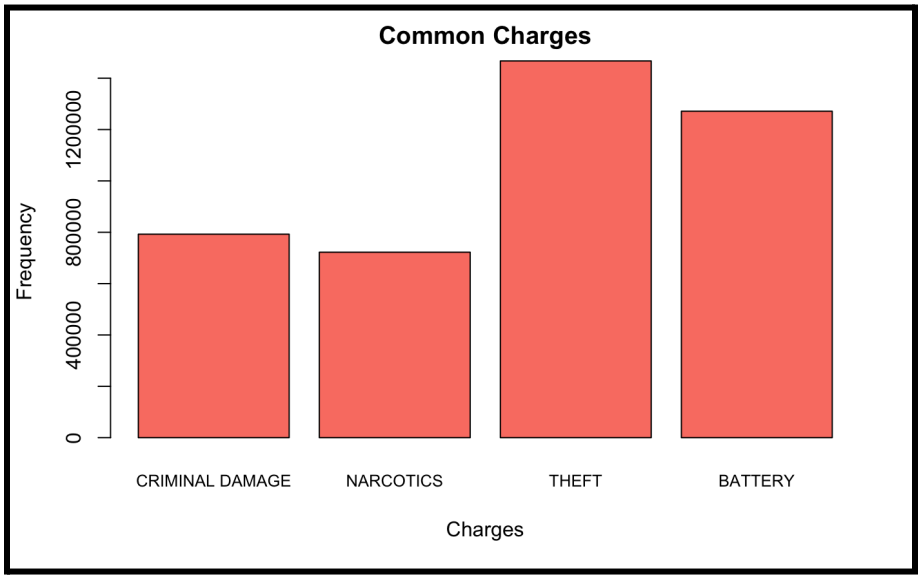
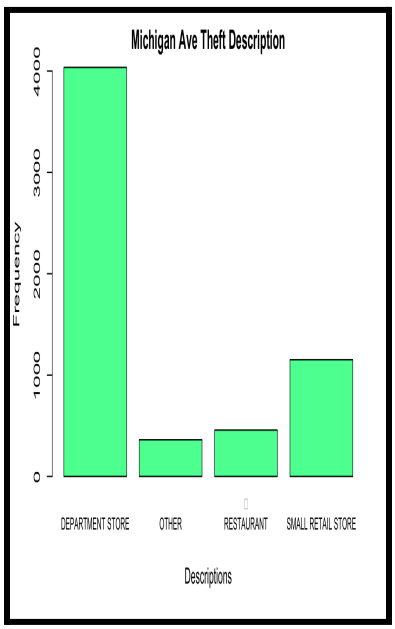
Capstone Project(Chicago Crime) report

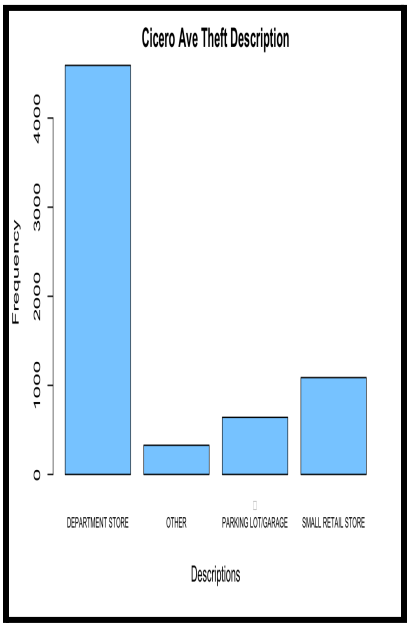
This dataset was important to me because I wanted to learn more about crime in chicago. I wanted to look specifically at theft because it is the most common type of crime. My main focus with this project was to see where exactly theft is happening in Chicago, and why that might be. I found my dataset from Kaggle. It is a data set of all crime in Chicago from 2001 - 2019. It was originally 6,954,967 rows by 30 columns (208,649,010 data points) before filtering/cleaning. First I selected only certain columns, because some of the columns were of no use for my project(ex. ID, Case Number, Police Districts). I then wanted to find out what was the most common type of crime. I used "table" to find out the frequency of the different types of crime, and found out that Criminal Damage, Narcotics, Theft, and Battery were the most common.

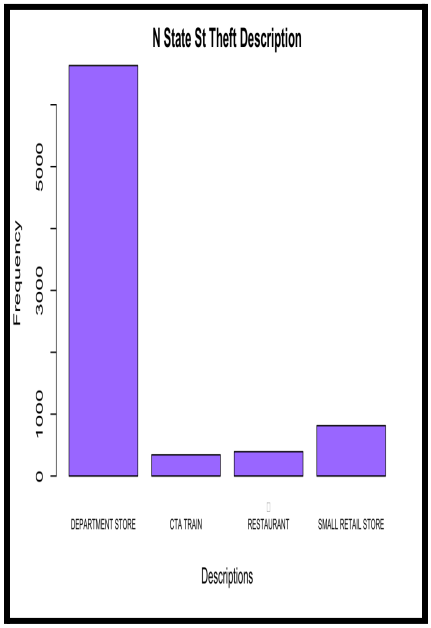


I wanted to look specifically at Theft for this project, so I then filtered the data set to only Theft. I then grouped it by Block, and used "tally" to see the frequency of each given street and filtered it to only the streets with 6000 or more in order to get the most common streets. I created a barplot to visualize the street names and their relative frequencies.

I also wanted to look at each individual street to see where exactly theft is occurring. After grouping by Location.Description, below were my findings summed up in barplots.







From these visualizations I found that Department stores and Small retail stores are the most common place of theft regardless of the street. This could be for many reasons including that it is a lot easier to steal from department stores, there are a lot of these stores, these stores also hold necessities(majority of thieves will steal necessities). I also saw a small difference specific to

each location. Michigan Avenue had numerous restaurant thefts, Cicero Ave had Parking lot/garage thefts, and N State street had CTA train thefts. After doing some research on each of these streets, I came to the conclusion that Michigan Ave had more restaurant thefts because it is an international tourist attraction filled with many fancy restaurants. This could be a possible reason why it had more thefts in restaurants compared to others. South Cicero Ave.'s thefts could have been due to the fact that it is near the Midway airport(which could have a higher frequency of parking lots/garages surrounding it), and ,finally, N State Street was due to the fact that it is the longest street in Chicago and it Runs through the heart of the Chicago loop, meaning that many cta trains and pace busses use the street to navigate chicago. Overall I found that the most common crime in Chicago is theft, and the streets that are most common are Michigan Ave, Cicero Ave, and N State Street. I also found that most theft occurs in department and retail stores, but there are also thefts in other locations specific to each location based on what they are most known for. I really enjoyed working on this project, and I think to further this project I could look at a different type of crime such as homicide or narcotics. I have found that you can do a lot with R when it comes to data sets, and I think that the next analysis that I would like to do would be something involving tennis,one of my biggest passions. I could possibly analyze data from the ATP tour and do something with serve speeds.