Gun Violence Uproar in the City of Chicago

Issac Li zl368

Final Project for BIS 623

Introduction

Sixty-one victims have been shot in Chicago in the first six days of 2017, and 4368 were shot in 2016, according to Chicago Tribune [1]. Now, Chicago has become one of the country's most violent major cities, while New York is now one of the safest. But if not for one factor — vastly different levels of gun violence — the murder rate of the two metropolises would be nearly the same[2].

In 2015, the most recent year for which data is available for both cities, the fatal shooting rate in Chicago was five times as high as it was in New York: 15.6 per 100,000 residents compared to 2.8 per 100,000. The difference as cities is one of the starkest findings of a new report due to be released later this month by the University of Chicago Crime Lab. According to the report, there were 762 homicides in 2016, which was about 58 percent more than in the previous year. But homicides were not the only category of crime to rise: other gun offenses, including nonfatal shootings and robberies, soared.

Some factors probably contribute to the increase in gun violence. One has to deal with the gun laws of neighbor states such as Wisconsin and Illinois. The state of Illinois and City of Chicago, in particular, have strict gun laws. However, it is feasible and common for people to buy guns from Indiana. In fact, Data from the Bureau of Alcohol, Tobacco, Firearms and Explosives from between 2010 and 2014 found that a remarkable number of crime guns in Illinois came from Indiana[3]. Another reason may be the decreasing presence of law enforcement on (some of) the streets in Chicago. As we can see from Figure 1A and B, the sharp decline in the no. of street stops coincides with the rapid climb in the no. of street stops.

President-elect Donald Trump has been targeting the gun violence in Chicago for his opinion on guns since the beginning of his campaign. On Jan. 2nd,2017, he tweeted that the city's murder rate was "record setting" and suggested that Mayor Rahm Emanuel should seek federal help if he "can't do it" himself. The underlying dynamics driving the change maybe too complicated to reverse and to quantitatively understand. For ordinary people, the best we can do is to stay away from troubles whenever we can. People who have stayed in a city should know where the troubles are. But what about new-comers such as newly arrived international students from overseas who have little knowledge about Chicago? As a student it might be beneficial to know how dangerous it is to be studying in Chicago and particularly what kind of common senses we should employ if we were to live in a crime-filled metropolis. **Gun crimes, being the extreme of violent crimes and the culprit of soaring homicides in Chicago, worth some scrutinizing. Thus the focus of this project is to quantitatively find the places and times where gun crimes are prevalent.**

Methods and Material

Chicago Crime Data since 2001

The dataset (1.37 Giga-bytes in .csv format) that is used in this study contains

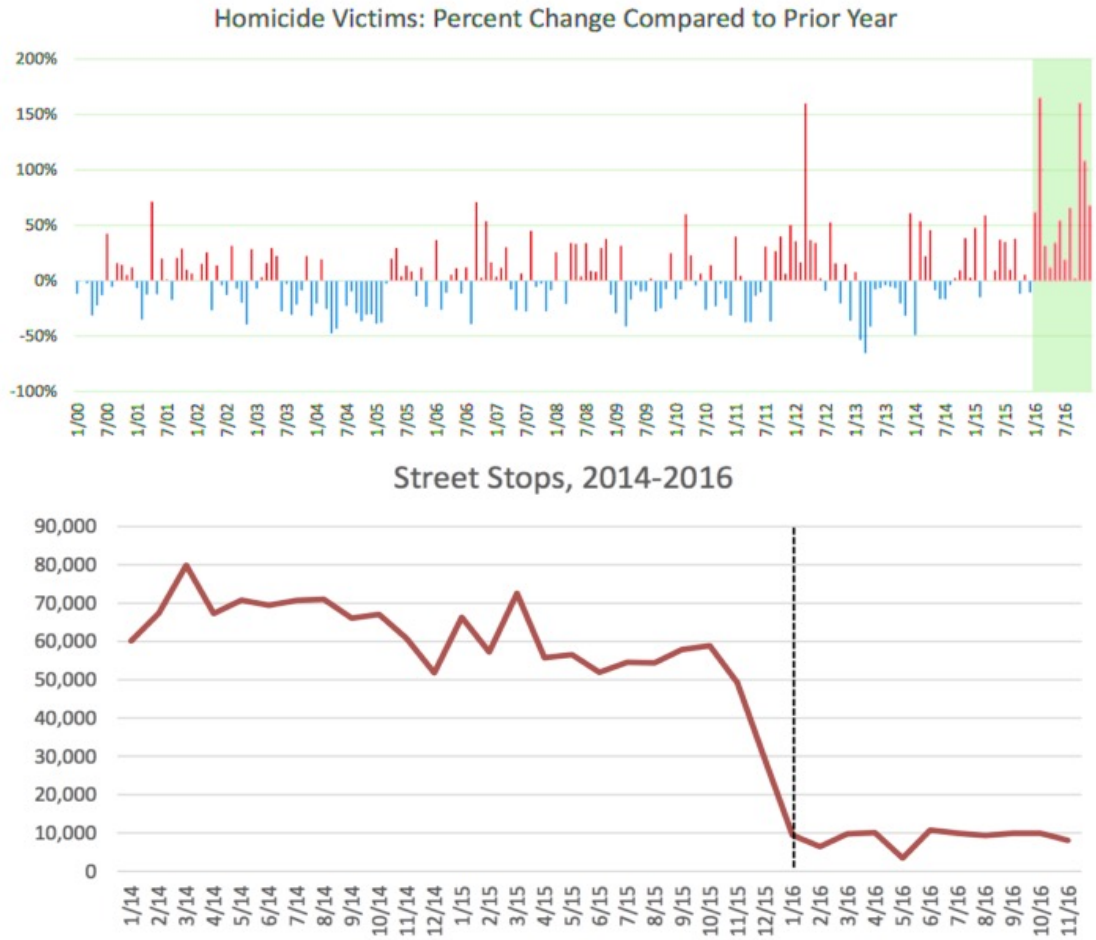


Figure 1A, Top Panel. Yearly Percent Change in Number of Homicide from 1900 to 2016. We can see that the homicides spiked since Jan. 2016.

Figure 1B, Bottom Panel. Occurrences of Street Stops Carried Out by Chicago police from 2014 to 2016. Yearly Percent Change in Number of Homicide from 1900 to 2016. There is significant decline in the number street stops after the third quarter of 2015 and that has been kept at low levels ever since. The sharp decline in the no. of street stops coincides with the rapid climb in the no. of homicide victims. Credit to: University of Chicago Crime Lab.

comprehensive information on more than 6,000,000 crimes reported by the Chicago Police Department (CPD) since 2001. The dataset can be accessed online via https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/__. The variables include information such as the GPS coordinates, the block, the police district, the community area, description of the location where the incident occurred, the date and time when it occurred and its category, etc. All the variables are discrete if not categorical. For this study, we assume every variable categorical when we implement regression modeling if not otherwise stated.

Exploratory Analysis

Exploratory analysis, mainly graphical analysis to visualize the distribution of (types of) crimes by district, community area, day of week, hour of day and so forth, is done with the "ggplot" and "ggmap" packages. Data cleaning and filtering are further performed based on these analyses with the "data.table" package.

Logistic Regression

Logistic regression for binary outcome is implemented with the 'glm' function in base R. Diagnostic plots and tests are used to determine goodness-of-fit and compare models.

Machine Learning Methods

Naive_decision tree is used to graphically illustrate the variables as nodes. Random forest is used to see the importance of variables. R packages "rpart" and "randomForest" are used for the implementation of these two methods. These two methods primarily serve as comparison to logistic regression.

Results

Number of Recent Crimes by Type

First, we subset the dataset by 'Year'>= 2007. This is because 1) data more than 10 years ago is likely to be irrelevant to current trends, and 2) the size of the data is simply huge and takes too much time to run. Number of reported crimes since 2007 in Chicago sums up to 3,420,383. More than 20% of crimes are of theft the amount of domestic battery is about the same of non-domestic (**Figure 2**). Using a heatmap with year as columns, we are able to see the trends of different types of crimes in the past 10 years (**Figure 3**). Darker red means more crimes and lighter means less. We can see clearly that occurrence of homicide in 2016. Deceptive practices and offense involving children have

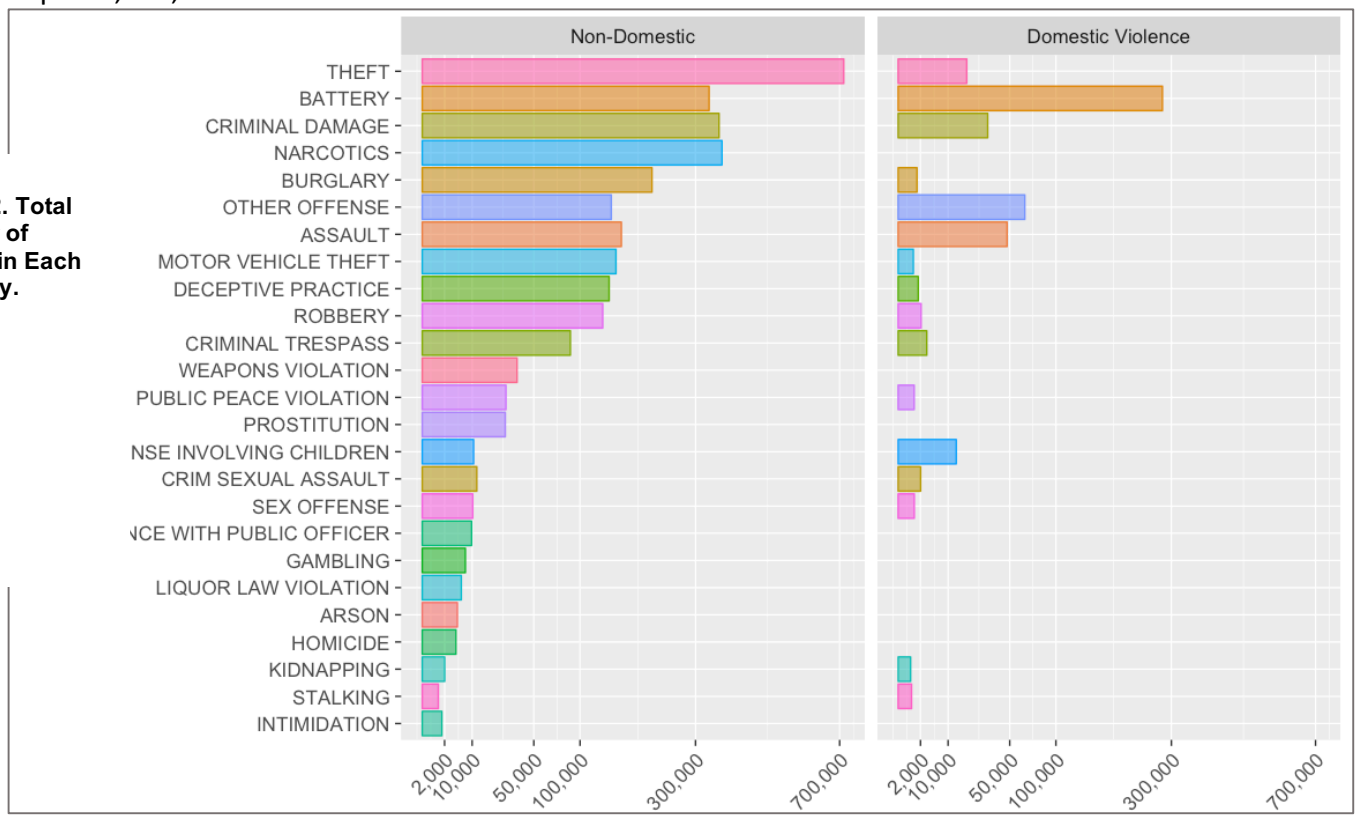


Figure 2. Total Number of Crimes in Each Category.

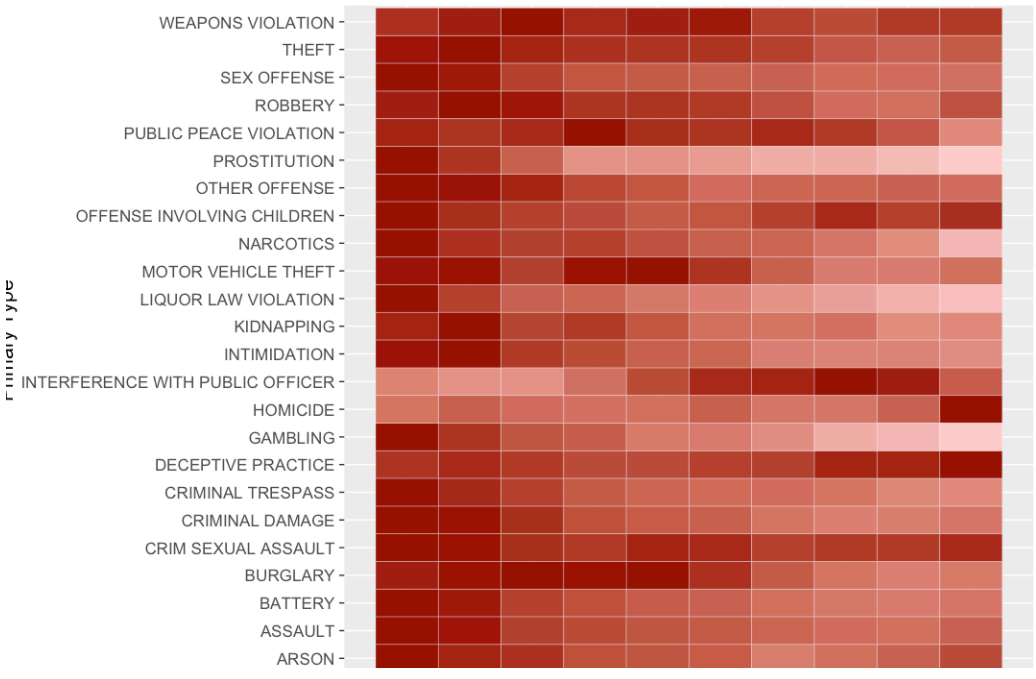


Figure 3. Trends in Occurrence of Crimes in the Past 10 Years. Each row is normalized to its maximum value. We can see from this plot that despite most crimes stay stable or decrease a little over years, number of homicides skyrocketed in 2016. Sexual assaults, interference with public officer, offense involving children and deceptive practices have also been increasing.

also been increasing in the recent five years. Number of reported common street crimes such as liquor violations and narcotics has been decreasing steadily since 2008 and there is a sharp drop in 2016. I suspect that this observation is due to less street stops and raids made by the CPD and thus less reported cases instead of all cases. Such is true for gambling and prostitution where there is usually no victim and the majority of such crimes are discovered by police raids. This observation may substantiate the findings by the Crime lab at the University of Chicago about the correlation of decrease in street stops and increase in homicide victims.

More Serious Crimes for University Students

Next we focus on more serious crimes that happen around locations where university students are likely to visit. For example, abandoned house, public housing, construction site or factories usually do not attract college students while you can expect to find students in stadium, movie theatres, grocery stores or simply on the streets. Offenses that are considered more serious crimes are index crimes[4]. Local law enforcements are required to report Index crimes to FBI and these crimes are composed of 1st and 2nd degree murder, criminal sexual assault, robbery, aggravated assault and battery, burglary, larceny over 500 USD, motor vehicle theft and arson. After this classification, we found that in Chicago, about 38 percent of all crimes are index crimes. We further define violent crimes which can result in bodily harm to victims and plot those crimes since 2007 on maps (**Figure 4 A and B**). In **Figure 4B**, we can see that it is comparatively safe within the periphery of University of Chicago, nothing too terrifying (murder and kidnapping) has happened since 2007. Also the two parks near the university seem to be safe.

Criminal Hop-spots in Chicago

There are 77 community areas in Chicago. Knowing what communities are dangerous can give us a rough guide to safety. Summarizing the number of violent crime by community areas and then dividing by the area of each community gives us the most "violent" areas in Chicago. The violent crimes in top 10 violent community areas (see table below) account for 38.9% of the violent crimes in Chicago yet the area of these 10 community areas combined is only 16.8% of Chicago.

Community Area	Density (case/mi^2·Year)
West Garfield Pk.	413
South Shore	310
West Englewood	272
Englewood	268
East Garfield Pk.	260
North Lawndale	250
Austin	231
Grand Crossing	223
Humboldt Park	222
Auburn Gresham	220

Table 1. Top 10 Most Violent Areas by Crime Density.

Temporal Criminal Patterns

Other than geographical patterns, temporal patterns also tend to exist among crimes. For exploratory purpose, we look at yearly, monthly patterns and the importance of day of week and hour of day (see section gun crimes). We find that all crimes speak in summer in Chicago (**Figure 5A**, especially homicides. Robberies, thefts and homicides are at lower levels during spring (or winter actually, since it is Chicago). Cold temperature (average temperature from November to March are below 20 F) and snow probably deter people from going on the streets and some criminal activities. Looking at the heatmap (**Figure 5B**), we find that not surprisingly, occurrences of alcohol-

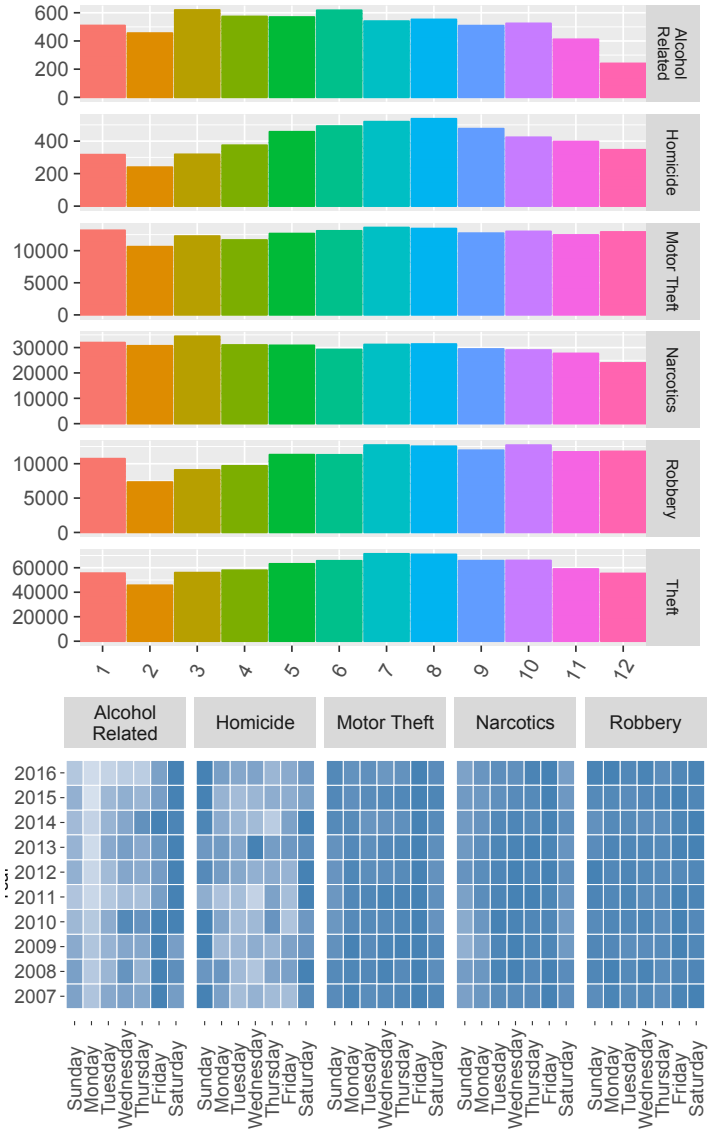


Figure 5A, Upper Panel. Distribution of Crimes across Months. **Figure 5B, Lower Panel.** Distribution of Crimes across Weekdays and Years.

related violation are lowest on Mondays and highest on Fridays and Saturdays. However, surprisingly, for not obvious reasons, more homicides happen on Sundays and then Saturdays. Other crimes are fairly evenly distributed within each week.

Gun Crimes

After we have explored the various violent crimes, now we want to focus on the crimes that include the use of guns, since we believe

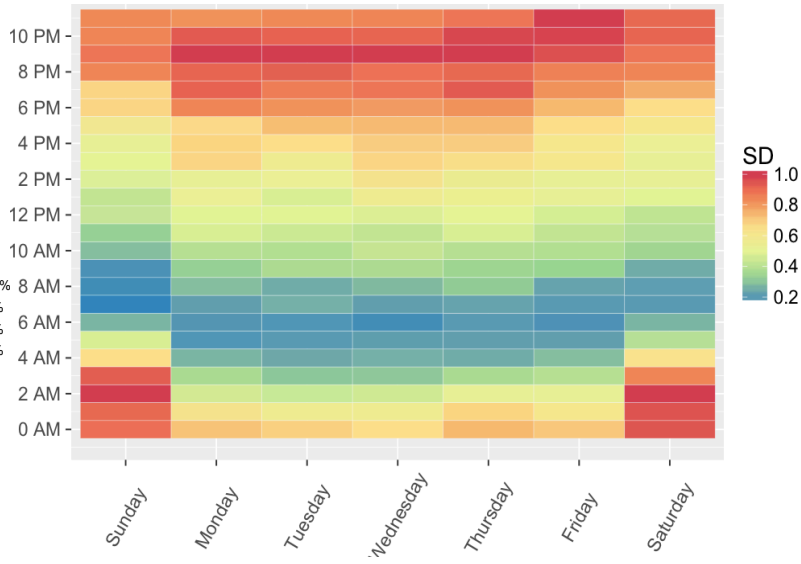
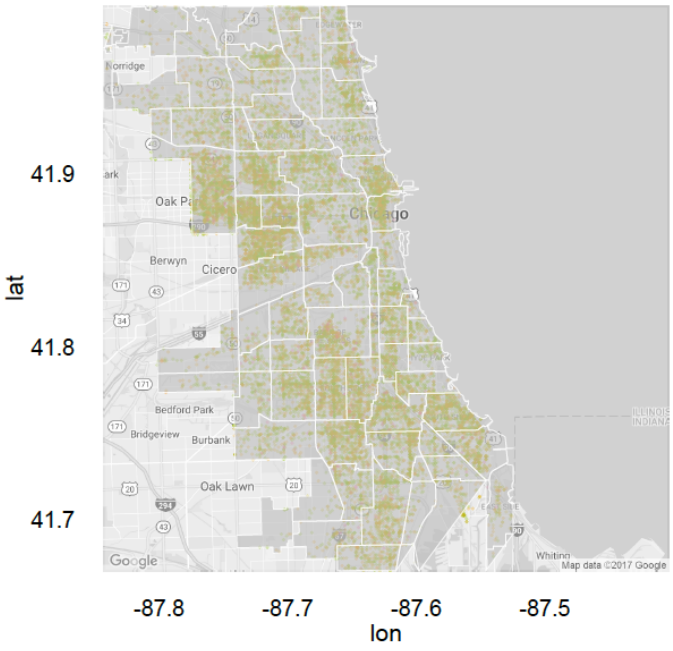
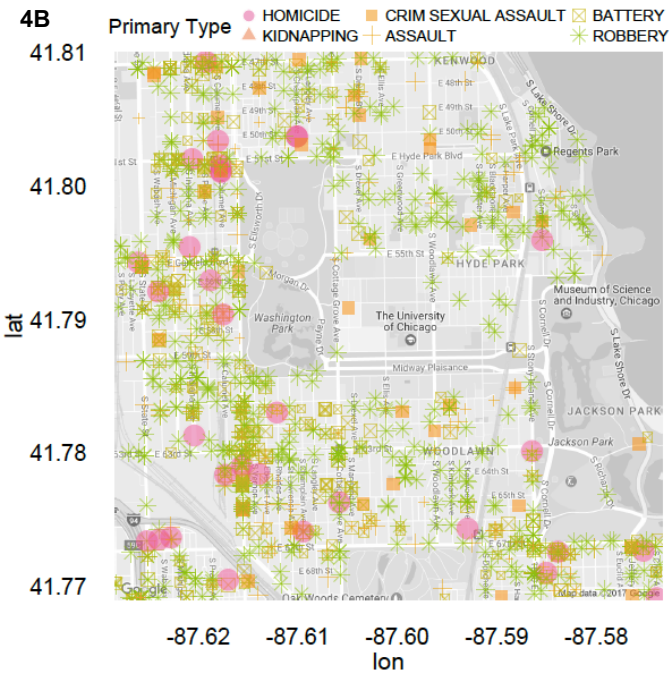


Figure 6. Heatmap Showing Distribution of Gun Crimes over Hour of Day and Day of Week. We see that there is an hour dependency but not so much weekday.

Figure 4A Violent Crimes in Chicago and Near UChicago



4B



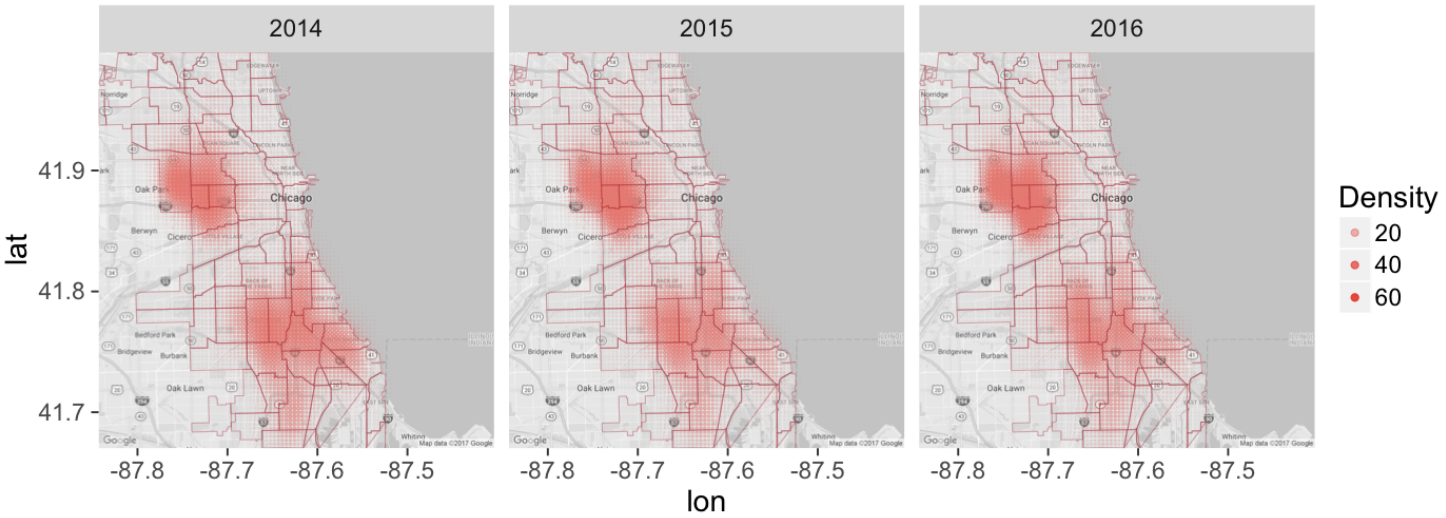


Figure 7 Density Plots of Gun Crimes from 2014 to 2016. The redness is darker in the west cluster in 2016 compared to 2014. The redness is lighter in the south cluster in 2016 compared to 2014.

that wide spread of illegal guns contributes to the escalation of violent crimes in Chicago. We can filter the dataset by the 'Description' of crimes, which is the subtypes of primary types of crime. The use of weapons is identified in this attribute so we can search for the crimes with handguns and other fire arms. In this data, about 36.6% violent crimes use guns. A density plot is generated (**Figure 7**) to visualize the hot spots of gun crimes. We can see that the south side and west side of Chicago are the most troublesome regions and in the recent three years more gun crimes have happened in the west part.

Logistic Regression

Based on the previous exploratory analyses about the temporal and geographical patterns that exist in crimes in general and violent crimes specially, we want to **quantify the effects of location and time on determining whether a gun is involved in a violent crime with regression models**. First, we engineered a new feature `Gun` with 1 indicating use of guns and 0 otherwise. Then, we randomly assigned crimes in `Community Area` 77 to 1 and 3 and 76 to 10 and 17, and set this attribute to numeric because smaller numbers are clustered in northeast region of Chicago while the numbers get bigger if we move to the south or west, so the numeric values of area code contain relevant information. Also we can collapse the months in two four season. 6-8 being summer, 9-10 being autumn, 11-2 being winter and 3-5 being spring. Notice that this is not the traditional classification of seasons but this matches the pattern observed in **Figure 5A**. We can also hours in to eight intervals, starting from 12-2 AM being the first interval **Figure 6**. The feature engineering also reduces computational complexity of the regression tasks for factor interactions. Then we fitted a general linear model (referred to as the "full model", though it is not) with the following formula:

$$Gun \sim District + \text{`Location Description`} + \text{`Community Area`} + hour + weekday + month + District*Hour + + \text{`Location Description`}*Hour + Season*Weekday$$

with AIC = 54780. 'Hour' is the engineered factor feature whereas 'hour' is the original factor. A null model with `Gun` ~ 1 was constructed for comparison. Next we used step() and the null model to do a forward selection and got another model:

$$Gun \sim District + \text{`Location Description`} + \text{`Community Area`} + month + Hour + Season + Weekday$$

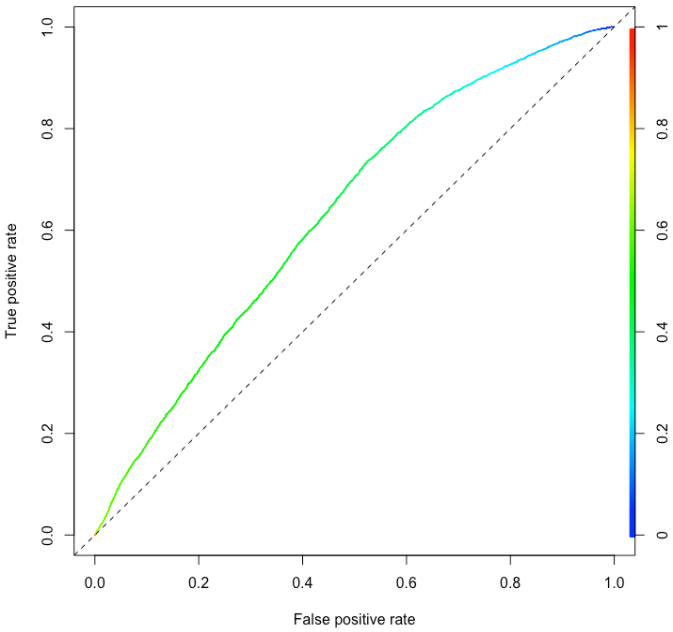
with AIC = 54609. We also built a model with no interaction terms or engineered terms. A likelihood ratio test between this model and

the full model yields P-value << 0.001, meaning the addition of engineered terms and interactions are indeed significant.

Neither of the regression model fits well as their residuals are far from normal (**Appendix I**) and their pseudo R^2 scores are small (0.589 and 0.527). However, their abilities to predict the use of gun in violent crimes were determined by contingency tables and summarized in the table below. ROC curve is also generated.

	Training Date		Test Date
Threshold	Sample π	0.5	0.5
Full model	57.6%	62.1%	60.57%
Step model	57.9%	61.1%	61.59%
Null model	58.9%	58.9%	59.71%

Table 2. Summary of Accuracy of Model Fit. Calculated from individual two-way contingency table. Probability of 0.5 is a better cutoff than sample proportion.



Based on the full model, community area, month and interaction term of `Location Description`:Hour all have significant effects on whether a gun is involved in violence with p-value 0.08,0.0015 and 0.0003. Relative odds can be easily calculated. For example, odds ratio between violent crimes happen in residential garage and in the alleys is 1.36 with 95% CI (1.13, 1.61). Odds ratio between Hour = 3-5 PM and Hour = 1-2 AM (inclusive) is 0.743 with 95% CI(0.69, 0.80).

Decision Tree and Random Forest

Because the logistic regression models we generated are not satisfactory, we also employed decision tree and random forest in the hope that we can achieve better predictive power. However, the decision tree (**Appendix II**) yields the exact same accuracy as the full model given by logistic regression and the random forest yields a slightly lower accuracy. This may be due to the limitation of number of factors its R implementation can deal with. The variable importance is given by the algorithm and is presented by a variance importance plot (**Appendix II**). The order of importance predicted by mean decrease accuracy agree with the results from logistic regression and simple decision tree.

Discussion

We are able to get some inference from the models but none of the model did a good job in fitting the training data. This could be due to serval reasons. First, the dataset is very big and the relative meaningful information from the variables is limited. There is no real meaningful continuous variable in the dataset. Secondly, basically we can only use the information on location and time to predict gun violence but logically there are other conditions that lead to whether a gun is used in a violent crime. Decision tree and random forest do not outperform logistic regression in this case but they offer better interpretability (**Appendix II**). Perhaps, the availability of guns in local areas or the presence of police etc. Also, the choice of weapon is determined by the predator, but there is no information on the suspects.

Classification of a crime, on the other hand, may be more feasible with such datasets [5].

Acknowledgement

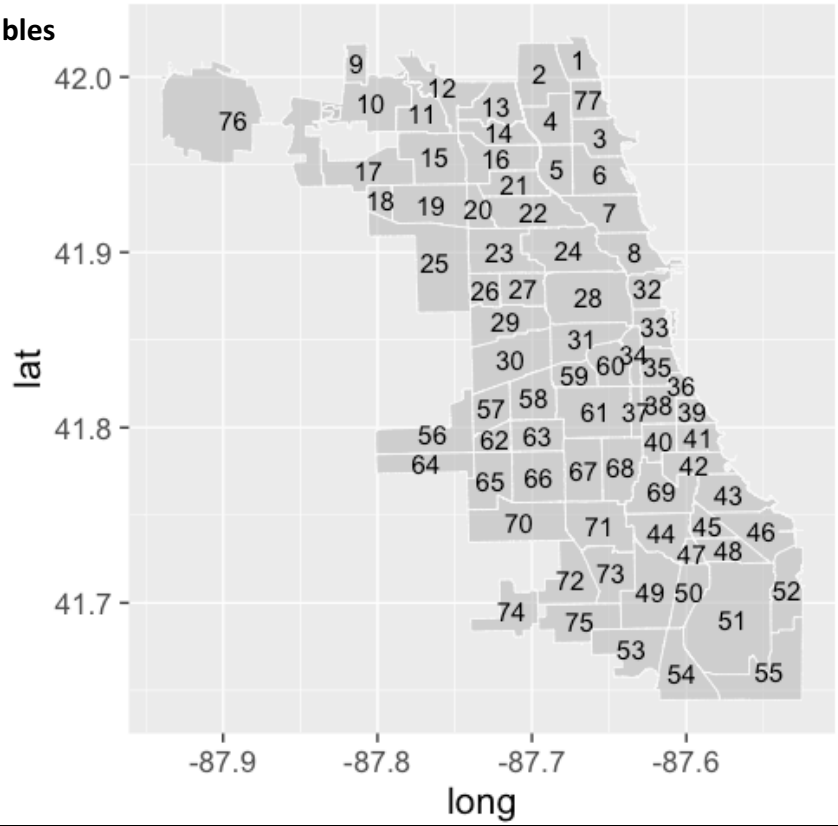
I referenced two online tutorials heavily for this project. A tutorial on ggmap (https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf) by David Kahle and Hadley Wickham. And another tutorial (http://trevorstephens.com/kaggle-titanic-tutorial/getting-started-with-r/) by Trevor Stephens. Many thanks to these authors and other contributors to Stack Overflow and Kaggle.

Reference

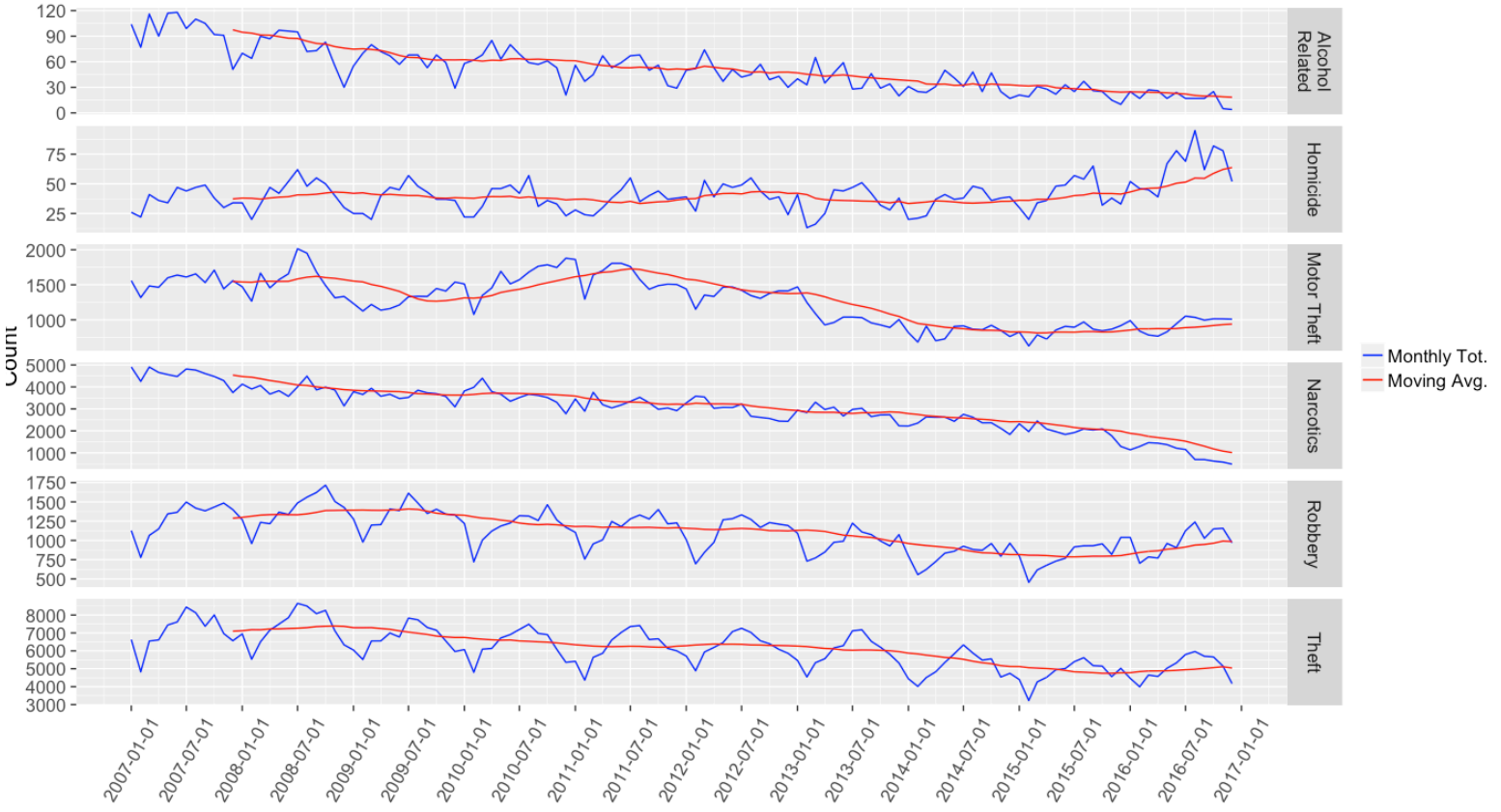
1. Chicago Tribune News Applications Team. Crime in Chicago -- Chicago Tribune. Available at: http://crime.chicagotribune.com/chicago/. (Accessed: 7th January 2017)
2. Chicago's Murder Rate is Typical for a Major Metropolis — Until Fatal Shootings are Factored In. *The Trace* (2017). Available at: https://www.thetrace.org/2017/01/chicago-murder-rate-fatal-shootings/. (Accessed: 7th January 2017)
3. Spies, M. & Fuhrman, E. Watch How Chicago Gets Flooded with Thousands of Crime Guns. *The Trace* (2015). Available at: https://www.thetrace.org/2015/11/chicago-gun-laws-shootings-trafficking/. (Accessed: 9th January 2017)
4. CLEARMAP Crime Type Definitions. Available at: http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html. (Accessed: 10th January 2017)
5. San Francisco Crime Classification | Kaggle. Available at: https://www.kaggle.com/vivekyadav/sf-crime/sfo-rmd-kaggle. (Accessed: 12th January 2017)

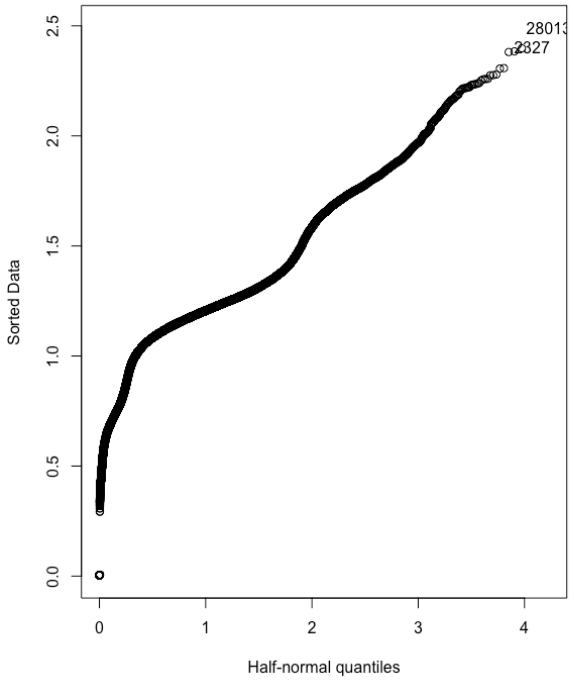
Appendix I Supplementary Figures and Tables

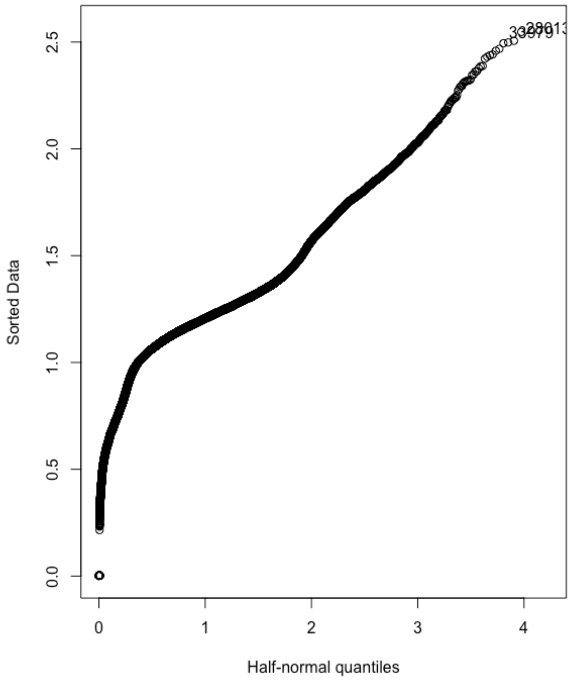
Figure to the right: boundaries of the community areas given by CPD. Table at the bottom: table showing the code and names for each community area. Gun crimes happened in 76 were divided In to 10 and 17 while those happened in 77 were re-assigned as either 1 or 3 in an effort to better the use the codes as numeric variables.



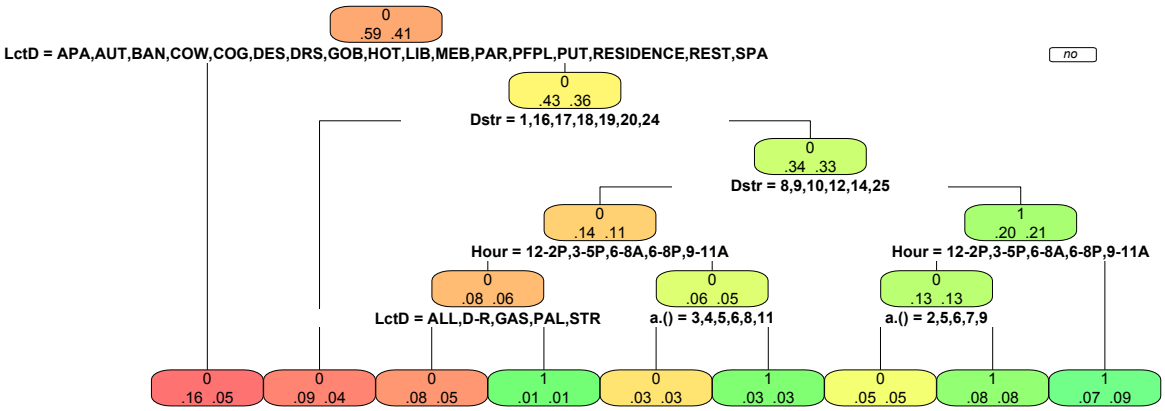
NAME	CODE	NAME	CODE	NAME	CODE
ROGERS PARK	1	EAST GARFIELD PARK	27	WEST PULLMAN	53
WEST RIDGE	2	NEAR WEST SIDE	28	RIVERDALE	54
UPTOWN	3	NORTH LAWNDALE	29	HEGEWISCH	55
LINCOLN SQUARE	4	SOUTH LAWNDALE	30	GARFIELD RIDGE	56
NORTH CENTER	5	LOWER WEST SIDE	31	ARCHER HEIGHTS	57
LAKE VIEW	6	LOOP	32	BRIGHTON PARK	58
LINCOLN PARK	7	NEAR SOUTH SIDE	33	MCKINLEY PARK	59
NEAR NORTH SIDE	8	ARMOUR SQUARE	34	BRIDGEPORT	60
EDISON PARK	9	DOUGLAS	35	NEW CITY	61
NORWOOD PARK	10	OAKLAND	36	WEST ELSDON	62
JEFFERSON PARK	11	FULLER PARK	37	GAGE PARK	63
FOREST GLEN	12	GRAND BOULEVARD	38	CLEARING	64
NORTH PARK	13	KENWOOD	39	WEST LAWN	65
ALBANY PARK	14	WASHINGTON PARK	40	CHICAGO LAWN	66
PORTAGE PARK	15	HYDE PARK	41	WEST ENGLEWOOD	67
IRVING PARK	16	WOODLAWN	42	ENGLEWOOD	68
DUNNING	17	SOUTH SHORE	43	GREATER GRAND	69
MONTCLARE	18	CHATHAM	44	ASHBURN	70
BELMONT CRAGIN	19	AVALON PARK	45	AUBURN GRESHAM	71
HERMOSA	20	SOUTH CHICAGO	46	BEVERLY	72
AVONDALE	21	BURNSIDE	47	WASHINGTON HEIGHTS	73
LOGAN SQUARE	22	CALUMET HEIGHTS	48	MOUNT GREENWOOD	74
HUMBOLDT PARK	23	ROSELAND	49	MORGAN PARK	75
WEST TOWN	24	PULLMAN	50	OHARE	76
AUSTIN	25	SOUTH DEERING	51	EDGEWATER	77
WEST GARFIELD PARK	26	EAST SIDE	52		


Count
Alcohol Related
Homicide
Motor Theft
Narcotics
Robbery
Theft
Monthly Tot.
Moving Avg.
2007-01-01
2007-07-01
2008-01-01
2008-07-01
2009-01-01
2009-07-01
2010-01-01
2010-07-01
2011-01-01
2011-07-01
2012-01-01
2012-07-01
2013-01-01
2013-07-01
2014-01-01
2014-07-01
2015-01-01
2015-07-01
2016-01-01
2016-07-01
2017-01-01


Sorted Data
Half-normal quantiles
2827


Sorted Data
Half-normal quantiles

Appendix II Additional Plots Pertaining to Decision Tree and Random Forest Method



Left: Half-normal of the residuals of the forward selection model. Right: The full model.