Sam Pennington and Brianna King

Data Mining – Fall 2018

Analysis of Potential Correlation in Atlanta

Weather And Crime

**Introduction**

It goes without saying that the city of Atlanta is a dangerous place. According to a recent article published by the AJC, the Atlanta police department is tracking up to 1000 different street gangs working within the limits of metropolitan Atlanta. While law enforcement attempts to combat this issue, they are at times simply overwhelmed by the sheer numbers of the gang population and as a result, the gangs continue to grow in size and number. This myriad of gangs is a large part of why Atlanta has seen a high rate of crime over the last several years. Our project will attempt to draw a correlation between crimes and weather data for the 2017 year in Atlanta. We will evaluate features of data including type of crime committed, type of weather on a given day, temperature, rate of crime committed on a day, and location of the crime committed.

**Prior Studies**

Many studies have been done attempting to correlate the weather with crime. A set of data published on a WeatherOps blog details statistics tracking various types of crime in Chicago. Temperatures measured ranged from -2 to 102 degrees fahrenheit. These graphs were split by crime type which included Assault, Battery, Narcotics, and Stalking. This study focused on crimes of opportunity and was unable to draw a definitive conclusion about temperature and crime other than that crimes of opportunity decrease significantly when the temperature is either really hot (>90 degrees) or very cold (<30 degrees). Crime is shown to steadily rise in the temperature range in between these two values for Assault and Battery while spiking around the 30 degree mark and staying rather steady until the 90 degree point.

Another study conducted through the Chicago tribune also focused on crime in Chicago. It looked at crime data from 2012-2017 including lesser crimes such as theft and more serious crimes such as homicide. The results showed that all measured crimes had a positive, linear correlation to temperature increases with the exception of homicides. Shootings had the highest rate of increase per degree temperature rise averaging an additional 5 shootings per degree rise, while narcotics had the flattest slope being almost negligible. Homicides were shown to not correlate linearly with temperature and were mistakenly graphed as a flat line.

Studies have also been done based on the premise of how humidity and rainfall correlate to crime. One such study, cataloged in the British Journal of Criminology, was unable to draw any definitive conclusions about how rain and crime were correlated as the experimented was performed on three separate datasets and drew three separate conclusions. One experiment showed a positive correlation, one showed negative, and one showed negligible correlation. Humidity was also studied with slightly more conclusive results. An experiment performed on multiple different sets of data showed a slightly negative correlation between humidity and crime rate.

**Data Set**

Our overall dataset consists of two subsets of data. The first subset is a comma separated value file detailing crimes that were committed within the Atlanta area. This dataset was found on Kaggle.com[[1]](#footnote-1) and is based off an original data set published by the Atlanta City Police Department. This ACPD data was compiled, cleaned, and republished for our usage. This dataset contains information about the type of crime committed, the location it was committed, the time is was reported, the number of subjects involved, etc.

Our second subset contains data about the weather. This data was measured by NOAA (National Oceanic and Atmospheric Association) from the Fulton County government station. Our data was obtained by submitting a request to the organization for daily weather data for Atlanta on each day in 2017. This information can be requested through their online portal[[2]](#footnote-2). This data includes maximum and minimum temperatures, precipitation information, windspeed, humidity, etc. Measurements were taken approximately every 30 minutes and the data can be averaged together on a day by day basis to do a more generalized analysis.

**Data Cleansing**

The data sets that we used came with a lot of information that we did not deem necessary. As a result, we decided it was best to not waste memory or load time within our Python notebook, and instead deleted the columns of unnecessary data entirely. After going through each of the two spreadsheets, we narrowed down the data to the following columns:

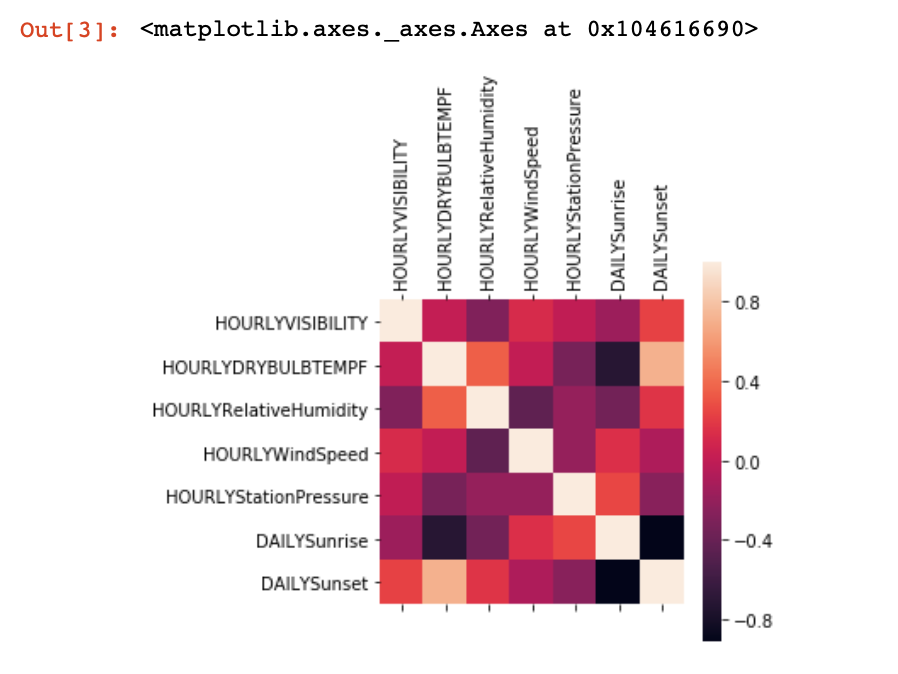
* Crime Data: Date, occurrence time, possible date, possible time, maximum number of victims, the day of the week, the string literal of the type of crime, the neighborhood, and the geographic coordinates where the crime occurred
* Weather Data: Data, visibility, dry bulb temperature in farenheight, relative humidity, wind speed, atmospheric pressure, and the sunrise and sunset times

The weather data needed further cleansing because the data was taken multiple times each day. To overcome this, we averaged the results of each category within each day. These results were then saved in place of the original data, resulting in averaged data for each day of the year.

Within the crime data set, we used a code creator to assign a code to each of the unique string objects within each data feature. This was done in order to perform our classification models and tests. There were also some blank entries within certain rows and columns. To smooth out the data, we decided to take the average of all other values in the column and fill in the missing values with that new value. This is a much safer way to handle missing data rather than filling in “0” values. It also maximizes the amount of data points in the set by not deleting all rows with a single missing value.

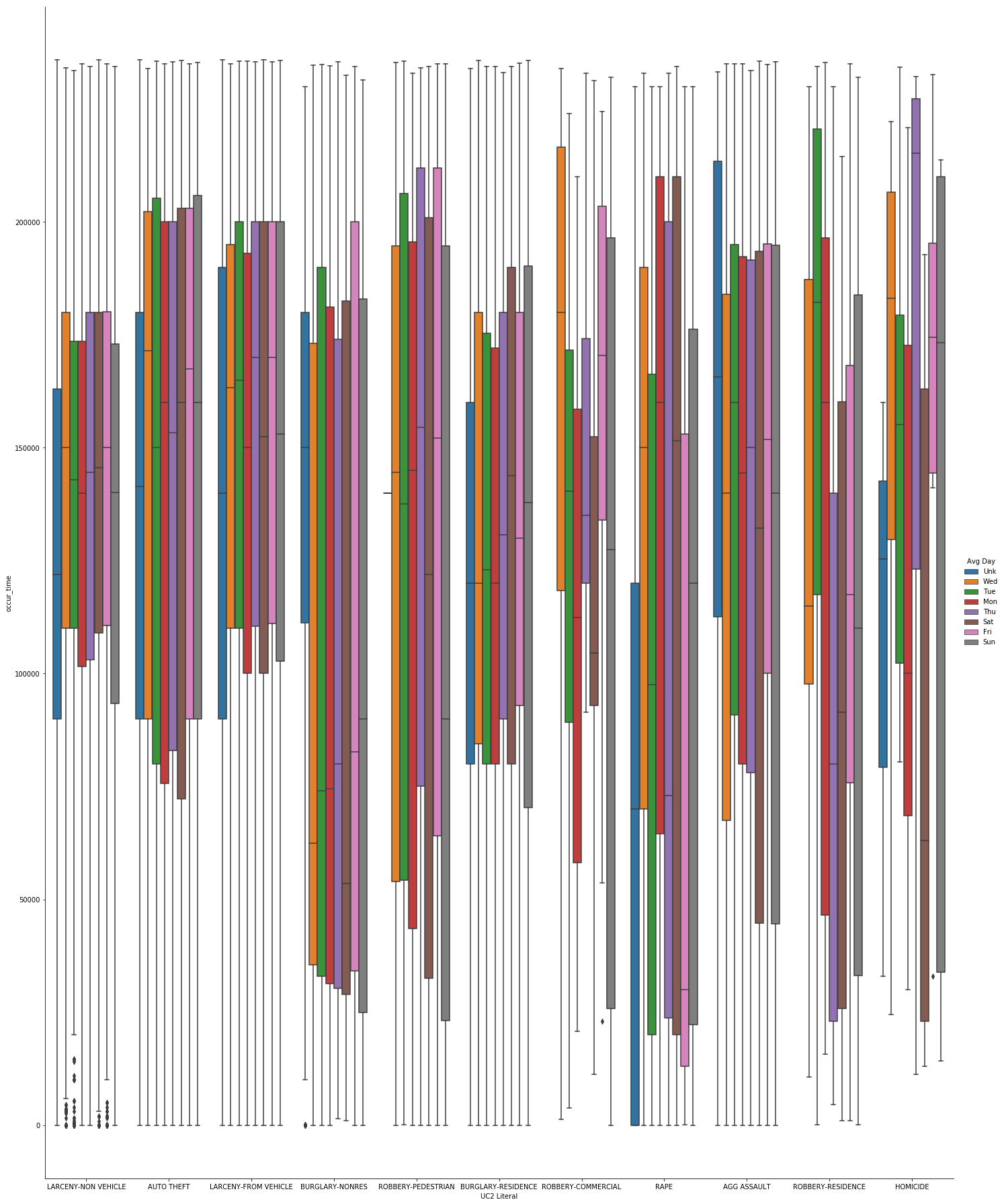
**Exploratory Analysis**

Before beginning to run analysis on a combined weather and crime data set, we decided to run exploratory analysis on each set individually. The aim of this process was to understand our data better in order to better understand the results of our analysis later on. We first decided to draw a correlation matrix based on the numeric attributes of the weather data.

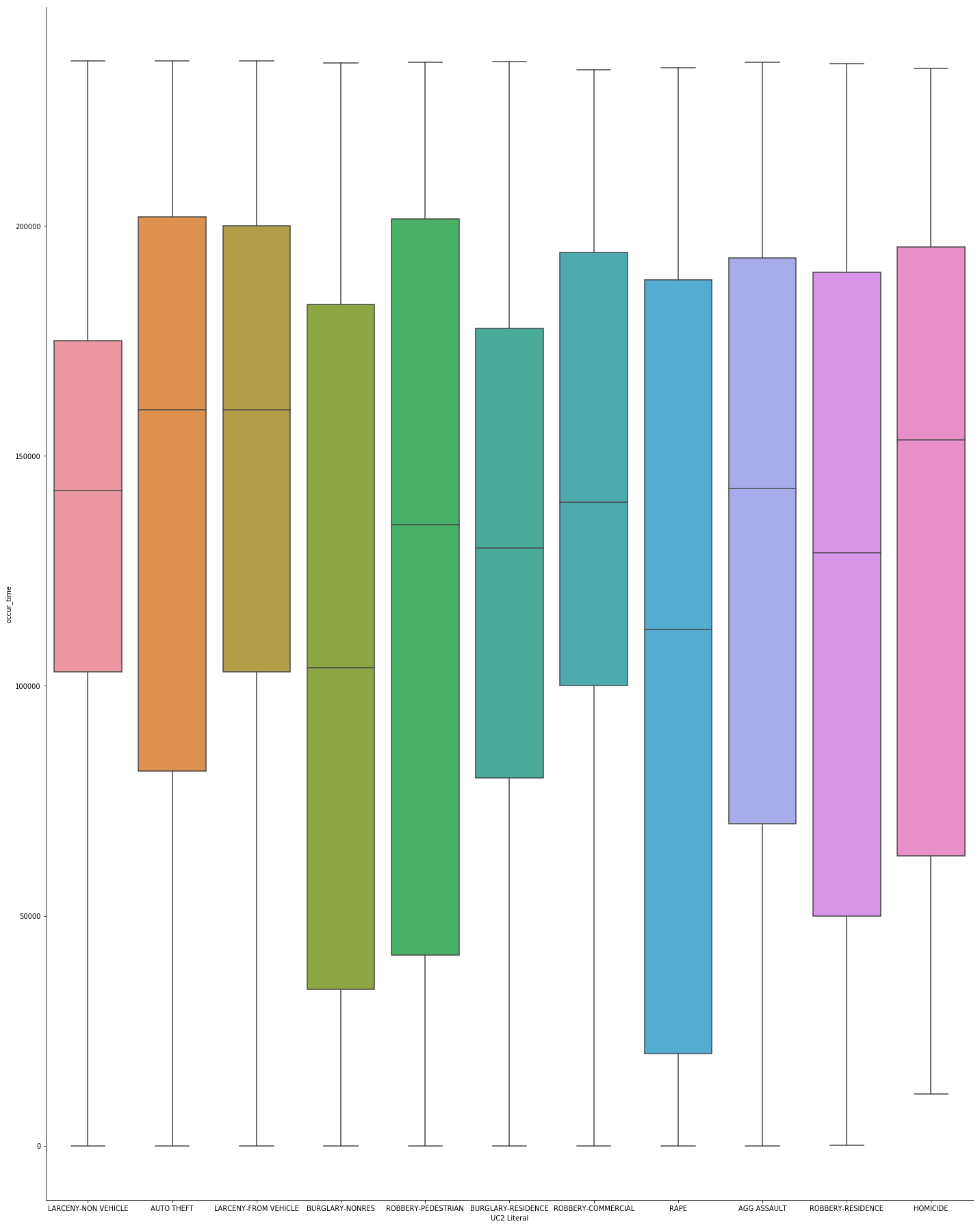


From this correlation matrix heat map, we were able to see that there are strong negative correlation ties between the sunrise and sunset. This would of course be because as days get longer, the sunrise is earlier while the sunset is later, and as the days get shorter, the sunrise gets later while the sunset is earlier. This shows inverse correlation which matches with our matrix. There are also strong positive correlation ties between the relative humidity and the hourly dry bulb temperature. This is to be expected as the temperature generally rises in the summer and cools in the winter. Based on the climate type of the state of Georgia, you can also expect wet, rainy summers, and cool, dry winters. This would seem to match the matrix as a hot summer would be positively correlated to a wet summer, while a cool winter would correlate to a dry winter. One final thing we can see from this matrix is that the dry bulb temperature is positively correlated to the sunset time and negatively correlated to the sunrise time. This would seem to be because as the sunrise gets earlier, and is therefore a smaller numeric value, and the sunset gets later, and is therefore a larger numeric value, the days are longer and the sun has more time to heat the Earth. This would of course result in a higher temperature and would likely be seen in the summer. The result of this matrix is that we should be able to conclude that if one feature in the crime set seems to predict a feature in the weather set or vice versa, then that same crime feature should be able to predict/be predicted by a strongly correlated feature to the corresponding weather feature.

When doing exploratory analysis, we were mostly interested in the average time of day and average day of the week that each type of crime was occurring. Pictured below are two boxplots that are broken out by crime type. The first plot shows both time based on day of the week while the second shows an average time regardless of the day of the week.

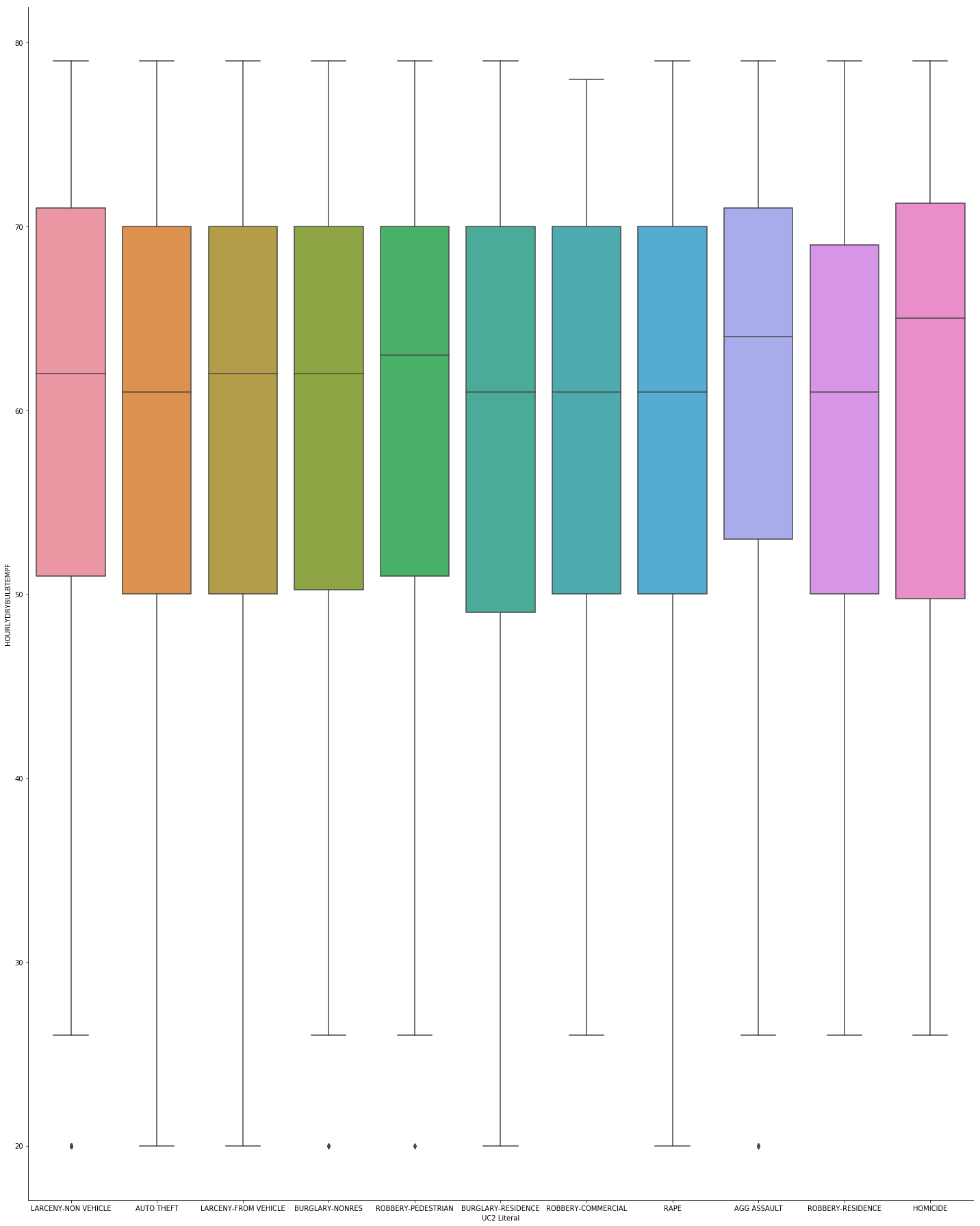


This graph shows a few interesting things. First, it does not seem to matter for most crimes what day it is, as most incidents of the crime will occur around the same time each day. Second, this graph shows that there is a lot of disparity between when a homicide is likely to occur depending on the day of the week. It appears that people are more likely to commit homicide in the evening on Friday, while being more likely to commit homicide in the early morning hours on Saturday. One speculative reason for this is that people are out and about in the beginning hours of the weekend as they have completed their work-week tasks and therefore there are more potential people to commit this crime. This however does not appear to be a trend with any of the other tracked crime types.

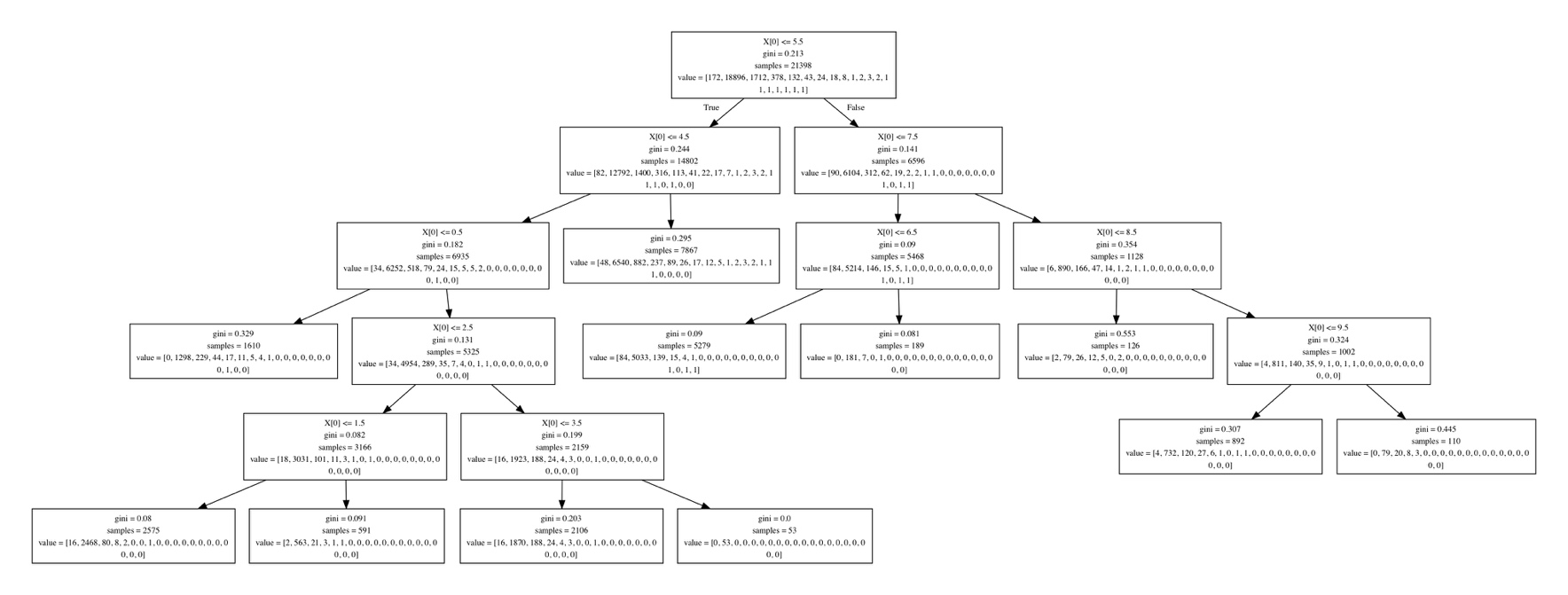
****

Compared to the other box plot, this graph shows us significantly less interesting information. One possible thing of note is that most larceny cases are observed between lunch and dinner. Auto theft incidents as well as aggravated assault incidents are also mostly concentrated during the workday hours.

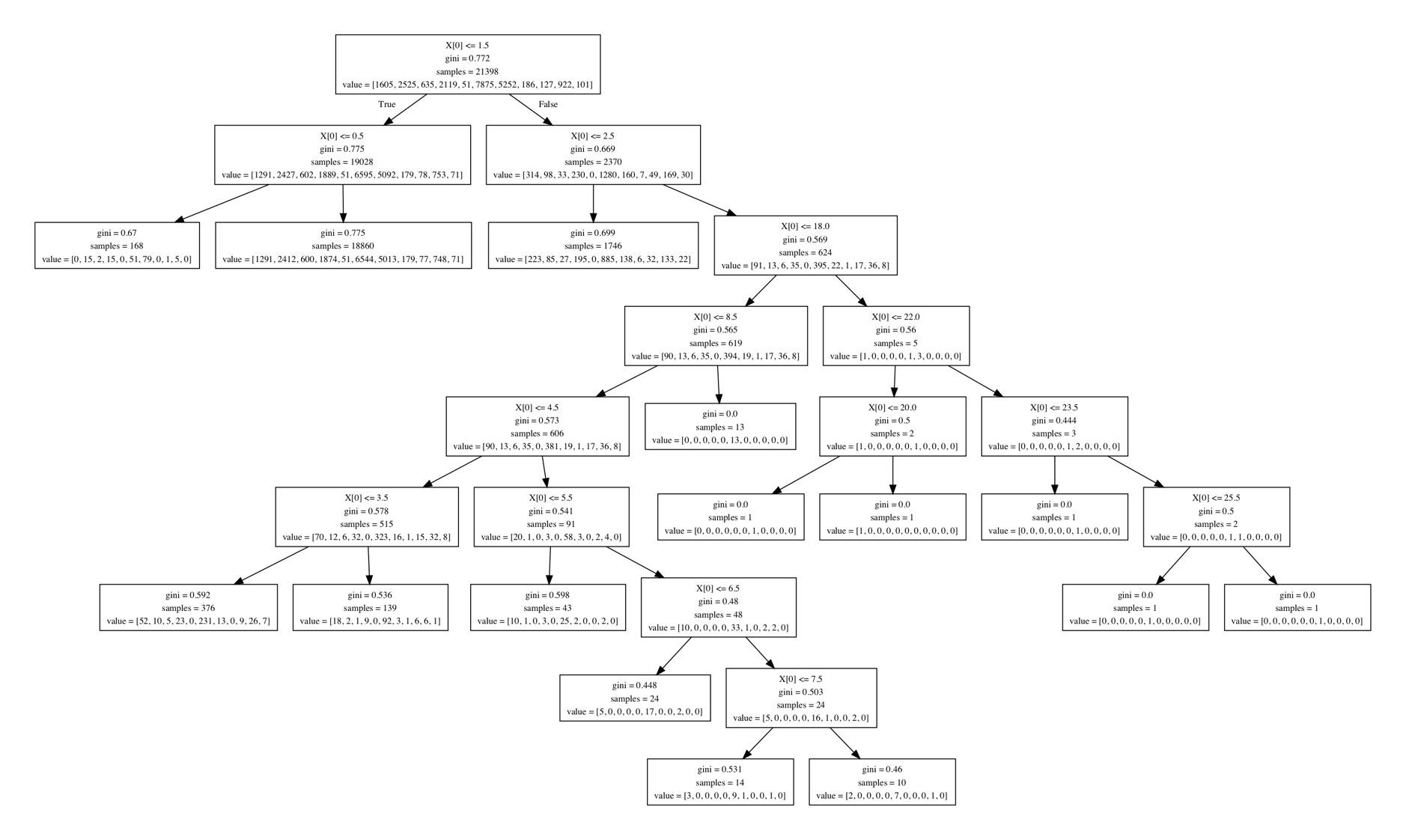
After combining our data into a larger Crime and Weather data set, we did exploratory analysis on the temperature in relation to the type of crime. The results shown below indicate that crimes tend to happen very infrequently in extremer temperature settings. Most crimes in this data set were committed when the temperature was in the range of fifty to eighty degrees Fahrenheit.

****

**Decision Tree Results**



X[0] =



X[0] =

\*Other decision trees were created but were too large to show clearly within a document.

Sources

<https://www.ajc.com/news/crime--law/georgia-gang-situation-getting-worse-investigators-say/hf4OxArZ9YL8UaUQ9yP6jJ/>

<https://blog.weatherops.com/how-does-the-weather-affect-crime-rates>

<https://www.chicagotribune.com/news/data/ct-crime-heat-analysis-htmlstory.html>

1. https://www.kaggle.com/priscillapun/crime-in-atlanta-2017?fbclid=IwAR2P\_xMijBLjw4UtxWwqKuoFtARy1WqCG5GqU\_n47iLV1SmLVAjGomVV9YQ#COBRA-YTD2017.csv [↑](#footnote-ref-1)
2. https://www.ncdc.noaa.gov/cdo-web/datatools/lcd [↑](#footnote-ref-2)