

MA615 Midterm Project

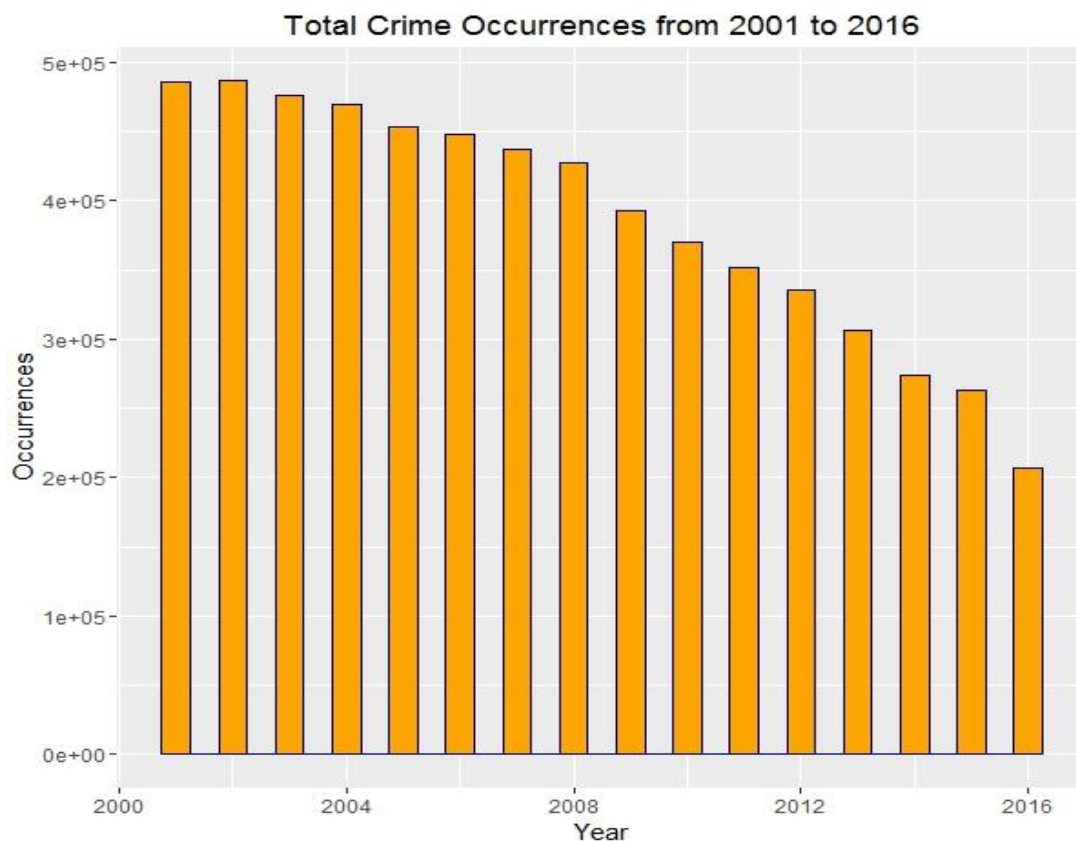
Chicago 2016 Crime Report Analysis

Ningxi Wei

October 23, 2016

In this project we select a data set that consists of crime reports in the city of Chicago from 2001 to present. Data is extracted from the Chicago Police Department's Citizen Law Enforcement Analysis and Reporting system and originally, contains 3,275,608 report records and 22 measures. Our task is to analyze the crime reports documented in Chicago within year 2016 and use R to generate some statistical plots.

To start, I used R to transform the original data set, a .csv file, into one data frame. First I did a histogram plotted the total occurrences of crimes from 2001 to 2016:



From the plot above, we could observe a significant decreasing trend of crime occurrences from year 2001 to 2016 in the city of Chicago. For this project I mainly focused on crimes happened within 2016 and among the 22 measures, I am only interested in the crime types, crime locations and the community area code where the crimes took place. To reduce the data size, I used command “subset()” to select observations happened in 2016, then I dropped the rest 18 measure columns to reduce the size of the data frame to a computational friendly size. Then I used the “filter()” command from R package “dplyr” based on community area code, to generate 77 smaller data frames represented crime reports recorded in 77 different community

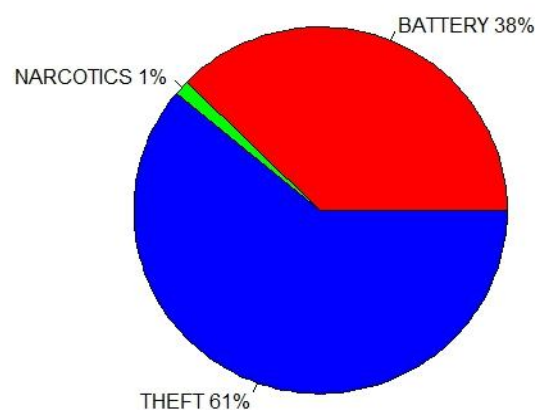
areas. Below is a code example of filtering out a sub-data frame from the original data:

```
#filter by community area
#try community area 1 as an example
df <- data.frame(data=read.csv("Crimes2016.csv",header = TRUE, stringsAsFactors = TRUE))
dfCA1 <- filter(df,df$Community.Area==1)
```

From here I had organized the original large data set and transformed it into smaller unit data frames for the next step: Analysis.

My first interest is to study the crime that has the highest occurrences among 77 community areas. In 77 data frames, I ranked the observations in descending order based on the occurrence of each crime. Then I used command “rbind()” to make a new data frame consisted the first row of 77 data frames represented the top occurrence crime types in each community area. My next step is to generate a pie chart to show graphically, the top occurrence crime types in 77 community areas.

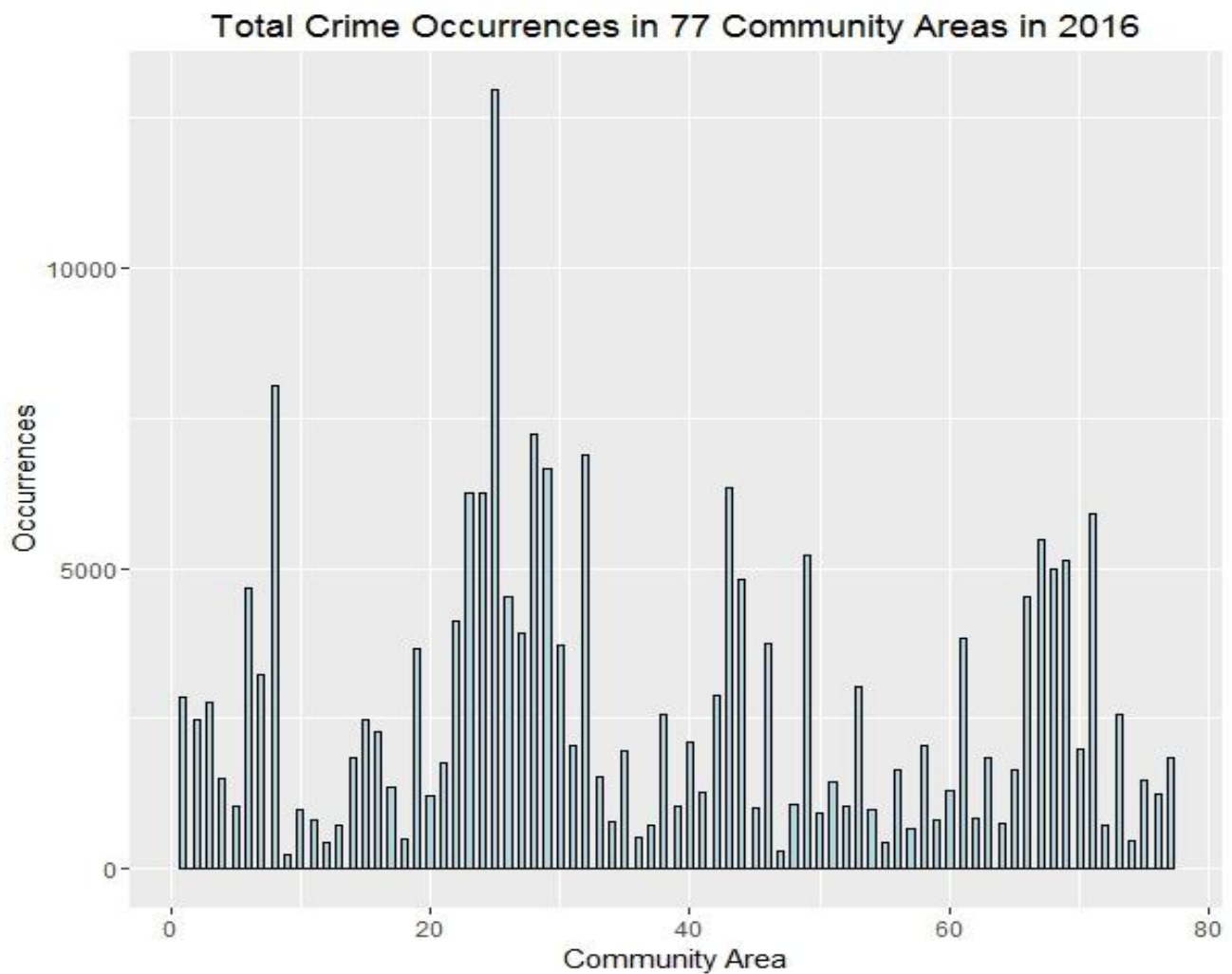
Pie Chart of Top Crime Type in 77 Community Areas



From the plot I observed that theft was the major crime type reported in the city of Chicago and followed by battery. Although narcotics had a high occurrence in some community areas but in a larger scale it only had a small percentage compared to theft and battery.

My next research interest is to look at the total crime occurrences in each community area and find out the safest and the most dangerous neighborhood in Chicago. Again, I used a histogram to plot crime occurrences in all 77 community areas (plot attached below). Clearly from the histogram, community area 25 has the highest crime occurrences in 2016. The top crime type occurred in community area 25 is battery (use the result from previous analysis) and I would like to analyze the crime locations where batteries in community area 25 took place (table attached below). From the table we can observe that about 27.650% battery took place in apartments in community area 25. One good suggestion to the local is to make sure you secure your apartment properties and lock doors and windows properly. Street battery also had a relative high occurrence rate in community 25. To solve this safety issue, Chicago Police Department should arrange more police force on the street to improve

neighborhood safety.



	Location	Occurance Percentage
1	APARTMENT	0.27649616
2	SIDEWALK	0.19993313
3	STREET	0.15546640
4	RESIDENCE	0.14309595
5	RESIDENTIAL YARD (FRONT/BACK)	0.03009027

Supplement Materials:

- <https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>
- <https://catalog.data.gov/dataset/crimes-2001-to-present-398a4>
- https://learn.bu.edu/bbcswebdav/pid-4532976-dt-content-rid-15561649_1/courses/16fallgrsma615_a1/Tidy%20Data%20-%20v59i10.pdf

Co-worker: Lu Li