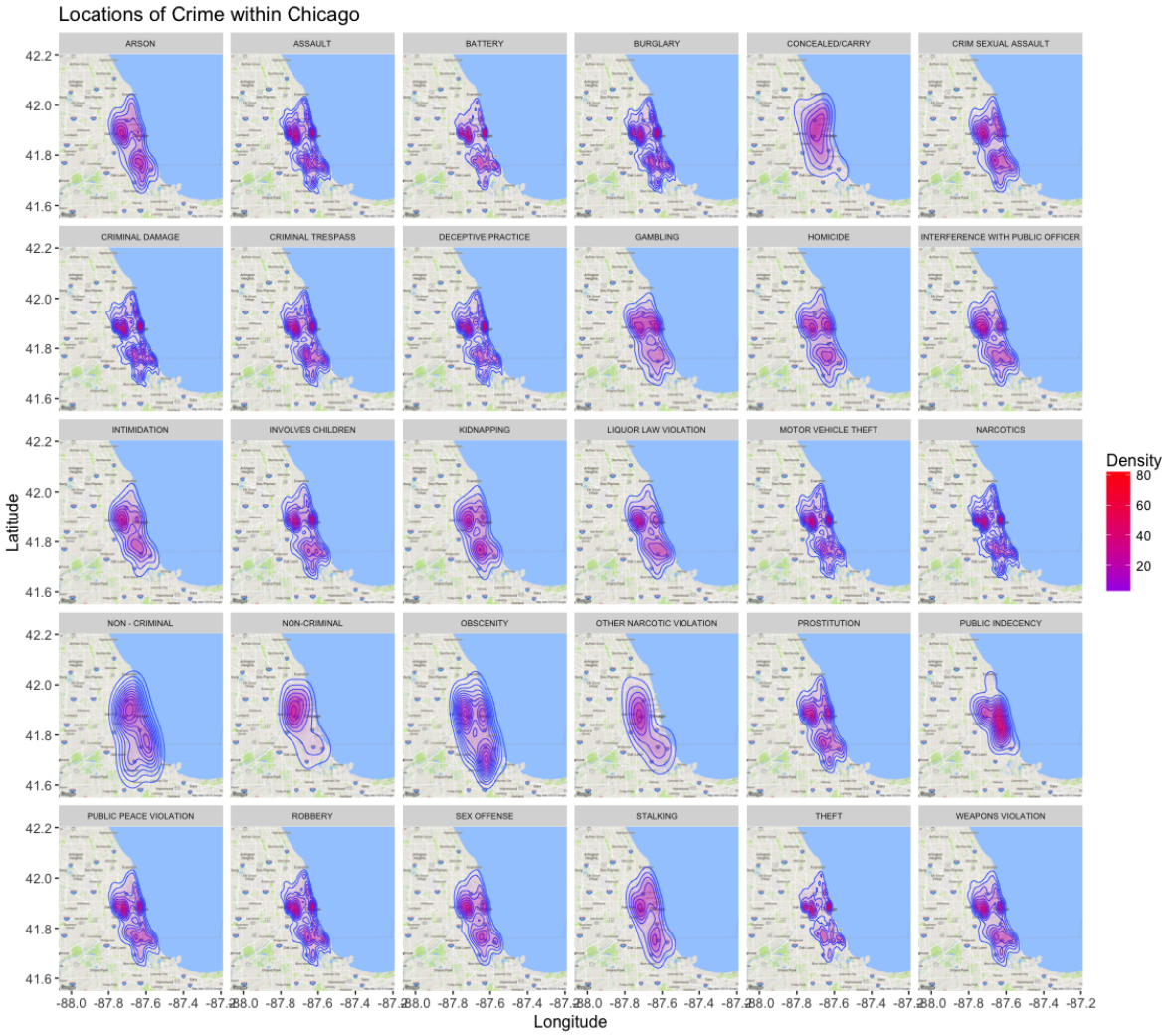

Figure 3: Heatmaps of every crime type.

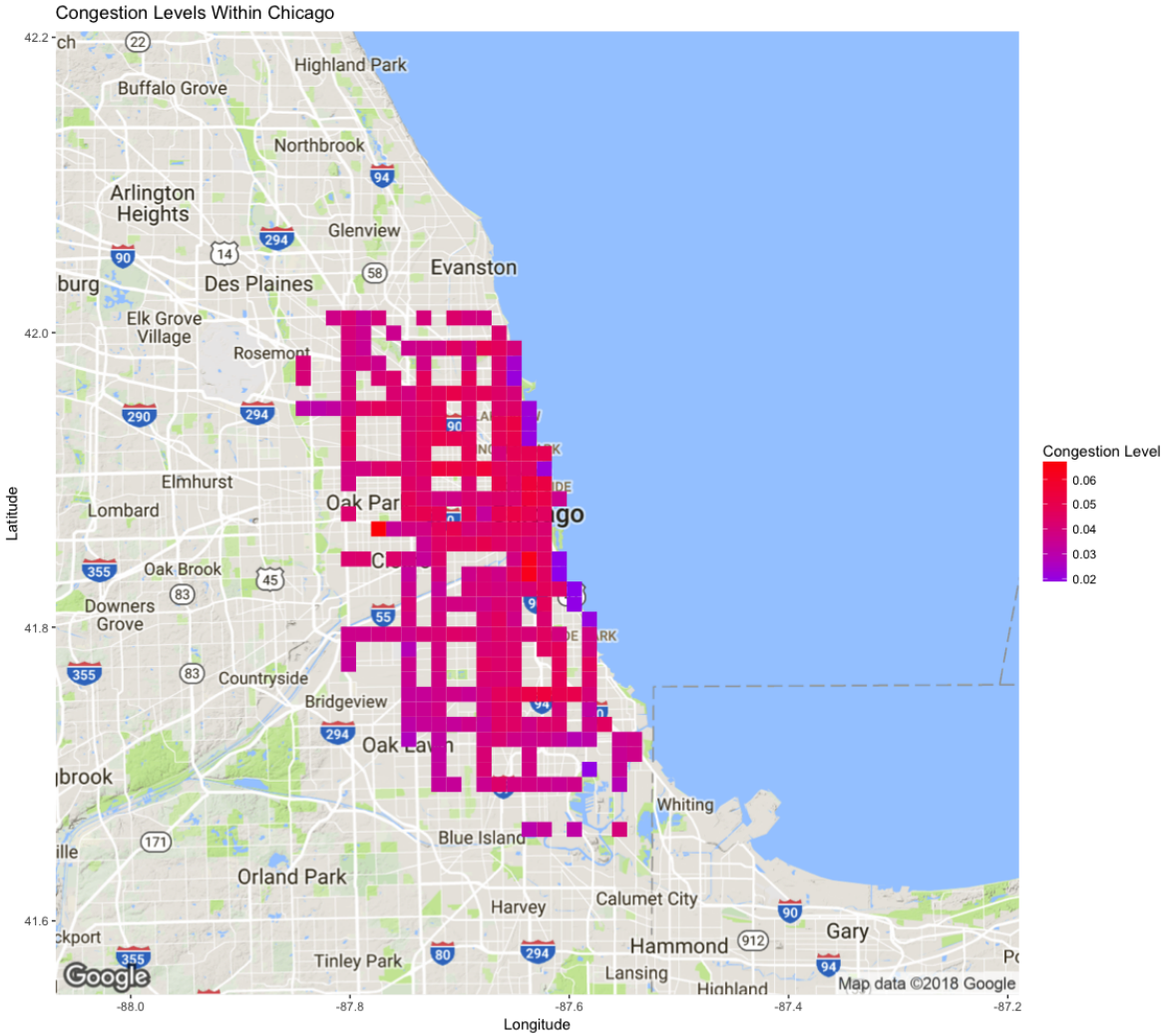

Figure 4: Spatial analysis of congestion levels.

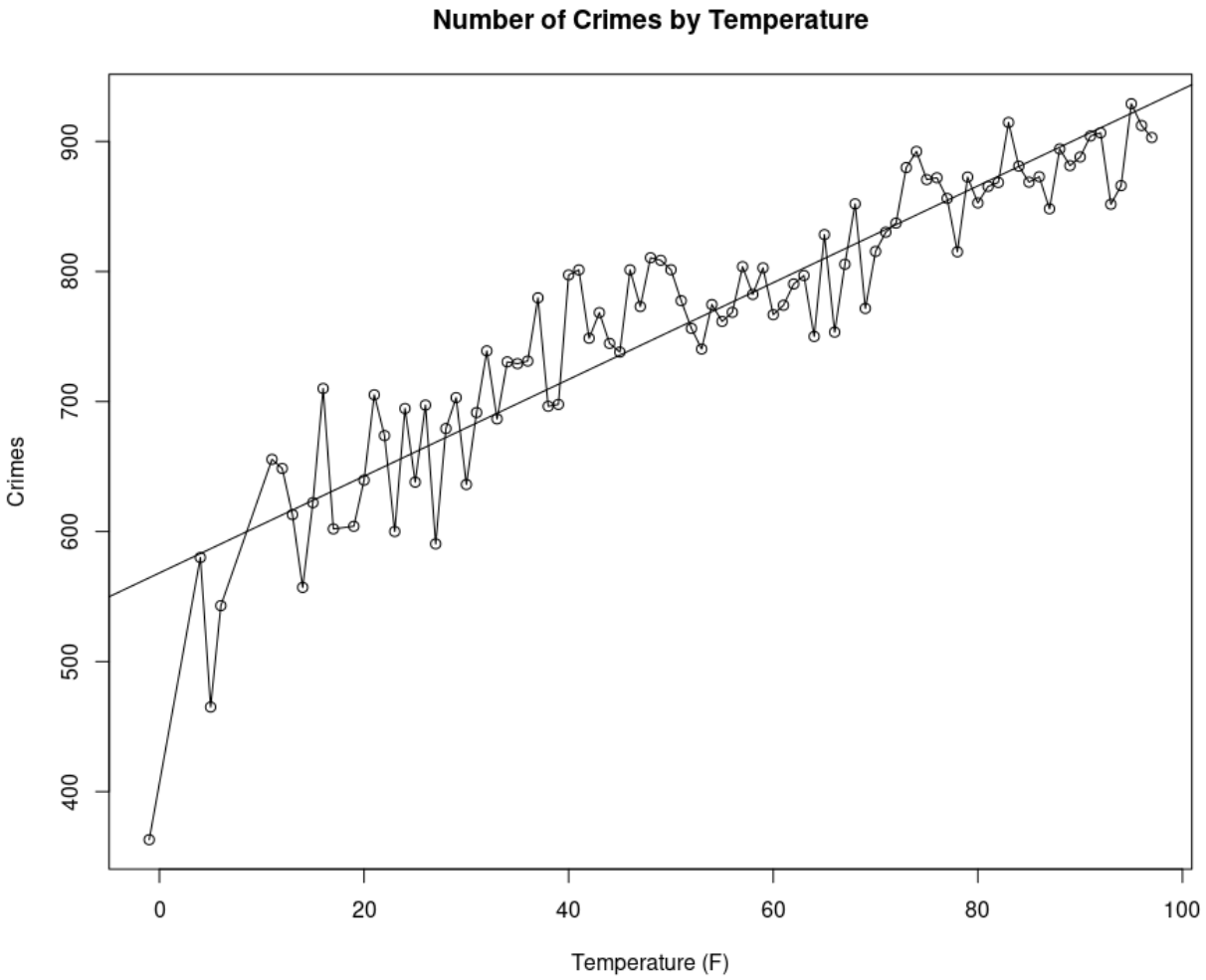

Figure 5: Model of the average number of crimes by daily high.

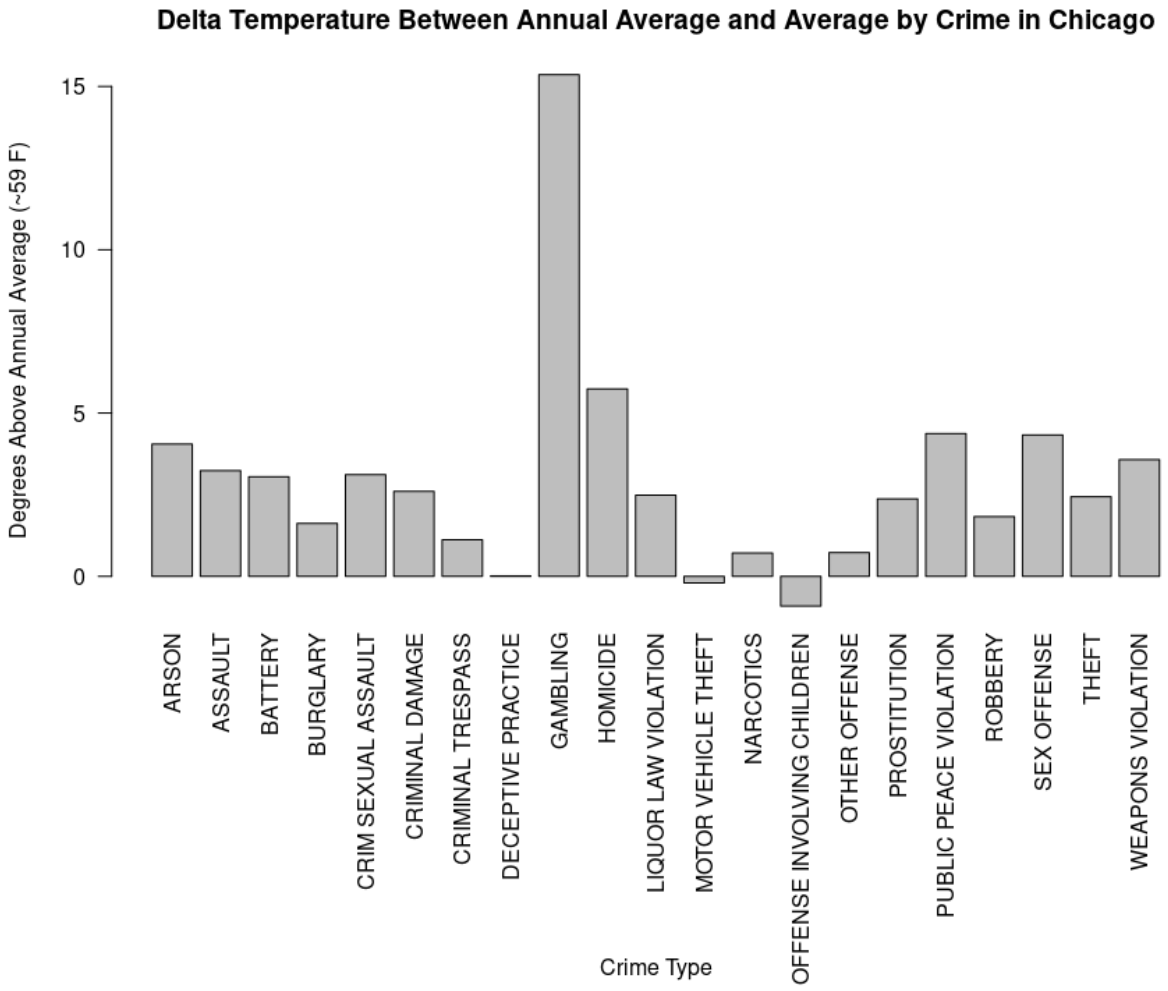

Figure 6: Average temperature above median by crime.

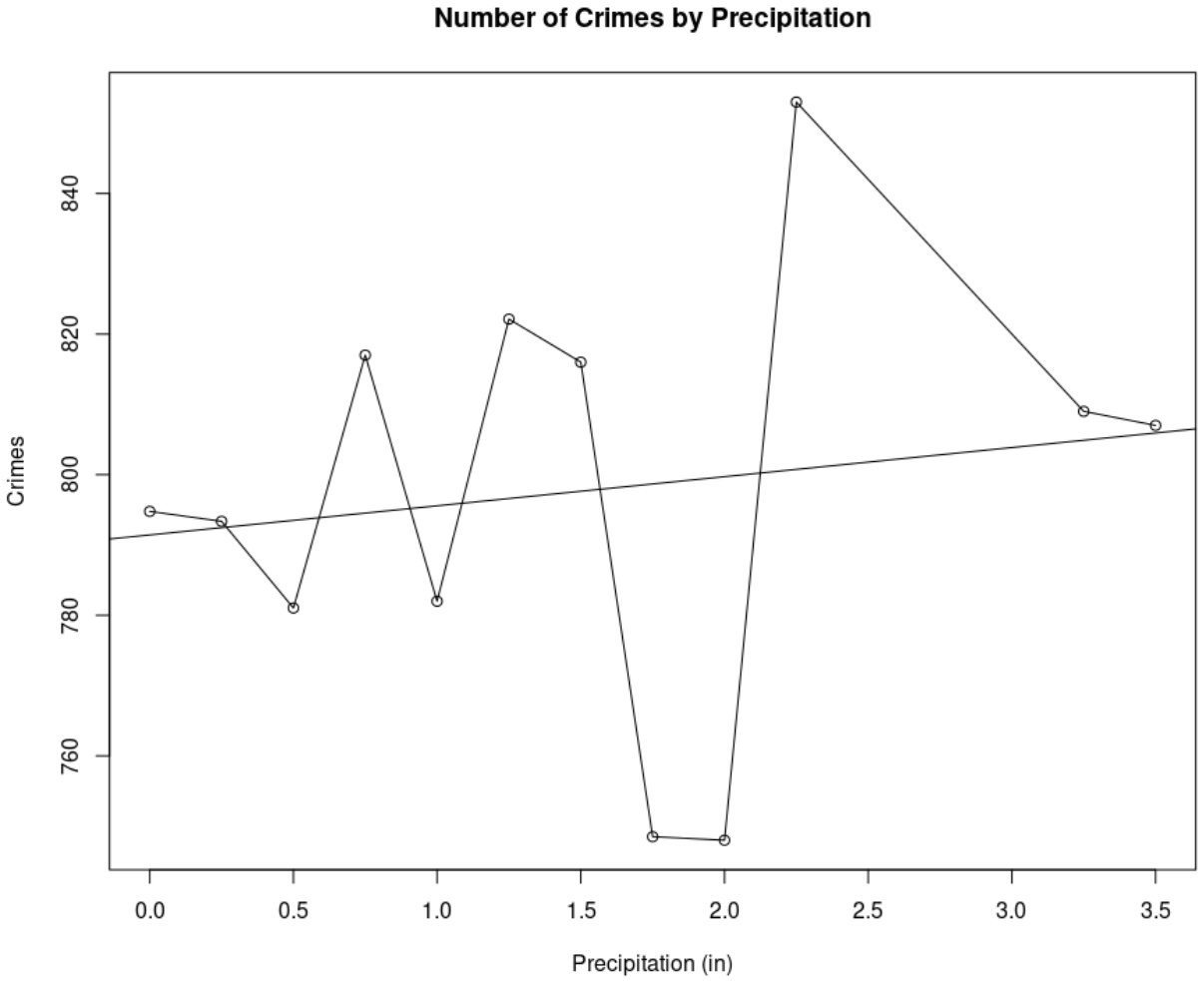

Figure 7: Average crimes per daily rainfall.

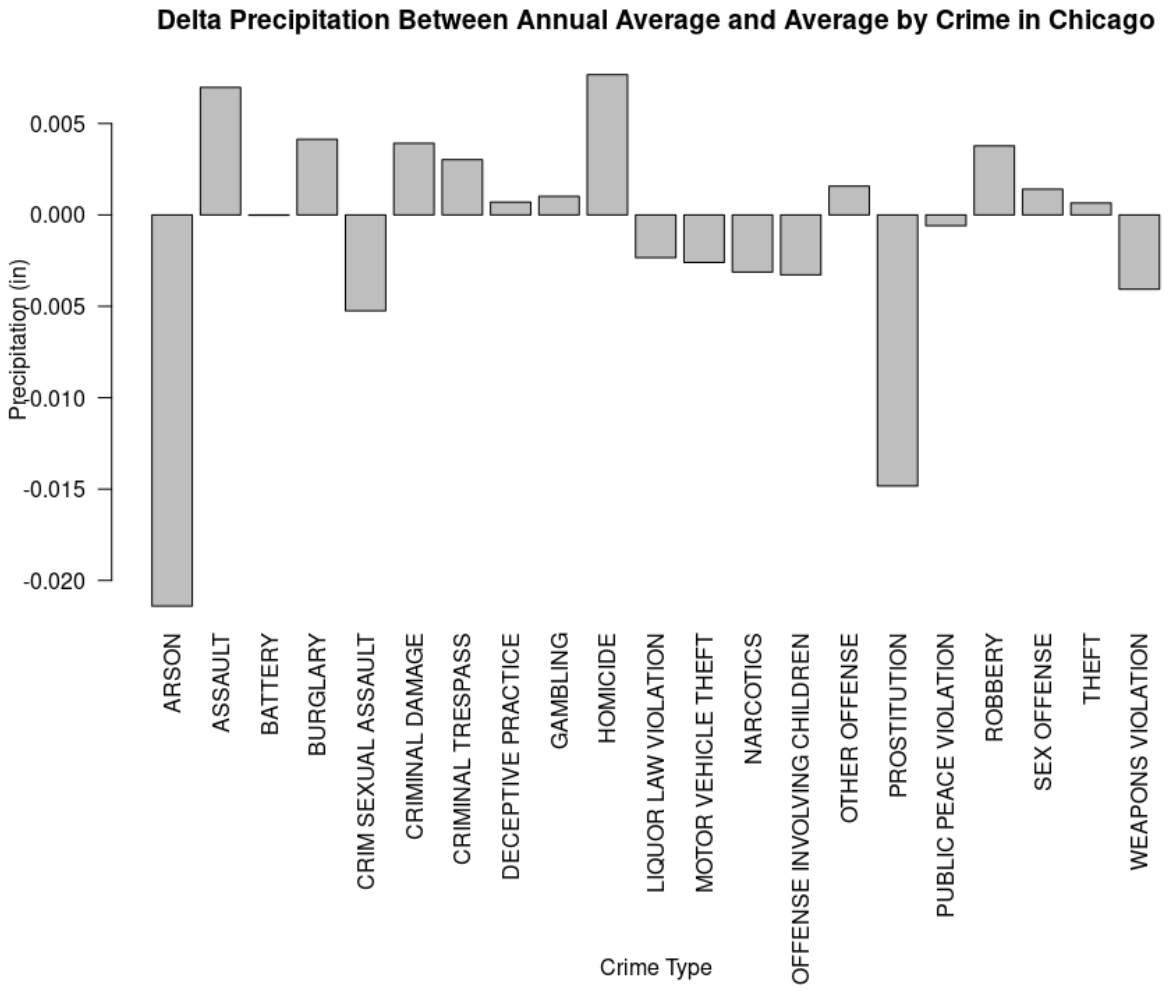

Figure 8: Average precipitation above median by crime.

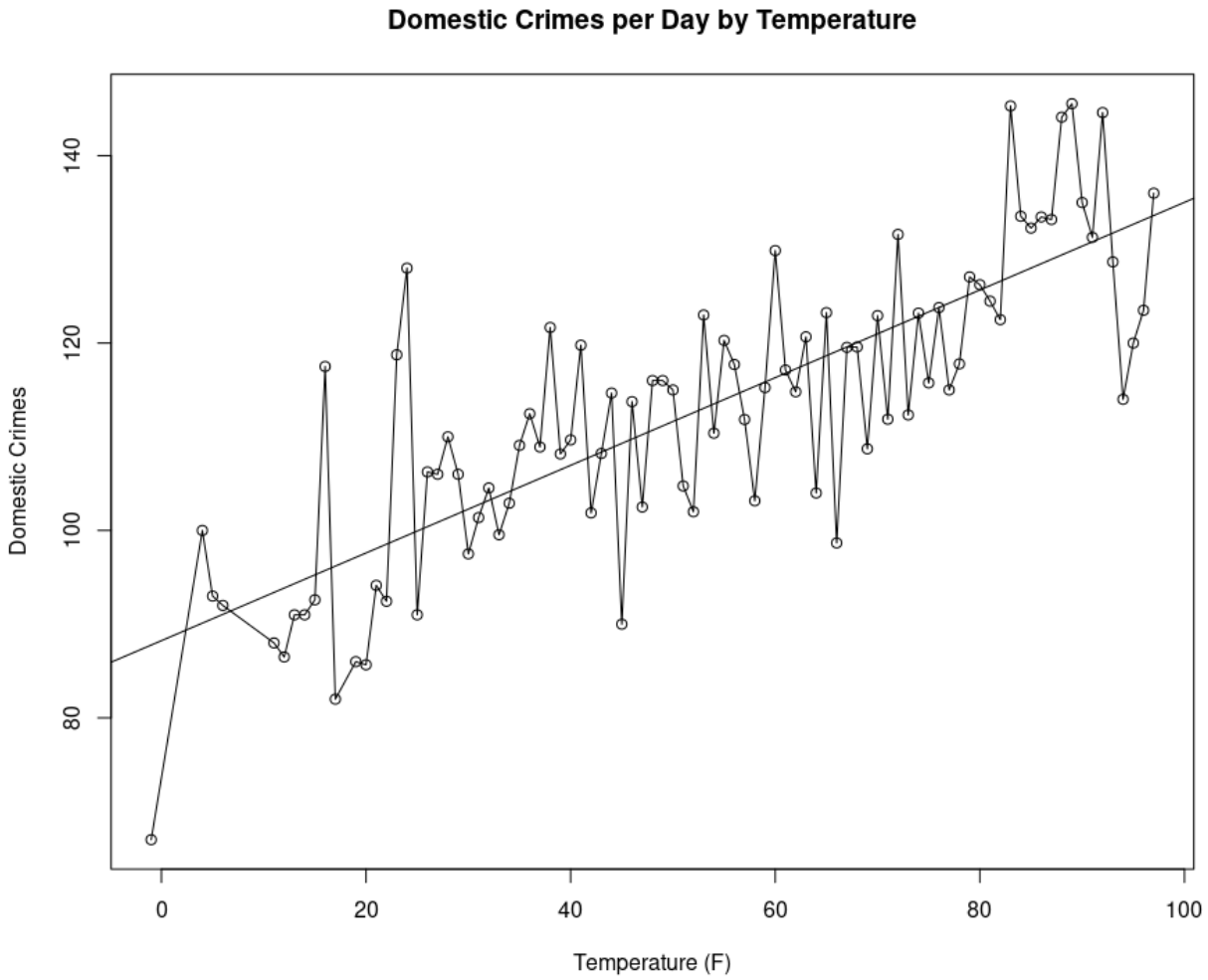

Figure 9: Average number of domestic crimes by temperature.

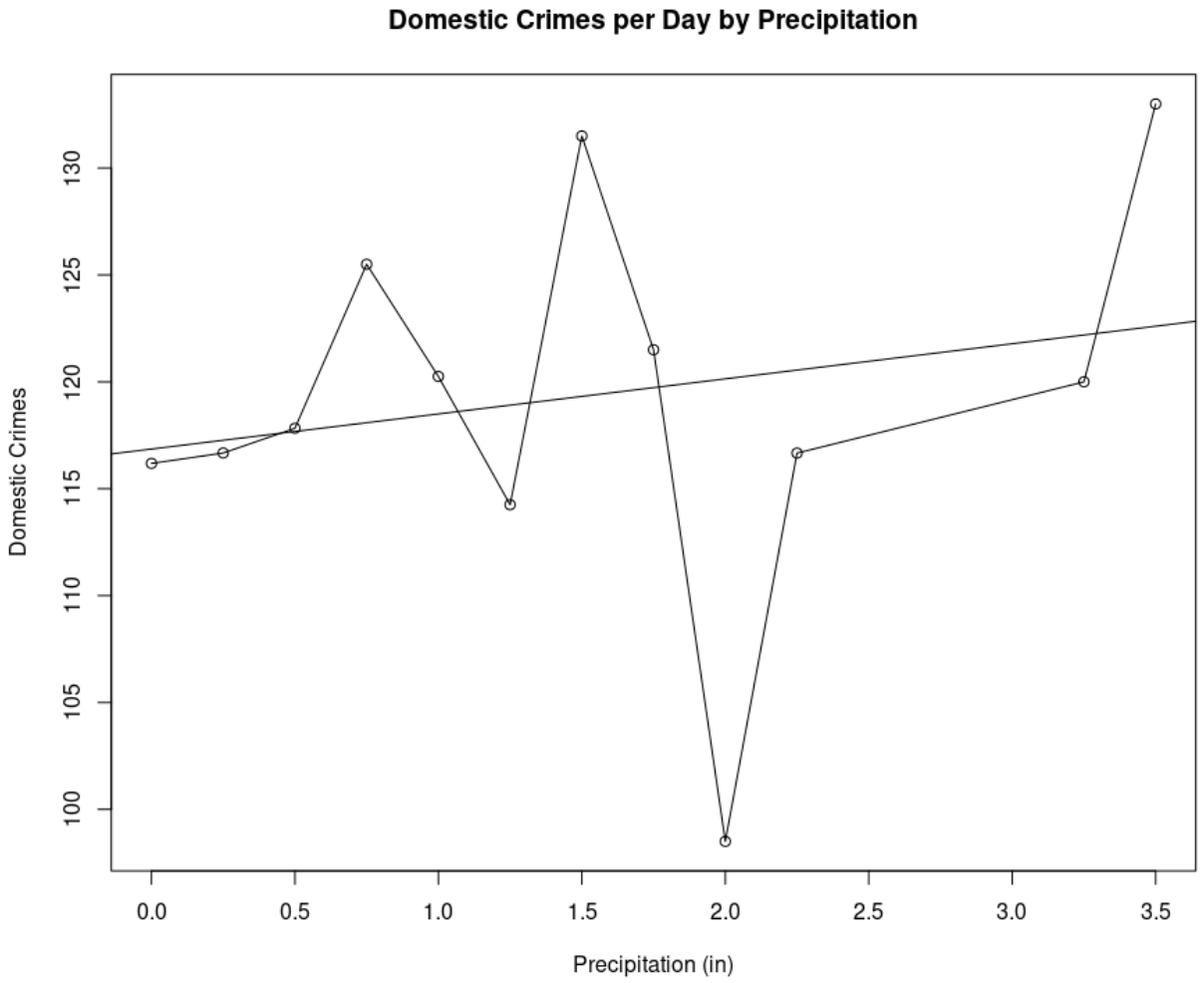

Figure 10: Average number of domestic crimes by rainfall.

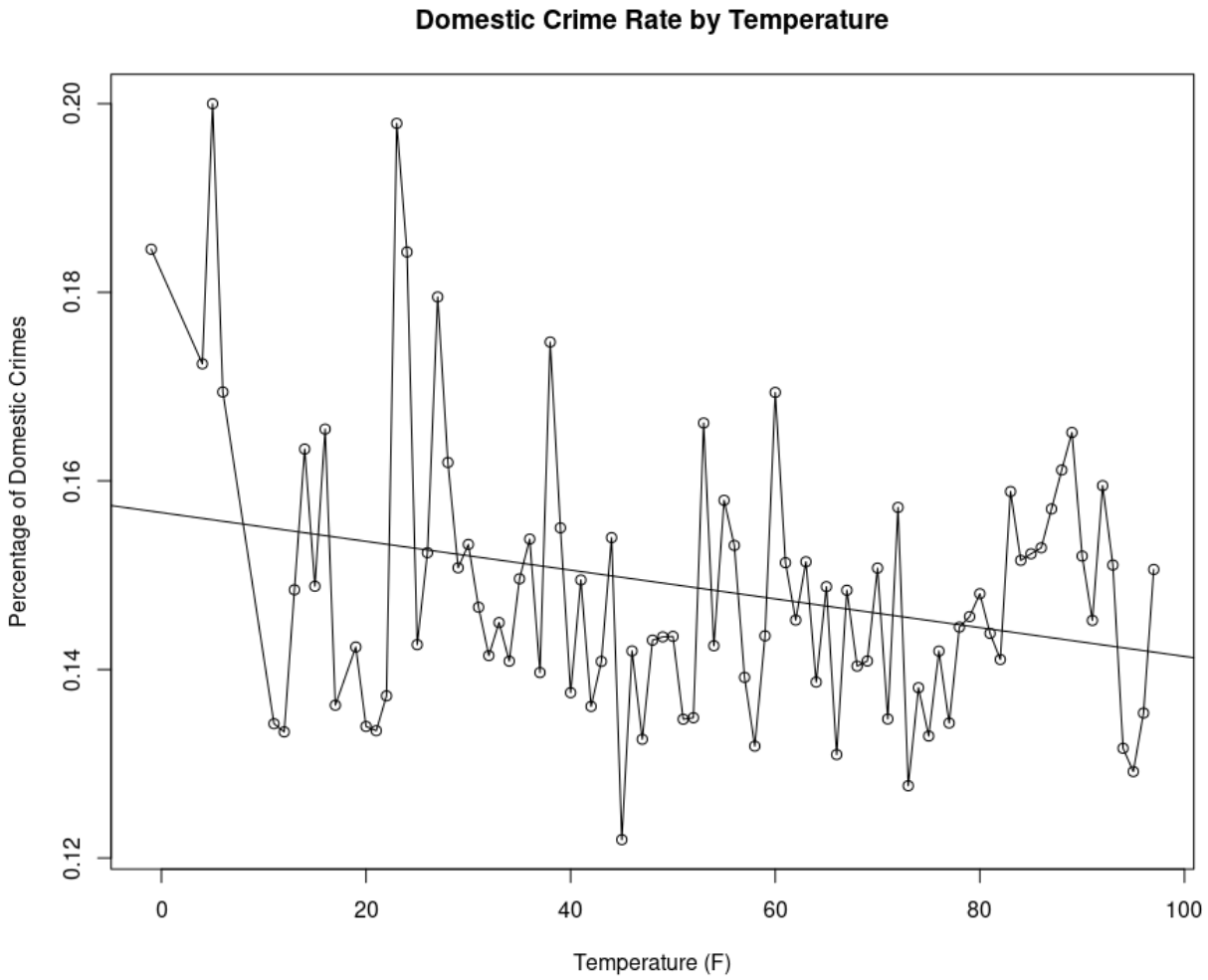

Figure 11: Domestic crime rate by temperature.