**Final Project: Austin Crime**

Who wouldn’t like to live in a safer city? Obviously, crime is a problem that every person would like to eliminate. Residents of Austin, tourists visiting the area, students at the University of Texas, and the Austin Police Department would be happy to live in a safer environment. With this in mind, we set out to investigate patterns in historical data on crime in Austin, hoping to find insights into how to make the city safer.

Our dataset contains every crime the Austin Police Department responded to in 2014 and 2015, pulled from data.AustinTexas.gov. We would have preferred additional years of data, but these were the only years available. There were approximately 79,000 instances of crime reported during these years, including aggravated assault, burglary, murder, rape, robbery, and theft. Each crime has 15 associated data points such as crime type, location, date, etc.

We dug into this data with the hope that it would provide insights about the types of crimes that are the most prevalent in each area of Austin. This valuable information can be used to caution ordinary people when they are spending time in areas where high rates of crime occur. It can also be used to advise Austin Police Department (APD) on staffing levels and patrol areas. We also sought to identify potential seasonality of crimes as well as any possible weekly cyclical trends. Lastly, we sought to outline the expected number of crimes that will be committed on any given day, in any given Austin zip code.

This expected crime count will be useful to the city of Austin and to the Austin Police Department as they plan for staffing requirements on a weekly basis. It may also be interesting to tourists and residents who are concerned about how many crimes are likely to be committed in their area or around their children.

**Preliminary Exploration**

We began our exploration in Python by reading in the crime datasets from 2014 and 2015. We then combined the two years into one dataframe to begin our exploratory analysis. Initially, we noticed that several of the 15 variables convey overlapping information. For example, location is described in multiple columns, including full address, zip code, council district, X coordinate, and Y coordinate (latitude and longitude using the NAD 1983 State Plane Texas Central FIPS 4203 projection, in feet). Because of this, we focused on a subset of attributes for our analysis: Offense Description (general and detailed), Report Date, Zip code, Council District, Clearance Status, Clearance Date.

**Offense Description Analysis**

We standardized the offense descriptions to align corresponding crimes across the combined dataset (*Figure 1*). APD classifies crimes into 6 categories: theft, burglary, aggravated assault, robbery, rape, and murder. Theft was overwhelmingly the most common offense with 32,000 and 30,000 crimes in 2014 and 2015 respectively. This was followed by burglary with 10,000 occurrences. Aggravated assault, robbery, and rape each had between 1,000 and 5,000 occurrences in the combined two years, but there were only 55 murders in Austin over the two year period.

When we broke theft by subcategory (*Figures 2 and 3*), we found that ‘burglary of vehicle’ was most prominent, followed by shoplifting and auto theft. However, for the purposes of future analysis, we kept ‘theft’ as one uniform category.

**Clearance Status Analysis**

After a crime has been recorded, there are three possible clearance outcomes - cleared by arrest, cleared by exception, or not cleared (*Figure 4*). ‘Cleared by arrest’ means at least one person has been arrested/charged, and ‘cleared by exception’ means the perpetrator has been identified and located but cannot be charged due to something such as death or lack of jurisdiction. The status ‘not cleared’ means that the APD was unable to identify the perpetrator but has closed the investigation.

We analyzed both the clearance status and clearance time for the general crime types in our dataset and found that 82% of all crimes in 2014-2015 have the status ‘not cleared’ (*Figure 5*). In specific, burglary, robbery, and theft had the highest proportion of not-cleared crimes. Interestingly, theft also had one of the fastest clearance times. This may imply that the APD puts more time, money, and resources into investigating more serious crimes. Murder had the highest proportion of crimes cleared by arrest (60%), and an average clearance time of about 28 days. Rape had the slowest clearance time, averaging almost 60 days. We next moved to analyzing the geographic distribution of crimes.

**Geographical Distribution Analysis**

Our initial analysis focused on finding which streets had the most crime (*Figure 6*), and so we filtered the addresses of crimes to extract the unique street name associated with each crime, including both in the analysis for the crimes indicated as occurring at the intersection of two streets, like Guadalupe and 29th Street. We found that I35 overwhelmingly experiences the most crime, with other major streets like Lamar Boulevard and Airport Boulevard also experiencing a large share of crimes. Something to note is we did not adjust for the street’s length when computing these numbers, so the I35 experiences the most total crimes, but does not necessarily have the highest crime density.

Initially, we looked at crime type patterns for more specific locations, including distinct zip codes (*Figure 7*) and district but found that the same general patterns applied. Theft was still by far the most prevalent crime in any given area of Austin, and we felt that this gave us no new insights.

We also plotted a heat map of the different crime types’ distribution over the city of Austin, for theft, burglary, aggravated assault, and robbery (*Figure 8*). Each distribution looks like a crescent moon over the city, with highest crime around downtown, with concentration trailing off to the west along 183 in the north and west along 290 in the south. As expected from the counts of each offense description discussed earlier, the intensity in each heatmap decreases from theft down to robbery. We decided not to plot the murders, thinking that the sample size of 55 would be too small to be illustrative. Additionally, for victim privacy the locations of rapes are withheld, so we did not map this crime type either.

We found there were too many points to plot each individually over the whole map of Austin, but we did zoom in on west campus, zip code 78705, and plot each crime in the two year period over the map, colored by crime type (*Figure 9*). We also noticed the number of crimes in our dataset at addresses on the UT Austin campus is suspiciously low, and upon investigation realized the University of Texas Police Department keeps separate records, and so our dataset does not include every crime reported to the UTPD, only those with an investigation assisted by the Austin Police Department.

**Seasonal Pattern Analysis**

Lastly, we examined the dataset for potential seasonal patterns. From 2014 to 2015, we found that most crimes decreased in overall count. However, assault and robbery had 1.4% and 7.0% increases, respectively. On average, crime was highest in the summer months with a peak of approximately 7,000 crimes in July and lowest in February with under 6,000 crimes (*Figure 10*). On any given week, crime rate tends to peak on Monday and gradually decrease throughout the week until Sunday (*Figure 11*). These weekly trends are weighted by the thefts, as it happens to be the most common type of crime.

**Solution and Insights**

Upon completion of the exploratory analysis, we set out to find an estimate for the weekly quantity of crimes committed in Austin, grouped by zip code and month.

Looking at a graph of month-by-month total crimes, we saw evidence for some seasonality, with lows in midwinter and highs in summer (*Figure 12*). However, with only 2 years of data to model with, we didn’t feel comfortable starting with a linear regression and playing with the residuals. Two points make a line, but without more years of data we doubted we’d accurately capture the seasonal trend.

Instead of modeling the seasonality, we opted for a granular average approach, where we summed the number of crimes reported, grouping by zip code, day of week, crime type, and month. We then divided these numbers by the corresponding number of that particular weekday that occurred in our 2 year period (for example, there were 8 Mondays in January 2014 and January 2015, but 10 Fridays). This gave us the average number of each crime type on each weekday, by zip code, averaged across all weeks in our two year sample (*Figure 13*).

For example, on an average week in May in zip code 78753, the Austin Police Department should expect 13 crimes on Monday, 10 on Tuesday, 14 on Wednesday, 11 on Thursday, 11 on Friday, 12 on Saturday, and 7 on Sunday. Conversely, the same area in December should expect 12 crimes on Monday, 9 on Tuesday through Saturday, and 7 on Sunday. Each day’s distribution varies slightly, with the majority of crimes expected being theft. With this in mind, the Austin Police Department should be able to better distribute the patrol cars and staffing requirements.

While we feel these values are the most accurate prediction of future crime counts, given the available information we worked with, we recognize that two years of data is insufficient to make a strong prediction for year-over-year trends. Were we to replicate this analysis with more data, we feel more complicated models such as linear regression with dummy variables for month and interaction terms, or decision/ensemble trees, would prove much more accurate. Additionally, with more years of data we could take into account special events and holidays, like Austin City Limits, UT Austin graduation days, or New Year’s Eve, and more accurately model the changes in crime around these occasions.

**Appendix**

Figure 1.

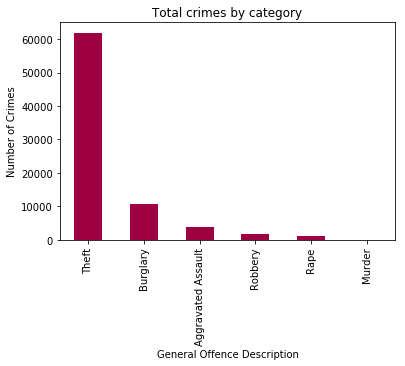


Figure 2.

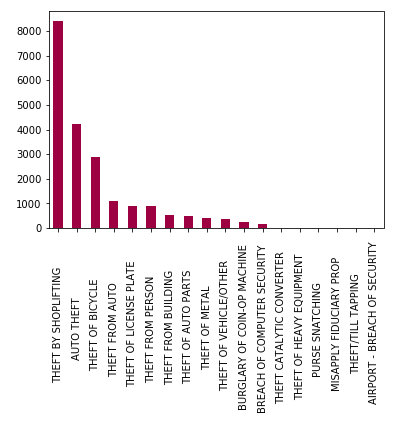


Figure 3.

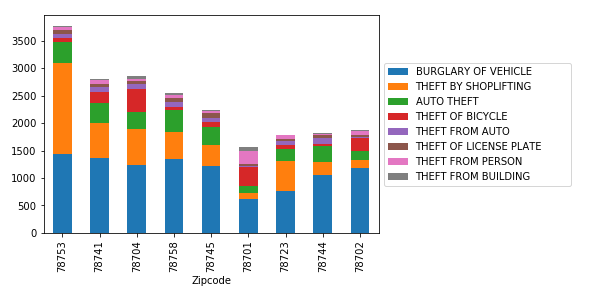


Figure 4.

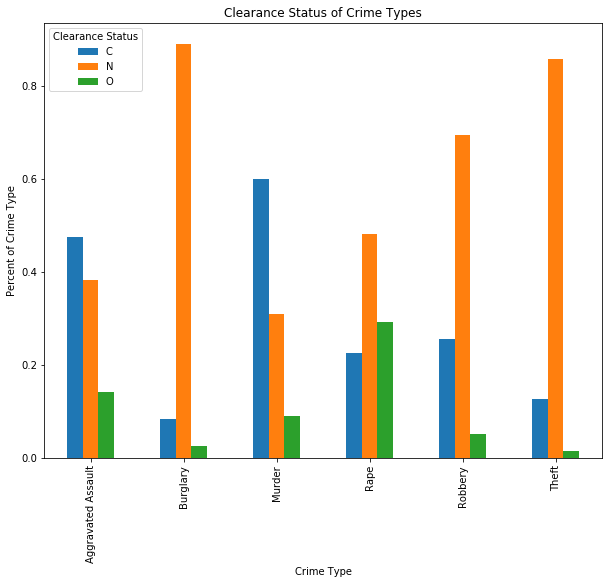


Figure 5.

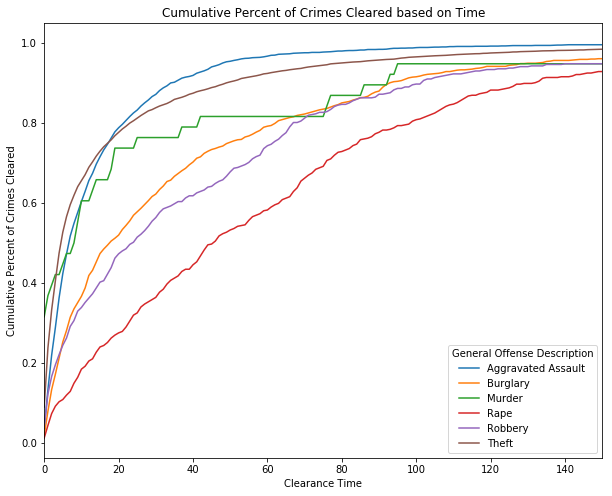


Figure 6.

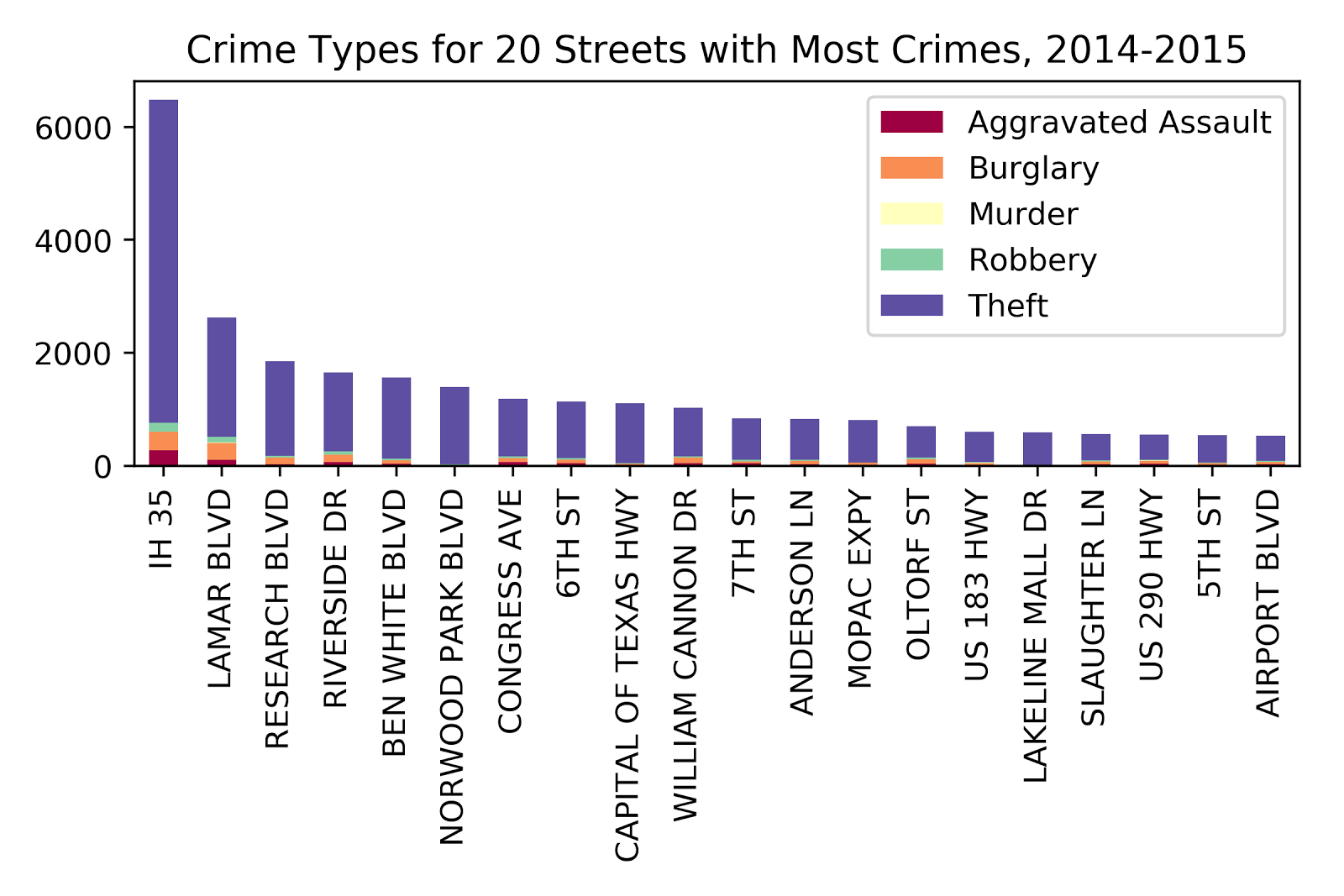


Figure 7.

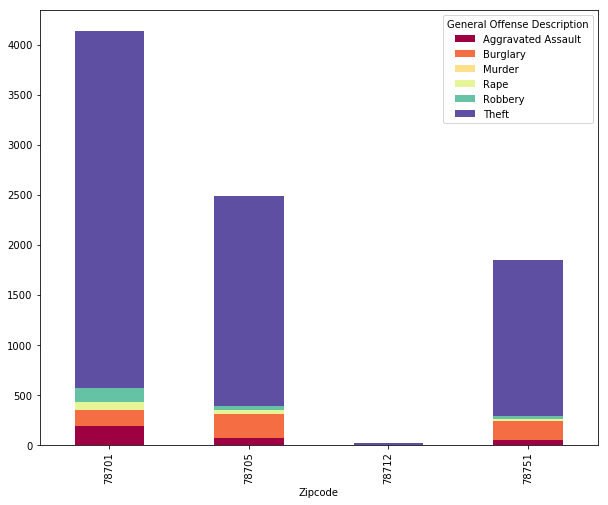


Figure 8.

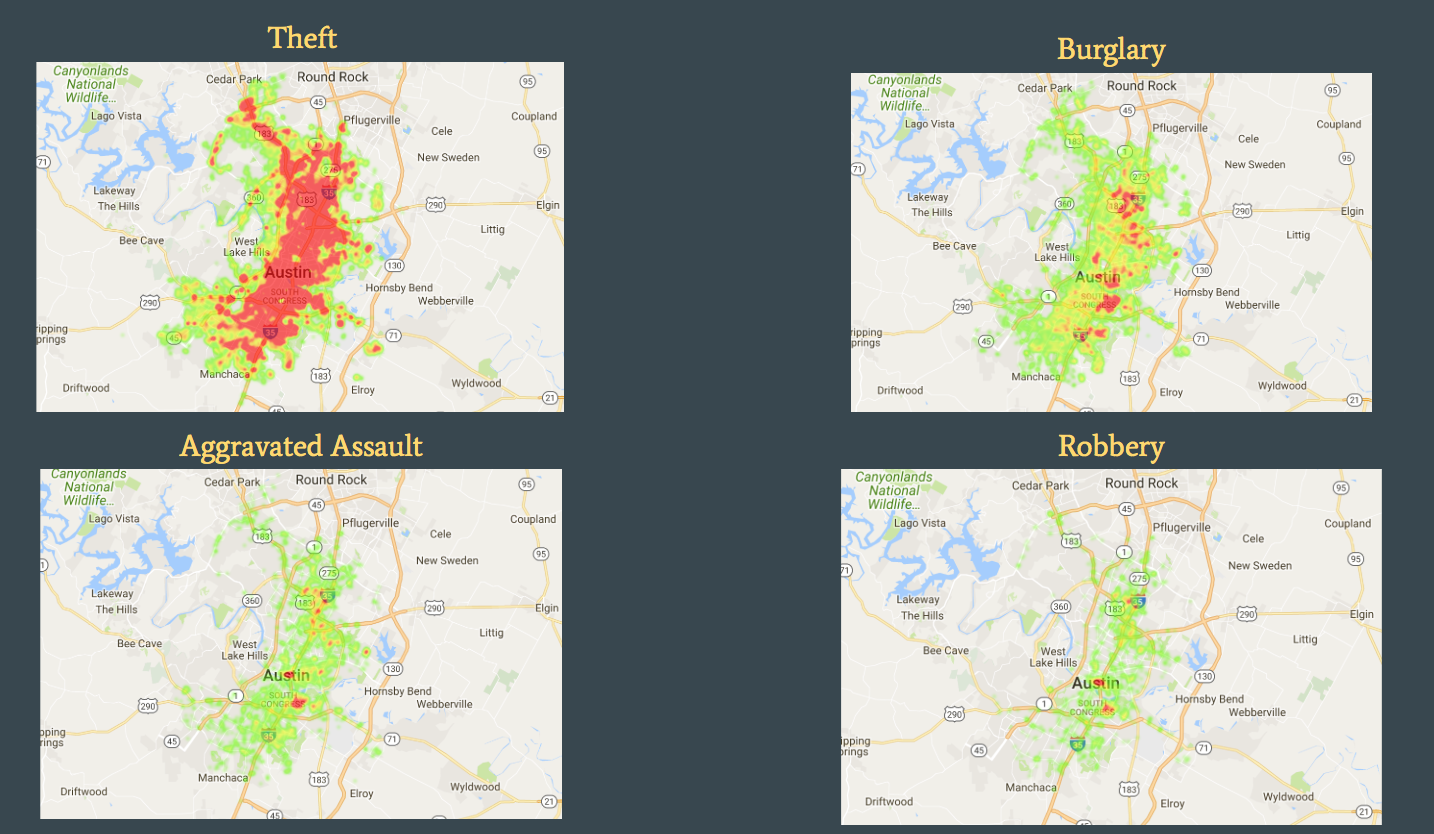


Figure 9.

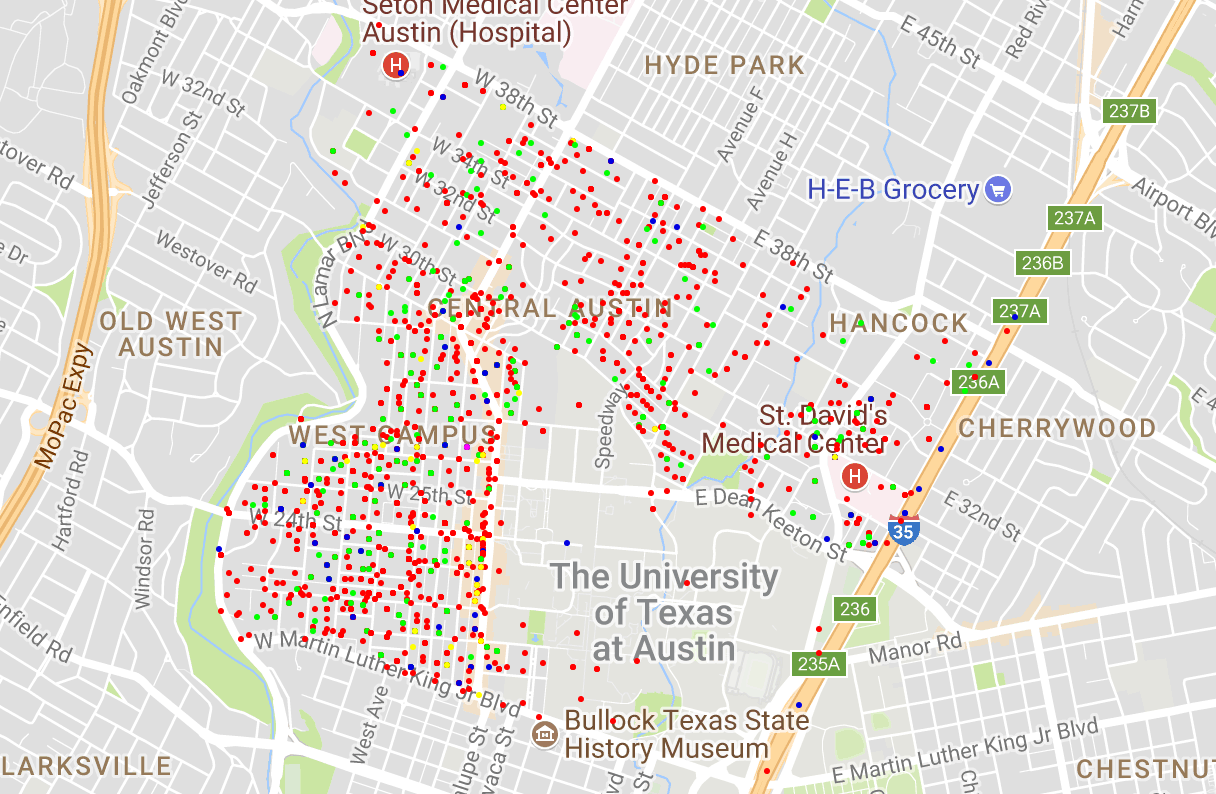


Figure 10.

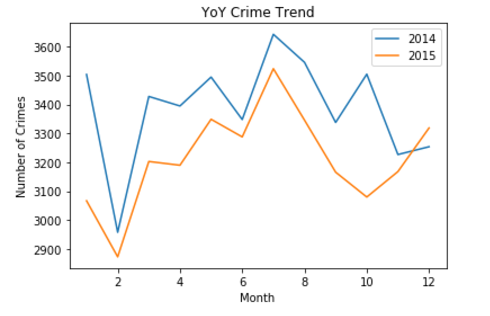


Figure 11.

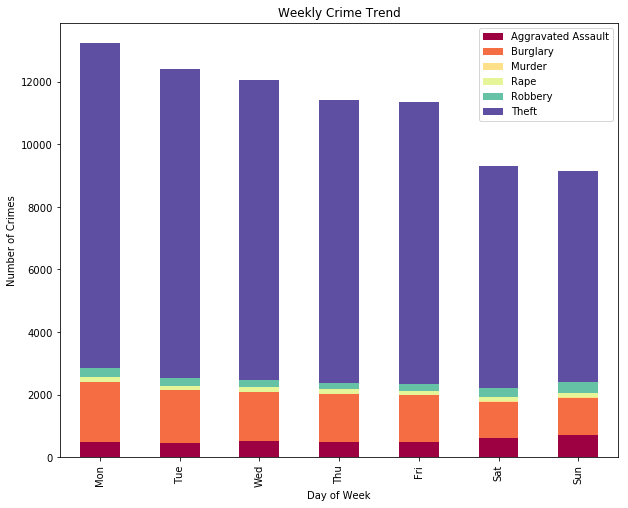


Figure 12.

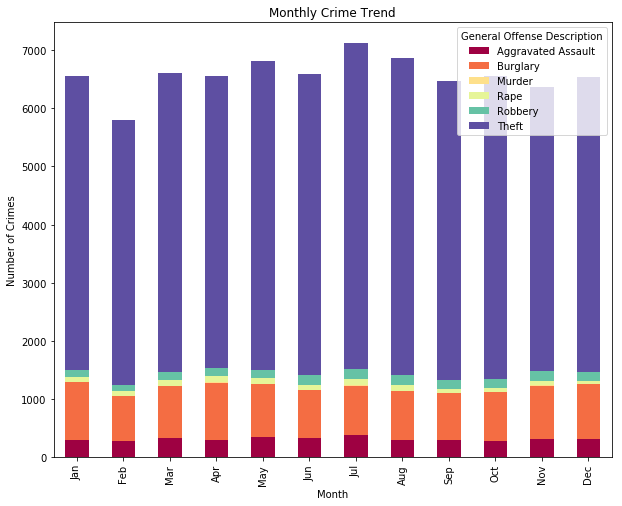


Figure 13.

