Chicago Crime Data Analysis Final Report

Group Name:

Chicago Crime Analysis Group

Group Members:

- Jiayi Chen (jchen246)
- Vetrie Senthilkumar (vetries2)
- Tae Kyung Lee (tkl2)

File Overview:

Crimes_-_2001_to_present (1).csv - the dataset used for the project
ChicagoCrimeDataVisualizations.Rmd - the rmd file where we wrote all the method and graphs
ChicagoCrimeDataVisualizations.html - the knitted html file for Rmd file
Chicago_Crime_Visualizations folder:
app.R - the file where we wrote the shiny application
ChicagoCrimeDataAnalysisFinalReport.pdf - final report for our project
Chicago Crime Data Analysis Shiny Application Slides.pptx - slides for demo video

Part1. Introduction

For our project, we chose to analyze crime data in Chicago to identify crime patterns and problematic regions that could potentially be used to reduce crime in one of America's most troubled cities. Chicago's reputation as a violent, lawless city has continued to grow in recent years to the point where many of us have become desensitized to acts of brutality such as assault, robbery, and homicide since we seem to constantly hear about it on the news. As a result, our group feels that this issue is not receiving the attention it deserves, especially when we consider that there are thousands of Chicagoan residents living in these dangerous, crime-riddled neighborhoods fearing for their safety and wellbeing. Furthermore, we have a more personal connection to this topic since many of our friends and fellow students on campus come from Chicago or communities near it and have been personally impacted by the high rates of crime. We hope that this project will be a step forward in making Chicago a safer place for them.

Link to dataset:
https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data

We obtained our data from the City of Chicago's publicly available datasets. Our data was stored inside a .csv file that consisted of the following variables: ID, Case Number, Date, Block, Primary Type, Description, Location Description, Arrest, Beat, District, Community Area, FBI Code, X Coordinate, Y Coordinate, Year, Latitude, Longitude, Location. Each observation in our data contains information about a crime committed. We then filtered our dataset using Year so we only had the crime statistics for 2017 since we wanted our project to focus on recent crime history. One of the goals of our project was to identify problematic regions with high crime rates. Variables like District, Community Area, and Block allowed us to narrow down the areas that required the most police attention. The Latitude and Longitude variables were also very helpful as they allowed us to pinpoint the exact location that each crime occurred. This information could then be used to create crime markers on a map of Chicago. Another variable that was essential was Primary Type which provided information on the type of crime committed (Assault, Arson, Battery, etc.). We used this variable to achieve another one of our project's goals: finding patterns in crime. Specifically, we were able to see which types of crime were prevalent and on the rise and which types of crime were occurring less frequently than before. Another important variable was Arrest, a boolean variable that indicated whether the person who committed the crime was caught. We used the Arrest variable to observe how the arrest rate fluctuated in comparison with the overall crime rate. We felt that this would allow us to see patterns in whether the police department was effective in apprehending the perpetrators of crimes.

To achieve the project goals we explained above, we used many of the statistical programming methods we learned in class to analyze, extract, and visualize our data. To observe trends in different types of crime, we used subsetting methods to filter our dataset by Primary Type. We used similar methods to identify problematic regions by subsetting based on location-related variables such as District, Community Area, and Beat. Since our dataset also contained variables that stored text based information, we used a combination of stringr methods and regex to extract key information from these variables. We used ggplot to create line graphs that show fluctuations in crime rate over time and bar graphs that display the most dangerous areas and the number of crimes committed in each area. Then we used plotly to add interactive components to these graphs. These are just a few examples of the different statistical programming methods we used. In short, the success of our project relied upon the statistical programming methods we learned in class.

Part2. Related Work

Link to project :
https://rstudio-pubs-static.s3.amazonaws.com/294927_b602318d06b74e4cb2e6be336522e94e.html

I've included a link to a project we used as inspiration above. This project is a good example of the types of analysis and types of visualizations we included in our project. One aspect of this project we found impressive was the time series analysis and the user interface that allowed users to interact with the time series graphs. These graphs made it easy to observe how crime rates varied over time. We also noticed how this project separated the different types of crimes into categories and provided visualizations on their frequencies. Thus, it was simple to understand which types of crime were more prevalent and problematic and which types were less of an issue. Another feature that caught our attention was the map of the Chicago region that showed what type of crime occurred where and the heatmap that accompanied it. We feel that including geographical overview makes the project feel more applicable and realistic and gives a different perspective on crime's impact on chicago.

Although we used some ideas from existing projects, such as this one, we also wanted our project to be original with its own unique components. Some features we added were an interactive map made with leaflet that would allow users to select different crime markers on the map. When clicked, the markers display information about the crime committed. Furthermore, our project identified problematic locations on a smaller scope. The project above shows which districts have the highest crime rates but our project extends this idea further and narrows down problematic regions to community areas, police beats, and even streets. In addition, our Shiny app contains interactive elements that allow users to select a type of crime which will then cause the app to display a map with markers for only that particular type of crime as well as a line graph that shows fluctuations in crime rate for that specific crime type. Similarly, users can also select a district and the app will update to display a bar graph of the crime statistics of the chosen district. We think that these interactive features combined with our focus on identifying problematic regions on a smaller scope make our project unique in its own way.

Part 3. Methods:

3.1 Data visualization

1. First, we used read.csv() to read our dataset into R and store it into a data frame
2. We converted the Date variable of our data frame from the character class into the Date class using as.Date.character(). This allowed us to more easily create time series graphs using the Date variable as our x -axis

3. We performed quantitative EDA on our dataset using methods such as dim(), head(), summary(), and str() before we began our analysis
4. We created a bar graph for the number of crimes committed per district using ggplot() and geom_bar().
5. We graphed the 20 most common settings (Street, Residence, Department Store, etc) that crimes occur in. To obtain the 20 most common locations, we used the dplyr method count() on our data to find the number of times each type of setting appeared in the data and then used arrange() and head() to obtain the 20 settings that appeared most frequently. Then we created a bar graph that displays the number of crimes committed for each of the 20 settings using ggplot() and geom_bar().
6. We graphed the number of crimes committed per each type of crime using ggplot() and geom_bar().
7. We graphed fluctuations in overall crime rate and arrest rate over time using ggplot() and geom_line(). To find the overall crime rate, we used count() on our data based on the Date variable to obtain the number of total crimes committed on each date. To find the arrest rate, we used the same technique but first had to use subset() to only include observations for which the Arrest variable was true.
8. We graphed the changes in crime rate for the 5 most frequent types of crime. To identify the most frequent types of crime, we used the count() method to find the number of observations per each type of crime and then used arrange() and tail() to obtain the 5 most frequent types of crime. Then, for each of the five most frequent types of crime, we used the following technique to produce a visualization. First, we used subset() to filter the data so it only includes the observations for that particular type of crime. Next, we used the count() method on our data based on the Date variable to find the number of incidents of that particular crime type per each date. Finally, we used ggplot() and geom_line() to graph the changes in crime rate for that specific crime type.
9. We graphed the 10 most dangerous community areas using ggplot() and geom_bar. To obtain the most dangerous community areas we used count() based on the Community.Area variable to find the number of time each community area appears in our data. Then we used arrange() followed by head() to isolate the 10 community areas that appeared most frequently and therefore have the highest crime rates.
10. We graphed the 10 most dangerous beats using ggplot() and geom_bar. To obtain the most dangerous beats we used count() on our data based on the Beat variable to find the number of time each beat appears in our data. Then we used arrange() followed by head() to isolate the 10 beats that appeared most frequently and therefore have the highest crime rates.
11. We graphed the 15 most dangerous streets using ggplot() and geom_bar(). Our data didn’t initially contain a street variable so we had to extract street names from the Block variable using regex and str_extract() and then add this information to our dataframe as a

new column. Then, to obtain the most dangerous streets we used count() on our data based on the new street variable we created to find the number of times each street appears in our data. Then we used arrange() followed by head() to isolate the 10 streets that appeared most frequently and therefore have the highest crime rates.

12. We used sum(month(as.POSIXlt(crime_dates$Date, format = "%y-%m-%d")) == i) to get sum for the number of crime happened each month and sum(month(as.POSIXlt(arrest_dates$Date, format = "%y-%m-%d")) == i) to get the sum for number of crime got arrested each month.
    And we use fit = lm(arrest_total ~ crime_total) to make general linear regression for arrest_total and crime_total.
    At last we use summary/anova/ and plot function to generate more info and plots with this linear regression model.

3.2 Shiny Application

3.2.1 Reactive and subset function

We use reactive() function to correspond to users' selection so as to filter and subset certain types of crime by the column "Primary.Type" in our dataset. When the primary.type changes, both the crime map and the overall crime rate line graph and scatter plot changes. And we use reactive to filter the certain district for crimes too.

3.2.2 Crime map

We made a crime heat map using renderLeaflet() function. leaflet() and addTiles() function generates an empty default map for use to add markers on. Users get to choose certain types of crime using a drop down button, and with their choices we generate a crime heat map, which is a great visualization to see the frequency and location of certain types of crimes happen in the city of Chicago. We use addCircleMarkers() function to navigate each marker's location by latitude and longitude info in the dataset. And for each colored markers, we use paste() function to add details such as crime type, date and arrest status so user can click on each marker to see the info of one specific crime.

3.2.3 Overall Crime Rate Line Graph and Scatter Plot

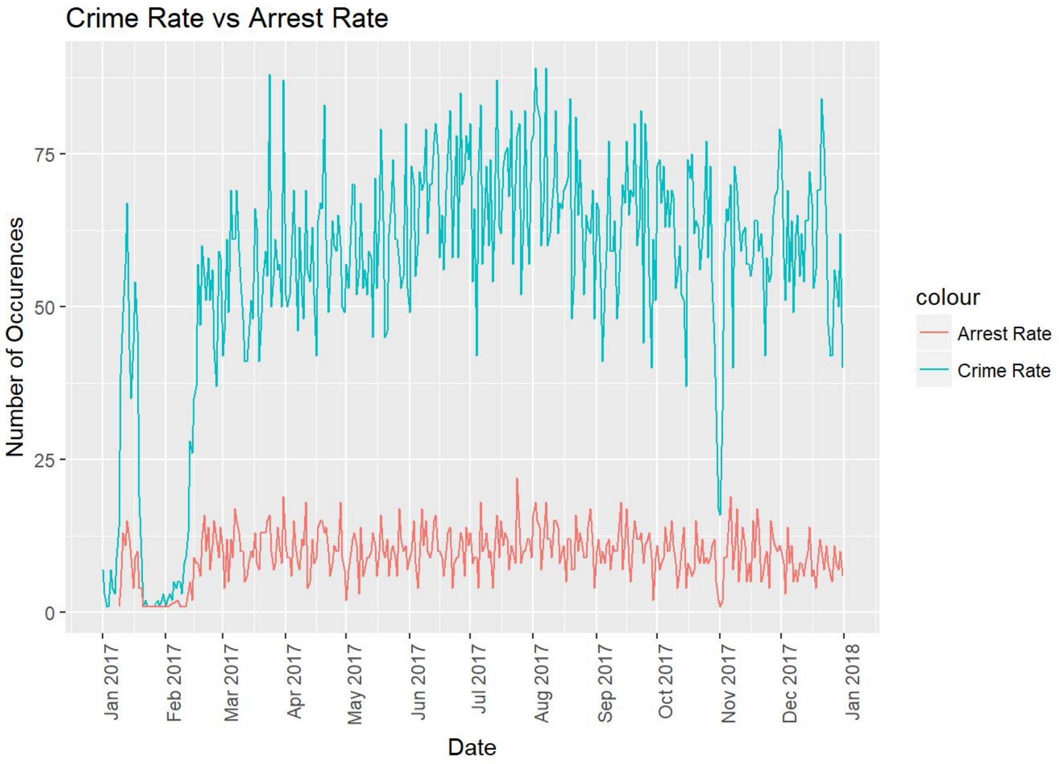
We made a line/scatter plot based on overall crime rate. renderplotly() function output and render functions for using plotly with Shiny and we generate lines and scatters using geom_line() and geom_point() with x axis being the "Date" and all the dates are separated by month using date_breaks. And y axis being "Number of Crime Incidents". We think the line graph is a great visual demonstration to see the overall picture of all the crimes happened based on each months. And users can click on each point in the scatter plot to see how many crimes happen at certain date.

3.2.4 Bar plot for Crime rates in different district

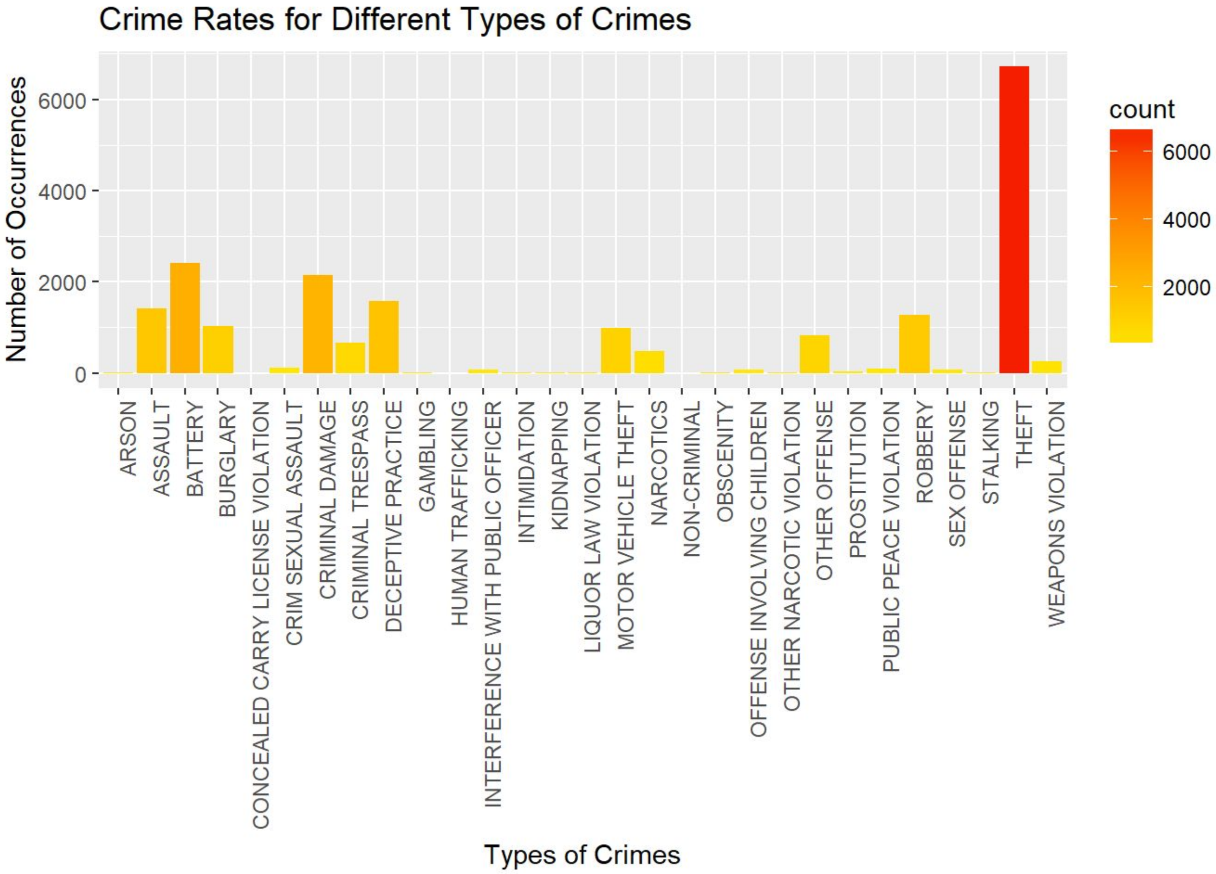
We made a bar plot for crime rates for different type of crime according to each district. We implemented a slider using sliderInput() for users to slide from district 1 to 25. And the bar plot shows each individual type of crime in different colours based on number of occurrences while lower ones are shown as lighter yellow and higher ones are shown as orange. We used ggplot and geom_bar function and the x being the primary type and x axis being the number of occurrences. We think it is a great demonstration for users to see what specific type of crime tend to happen more often in certain district using the bar plot.

Part 4. Results & Discussion:

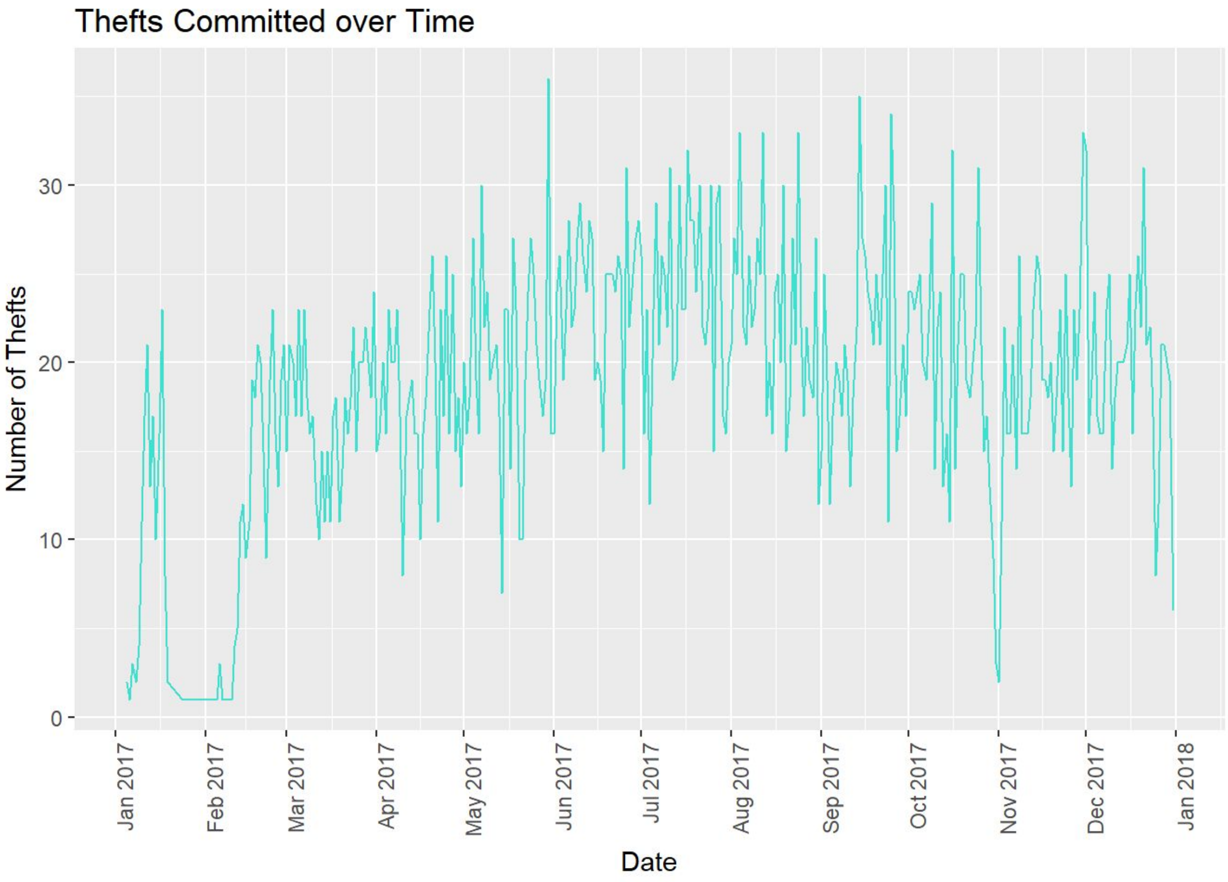
4.1 Data visualization



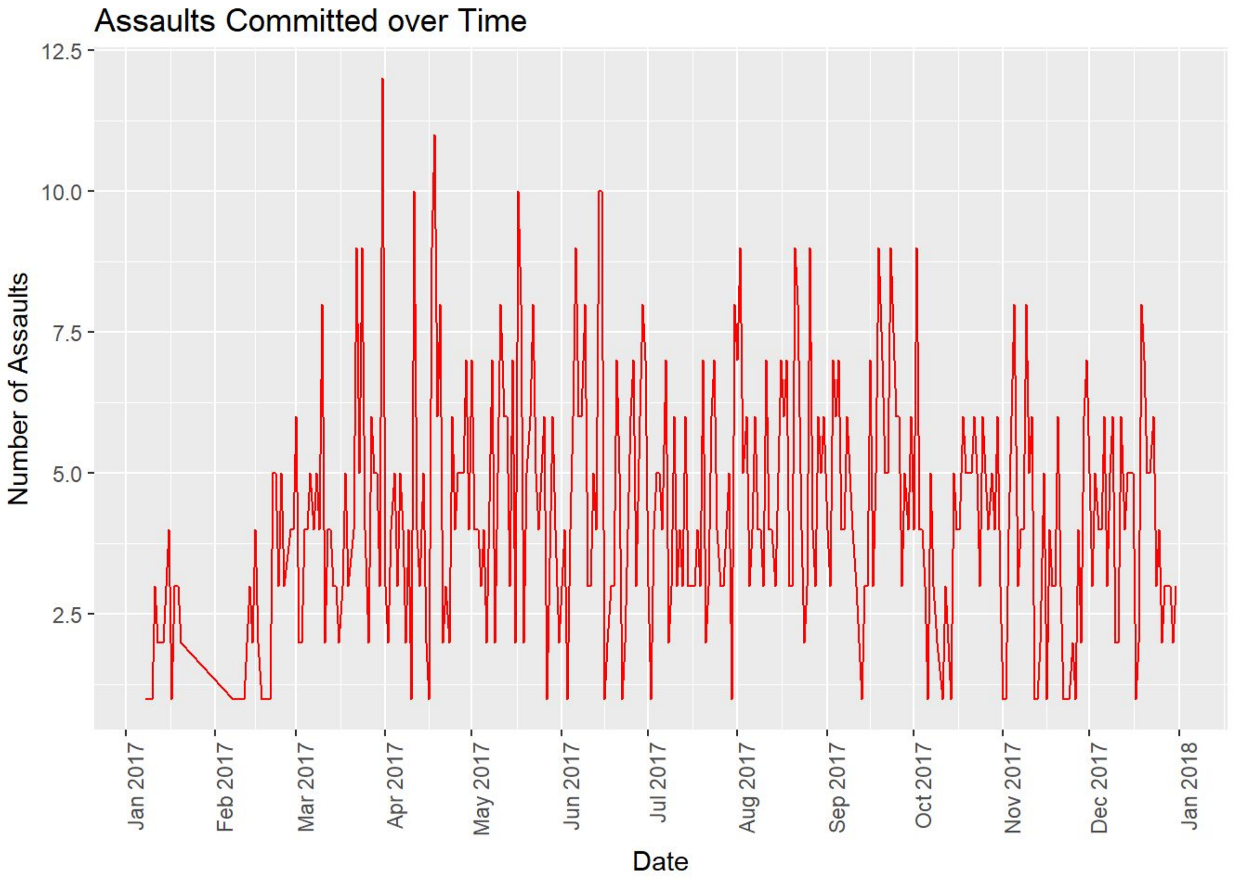
This line graphs displays the fluctuations in the overall crime rate against the fluctuations in the arrest rate for 2017. It seems that crime generally increases from spring to summer and then starts to decrease during early fall. Furthermore, it seems that there is a time period in mid January to mid February where crime experiences a dramatic decrease. Thus, the crime rate could be influenced by the current season. Ideally, upward trends in the overall crime rate would be followed by upward trends in the overall arrest rate since this would indicate that as more crimes are committed, arrests also go up as the police department catches the perpetrators of the crimes. Looking at this graph, it seems that spikes in the overall crime rate are matched by upward increases in the arrest rate. However, the increases in the arrest rate are very small compared to the increases in the crime rate which seems to suggest that many criminals go unpunished. In fact, one of the most noticeable aspects of this graph is how small the arrest rate is in comparison to the crime rate. This indicates that many criminals face no consequences for the crimes they commit, allowing them to keep breaking the law and threatening other people's safety. This is a very alarming crime pattern that once again reinforces Chicago's reputation as a lawless city and emphasizes why we should be concerned about Chicago's crime issues.



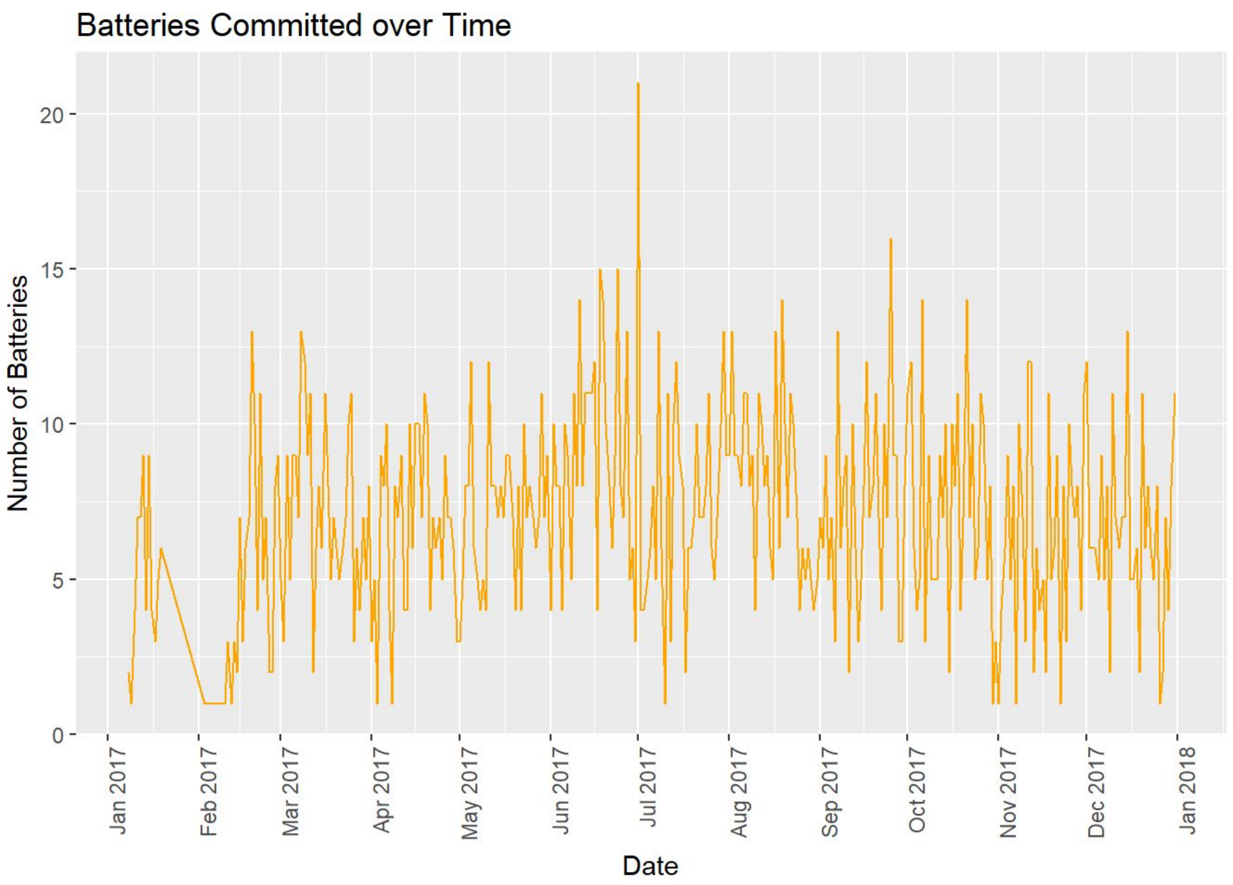
This bar graph displays the number of crime occurrences for all the different types of crimes. Theft is by the far the most frequent type of crime. Other frequent types of crime include assault, battery, criminal damage, and deceptive practice. In general, there seems to be a large number of theft-related crimes such as burglary, robbery, and motor vehicle theft. Thankfully more serious crimes such as human trafficking, kidnapping, and criminal sex assault are far less common. This graph suggests that to improve Chicago neighborhoods, police need to focus on preventing theft-related crimes since such crimes account for the vast majority of Chicago's overall crime incidents.



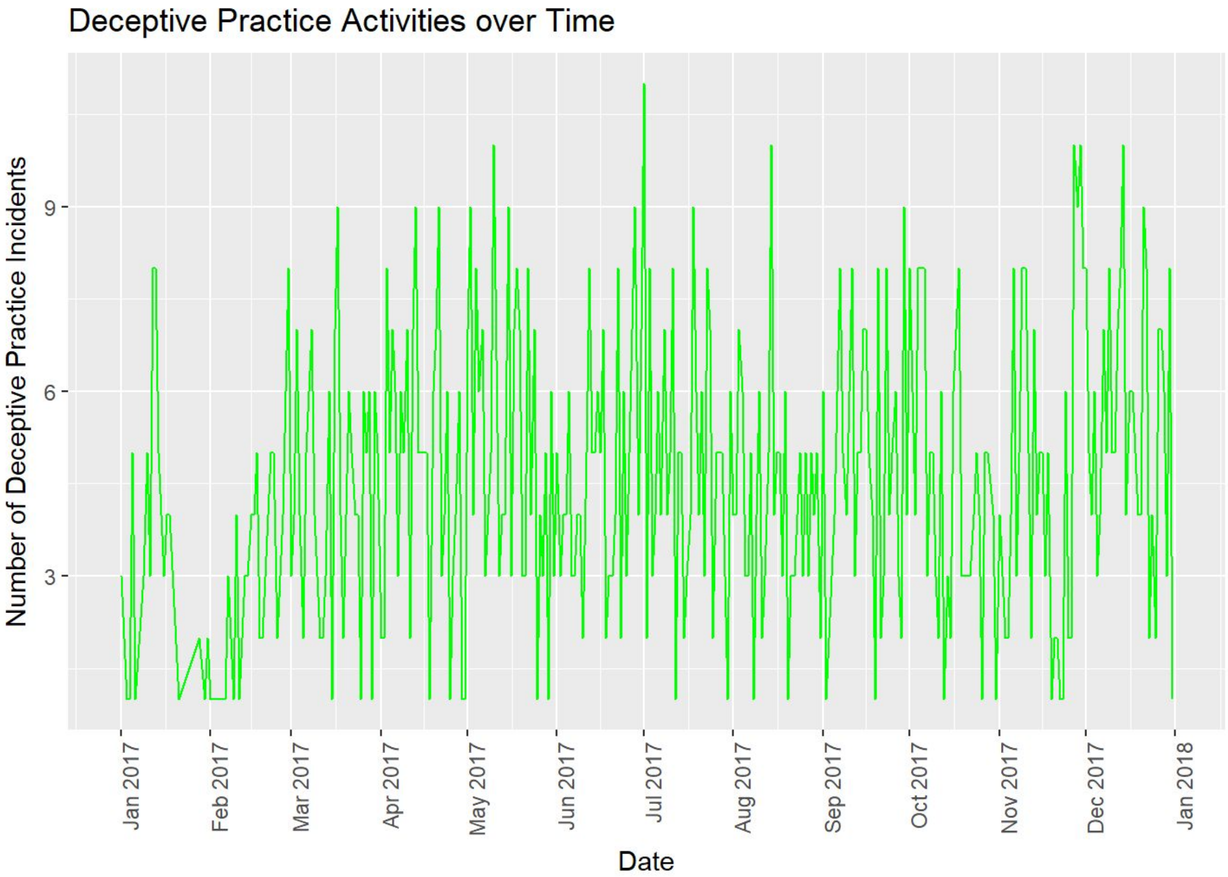
This line graph displays the fluctuation in thefts committed over the course of 2017. Thefts are the most common type of crime by a large margin and based on the results of this graph, that trend shows no signs of stopping soon. The number of thefts committed has generally risen over the course of 2017. It seems that thefts are most frequent during summer and early fall since the number of thefts hit a peak during the months of May through October. As the holiday season approaches, the number of thefts gradually started decreasing. Police could use these trends in theft rate to combat theft more effectively. For example, during summer months, increased police presence in high theft areas could deter criminals and reduce theft



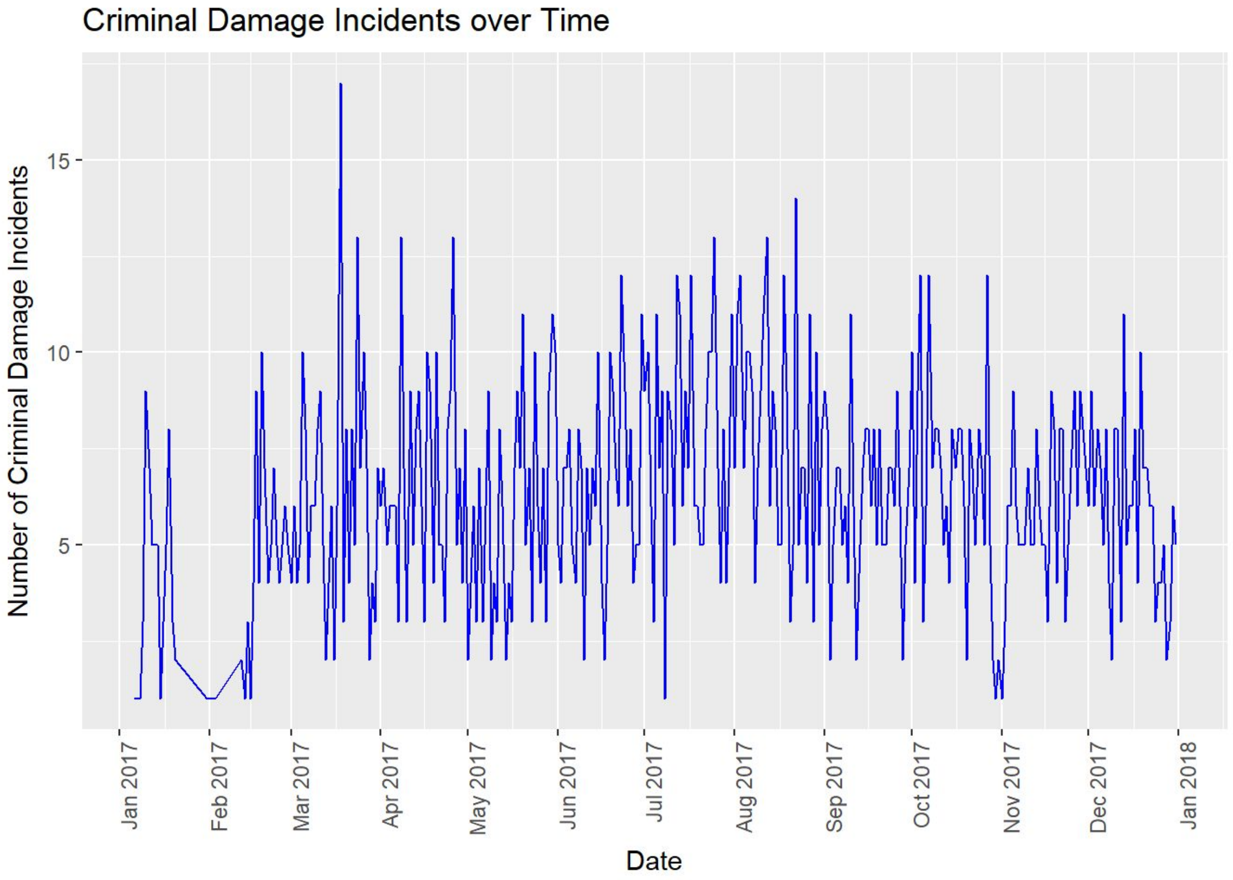
This line graph displays fluctuations in assaults committed over the course of 2017. Assault are another common type of crime in Chicago. Based on this graph, we can see that assaults rose dramatically starting around March but have gradually decreased towards the end of 2017. Furthermore, there were very few assaults during late January and early February. However, the assault rate at the end of the 2017 is still higher than what it was during the beginning of 2017, indicating that assaults are still a concerning issue in Chicago that deserves police attention.



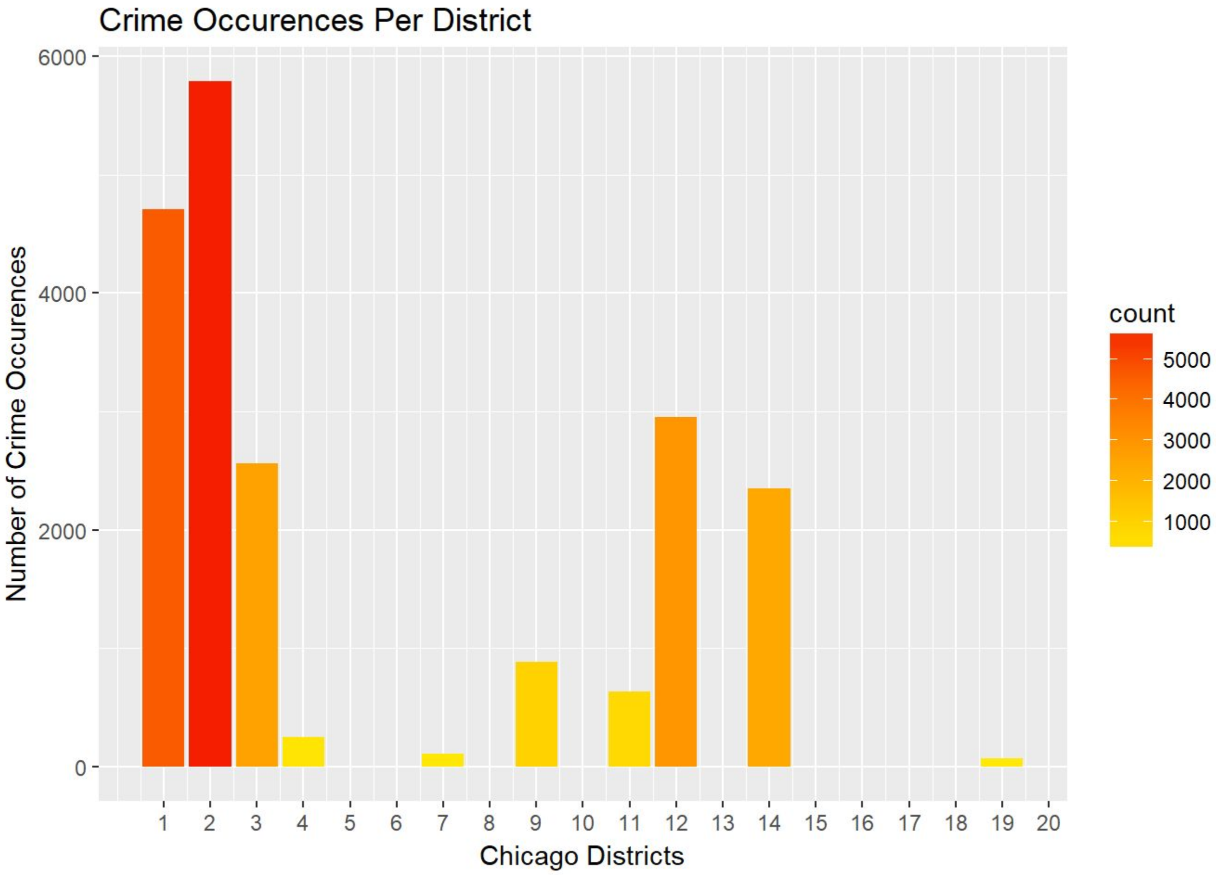
This line graph displays fluctuations in batteries committed over the course of 2017. Like assaults and theft, batteries are one of the most frequently occurring crimes in Chicago. This graph indicates that number of batteries committed has remained relatively consistent throughout 2017 apart from one large spike in late June and a period of time in February where there no such incidents. Therefore, while batteries are still a prevalent issue, at least there doesn’t seem to be an upward trend for this particular type of crime.



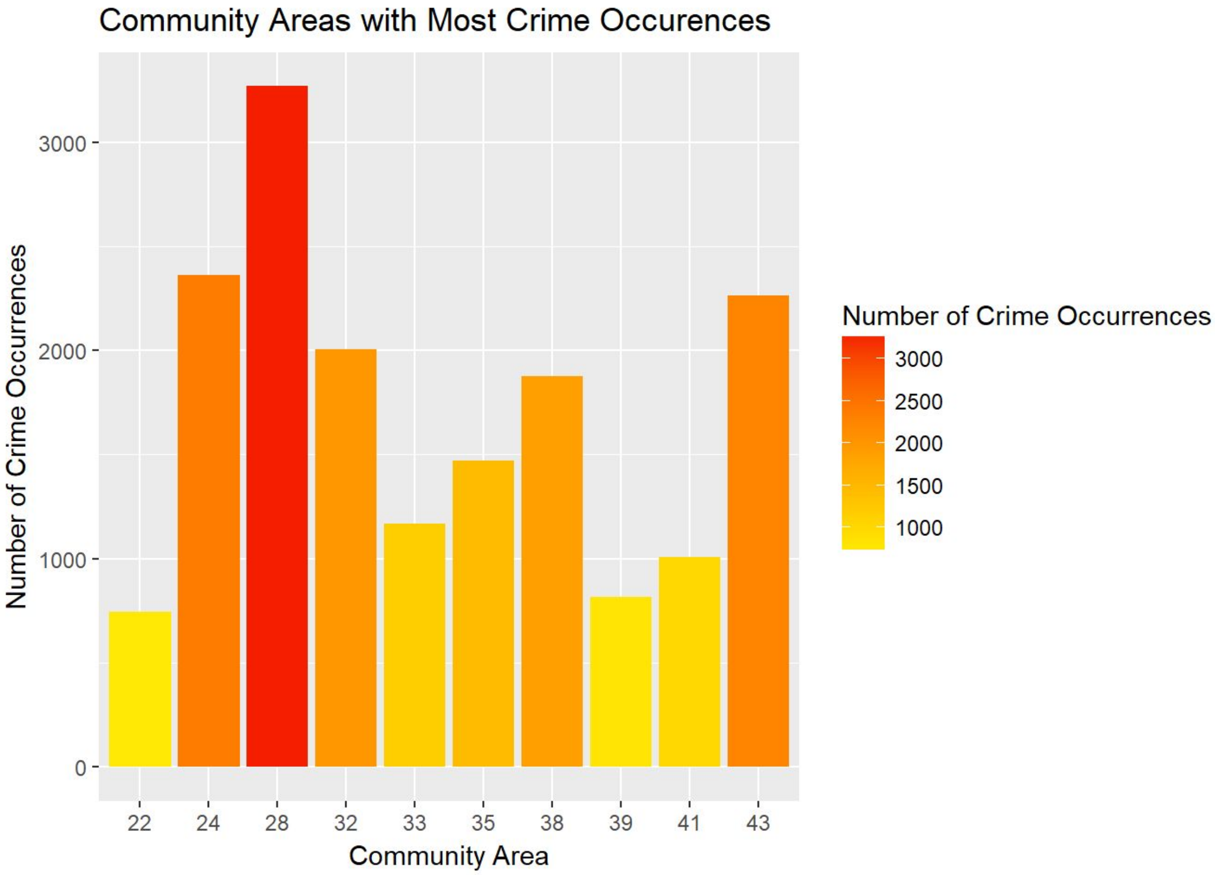
This line graphs shows fluctuations in deceptive practice incidents over the course of 2017. Based on this graph, there is no real upward or downward trend in the number of deceptive practice incidents. The number of deceptive practice incidents is pretty consistent throughout 2017 except for a time period in February where there were very few incidents. Similar to battery, deceptive practice incidents remain concerning but it is a positive sign that they are not increasing.



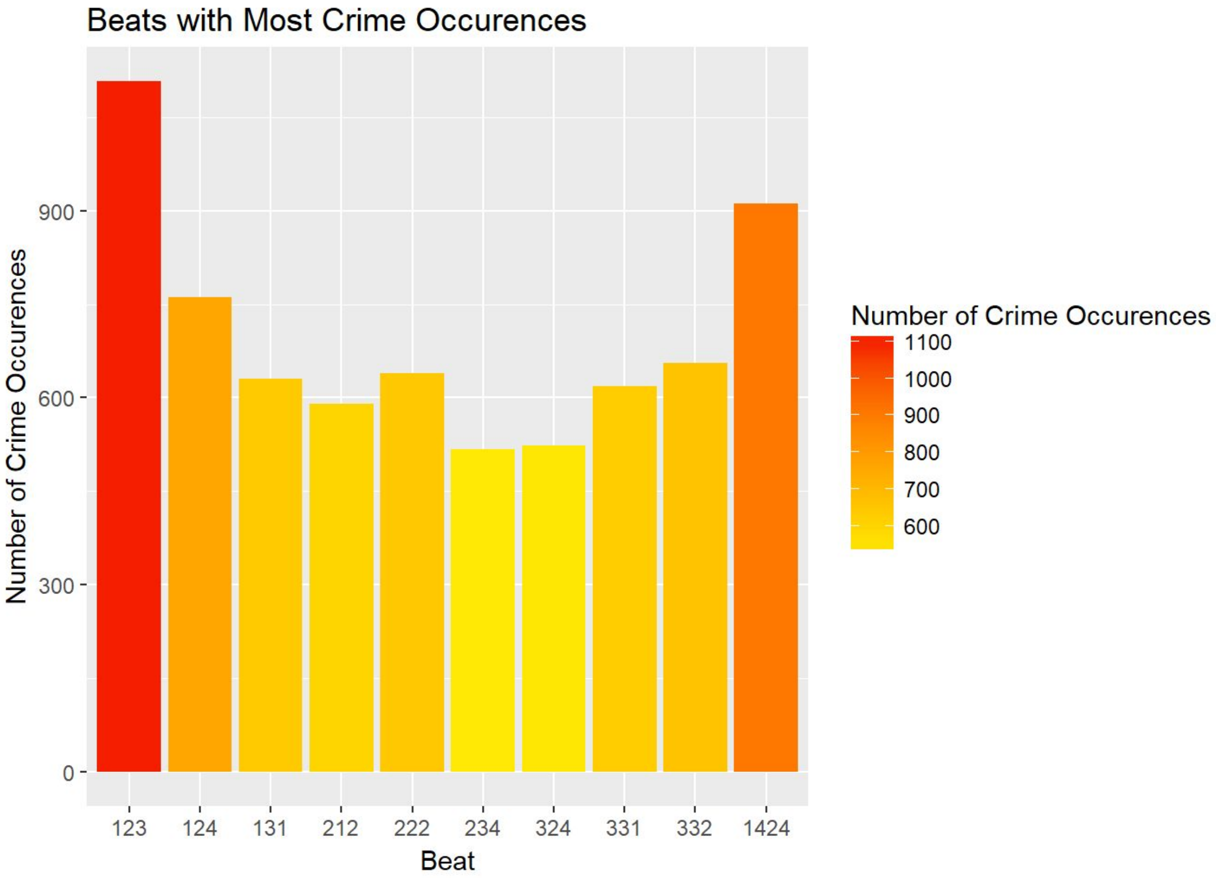
This line graph displays fluctuations in criminal damage incidents over the course of 2017. There were were very few criminal damage incidents in February but the number of criminal damage incidents spiked up suddenly during mid March and generally rose throughout the summer. However, the number of criminal damage incidents at the end of 2017 was at a similar level to the number of criminal damage incidents at the beginning of the year which suggests that criminal damage incidents are not experiencing an upward trend but are just influenced by the current season instead.



This bar graph examines the number of crime incidents per each of Chicago's districts. Surprisingly, many of Chicago's districts are safe, having little to no crime . Districts 1 through 3 in addition to districts 12 and 14 account for almost all of Chicago's crime issues. District 2 in particular has experienced approximately 5800 crime incidents, an astonishing number. Thus, these districts could really benefit from increased police presence as they are by far the most problematic regions within Chicago.

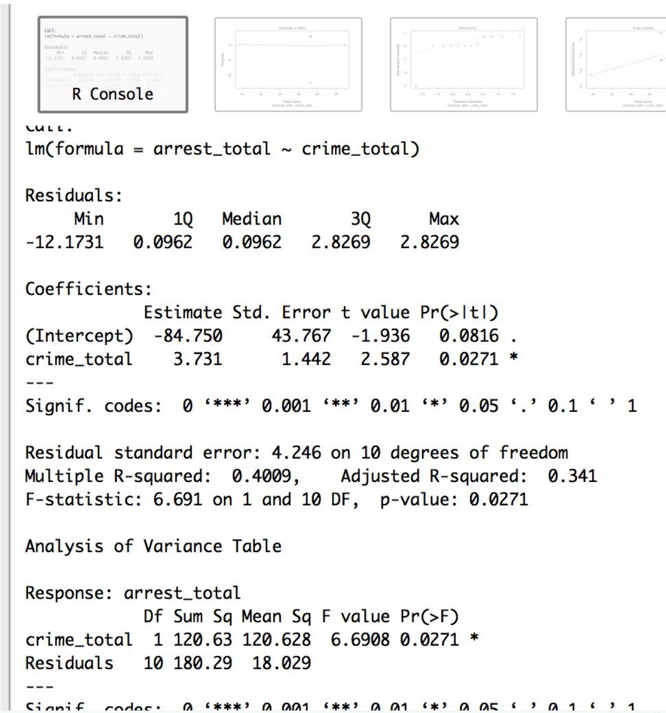


This bar graph displays the 10 most dangerous community areas and the number of crime incidents in each one. With over 3000 crime incidents, Community area 28 is the most dangerous community area in Chicago. To solve crime within Chicago, these community areas must be the first regions targeted by the police.

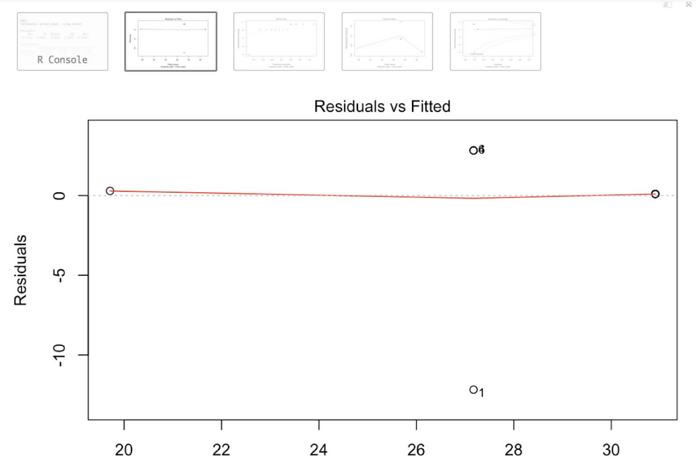


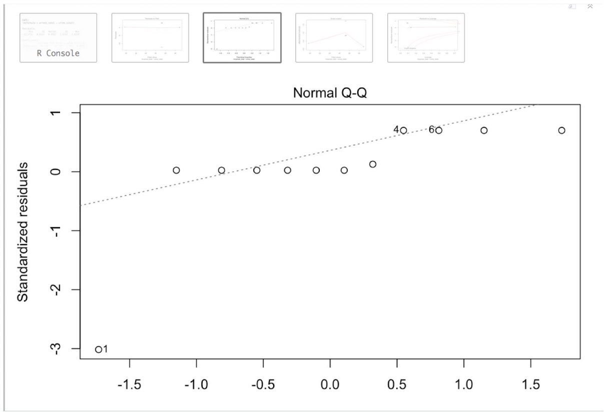
This bar graph displays the 10 most dangerous beats and the number of crime incidents that occurred in each one. Beat 123 seems to be the most dangerous beat in Chicago with approximately 1100 crime incidents. Beat 1424 is next with around 900 crime incidents. Thus, to solve Chicago's crime issues, the police department should direct their focus toward preventing crime in the beats shown above.

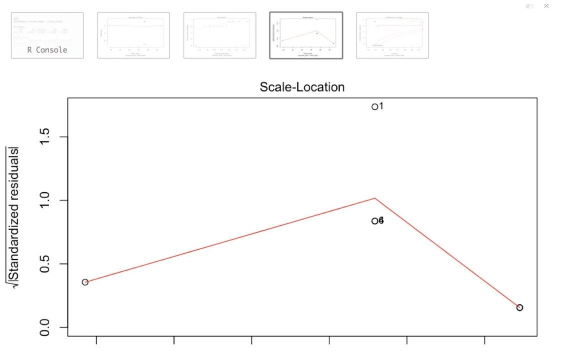
Linear Relationship Plot between Number of Crimes and Number of Arrests:
The residual vs fitted plot shows that there is no relationship in between them.
And the normal QQ plot does not follow the straight line so it is not a normal distribution.
The (residual vs leverage) vs residual plot shows that there is no outlining points in the graph.

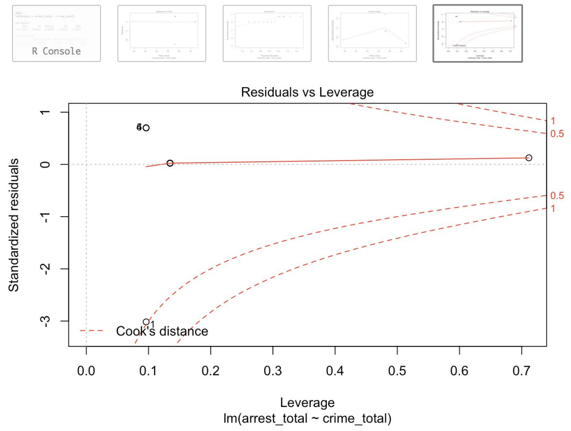


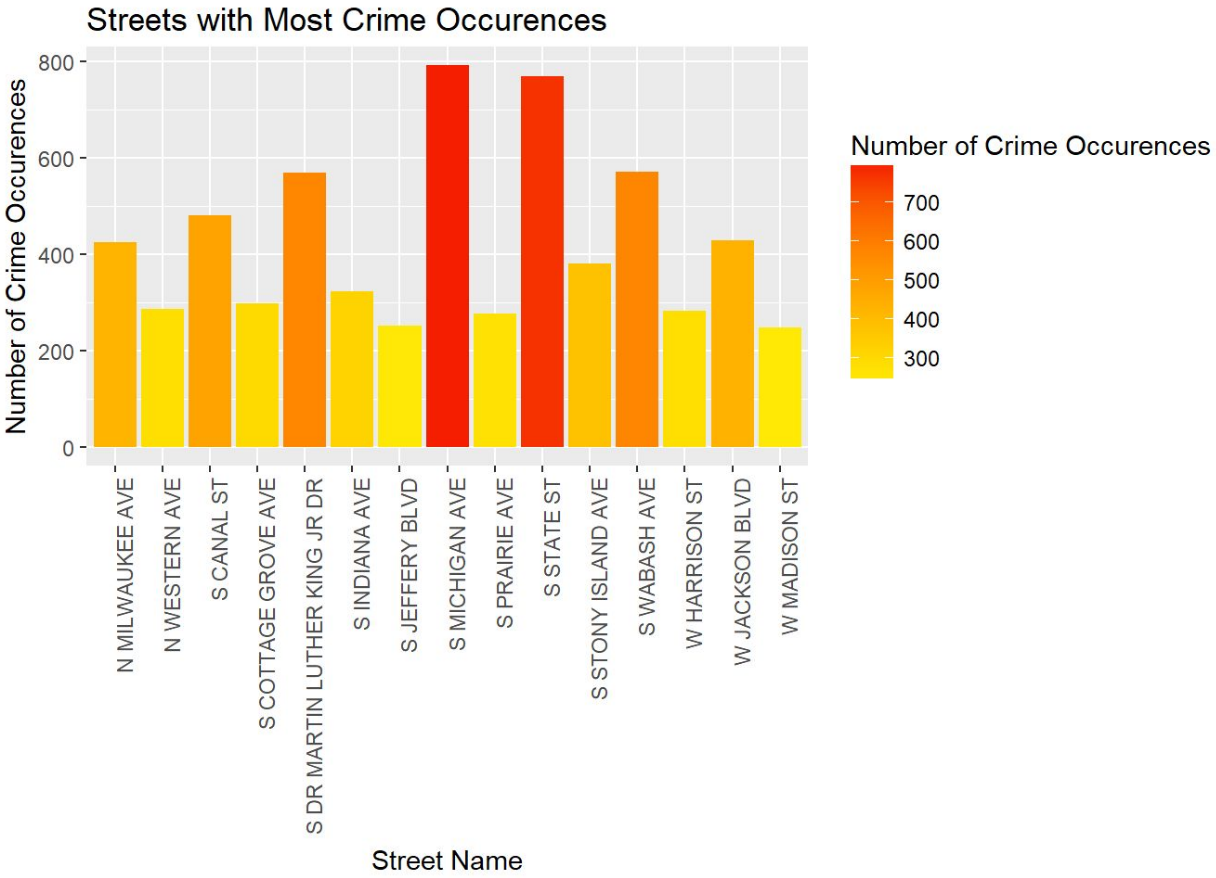
```
Call:
lm(formula = arrest_total ~ crime_total)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.1731   0.0962   0.0962   2.8269   2.8269 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -84.750     43.767  -1.936   0.0816 .
crime_total    3.731      1.442   2.587   0.0271 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.246 on 10 degrees of freedom
Multiple R-squared:  0.4009,	Adjusted R-squared:  0.341 
F-statistic: 6.691 on 1 and 10 DF,  p-value: 0.0271

Analysis of Variance Table

Response: arrest_total
            Df Sum Sq Mean Sq F value Pr(>F)  
crime_total  1 120.63 120.628  6.6908 0.0271 *
Residuals   10 180.29  18.029                 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```











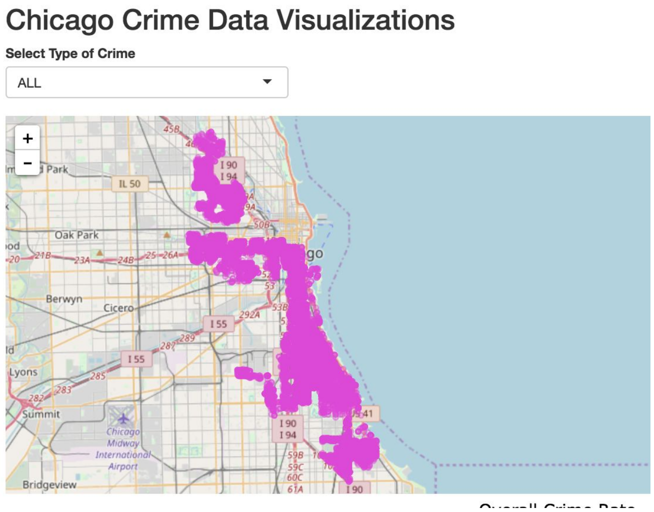
This bar graph displays the 15 most dangerous streets and the number of crime incidents that occurred on each of the streets. South Michigan Avenue and South State Street both experienced almost 800 criminal incidents and were the two most dangerous streets in Chicago during 2017. To make Chicago a safer place, police can patrol these crime-riddled streets more often and in increased numbers, which will hopefully deter people from committing criminal acts.

4.2 Shiny Application

4.2.1 Reactive and subset function

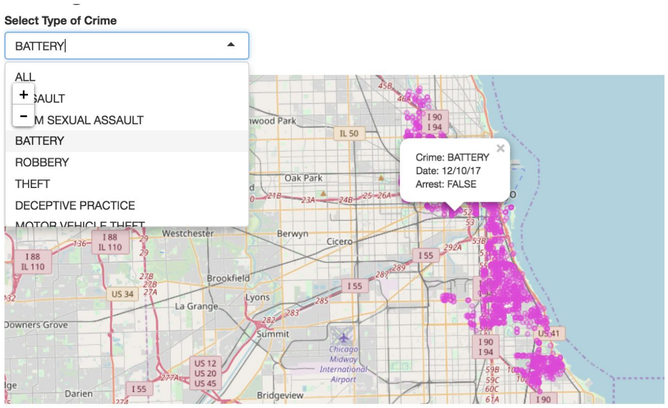
reactive() function to correspond to users' selection using dropdown menus for different types of crime or sliders for certain districts.

4.2.2 Crime map


Chicago Crime Data Visualizations

Select Type of Crime: ALL

Overall Chicago map for all types of crime and the dropdown menu

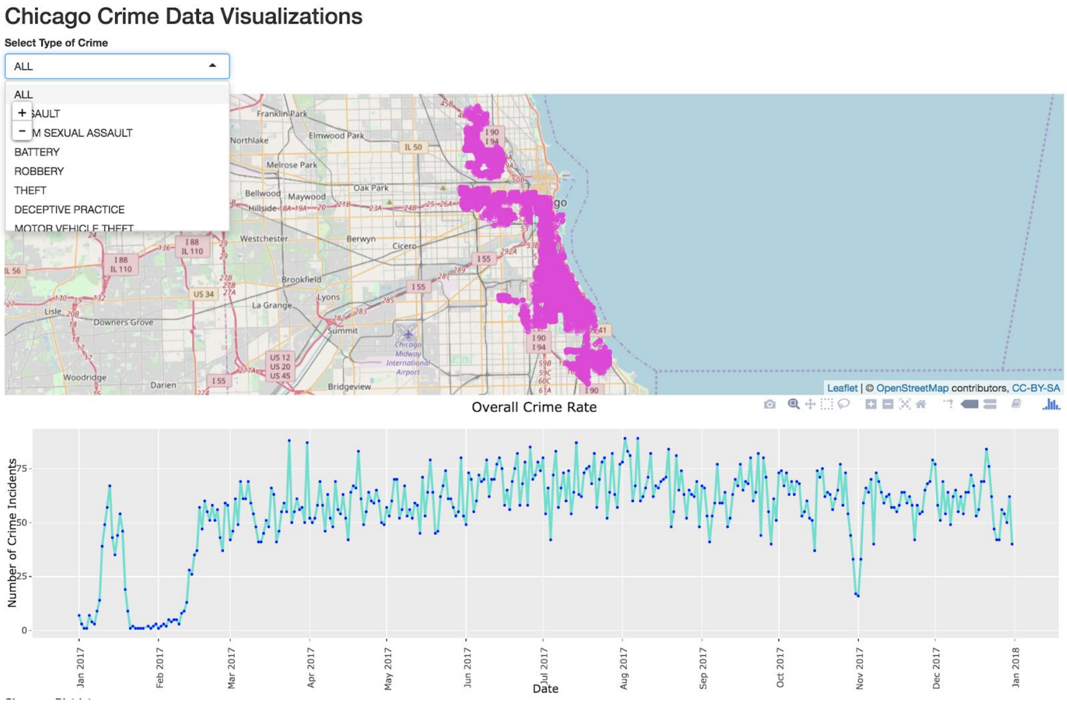

Select Type of Crime: BATTERY

Crime: BATTERY
Date: 12/10/17
Arrest: FALSE

Filtered crime map for one certain type of crime

Conclusion: We can see clearly from the crime map that south east part of chicago is where most of the crime happened. Expect thefts and robberies tend to happen around north of chicago, and we assume that it might because there are downtown areas. Battery happens a lot at the central part of Chicago. And it seems like there is no specific pattern of locations for crim sexual assault and assaults in general.

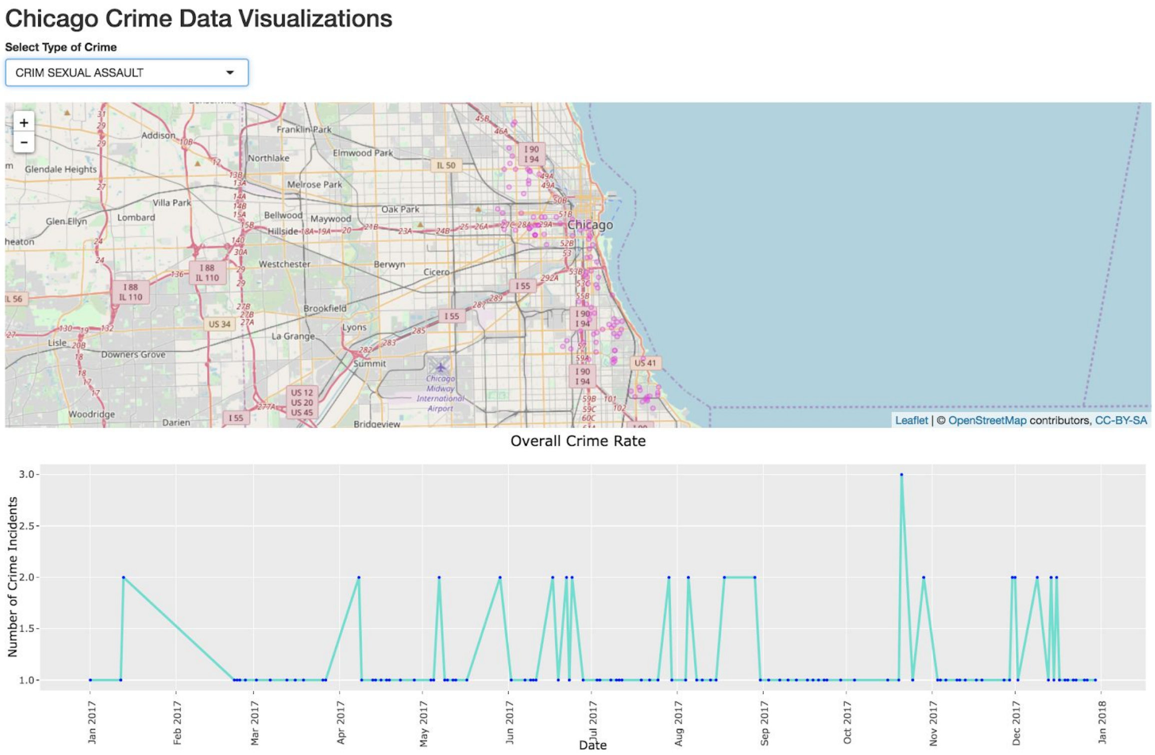
4.2.3 Overall Crime Rate Line Graph and Scatter Plot



Chicago Crime Data Visualizations

Select Type of Crime: ALL

Overall crime crate for all crimes:



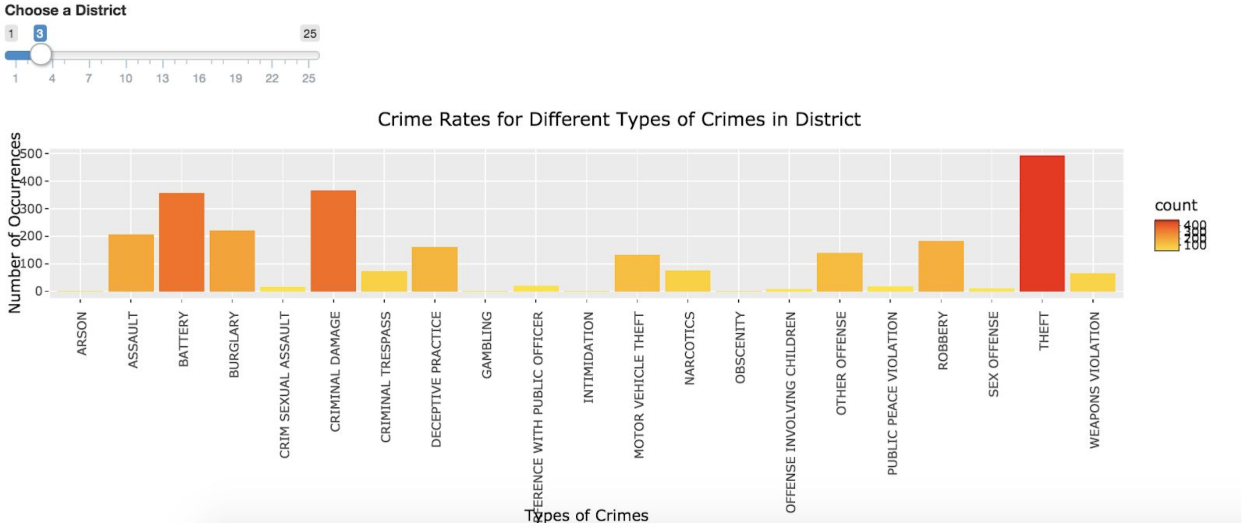
Chicago Crime Data Visualizations

Select Type of Crime: CRIM SEXUAL ASSAULT

Crime rate for for sexual assault:

Conclusion: We can see that february 2017 is the month when crime rate is incredibly low compared to other months. And we can see that crime rates are relatively high during the summer months from june to august. And it dropped dramatically since september and it remains low until the end of November. As far as specific crime is concerned, sexual assault and weapon violation are more likely to occur during summer. We had a hypothesis that crimes may be increasing with high temperatures and the research we did proved it to some extent. (research citation: http://www.chicagotribune.com/news/data/ct-crime-heat-analysis-htmlstory.html)

4.2.4 Bar plot for Crime rates in different district



Conclusion: We can see that thefts are high in almost all neighbourhoods, especially for neighborhood 14 that has 900, neighborhood 12 has 1000 theft incidents while all other neighborhoods only has 100-200 on average. And in neighbourhood 2,4, 7-11 assault, battery and criminal damage, including motorcycle theft is noticeably higher than other districts. And there are almost no crime happened in 2017 in neighbourhood 15-18, and 20-25. In general, neighborhood 2-14 are more dangerous in 2017.

Part 5: Conclusion

In conclusion, we feel that we've accomplished the main goals of our project which were identifying patterns in crime and locating the most problematic regions in Chicago. One of the key observations we made were that the most prevalent forms of crime in Chicago are theft, deceptive practice, criminal damage, assault, and battery. In general, Chicago experiences a lot of theft-related incidents, including robberies and burglaries. Furthermore, we noticed that overall crime tends to increase during summer but often experiences a sharp drop during February. This phenomenon holds true for many specific types of crime as well so we believe that crime in general is influenced by the season. We also accomplished our other goal of locating problematic regions within Chicago by finding the districts, community areas, beats and streets with the highest number of crime occurrences. Districts 1 and 2 are the most dangerous districts in Chicago. Similarly, community area 28 is the most dangerous community area. Likewise, beats 123 and 1424 are the most dangerous beats. Lastly, South Michigan Avenue and South State Street are the most dangerous streets. We hope that all of this information we've gathered through this project can be used to reduce crime in the Chicago and make the city a safer place for its residents.

Part 6: References:

5.1 list (5+) of papers or items read to write this proposal:

- Azadeh Ansari and Rosa Flores, "Chicago's 762 homicides in 2016 is highest in 19 years", 2017, https://www.cnn.com/2017/01/01/us/chicago-murders-2016/index.html
- Garrett Grolemund, "Quick list of useful R packages", 2018, https://support.rstudio.com/hc/en-us/articles/201057987-Quick-list-of-useful-R-packages
- Vivek Mangipudi, "ANALYSIS OF CRIMES IN CHICAGO 2001 - 2017", 2017, https://rstudio-pubs-static.s3.amazonaws.com/294927_b602318d06b74e4cb2e6be336522e94e.html
- Andrew V. Papachristos, "48 YEARS OF CRIME IN CHICAGO: A Descriptive Analysis of Serious Crime Trends from 1965 to 2013", 2013
- Sai Krishna Vithal Lolla, "Crime Occurrence Analysis in Chicago City" , 2013, https://www.jmp.com/about/events/summit2013/resources/Poster25_Lolla_Liu.pdf
- Udeh Tochukwu, "CRIME MINING AND INVESTIGATION USING R" , 2015, https://www.researchgate.net/publication/277020625_CRIME_MINING_AND_INVESTIGATION_USING_R

5.2 list all R packages or software referenced:

dplyr

adley Wickham, Romain Francois, Lionel Henry and Kirill Müller (2017). dplyr: A Grammar of Data

Manipulation. R package version 0.7.4. https://CRAN.R-project.org/package=dplyr

stringr

Hadley Wickham (2018). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.3.0. https://CRAN.R-project.org/package=stringr

ggplot2

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

lubridate

Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of

Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/.

Shiny

Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2017). shiny: Web

Application Framework for R. R package version 1.0.5. https://CRAN.R-project.org/package=shiny

Leaflet

Joe Cheng, Bhaskar Karambelkar and Yihui Xie (2018). leaflet: Create Interactive Web Maps with

the JavaScript 'Leaflet' Library. R package version 2.0.0.

https://CRAN.R-project.org/package=leaflet

Part 7: Appendix

6.1 Functions Overview

- read.csv() takes in a .csv file as input and returns a data frame that contains the data of the .csv file. Since our dataset is contained in a .csv file, this method is essential to our project and will be the first method called
- head() accepts a data structure as input and returns the first six elements. We will use the head() function on our data frame created above to gain some more insight into the structure of our data, especially the first couple observations, before we begin a more in-depth analysis
- tail() accepts a data structure as input and returns the last six elements. Similar to head(), we will use tail() to gain more insight about our data before we begin analyzing it
- summary() is a generic function that accepts an object as input and returns a statistical summary of the data the object contains. We will use summary() on our data frame to get

a better sense of the distributions of different variables the data contains. Like head() and tail(), summary() will help us get a better understanding of our data before the analysis phase.

- count() is a dpylr method that counts the number of times a unique element appears in a data structure
- str_extract() accepts the same two parameters that str_detect() does. If the regex pattern is found in a string, it will return the portion of the string that matches the regex pattern. We used str_extract() to extract street names from our data set so we can identify which streets are the most dangerous.
- geom_bar() when used with ggplot() takes in a column of a data frame as input and produces a bar graph based on the categorical variable. We plan on producing graphs for the text-based variables that were derived using str_detect and str_extract above using geom_bar()
- geom_line() is another important ggplot function. It accepts an x variable and a y variable and produces a line graph based on these inputs. We plan on using geom_line() to produce our time series analysis graphs, using the date as our x variable and number of crime occurrences as our y variable
- sliderInput() is a Shiny UI input function that creates a slider for our user interface. It accepts parameters that allow us to set a label, set a default value the slider will be positioned at, and max and min values the slider cannot go beyond, in addition to numerous other optional parameters. We used this feature to let users select a district on our shiny app. The app then displays crime statistics for that particular district
- selectInput() is a Shiny UI function that allows users to select an option from a drop down menu that contains a list of possible choices. It accepts parameters for its label, list of choices, and default choice, in additional to other optional parameters. We will add this feature to our user interface to allow users to choose a specific type of crime they wish to analyze. The user interface would then display statistics and graphs for that particular type of crime.
- renderPlot() is a Shiny rendering function that allows us to add graphs and plots to our Shiny app. This function takes in an expression parameter that is used to generate the graph as well as height and width parameters that are used to set the dimensions of the plot. This function is essential since it will be used to add our data visualizations, especially our time series analysis plots, to our Shiny app.
- ggplotly() converts a ggplot object to plotly object. We used this to add interactive components to our ggplot graphs
- renderPlotly() allows Shiny to display plotly graphs
- as.Date.character() converts a character string of a specified format into a Date object. We used this method to convert our Date variable from a vector of strings to a vector of Date objects which was useful for our time analysis graphs

6.2 Code Book

- ID : Unique identifier for the record
- Case Number: The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.
- Date: Date when the incident occurred
- Block: The partially redacted address where the incident occurred, placing it on the same block as the actual address
- Primary.Type: Type of crime committed
- Description: Provided further details about the nature of the crime
- Location.Description: Description of the location where the incident occurred.
- Arrest: Indicates whether an arrest was made
- Beat: Indicates the beat where the incident occurred. A beat is the smallest police geographic area
- District: Indicates the police district where the incident occurred
- Community.Area: Indicates the community area where the incident occurred
- FBI.Code: Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS)
- X.Coordinate: The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block
- Y.Coordinate: The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block
- Year: Year incident occurred
- Latitude: Latitude of location the incident occurred
- Longitude: Longitude of the location where the incident occurred
- Location: Contains pair of coordinates representing latitude and longitude of location of crime