Preliminary Results

Adam Shelton

2/21/2020

Introduction

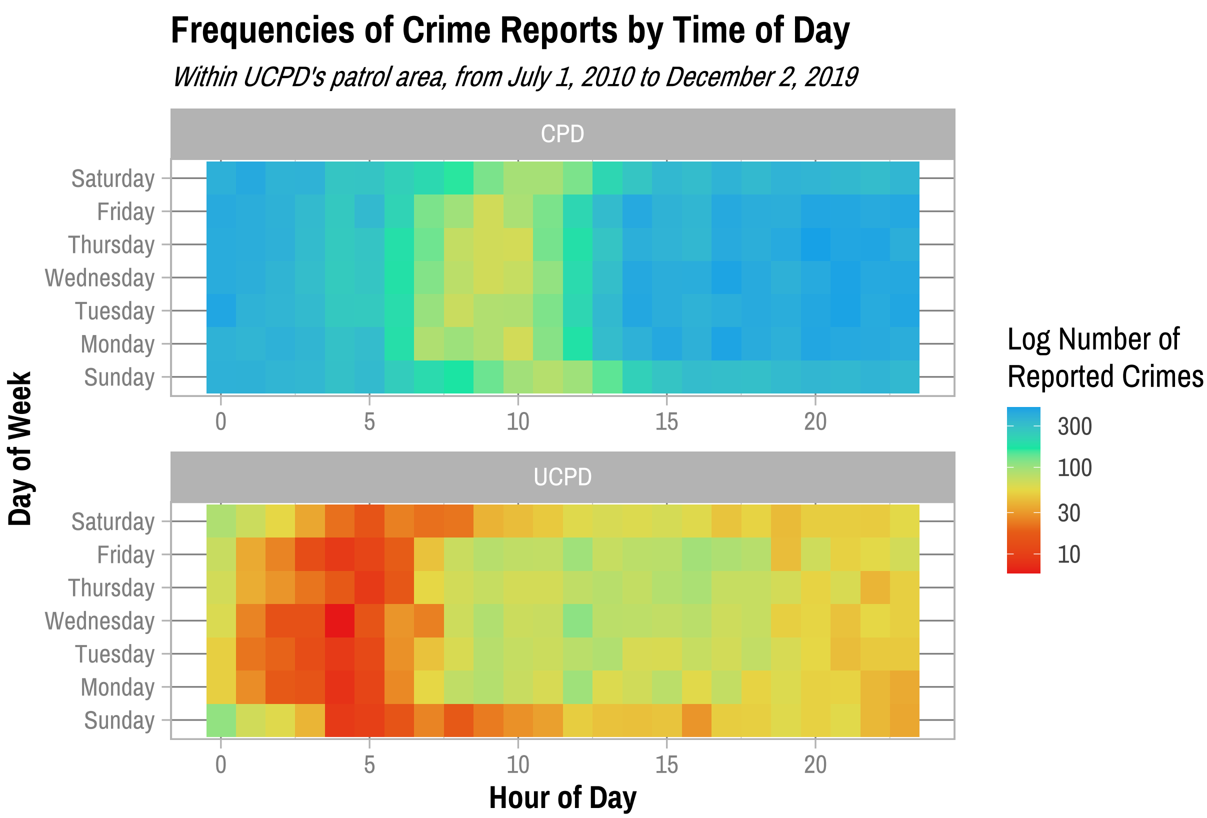
While the University of Chicago Police Department (UCPD) provides significant technical and human resources to police Hyde Park and the surrounding area, their jurisdiction is shared with the Chicago Police Department (CPD). This necessitates an understanding, not only of how the UCPD operates, but also of the CPD and the interaction of private police officers employed by the University with public police officers employed by the city. Therefore, in addition to understanding the reports of crimes that both UCPD and CPD receives and responds to, and their outcomes, this project aims to answer how CPD and UCPD crime reports and outcomes vary within the UCPD's jurisdiction, and directly outside of it. Additionally, as some reports of crimes are handled by both the UCPD and CPD in tandem, we also attempt to discern departmental differences from how each agency may handle or report the same "case" in a different manner.

This project combines data from three sources, the University of Chicago, the Chicago Police Department, and the United States Census Bureau. The University of Chicago publishes data on incidents responded to by the University of Chicago Police Department on the website for the university's department of Safety and Security. While this data is publicly view-able, the university does not provide an option to download the data in an easy to analyze format, so the data was web-scraped using the `rvest` package in R. Data on reported incidents published by the university and the city include time, location, descriptions, and outcomes of crimes reported to both agencies. However, as neither agencies publish demographic data of suspects or victims involved with a reported crime, this is inferred from the characteristics of the census block where the crime was reported to have happened. The University of Chicago publishes the UCPD's "area of patrol" via a PDF map online. This was copied by hand into a shape-file via Google Maps, to geocode whether a crime was reported to have happened in the area the UCPD regularly patrols.

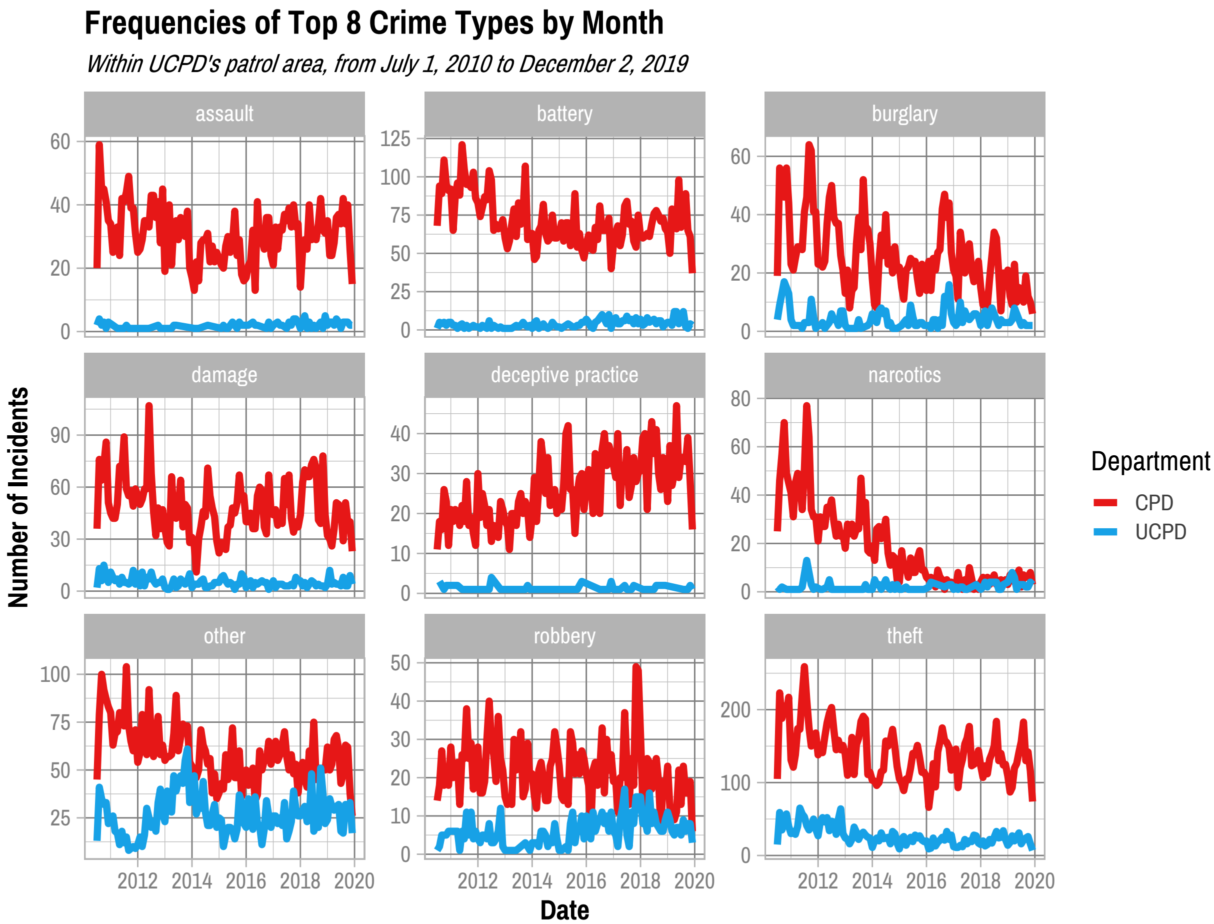
Data from the University of Chicago included incidents that the UCPD responded to that were not possible crimes, such as medical emergencies, which were removed from the data-set used for analysis. However incidents that could be related to a crime, such as "missing property" or "information [related to a possible crime]" were kept in the data-set for analysis. Categories were hand coded and collapsed into as few matching categories between data from the University and the city as much as possible, with any categories that only appeared in the data from one department re-categorized as "other".

Visual Analysis

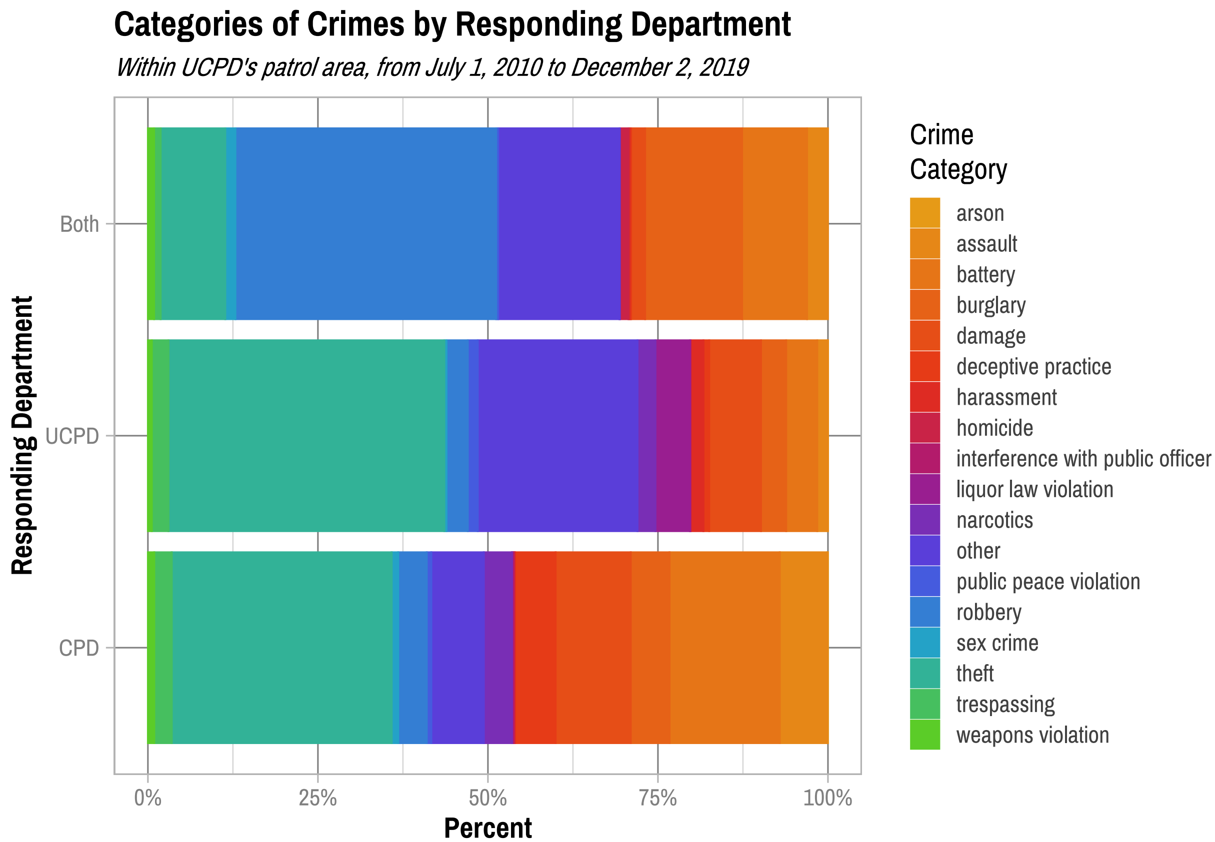
Especially in the case of geo-temporal data, such as crime reports, a visual analysis is an essential tool to study the data. Finding relationships and patterns in a data-frame of numbers is a much harder task for humans compared to performing the same task with numbers and shapes [citation needed]. Therefore, we broke down several dimensions of our crime report data into much easier to understand visualizations using the `ggplot2` package in R.



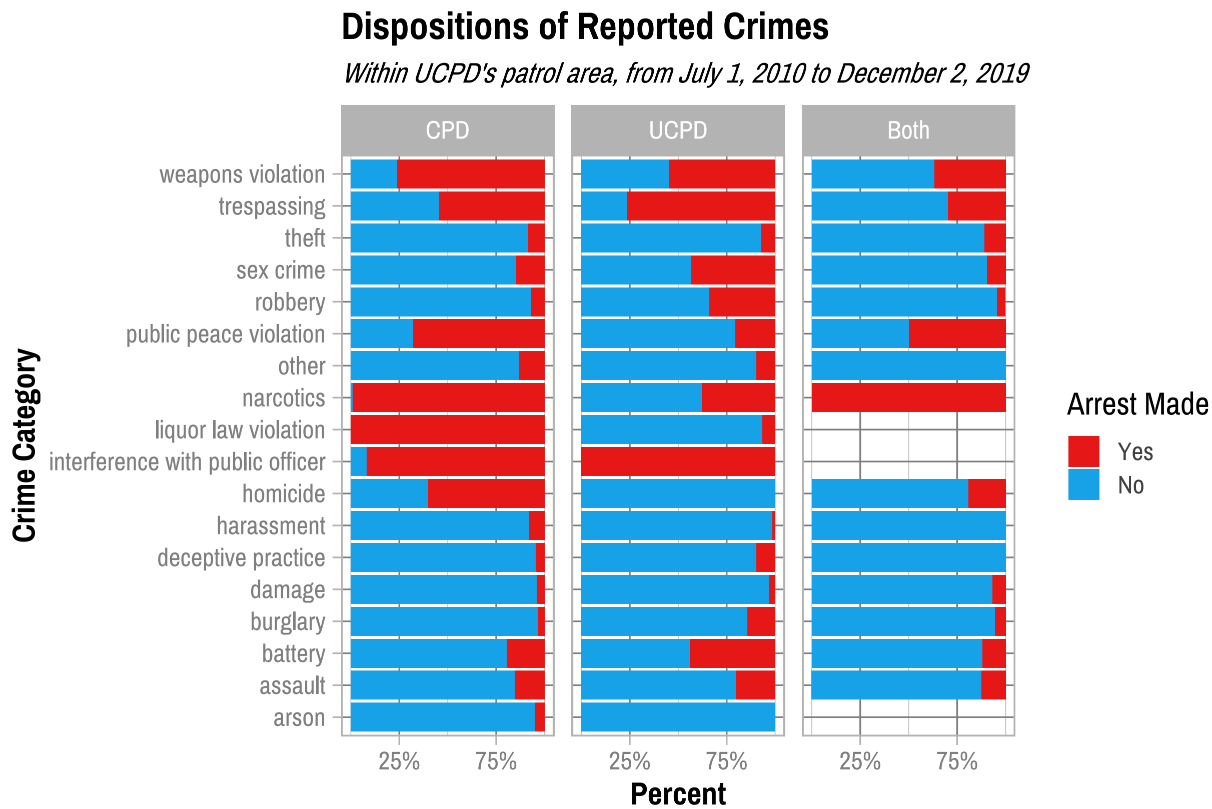
Crime frequently ebbs and flows temporally, with some relation to the natural day-night rhythms of people. Naturally overall crime are more likely to occur when more people are out and about, although certain types of crimes may contradict these patterns. For example, burglaries frequently happen at night, as the darkness and lack of people about gives cover to a criminal attempting to surpass theft prevention devices. For all crimes in UCPD's patrol area we see a distinct offset in the peak times of reported crimes, even though both departments are responding to the same exact geographical space. The lull in reported crimes for UCPD, while covering a similar interval, starts much earlier than for city police. On weekdays, this nightly lull starts around 1 AM and lasting until about 5 AM for the University, but goes from about 6 AM to 10 AM for CPD. This drop in reports is much shorter and later on the weekends for city officers, from about 10 AM to 12 PM, but only slightly later for UCPD from about 4 AM to 8 AM.

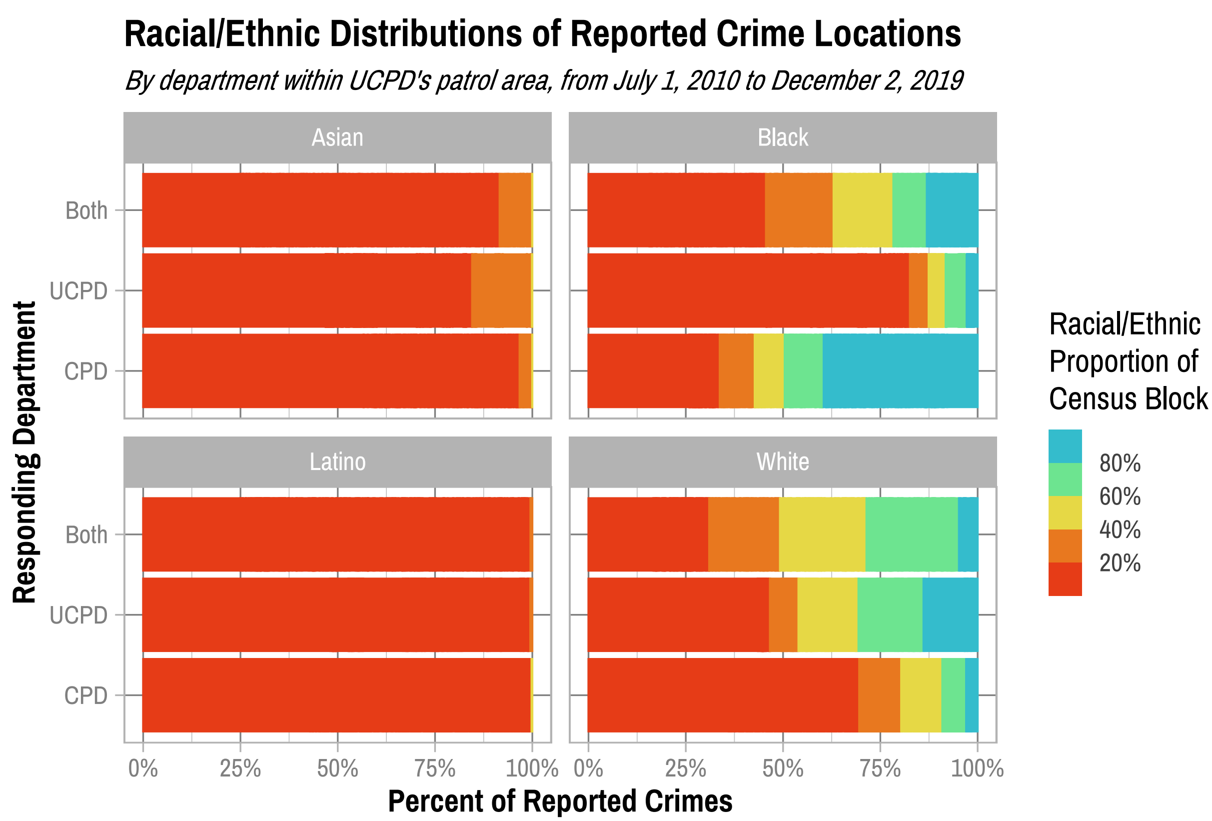


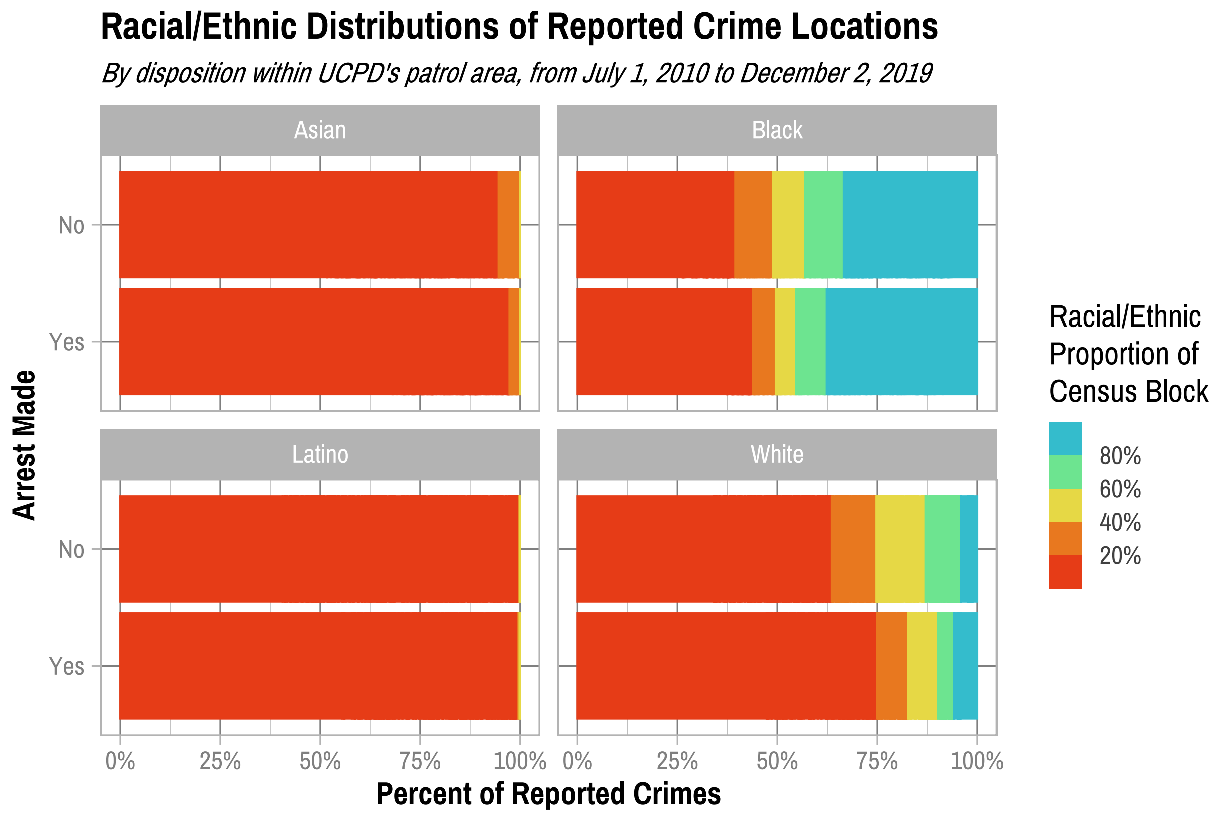
Over a larger temporal scale, crimes reported to the city show a stronger seasonal effect than those handled by the University. While most types of crimes seem to increase and decrease similarly for both departments, incidents involving narcotics or robbery appear to be more independent. While the CPD has seen a large steady decrease in narcotics related crimes, the UCPD has actually seen a slight uptick in recent years, with robberies following a similar but less divergent pattern. As expected, the University handles a much smaller volume of reported crimes across the board than CPD does in the same area.



Within the same area, departments proportionally respond to different types of crimes with varying frequencies. The CPD alone responds to a larger proportion of violent crimes, such as burglary, battery, and assault in contrast to University officers. The UCPD alone responds to a large proportion of thefts and "other" crimes, the latter of which is likely more an artifact of the data collection process than the incidents themselves. Interestingly, despite each department alone responding to a similar proportion of thefts and robberies, when both departments respond, it is overwhelmingly likely to be a robbery instead of a theft, in stark contrast to the reports each department receives on their own.

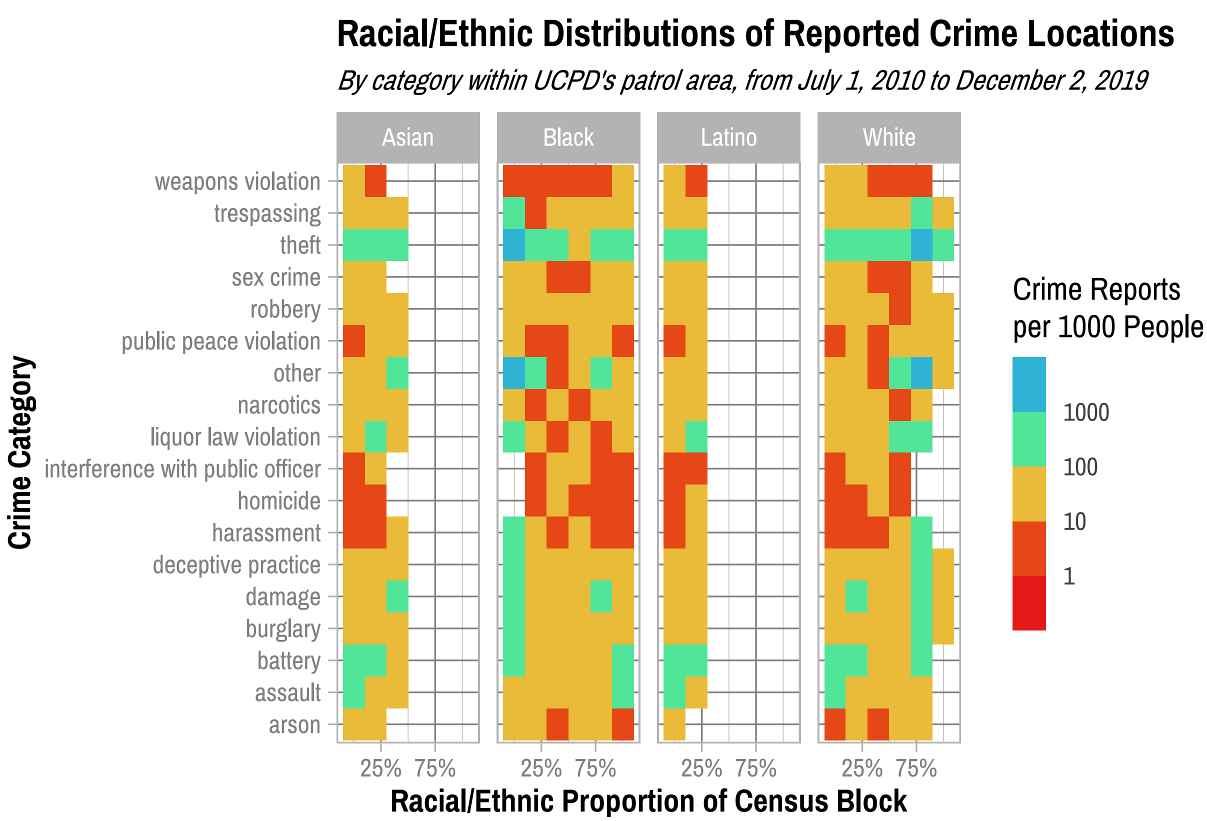

Dispositions of Reported Crimes
Within UCPD's patrol area, from July 1, 2010 to December 2, 2019
CPD
UCPD
Both
weapons violation
trespassing
theft
sex crime
robbery
public peace violation
other
narcotics
liquor law violation
interference with public officer
homicide
harassment
deceptive practice
damage
burglary
battery
assault
arson
Crime Category
25%
75%
Percent
Arrest Made
Yes
No

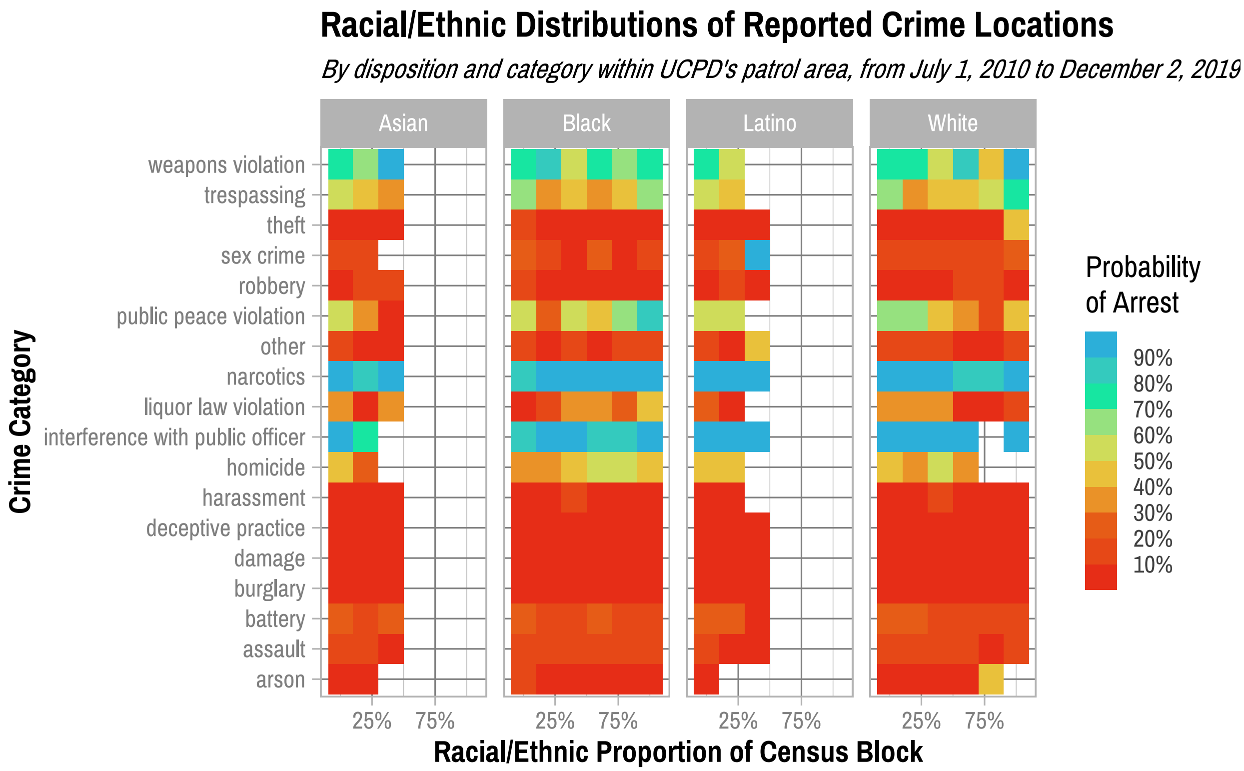

Racial/Ethnic Distributions of Reported Crime Locations
By department within UCPD's patrol area, from July 1, 2010 to December 2, 2019
Asian
Black
Latino
White
Both
UCPD
CPD
Responding Department
0%
25%
50%
75%
100%
Percent of Reported Crimes
Racial/Ethnic Proportion of Census Block
80%
60%
40%
20%


Racial/Ethnic Distributions of Reported Crime Locations
By disposition within UCPD's patrol area, from July 1, 2010 to December 2, 2019
Asian
Black
Latino
White
Arrest Made
No
Yes
0%
25%
50%
75%
100%
Percent of Reported Crimes
Racial/Ethnic Proportion of Census Block
80%
60%
40%
20%


Racial/Ethnic Distributions of Reported Crime Locations
By disposition and department within UCPD's patrol area, from July 1, 2010 to December 2, 2019
Asian
Black
Latino
White
CPD
UCPD
Both
Arrest Made
No
Yes
25%
75%
Percent of Reported Crimes
Racial/Ethnic Proportion of Census Block
80%
60%
40%
20%


Racial/Ethnic Distributions of Reported Crime Locations
By disposition and department within UCPD's patrol area, from July 1, 2010 to December 2, 2019
Asian
Black
Latino
White
Responding Department
Both
UCPD
CPD
0%
25%
50%
75%
100%
Racial/Ethnic Proportion of Census Block
Probability of Arrest
90%
80%
70%
60%
50%
40%
30%
20%
10%

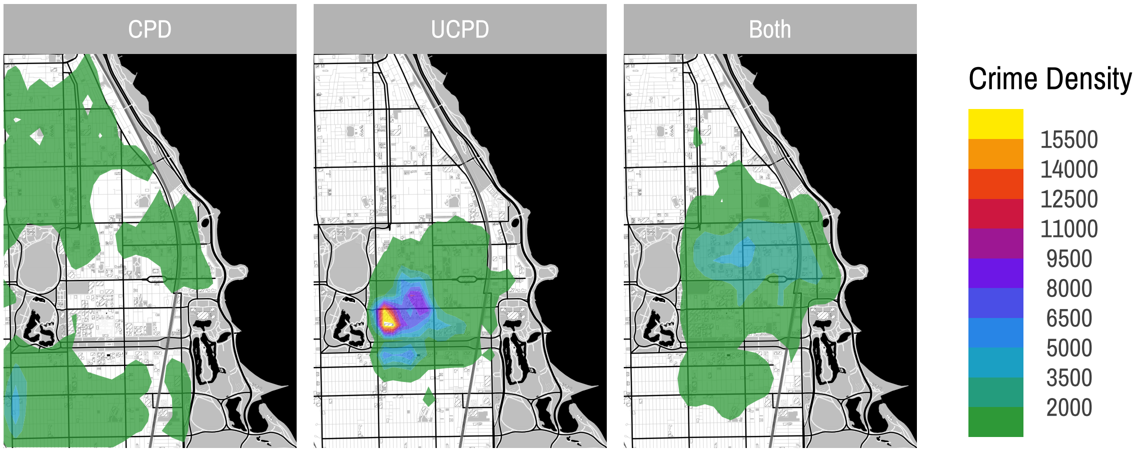

Racial/Ethnic Distributions of Reported Crime Locations
By category within UCPD's patrol area, from July 1, 2010 to December 2, 2019
Asian
Black
Latino
White
Crime Category
weapons violation
trespassing
theft
sex crime
robbery
public peace violation
other
narcotics
liquor law violation
interference with public officer
homicide
harassment
deceptive practice
damage
burglary
battery
assault
arson
25%
75%
Racial/Ethnic Proportion of Census Block
Crime Reports per 1000 People
1000
100
10
1

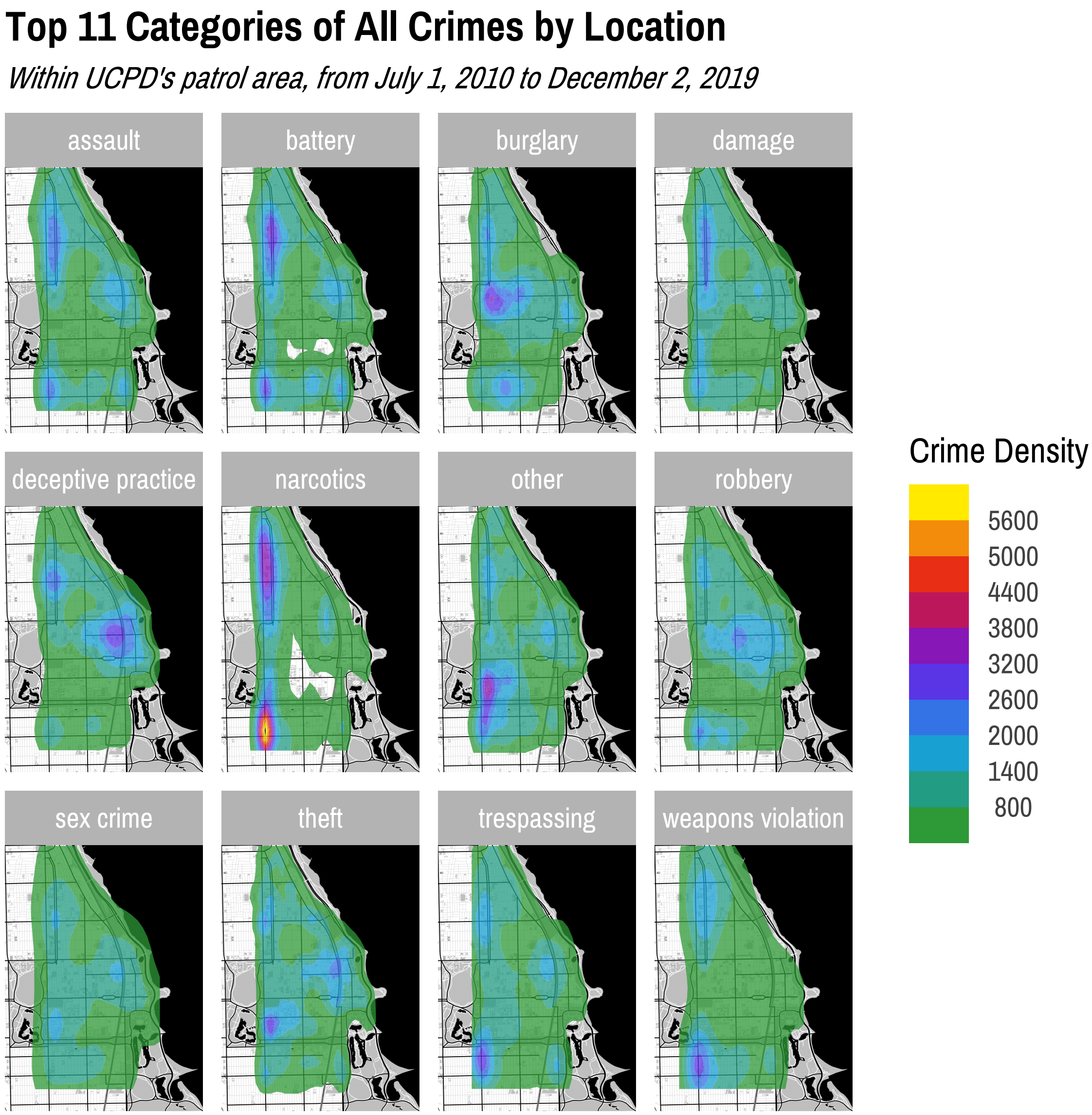


Overall, both departments respond very similarly to crimes, when it comes to making an arrest. Thefts, harassment, damage to property, burglary, and deceptive practice, all have similarly low arrest rates among both departments, while interference with a public officer is very likely to end in arrest regardless of who responds. However, the UCPD appears to be much more lenient with crimes involving, substance abuse, as for both narcotics crimes and liquor law violations, the CPD has quite a high arrest rate, while University officers are much less likely to perform an arrest. However, the UCPD does have a slightly greater tendency to make an arrest for trespassing and battery than officers from the city.

Location of Crimes by Responding Department

For all crimes, from July 1, 2010 to December 2, 2019





As we might expect, responses involving just University officers are tightly confined to campus, while those involving city officers typically do not. However, reported crimes where both departments responded are

Table 1: Frequencies of Observations by Responding Department

responding_dept	n
both	2083
cpd	53809
ucpd	7328

Table 2: Frequencies of Observations by Arrest

arrest	n
FALSE	53152
TRUE	10068

spread evenly across the entirety of the UCPD's patrol area. Similarly most types of crime are reported to have occurred all across the UCPD's patrol area. Reports of deceptive practice occur with much higher density around "downtown" Hyde Park on the east end of 53rd street, with burglaries happening on the west end of 53rd by Washington Park. Violent crimes and narcotics offenses are commonly reported north and south of Hyde Park, concentrated along Cottage Grove Avenue. Overall, the campus seems to have much lower reported crime rates than the surrounding area, especially with reports of violent crimes.

Modeling Process

Due to the large number of variables in the data-set, a modeling method with built-in feature selection was essential. Lasso regressions, and more recently, Random Forest models have proved very effective for feature selection (Tibshirani 1996; Bühlmann and Zürich 2006) on top of modeling relationships in crime report data (Kadar et al. 2018; Chalfin et al. 2019). Their popularity not only stems from their performance, but also their ease of use and compatibility with a wide selection of different categorical outcomes. However, there can be variance between models fit on the same data, which for significant robustness, necessitates the use of boot-strapping with lasso regression models (A Alves, Ribeiro, and Rodrigues 2018). Similarly, while lasso regression models perform well with feature selection, other linear-based models such as General Additive Models (GAMs) are superior for predictive performance (Kadar et al. 2018). For this initial analysis, only single lasso regression models were built and interpreted, but boot-strapping, Random Forests, and GAMs are all additional components planned for the final paper.

The lasso models were built using the `glmnet` package in R, specifically using the `cv.glmnet` function which performs k-fold cross-validation to determine the optimal lambda value. The default value of 10 folds for cross validation were used. For binary outcomes logistic lasso regression was used, with cross-validation optimizing on the area under the receiver operator characteristic (ROC) curve, while for multiple-categorical outcomes multinomial lasso regression was used, optimizing on the miss-classification rate. Data was partitioned into a 75/25 training/testing split using the `caret` package.

Initial Results

Arrests

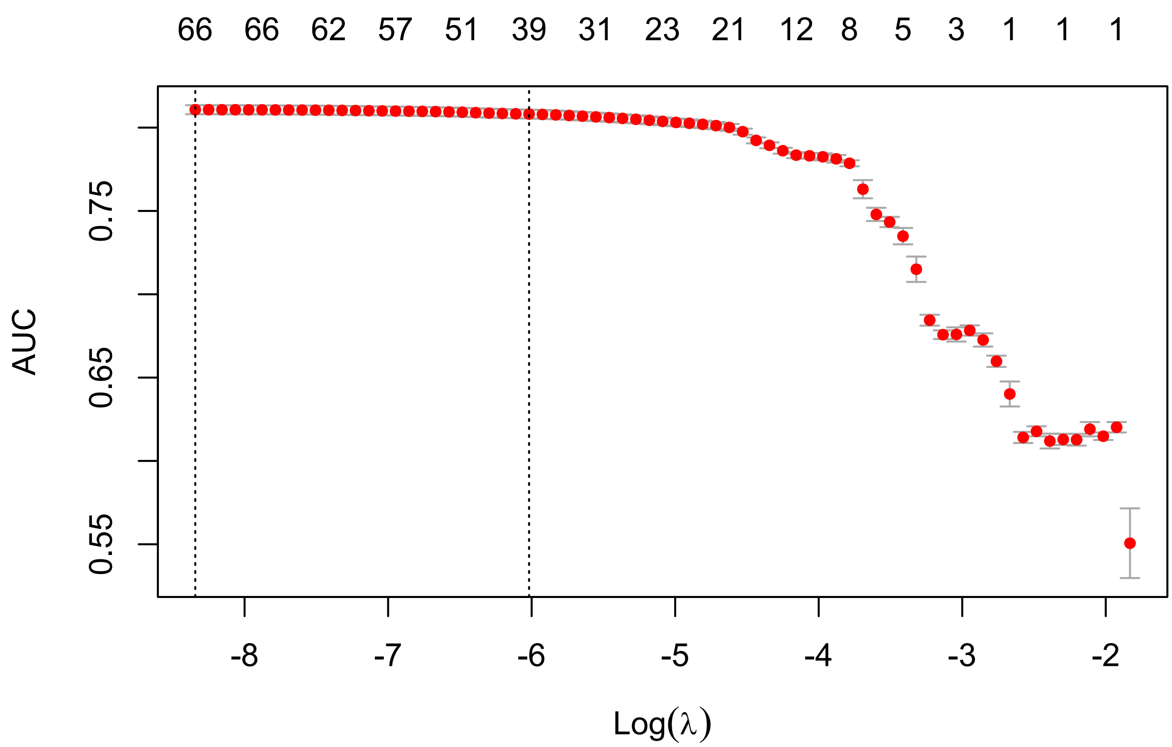
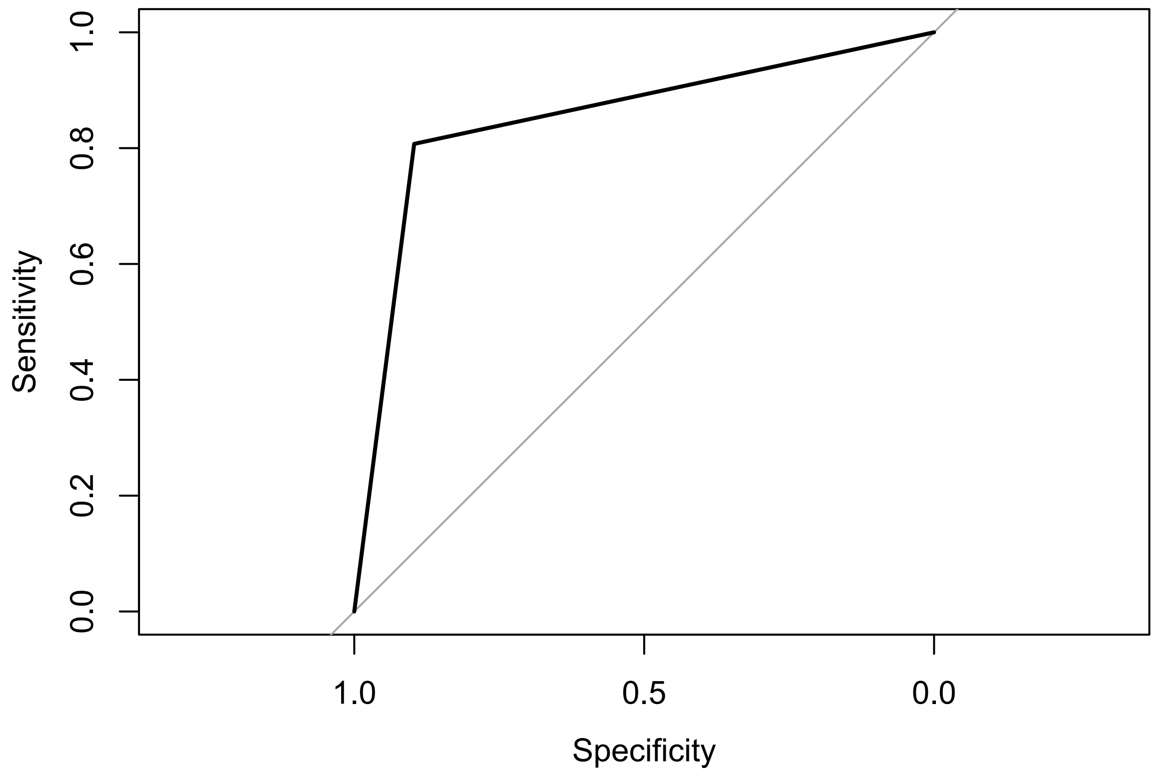


Table 3: Arrests Model Coefficients

Variable	**Est.**
(Intercept)	144.792
responding_deptcpd	0.019
date	0.000
primary_typeassault	0.215
primary_typebattery	0.602
primary_typeburglary	-0.768
primary_typedamage	-0.919
primary_typedeceptive practice	-0.693
primary_typeharassment	-0.363
primary_typehomicide	1.582
primary_typeinterference with public officer	3.590
primary_typeliquor law violation	0.133
primary_typenarcotics	4.638
primary_typepublic peace violation	2.099
primary_typerobbery	-0.127
primary_typesex crime	0.096
primary_typetheft	-0.363
primary_typetrespassing	2.233
primary_typeweapons violation	2.846
armedTRUE	-0.296
aggravatedTRUE	0.298
time	0.004
monthmar	0.088
monthsep	0.001
monthnov	-0.002
hu_occupied	-0.381
hu_owned	-0.786
hu_owned_loan	-0.186
total_pop	-0.001
pop_asian	-1.496
pop_family_hh	0.757
pop_husb_wife_fam	2.075
pop_latino_black	-0.345
pop_latino_other_race	0.788
pop_latino_white	-0.860
pop_male	-0.072
pop_one_pers_hh	-0.064
pop_other_race	1.079
pop_two_plus_races	-1.497
lat_y	-3.466

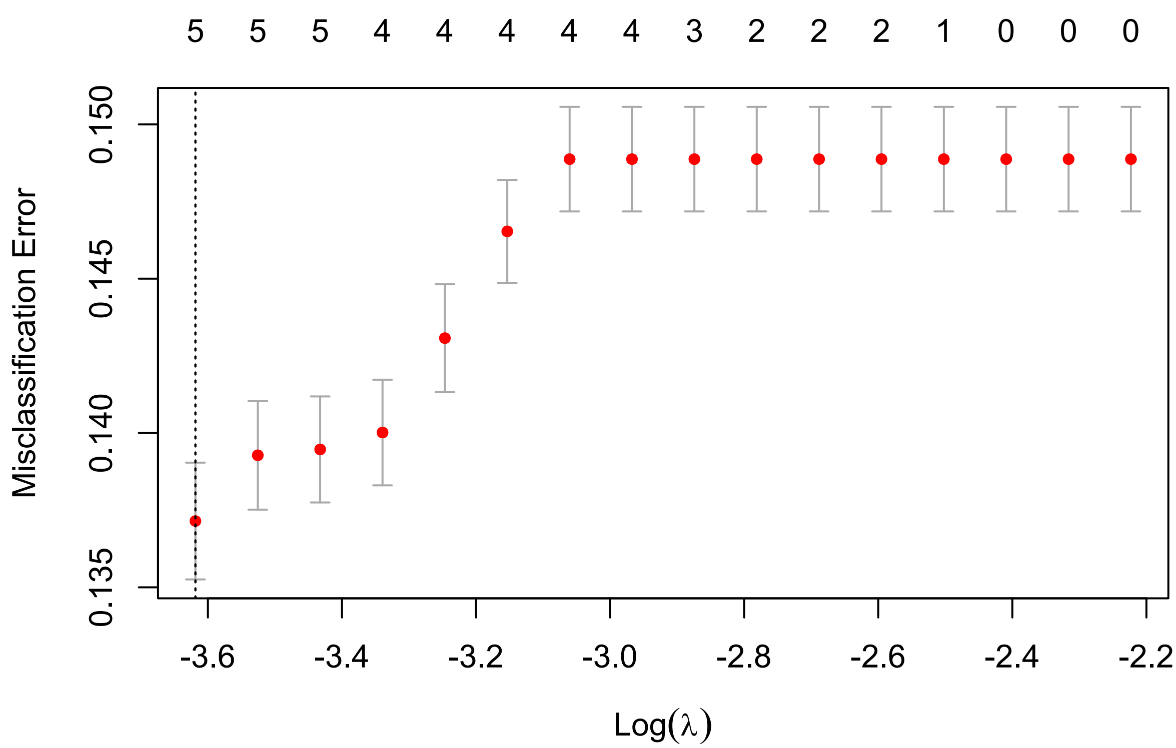


This model performed very well, with an AUC of 0.852. As expected, many of they types of crimes that had high arrest rates such as narcotics crimes and interference with a public officer result in a large increase in the log-odds of a reported incident ending in arrest. A higher proportion of people who identified as an "other race" or having a two family household (husband and wife) on the US census where the crime was reported is also associated with a large increase in the outcome. A higher proportion of people identifying as two or more races, white Latino, Asian, and owned housing units were associated with a lower log-odds of a reported crime ending in arrest.

Table 4: Responding Department Model Coefficients

Variable	**both**	**cpd**	**ucpd**
(Intercept)	657.819587743414	-206.253990835094	-451.56559690832
primary_typeliquor law violation			1.339122659418
primary_typeother		-0.793419006916074	
primary_typerobbery	1.59777155420003		
domesticTRUE		0.333032741816903	
avg_family_size			-0.0644496178418741
pop_asian		-3.04264350234911	
pop_black		0.334681428017972	
pop_family_hh			-2.65704403977373
pop_in_hus			-0.595858530891377
pop_white		-1.10109451661344	
lon_x			-12.6893799693881
lat_y		20.7635859355577	

Responding Department



While this model appears to perform well by its 86.219% classification accuracy rate, this does not show the whole story. Due to the class imbalance, the model is performing well only when predicting incidents responded to by CPD, not those handled by UCPD or both departments. This necessitates a different modeling approach to obtain better prediction performance, perhaps by creating two logistic models that can be cross validated using the ROC AUC. However, while this model is ultimately inconclusive due to its poor performance, it does suggest that there are strong geographical and racial components to the department receiving and responding to a reported crime.

Discussion

Despite the need to build a more diverse group of robust models for this data, with the work completed so far, distinct differences in how the University of Chicago Police Department and Chicago Police Department react and interact are already becoming more clear. The culture of the University appears to impact the policing culture of the UCPD, like the times at which they respond to reported crimes, and their leniency on crimes of substance abuse, that are common at colleges across the country. Despite operating in the same geographical space, and even receiving much of the same training (Sherman 2019), these departments clearly operate in different manners, seemingly because of the different stakeholders they represent.

References

A Alves, Luiz G, Haroldo V Ribeiro, and Francisco A Rodrigues. 2018. "Crime prediction through urban metrics and statistical learning." http://arxiv.org/abs/1712.03834v2.

Bühlmann, Peter, and Eth Zürich. 2006. "HIGH-DIMENSIONAL GRAPHS AND VARIABLE SELECTION WITH THE LASSO." *The Annals of Statistics* 34 (3): 1436–62. https://doi.org/10.1214/009053606000000281.

Chalfin, Aaron, Benjamin Hansen, Jason Lerner, Lucie Parker, Roseanna Ander, Robert Apel, Monica Bhatt, et al. 2019. "Reducing Crime Through Environmental Design: Evidence from a Randomized Experiment of Street Lighting in New York City."

Kadar, Cristina, Neue Zürcher Zeitung, Irena Pletikosa, Eth Zurich, Ckadar@ethz Ch José, Iria Mobiliar, and Irena Pletikosa Cvijikj. 2018. "Exploring Foursquare-derived features for crime prediction in New York City Crime Prediction View project Mobiliar Lab for Analytics View project Exploring Foursquare-derived features for crime prediction in New York City." https://doi.org/10.1145/1235.

Sherman, Stephen Averill. 2019. "From Revanchism to Inclusion: Institutional Forms of Planning and Police in Hyde Park, Chicago." *Journal of Planning Education and Research*, 0739456X1987768. https://doi.org/10.1177/0739456x19877683.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.