The Habitat of Crime

Examining the Spatial and Demographic Characteristics of Chicago Crime to Inform Resource Deployment

Author: Jonathan Morken

April 2016

Abstract

The "client" has requested analytic support to determine how to invest efficiently in new police infrastructure. The purpose of this analysis was to determine significant spatial, temporal, and demographic crime patterns that contribute to Chicago community crime rates and then use those patterns to guide investments and/or inform policing patterns. The initial spatial hypothesis was as follows:

1. *Crime rates (as measured by* crime count/km^2*) decrease as one moves closer to a police department.*
    a. *This assumes police presence is a deterrent to crime. It also assumes that its deterring effect is strong enough to overcome any increase in proactive arrests due simply to increased patrol frequency.*
2. *Crime rates increase as one moves closer to a transit station.*

The data from 2013-01-01 through 2014-12-31 indicates that crime rates do not decrease but actually <u>*increase*</u> *as one moves closer to a police station. This evidence is not enough to invalidate the hypothesis that police presence is a deterrent. Instead, given the intuitive nature of that assumption, the pattern may simply underscore the effectiveness of proactive policing. E.g. that patrols are effective enough to overcome the natural deterrent one would otherwise expect from increased police presence. Part 2 of the hypothesis was supported by the data. Crime rates increase exponentially as one approaches a transit station. The final recommendation as it concerns police infrastructure can be found in the conclusion.*

1 Table of Contents

2 Introduction

As has been reported in recent news [spring of 2016], Chicago is "off to [its] deadliest start in nearly two decades." Homicides and shootings are up by more than 70% compared to 2015, while arrests and investigative stops have fallen. As part of its response, the City Police Authority has requested analytic support to determine how to invest efficiently in new police infrastructure and increase patrol effectiveness. The purpose of this analysis is to determine significant spatial, temporal, and demographic crime patterns that contribute to Chicago community crime rates and exploit those patterns to guide investments and/or inform policing patterns. This paper examines Chicago Crime Data from several different angles and utilizes multiple visual methods including the following:

- Histograms; Frequency of crime by type relative to
  - Police presence [proximity to PD's]
  - Proximity to Transit stations,
  - Hours of the day
  - Days of the week,
  - Community demographics
- Choropleth maps showing crime density by community
- Point analysis for the city as a whole that will include spatial standard deviation ellipses.
- Crime heat-maps
- Scatterplots and linear model analysis to determine the efficacy of a basic linear model in predicting crime density. This model could be used to inform policing patterns and/or future station locations.

The initial, spatial hypothesis:

1. Crime rates (as measured by crime count/km^2) decrease as one moves closer to a police department
   - This assumes police presence is a deterrent to crime. It also assumes that its deterring effect is strong enough to overcome any increase in proactive arrests due simply to increased patrol frequency.
     - *Note: for the sake of simplicity I assume that each police department employs the same number of officers. This could be tightened with actual weights.*
2. Crime rates increase as one moves closer to a transit station.

The analysis herein focuses on Violent Crimes and Non-violent Property Crime as defined below.

3 Data

3.1 Background on Data

The superset of data is a combined set of files including Chicago crime data, geo-spatial polygons, mass transit station locations, police department locations, demographics, categorization tables, and weather data. All requisite data files can be found in the appropriate folder on Github.

Note for analysists from Springboard: While the crime data currently covers dates from 2013-01-01 to 2014-12-31, this could be updated at will. I have intentionally left 2015 data out, as I am using it as test data for a separate analysis.

3.2 Sources

Base crime data, shape files, coordinate files, and census data come from the data repository at the City of Chicago website. Although not used in this iteration of analysis, weather data was downloaded from the Weather Underground and is available in the project Github drive. This analysis also uses a community-census tract mapping table courtesy of Rob Paral.

The base sources are then subsequently scrubbed, reformatted, joined, etc. in R. Steps taken can be viewed in the RMarkdown file "CreateSuperDataSet" and "MappingTheData" on Github.

3.3 General Data Structure Notes

Each row in the base crime dataset equates to 1 crime. While the base crime data is of high quality, this analysis still required mutation, cleansing, etc. The crime data does include some null fields, duplications where a given crime may have had more than one victim, and some erroneous location data such as crimes occurring within a police station, outside of city limits, or within a different state, etc . Missing spatial data and data outside of the city of Chicago is dropped. You can see more detail and follow along via the files on Github.

Most of the **demographic data** is provided by the city of Chicago's Department of Public Health as underlying pieces of their "Hardship Index." It includes the following:

- Percent of crowded housing
- Percent of households below poverty
- Unemployment rate, 16+
- Percent aged 25+ without a high school diploma
- Percent aged under 18 or 65+
- Per capita income

In order to retain as much data as possible but make the content of this analysis more intuitive, each **primary type of crime has been mapped to a "Meta Category**." These meta categories were created by grouping crime descriptions as closely to federal definitions as possible with provided data. Categories are as follows:

- Violent Crimes
    - Assault
    - Battery
    - Criminal Sexual Assault
    - Homicide
    - Etc.
- Non-Violent Property Crime
    - Arson
    - Burglary [ex Home invasion]
    - Motor Vehicle Theft
    - Etc.
- Financial Crime/Fraud
    - Deceptive Practices [embezzlement, etc.]
    - Intimidation - Extortion
    - Money Laundering
    - Etc.
- Other
    - Traffic violatons
    - Noise violations
    - Non-violent prostitution
    - Non-violent narcotic violations
    - Etc.

Note: The full mapping table is too large to print here, but is viewable in Github. This analysis focuses on Violent Crimes and Non-Violent Property Crime.

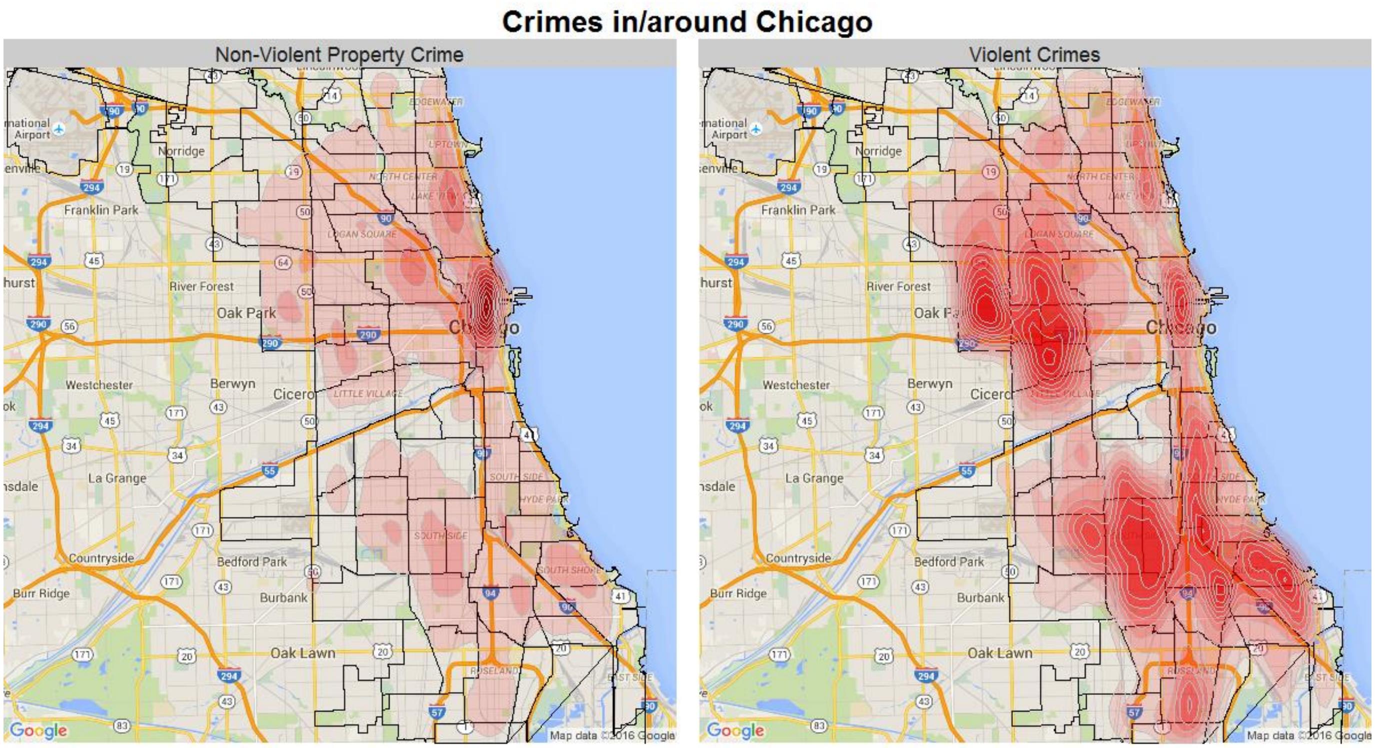
Hour Categories are split into 4 hour increments as follows:

- 7am - 11am is "AM Commute"
- 11am - 3pm is "Lunch"
- 3pm - 7pm is "PM Commute"
- 7pm to 11pm is "Evening"
- 11pm - 3am is considered "Night"
- 3am – 7am is considered "Pre Dawn"

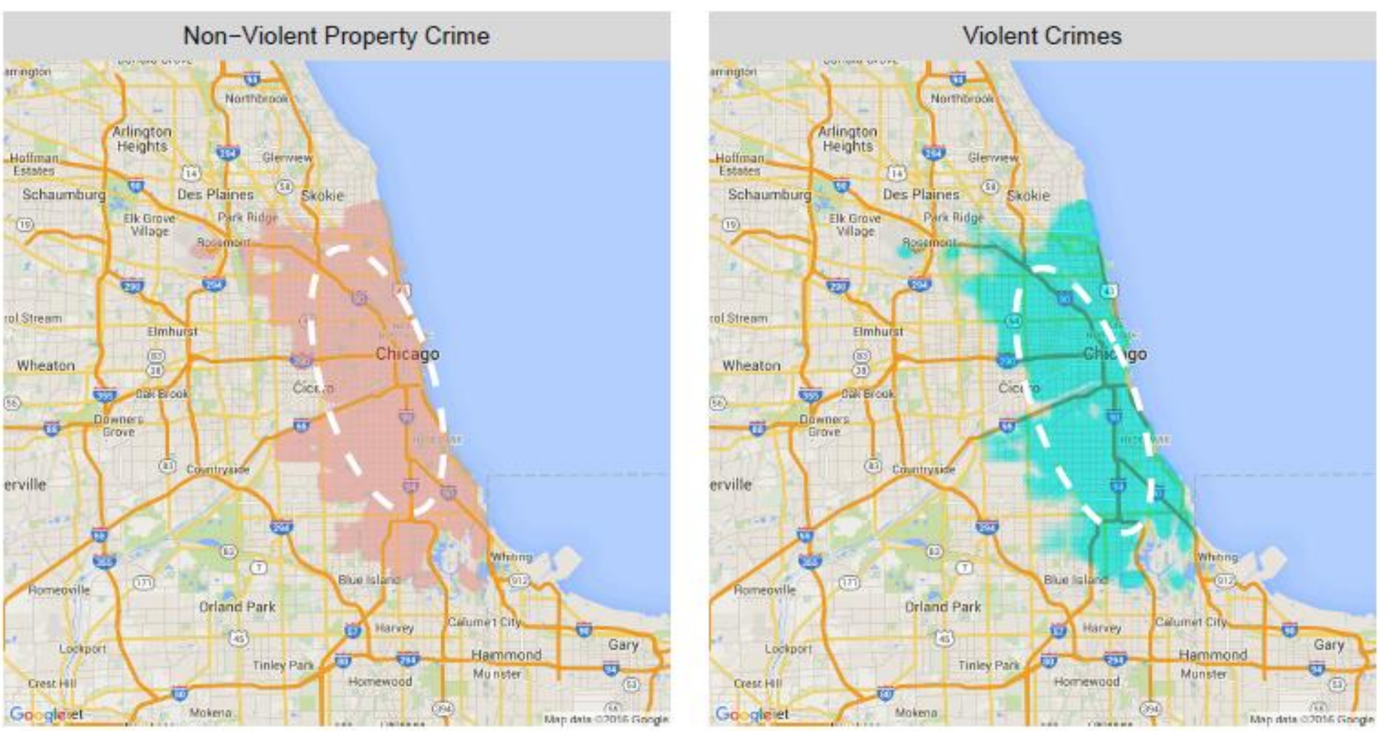
4 Overall View of Chicago Crime

4.1 Heat-maps and Dispersion

The following is an overall image of Non-violent Property Crime and Violent Crimes around Chicago. Property crime is highly focused downtown with a few moderate hot spots elsewhere. On the other hand, violent crime is widely dispersed with multiple hot zones.



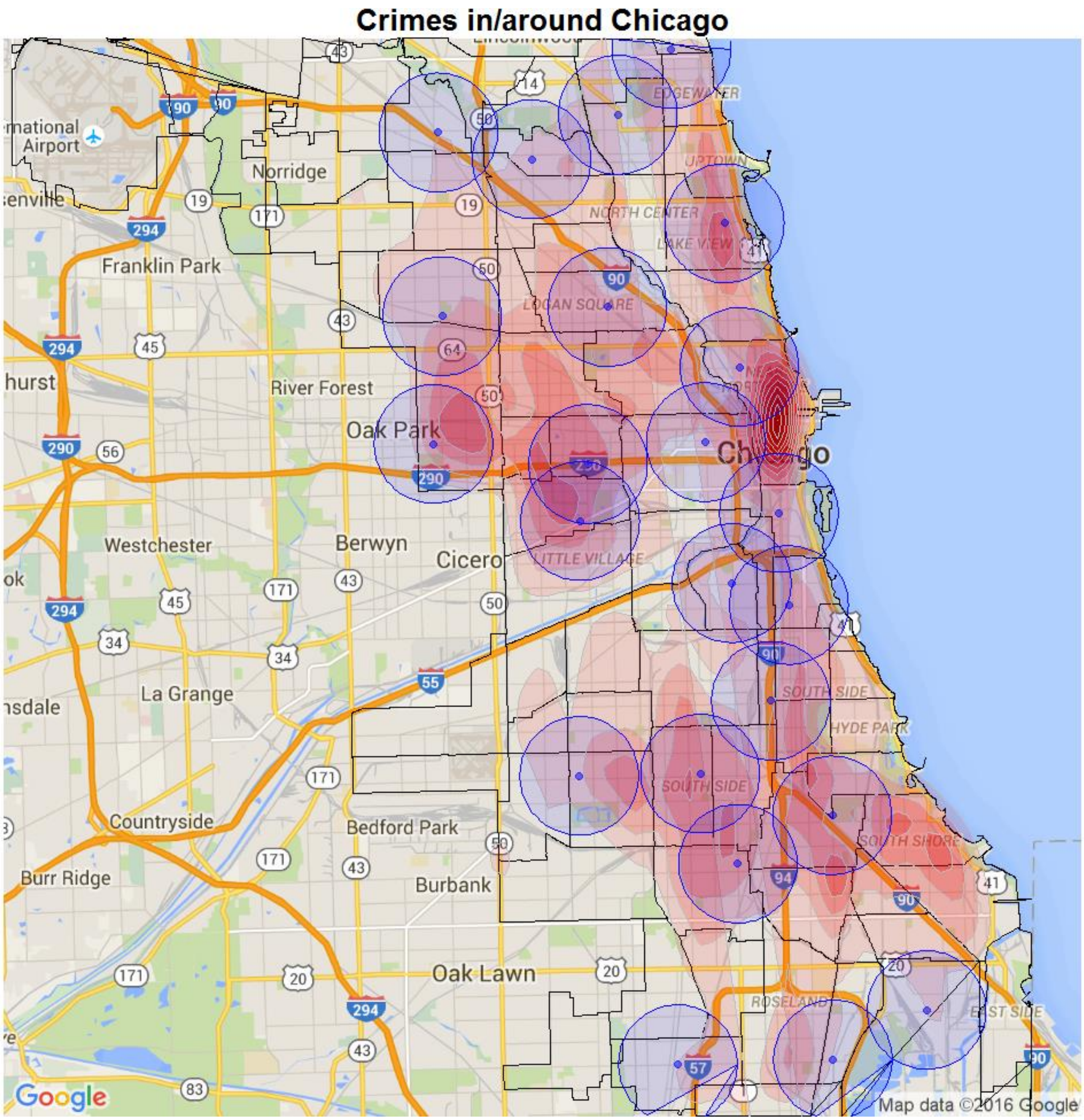
This creates a multi-front battle against crime that requires a targeted approach. If patrols were to evenly canvas the city, the force would be spread too thin. For illustration, the dotted lines on the following graphs encompass just 68% of all Chicago crime:



4.2 Heat Map with Police Departments and Transit Location Overlays

4.2.1 Crime Heat Map with Police Station Overlay

If the hypothesis regarding proximity to police departments is correct, there should be dips in a localized heat map around police markers. The following heat map shows police departments as markers with 2km patrol zones drawn around each one. Any proximity effects would likely be de minimis beyond that range [if there are any at all]. The black borders denote community boundaries.

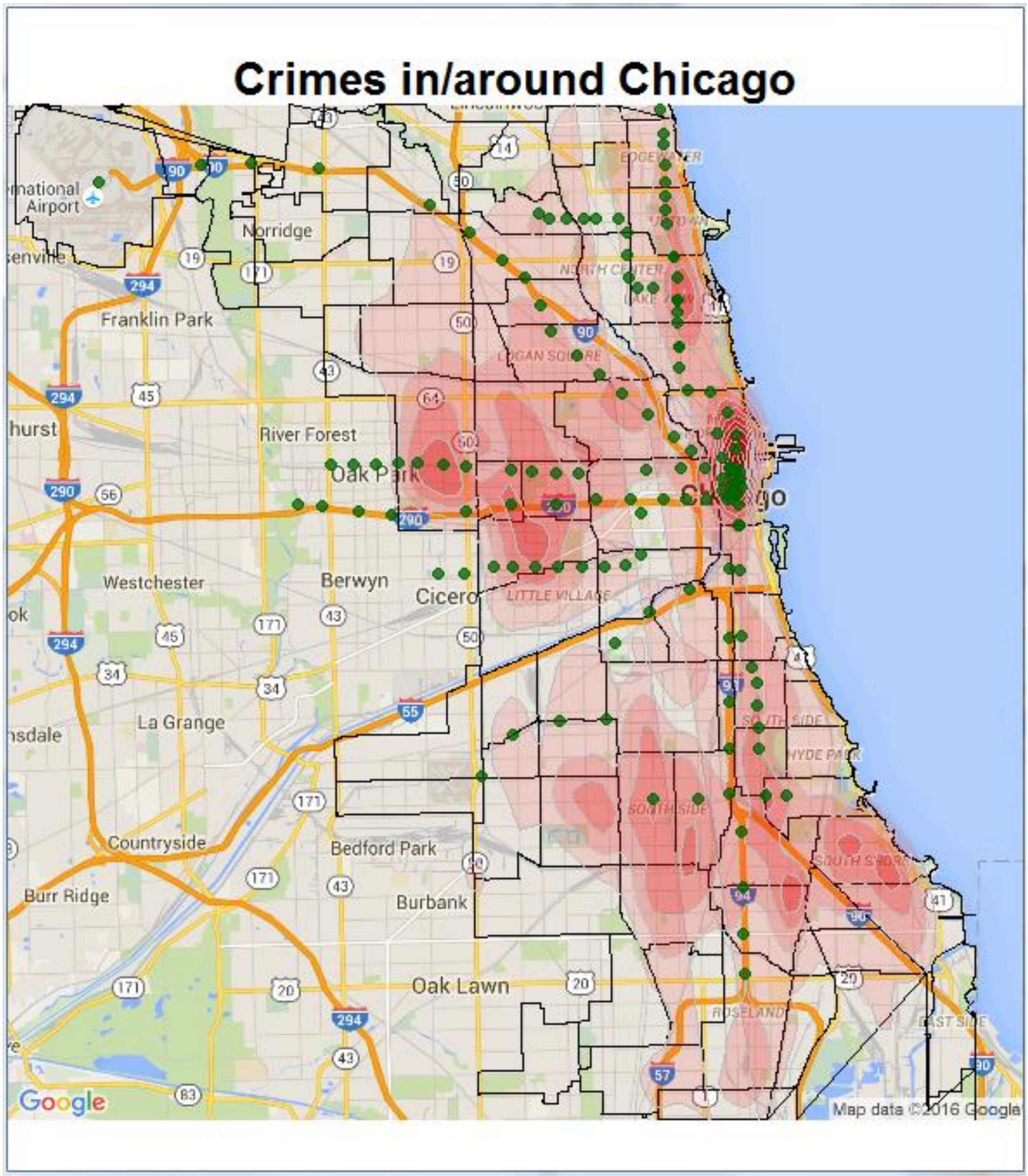


Chicago Crime Density with Police Department Markers and 2km Patrol Zones.

Close inspection reveals that PDs are typically not located at local peaks. It would be interesting to contrast crime patterns in these areas before and after each PD was built. Assuming one controlled for any changes in population/demographics, changes in local crime rates could corroborate patrol effectiveness and/or the hypothesized deterrent effect. Without that data one can only examine the current spatial characteristics. If police presence were a strong deterrent, it would likely create more prominent dips in the areas immediately surrounding stations than seen above.

4.2.2 Crime Heat Map with Transit Station Overlay

On the other hand, a transit station overlay (as represented by the [L train]) shows distinct overlap between common transportation avenues and crime rates. This is not enough to confirm the hypothesized relationship between transit proximity and spatial crime rates, but it is a promising initial picture.



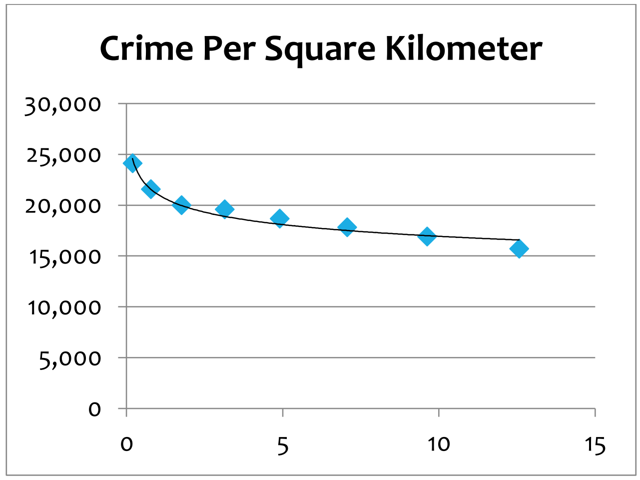
Chicago Crime Density with Transit Station Markers

5 Proximity Analysis

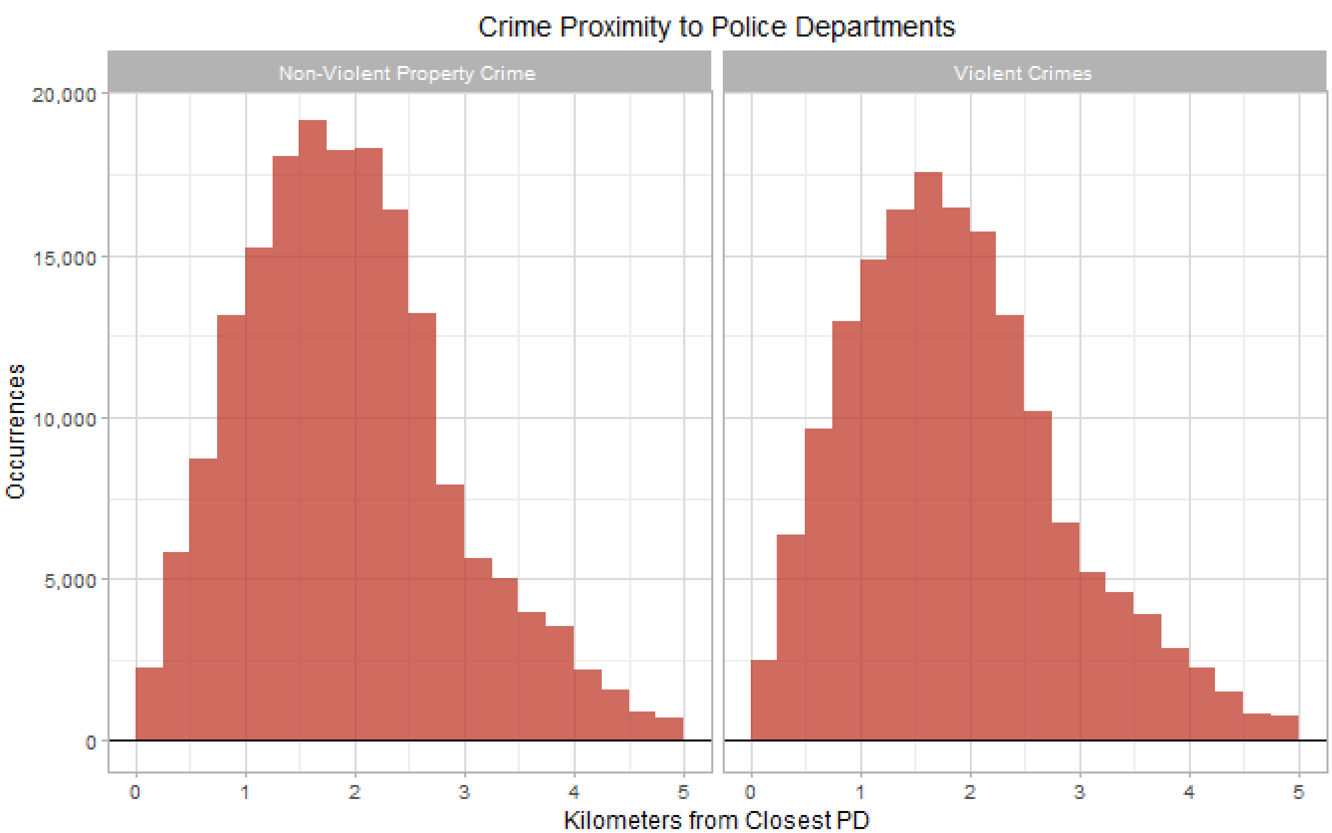
5.1 Police Departments

Contrary to the initial hypothesis, crime rates do not decrease as one gets closer to a given police department. While the effects are not extreme, Violent Crime and Non-Violent Property Crime rates actually increase:

Crime Proximity to PD				
Radius in km	Count	Area in km^2	Crime Per km^2	Change in Crime Density
0.25	4,733	0.196	24,105	12%
0.50	12,186	0.785	21,542	8%
0.75	18,377	1.767	19,973	2%
1.00	26,122	3.142	19,550	5%
1.25	30,083	4.909	18,640	5%
1.50	34,416	7.069	17,814	5%
1.75	36,694	9.621	16,901	8%
2.00	34,707	12.566	15,702	



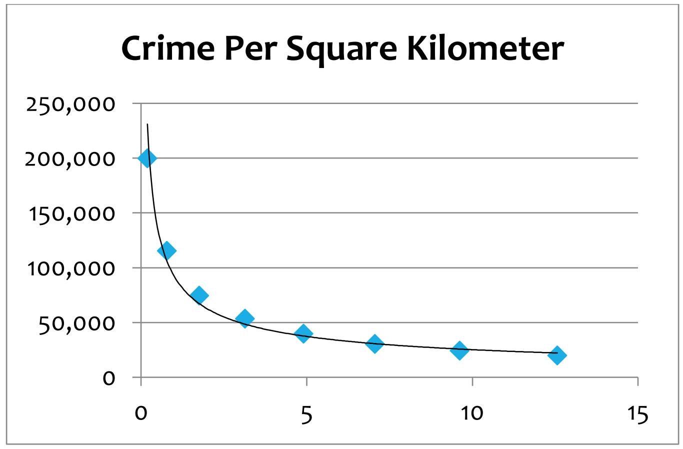
Some of the increase in the first ¼ kilometer is likely due to incorrect location entries such as crimes occurring within police stations. However, the line's overall slope is meaningful. While area shrinks 98% from a 2km radius to a ¼ km, density increases by 54%. If one ignores the first 1/4 kilometer, density still increases by 37%. As mentioned earlier, this does not negate the theory that police presence acts as a deterrent. However, if it is a deterrent as one would expect, that effect is not strong enough to offset more frequent patrols. This could be viewed as a success. I.e. patrols have been a relatively effective approach to confronting crime. To prove or disprove deterrent effects, one would need to contrast crime patterns around a given PD before and after being built.



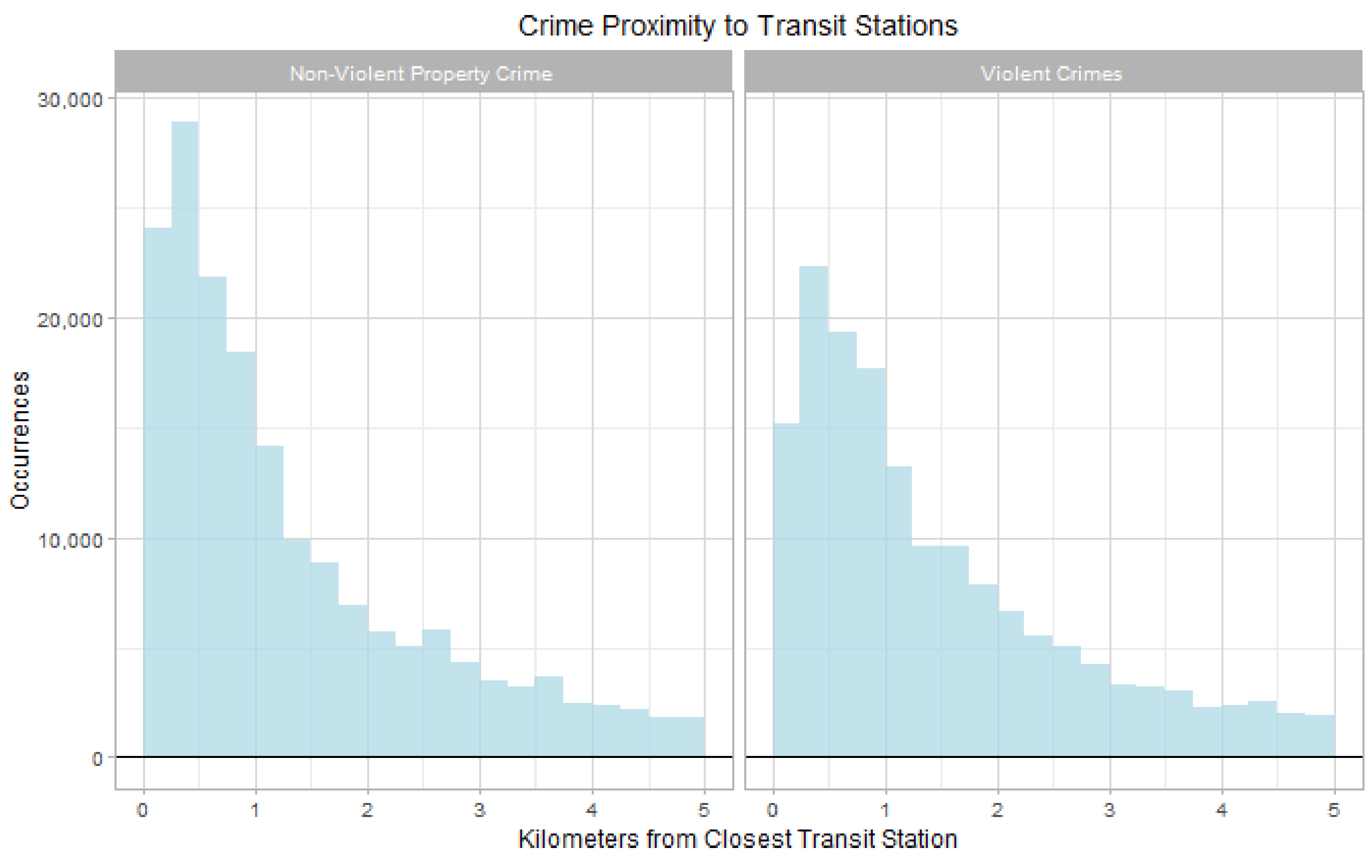
5.2 Mass Transit Locations (as represented by the L)

For mass transit locations the initial hypothesis appears to be correct. Crime rate (density) clearly increases as one moves closer to a given transit station. Density increases by 913% as one moves from 2km to a 1/4 km from a given transit station. This effect starts to flatten out beyond 2 square kilometers.

Crime Proximity to Transit				
Radius in km	Count	Area in km^2	Crime Per km^2	Change in Crime Density
0.25	39,219	0.196	199,741	73%
0.50	51,247	0.785	115,185	55%
0.75	41,169	1.767	74,490	40%
1.00	36,091	3.142	53,389	34%
1.25	27,399	4.909	39,751	31%
1.50	19,475	7.069	30,360	25%
1.75	18,488	9.621	24,227	23%
2.00	14,659	12.566	19,715	



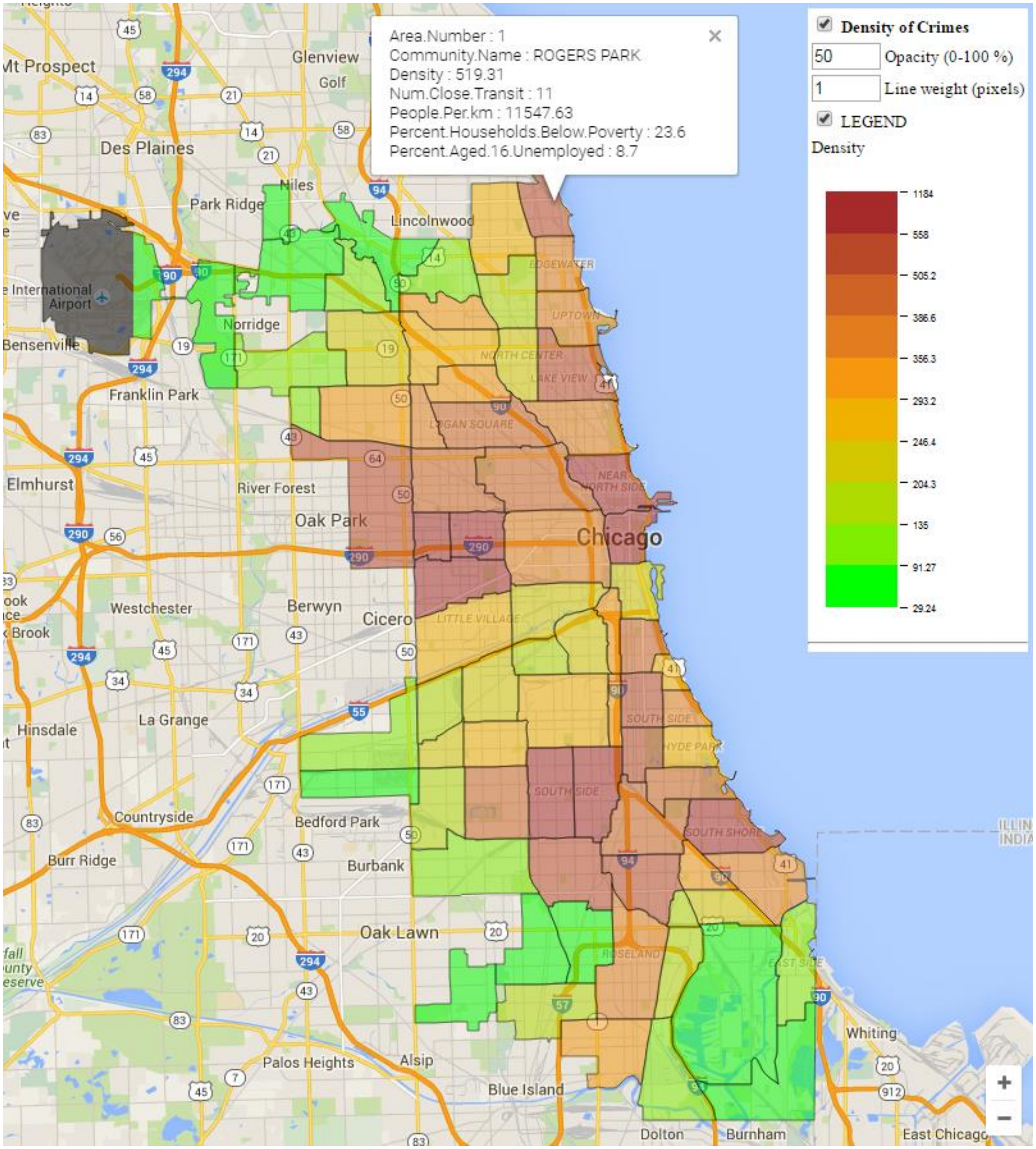
Note: This could be further refined with more transit related data such as bus stops and/or volume estimates. Regardless, it is clear that any future infrastructure investments and patrol assignments must account for a city's transit characteristics.



6 Community Level Analysis

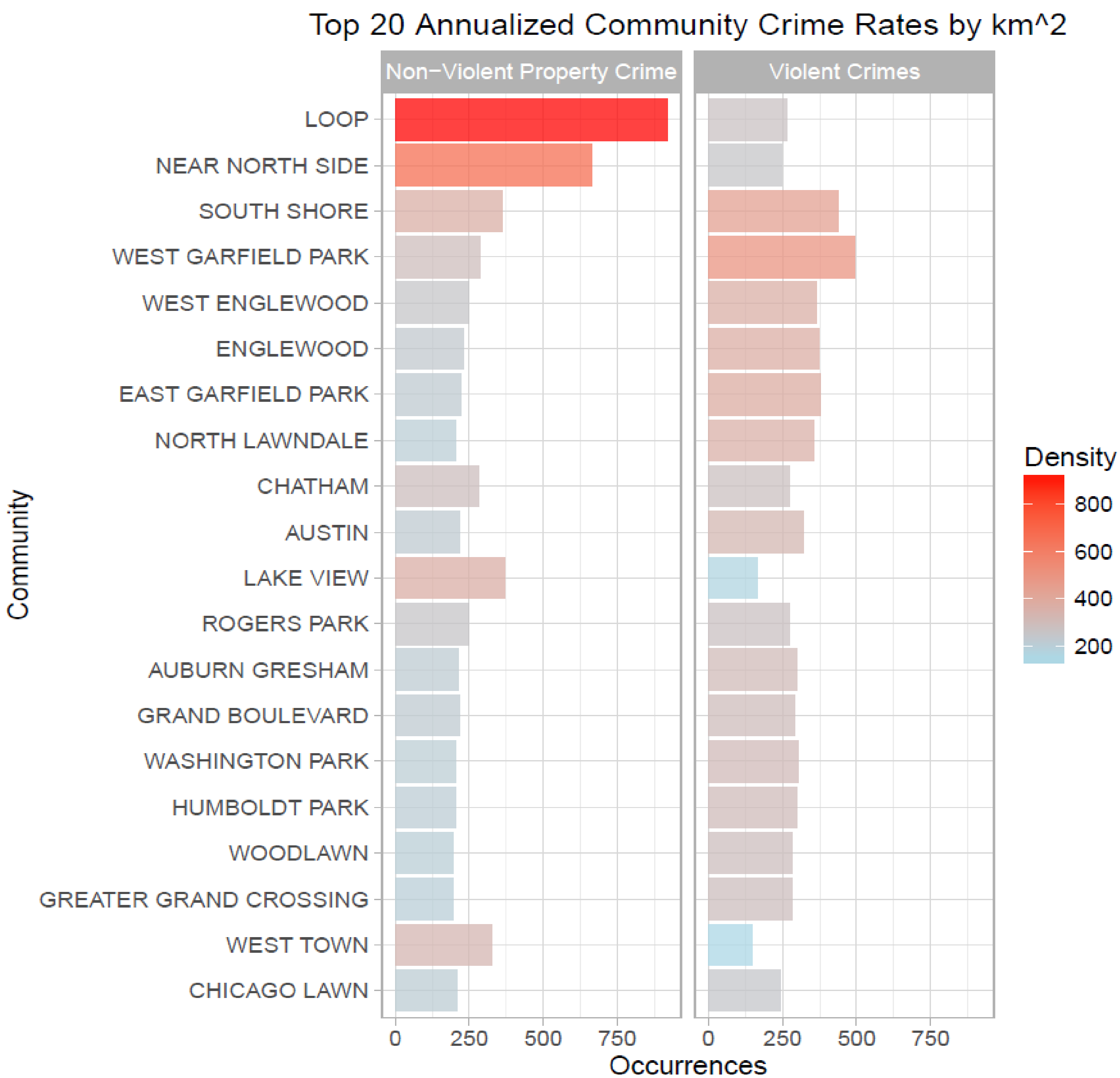
6.1 Density Choropleth

Given the structure and political influence of community organizations, it may prove equally useful to view crime rates and any resource deployment recommendations bounded by community borders. The following choropleth displays Chicago communities according to increasing crime rate as measured by annualized crime rate per square kilometer.



1This is an image of the interactive html file available via the project's Github folder

As an addendum to the above map, the following bar graph shows the 20 communities with the highest density ratings. It is ordered by a community's overall crime density and then split out to show Non-Violent Property Crime separate from Violent Crimes.

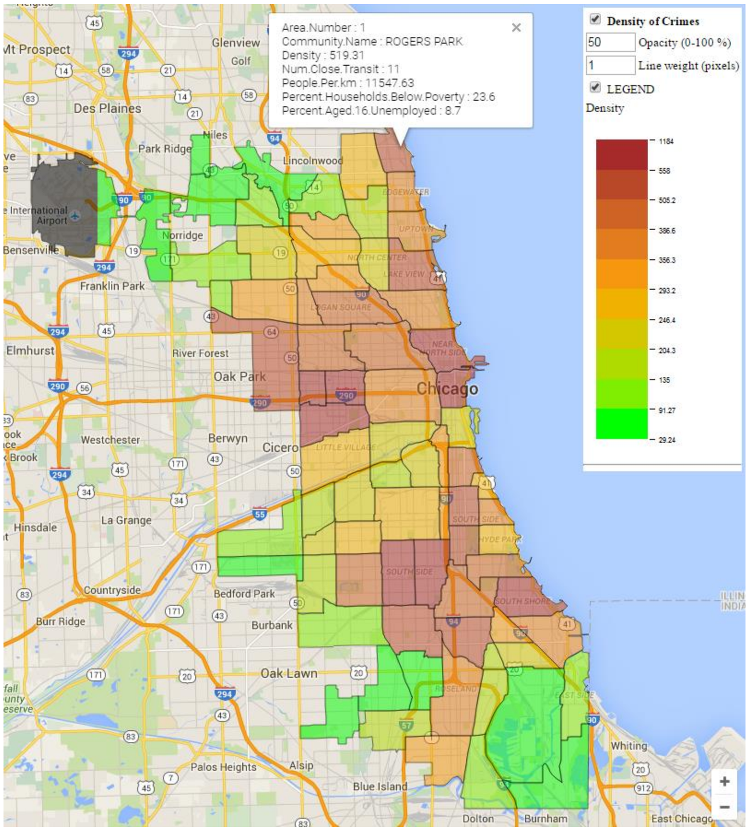


Some communities, like the Loop and the Near North Side, are very crime dense and have a markedly one-sided make-up of crime that traditional demographic factors would not have predicted. Both of these communities in particular are relatively affluent with per capita incomes of $65,526 and $88,669, respectively; however, they have a very large number of accessible mass transit locations within their communities. Per the section on community demographics below, mass transit locations may be a better predictor of crime rates than traditional demographic factors.

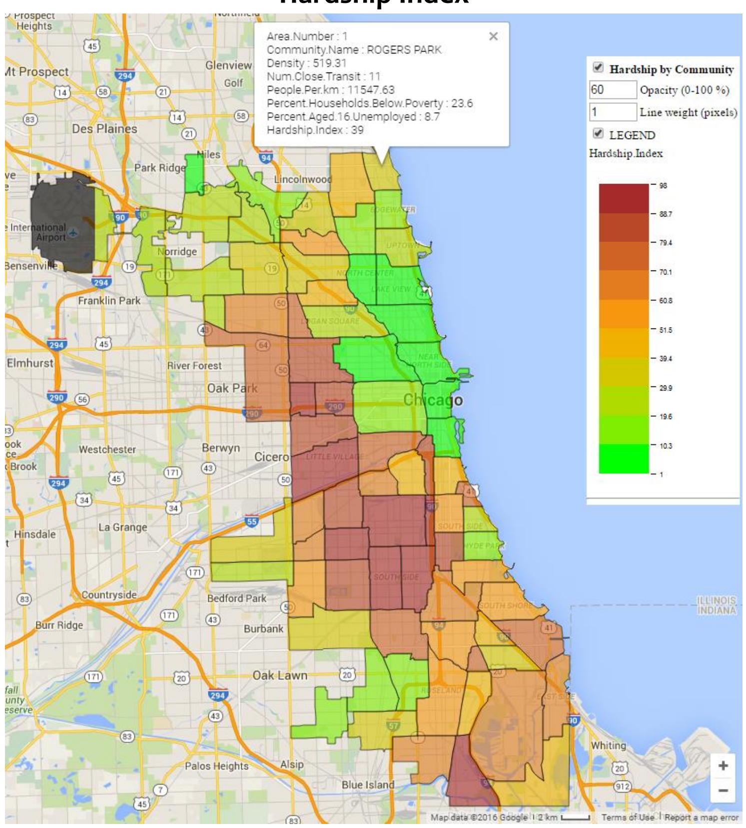
Community	Per Capita Income	Number of Closely Accessible Transit Stops
Loop	$65,526	41
Near North Side	$88,669	22
Median	$20,956	5

Using another metric, one could contrast crime density by community to the same map colored by community hardship. [See the section on data for information regarding the hardship index.] Note how hardship is not a tremendous predictor of crime. See the next section for a deeper dive on community demographic factors.

Crime Density



Hardship Index

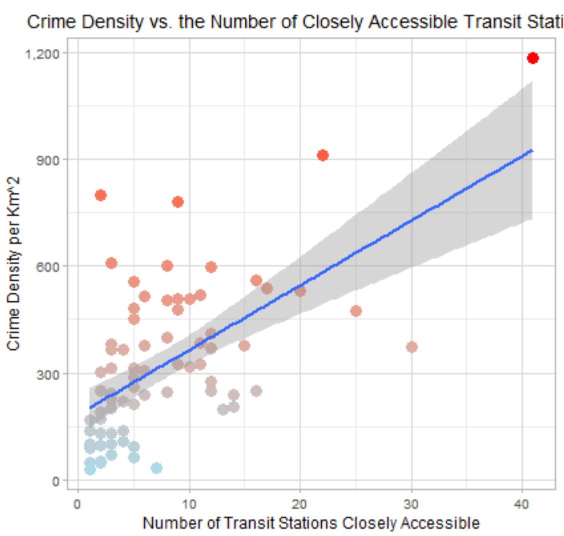


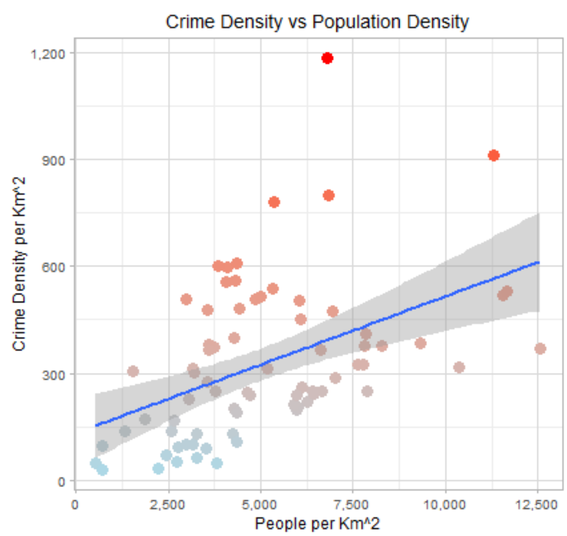
6.2 Demographics and Other Community Level Characteristics

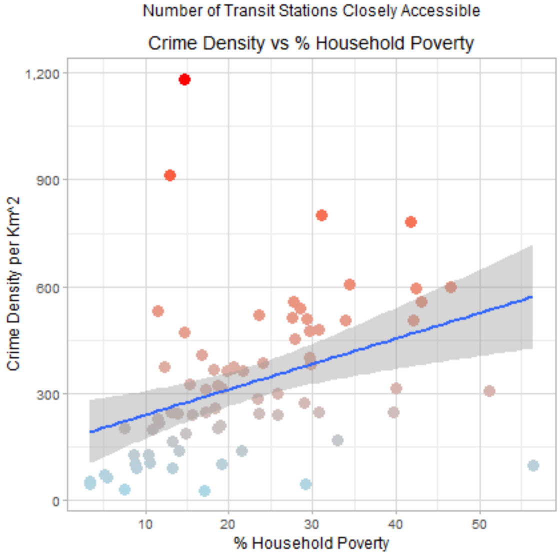
There are a number of demographic factors typically thought to cause higher crime rates. These include increased population density, lower relative income, unemployment, and poor or incomplete education. This analysis explored each of these metrics at the community level. To summarize findings, only 3 of the chosen demographic factors were significantly correlated to a community's spatial crime rate, and the number of transit stations accessible to a given community (as a proxy for individual crime proximity) is more highly correlated with spatial crime rates than any of the demographic variables. See the table and graphical correlation matrix below.

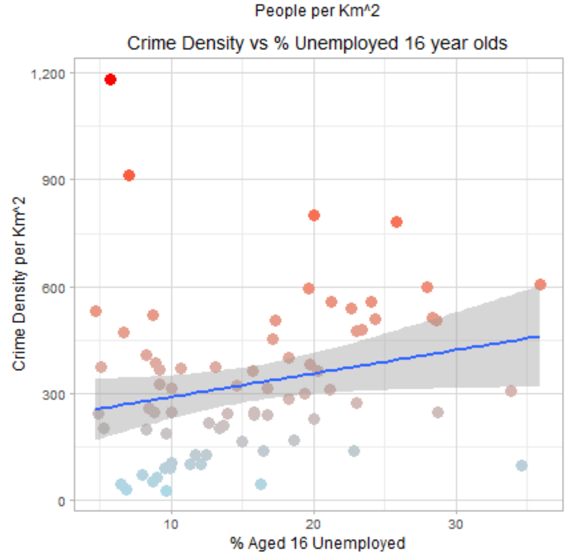
Community Characteristic	Correlation to Crime Density	P-value
# of Transit Stations Closely Accessible	0.598	0.000000019
Community Population Density	0.459	0.000039
% Households Below Poverty Level	0.386	0.00069
% Aged 16 Unemployed	0.231	0.048
Per Capita Income	0.171	0.15
Hardship Index	0.121	0.30
% Aged 25 Without High School Diploma	-0.074	0.53
% of Crowded Housing	0.042	0.72

The following graphs illustrate these relationships in order of decreasing correlation and significance:

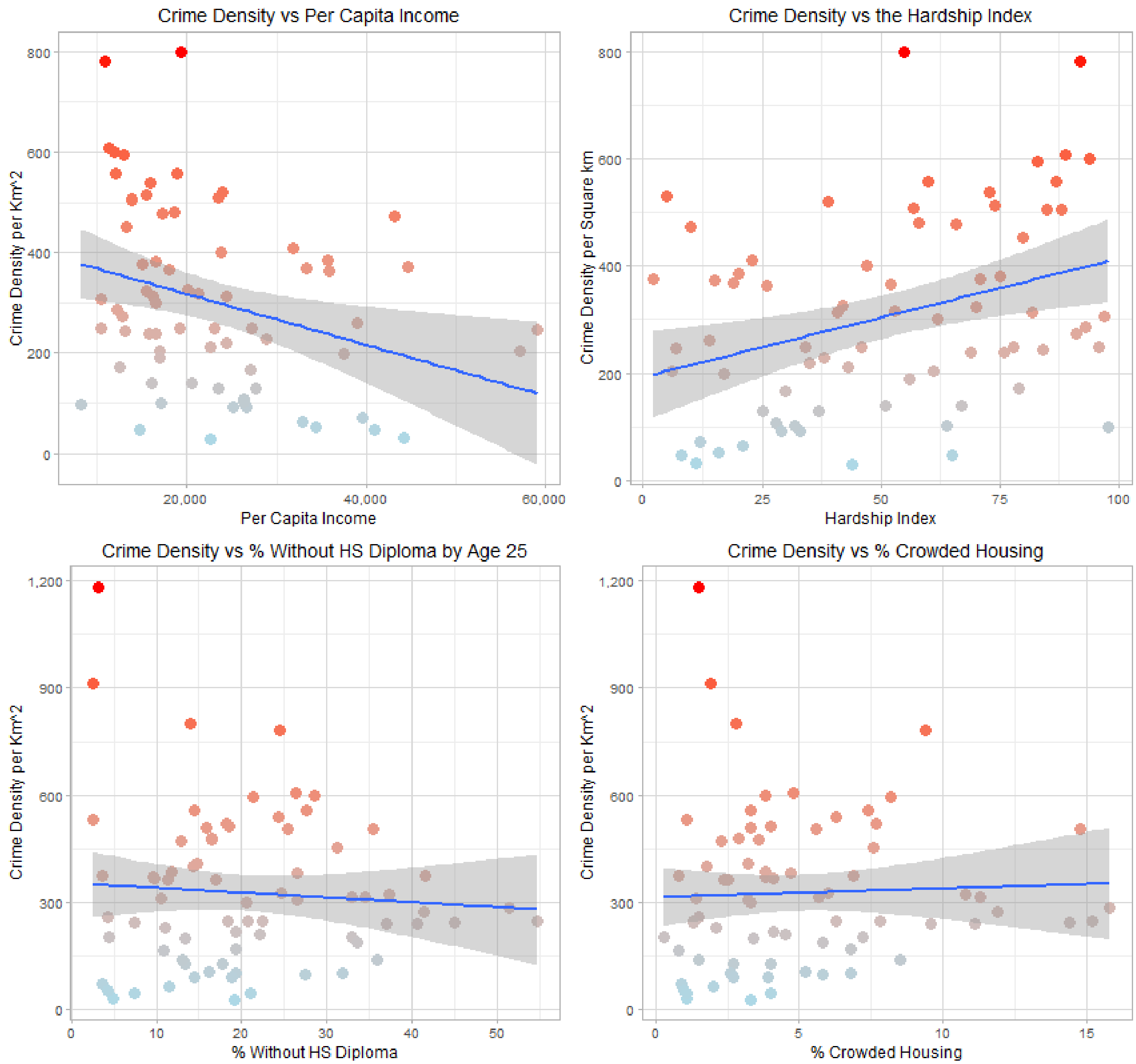








Demographic Factors (Continued)

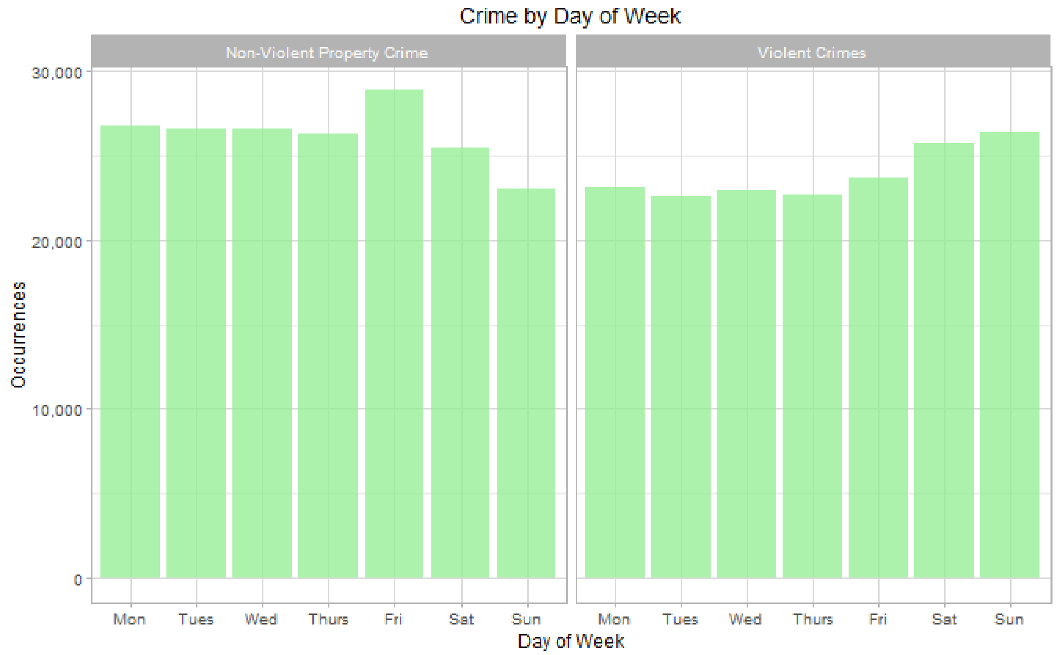


There is some co-linearity between "% Households Below the Poverty Line" and "% Aged 16 Unemployed" [corr = 0.81] as well as "#of Transit Stations" and "Population Density" [corr = 0.48]. This will dictate choosing the more significant factor of each pair for regression modeling. I.e. "% Households Below the Poverty Line" and "#of Transit Stations." See the modeling section below.

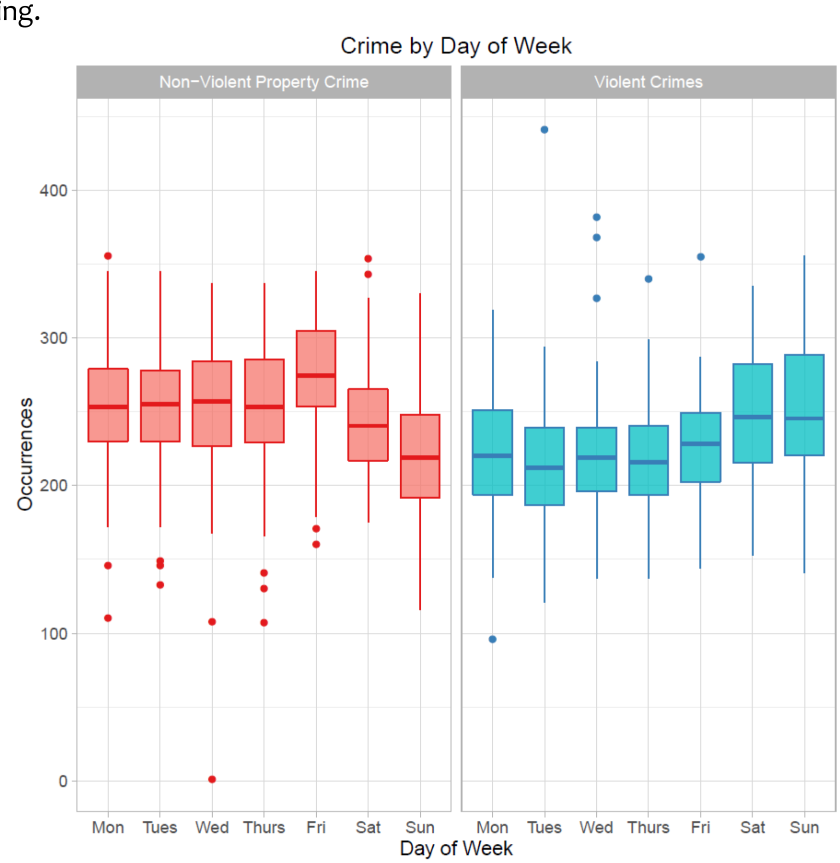
7 Temporal Analysis

7.1 Distribution by Day of Week

Beyond the spatial characteristics of crime, the city must efficiently deploy its resource on a daily and even hourly basis. As such, it would also be useful to examine the temporal characteristics of Chicago crime. While there is some variance within meta-categories, it does not appear that overall crime rates differ substantially by day of week.

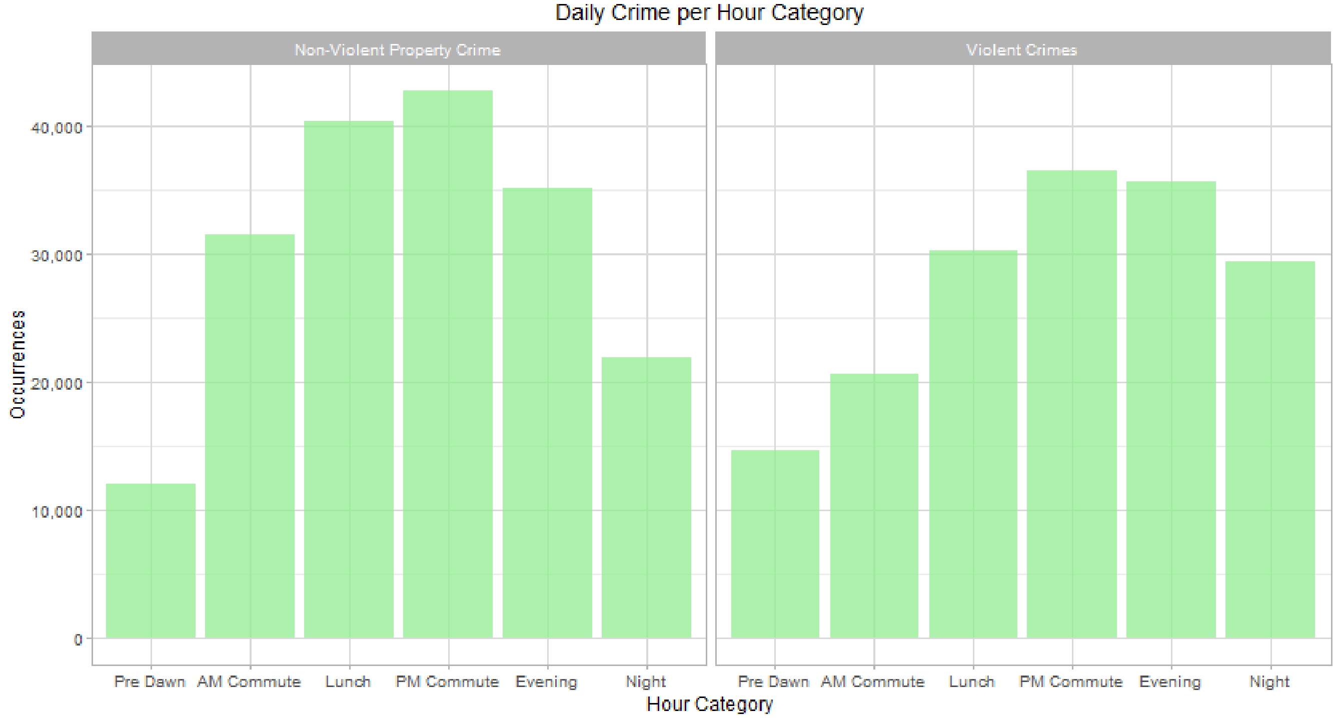


The following boxplots show greater detail. It is interesting to note that property crimes fall on Saturday and Sunday while violent crimes rise. If enforcement teams are organized by type of crime, it may be worth factoring this into resource planning.

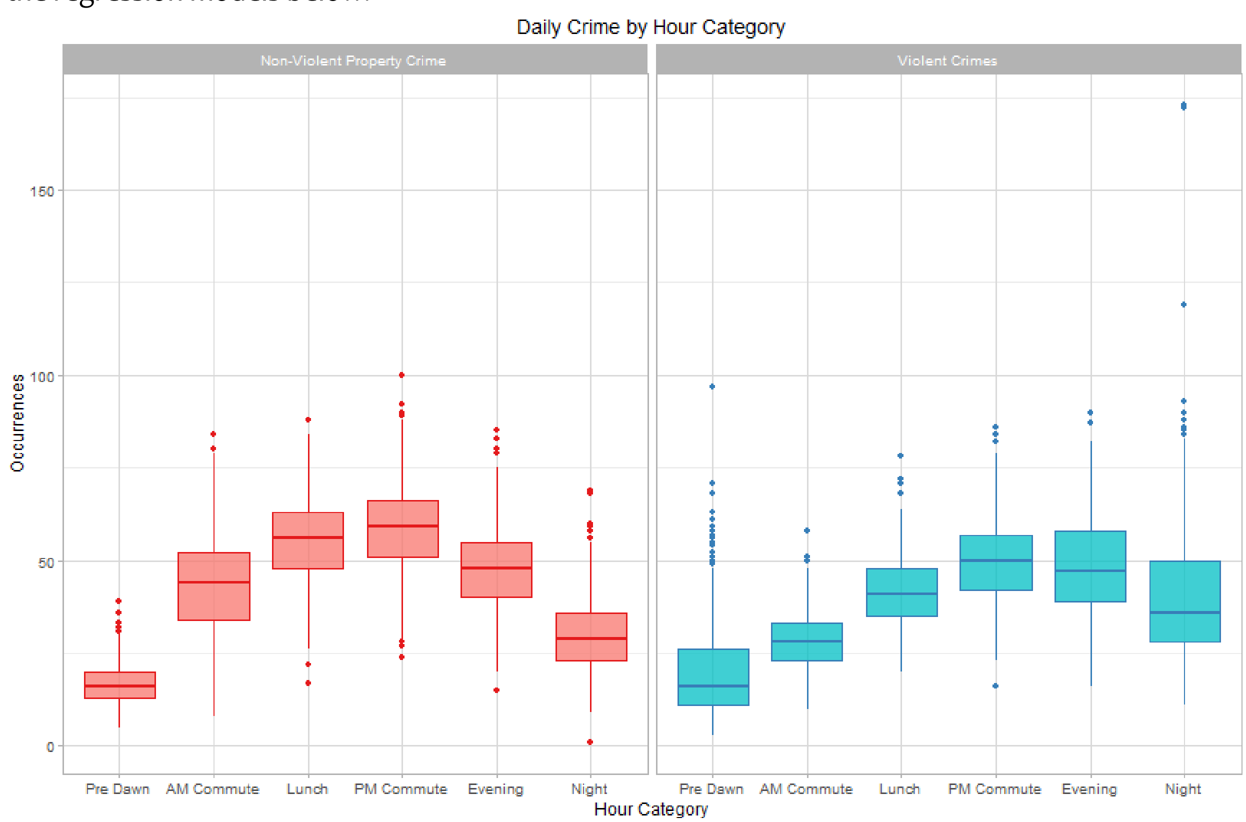


7.2 Distribution by Hour

Unlike day of week, there are large variances in crime frequency between hour categories:



Again, the following boxplots show greater detail. Note how property crimes peak over lunch and the PM commute, while violent crimes are more common in the evening. This distribution is statistically significant and will be explored more in the regression models below.



8 Modeling

The following model examines the relationship between a given community's crime density and those factors from the above analysis that were most significant after taking co-linearity into account. These are 'Hour Category,' the '# of Transit Stations in Close Proximity to a Given Community,' '% Households Below the Poverty Level,' and 'Minimum Distance to a transit Station' (as a proxy for access to any public transportation).

8.1 Model with Crime Density as Dependent Variable

Within a given year and hour category, a model based on these few factors can explain 51% of the annual variance in community spatial crime rates.

```
lm(formula = Density ~ Hour.Category + Num.Close.Transit + Percent.Households.Below.Poverty +
    Minimum.Transit.Dist, data = crime_by_community4)

Residuals:
   Min     1Q Median     3Q    Max
-68.22 -16.76  -1.79  13.92 229.05

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -21.06111    3.65400  -5.764 1.14e-08 ***
Hour.CategoryAM Commute           24.42560    3.51890   6.941 7.55e-12 ***
Hour.CategoryLunch                42.76953    3.51888  12.154  < 2e-16 ***
Hour.CategoryPM Commute           50.40680    3.51899  14.324  < 2e-16 ***
Hour.CategoryEvening              41.45542    3.51885  11.781  < 2e-16 ***
Hour.CategoryNight                22.28764    3.51884   6.334 3.81e-10 ***
Num.Close.Transit                  2.77874    0.15852  17.529  < 2e-16 ***
Percent.Households.Below.Poverty   1.19176    0.08804  13.536  < 2e-16 ***
Minimum.Transit.Dist              -2.81023    0.65185  -4.311 1.81e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30.27 on 879 degrees of freedom
  (55 observations deleted due to missingness)
Multiple R-squared:  0.5165, Adjusted R-squared:  0.5121
F-statistic: 117.4 on 8 and 879 DF,  p-value: < 2.2e-16
```

Each of the above factors is statistically significant. As one might expect given the correlation graphs in section 6, a community's transit characteristics appear to have a greater effect on crime density in Chicago than more traditional demographic predictors.

9 Conclusions

9.1 Summary of Findings

As an outgrowth of the baser human instincts, crime can be equated to a living and malignant organism. Similar to other parasites, there are specific factors that nourish its growth and structurally unique environments that it will prefer. While one could explore a myriad of additional factors to those appearing above, this analysis has clearly demonstrated several meaningful characteristics:

1. Using data from 2013-01-01 to 2014-12-31, crime rates do not decrease but actually increase as one gets closer to a police station.
    a. This evidence is not enough to invalidate the hypothesis that police presence is a deterrent. Given the intuitive nature of that assumption, the pattern may simply underscore the effectiveness of proactive policing. E.g. that patrols are effective enough to overcome the natural deterrent one would otherwise expect from increased police presence.
2. Crime rates increase exponentially as one approaches a transit station, and a community's transit characteristics may be more effective at predicting crime rates than traditional demographics.
    a. Crime rates increase more than 900% from a 2km radius to a smaller ¼ km radius around a given transit station. Most of that increase takes place within the kilometer nearest the station. As such, patrols should frequent these areas.
3. Spatial crime rates do not necessarily follow the traditional demographic factors in Chicago.
4. The distribution of crime does not vary dramatically by day of week, but it does by hour. Patrols should be organized accordingly

9.2 Infrastructure and Patrol Proposal

While it may or may not reduce crime rates, increasing police presence near transit stations will give police the best chance to affect crime in their areas. The city could accomplish this with frequent patrols tightly centered on transit stations or by setting up small, permanent outposts at their most affected stations. Either strategy could increase patrol efficacy while having a targeted effect on demonstrated hot spots. While permanent outposts would be the more capitally intense option, regular assignments to these posts could have the added benefit of improving community relations through personal familiarity. It would be worth running a few trials in the city's most crime dense communities to determine effect. Per the choropleth and bar graphs in the community section of this paper, a trial in West and East Garfield could be used to gauge effects on Violent Crimes while a trial in the Loop and the Near North Side could be used for Property Crimes. Trials in these areas should help isolate any effects that the strategy has on Violent Crimes vs. Non-Violent Property Crime.

9.2.1 Proposal for further analysis

As noted in the introduction, Chicago is "off to [its] deadliest start in nearly two decades." Homicides and shootings are up by more than 70% compared to 2015, while arrests and investigative stops have decreased. The reasons for these shifts are beyond the scope of this analysis. However, a time lapse analysis could shed some light on growth patterns. If the client wishes to move forward, I would suggest the following:

- Add additional transit station locations and volume estimates for each location. E.g. bus routes
- Add Police Station employee data
- Add additional community and/or demographic data
- Perform temporal analysis showing acceleration/deceleration of crime rates per community to understand trends.

9.3 Temporal (Schedule) Conclusions

Unless patrol teams are divided by type of crime, patrol schedules do not need to adjust for day of week volumes. There are, however, strong shifts in hourly volume and staff should be scheduled accordingly.

9.3.1 Proposal for further analysis

As a follow up to this analysis, I would suggest the following:

- Perform more in depth resource deployment analysis using to determine workforce requirements on an hourly basis per police district.
- Examine Holiday effects on crime.
- Expand prediction models to include other variables that could assist with scheduling decisions such as weather.