Is Predictive Policing Preventive in Crime?

Stephanie Moore, Luis Diaz, Julian Yeh

Table of Contents

Abstract

In our study we look at the effects of predictive policing systems and whether it reduces crime. In our paper we look specifically into Predpol, a predictive policing system that has been used by police departments around the United States and has spurred many controversies. We will investigate whether Predpol has a causal effect of reducing crime and if there is any bias in the system. Our study focuses on data from LAPD who has used Predpol since 2013. The results of our study will determine whether Predpol has an effect on crime and explore which crimes, areas, and people are most affected by the system.

Introduction

As many cities move toward using machine learning in their policing systems, we want to explore how these systems affect the people being policed. Several cities have implemented the use of predictive policing algorithms including New York City NY and Santa Cruz CA (7). In New York City, the algorithm used sensor alerting (gaudio gunshot detectors, television cameras, and license plate readers (LPRs)) and pattern recognition to alert officers to potential criminal activity and allow for a more informed police response (7). In Santa Cruz, an earthquake forecasting algorithm was adapted to predict spikes in criminal activity, much like how scientists predict earthquake “aftershocks” (8). For both of these models, the goal is not to predict *who* will commit crimes, rather *where* there will be a higher chance of criminal activity. The same algorithm used in Santa Cruz was later adopted in Washington, South Carolina, Arizona, Tennessee, Illinois, and other California cities, including Los Angeles (8). The Los Angeles Police Department (LAPD) adapted this algorithm and called it PredPol, and had been using it for a few years with varying degrees of success (8).

These predicted policing models claim to have reduced crime and increased efficiency in police resources. They have led us to investigate the question on how predictive policing is being used and its effect on common police interactions such as crimes and traffic stops. Our goal is to analyze whether PredPol has a causal effect on reducing crime and assess possible flaws of the model. Although we are only looking at data from the LAPD and the results cannot necessarily generalize, the structure of the predictive policing algorithms being used is similar: optimizing police resources to *deter* crime before it happens.

Literature Review

The PredPol model uses the algorithm to identify neighborhoods where criminal activity is more likely to be higher. It does this by generating 10 “boxes” on a map every 12 hours, approximately 500 feet by 500 feet (11). These “hotspots” represented the places that had the highest average arrests previously, or “high risk” areas. The maps were also published publicly to try and deter crime even more (11). Officers were expected to patrol these areas during times of “slower” activity in their assigned areas or sometimes assigned to only patrol there. The goal was to predict vehicle related crime and direct police resources to those hotspots (10).
It is important to note that this task is not trivial, as there are many confounding factors to consider for criminology data. “For instance, some communities are more likely to call the cops than others, and some crimes are more likely to go unreported than others. Additionally, whether or not an officer actually makes an arrest (i.e. the outcome of the crime) is determined largely by the officer. In cities that have operated using a “broken windows" ideology—including New York, Los Angeles, Boston, and many others—police are explicitly encouraged to look for and harshly penalized petty crime that may go unnoticed in other neighborhoods”(2). And according to Venkatasubramanian, “ when a tool like PredPol tells police where to go, crime data starts to be affected by PredPol itself, creating a self-reinforcing feedback loop” (3). The paper suggests that these systems switch from supervised machine learning to a reinforcement learning system to account for the action affecting the outcome.

The implementation of predictive policing has led to many questions and studies related to measuring the bias and “success” of these algorithms, specifically if they are more reflective of police actions rather than actual reported crimes. A study done by UCLA in 2015 claimed that Predpol reduced crime by showing that in the areas that Predpol determined were of high criminal activity risk, less crime was committed (4). However, these studies don’t take into account the effect on the areas from where officers were reassigned from. It is possible that crime could have been reduced in the target area but increased in the reassigned area. Additionally, many of the studies did not seem to study the datasets that were actually used in predictive policing algorithms. They stated conceptual flaws in the algorithms but didn't analyze the data for feedback loops over the years or look for bias in the models . Our project is necessary because we will not only look at how crime and stop rates changed in areas that had

Predpol deployed and the areas that those officers came from, but also assess the possible self- reinforcing feedback loop PredPol creates.

Data

The study aims to assess officer-controlled interactions such as traffic stops, as well as non-controlled interactions like reported crimes. The former allowed us to look at how being reassigned may affect an officer's behavior when initiating a stop. The latter is taken into account since we wanted to look at how the increase (or decrease) in police presence might affect criminal behavior. In order to deduce whether PredPol reduced crimes, we also wanted to analyze crime and stop rates before and after its deployment. The LAPD began to deploy PredPol in only three of its divisions (Foothill, Southeast, North Hollywood) back in 2013 before extending it to all 21 divisions in 2015 (6). Thus, we wanted to compare trends pre-2013 and post-2015. For observations made between 2013 and 2015, we wanted to look at how the numbers differed between the three divisions and the rest of the divisions. We defined the treatment for this study to be the reallocation of a police officer by PredPol. In other words, a crime or stop that occurred in an area where PredPol was active was marked as a treatment instance if the officer's home division was different from the actual location.

Fortunately, the LAPD published several datasets online for public use. For our study, we used datasets containing crimes, arrests, and pedestrian and vehicle stops that occurred between 2010 and 2019. Thus, they contained observations that were made before and after PredPol's deployment. Furthermore, all three datasets contained location and time of the observations which allowed us to see which ones were observed with PredPol active and which ones were not. However, only the stops data contained officer features such as reporting district so our defined treatment was only applied to stops. This was one of the drawbacks of the data since any peculiar observations made on the crimes and arrests data may not have been a direct result of PredPol and its reassignment of officers.

The stops data also contained a binary indicator on post-stop actions. However, it does not specify *what* actions were taken, so conclusions on change in officer behavior were limited to *if* it changed, but not *how* it changed. The crimes data included features which allowed us to analyze how crime types and charges changed over the years and to see if PredPol reduced all crimes equally or not. The arrests data was used in parallel with the crimes data to see whether arrests rates of a specific type of crime followed the same trend as the overall numbers.

Additionally, the stops and arrests data also allowed us to compute their rates with respect to different ethnicities to see if there were any biases. The 2010 census data was used to calculate such rates.

A few studies addressing predictive policing have also used similar data in the past. In Lum and Isaac’s case study on PredPol, they used crime data in Oakland, California to discuss potential consequences that the predictive algorithm generates (12). Ensign et al. also described how predictive policing systems are updated with discovered crime data rather than the true crime rate (3). Both studies showed how such a process may end up allocating police officers back to the same neighborhoods only. Another study also used crime data from the LAPD, as well as the Kent Police Department (United Kingdom), to analyze the extent to which predictive policing, specifically ETAS-based models such as PredPol, can affect crime hotspots (13). These past studies justify the reason behind using LAPD data on crimes, stops, and arrests to analyze the impact of PredPol in Los Angeles.

As aforementioned, this study used data published by the LAPD. They are available on the Los Angeles Open Data Portal and licensed under the CC0 1.0 Universal Public Domain Dedication. This license states that the owner has waived all rights regarding the work as far as the law allows. In other words, the public is allowed to use the work without permission (14).

The data used in this study are also anonymous in terms of data privacy concerns. For the Stops data, only the serial ID and division ID are available to identify a police officer. As for a stopped individual, only race and sex are provided. In both the Crimes and Arrests data, no officer information is available, and only the victim’s age, sex, and race are given. Furthermore, both datasets include a location field that has been provided to the nearest hundred block to maintain privacy.

The schema for the three processed datasets are laid out in the following tables:

Table 1. Schema for Vehicle and Pedestrian Stop Data

Column Name	**Description**	**Data Type**
Stop Number	Unique identifier	Integer
Stop Division	Division where the stop was made	String
Sex Code	F/M	String
Descent Description	Race	String
Stop Date	MM/DD/YYYY	DateTime

Stop Time	24 hour format	DateTime
Year	Year	Integer
Officer 1 Serial Number	Unique officer identifier	Integer
Officer 1 Division Number	Division of officer	Integer
Reporting District	Sub-area within division	Integer
Post Stop Activity Indicator	Whether there was further activity after a stop was made	Boolean
Reassigned Officer	Whether the officer belongs in another division	Boolean

Table 2. Schema for Crime Data

Column Name	**Description**	**Data Type**
DR_NO	Official file number	String
Date Rptd	Date crime was reported	DateTime
Year		
AREA NAME	Police division	String
Rpt Dist No	Sub-area within division	Integer
Crime Type	Type of crime	String
Crime Charge	Charge of crime	String
Crm Cd Desc	Detailed description of crime committed	String
Arrested	Whether the suspect was arrested	Boolean
LAT	Latitude coordinate	Integer
LON	Longitude coordinate	Integer
PredPol Deployed	Whether it occurred when PredPol was active	Boolean

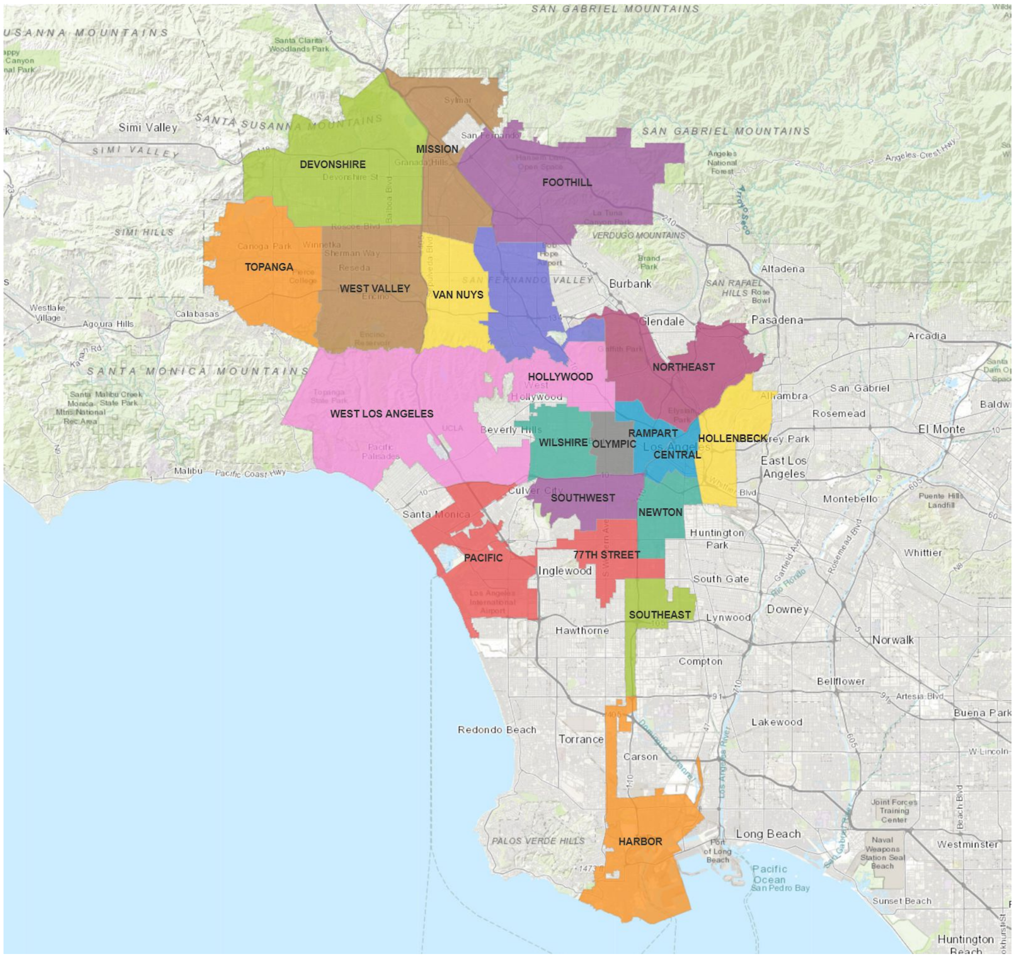
Table 3. Schema for Arrest Data

Column Name	**Description**	**Data Type**
Report ID	ID for the arrest	Integer
Arrest Date	MM/DD/YYYY	DateTime
Time	24 hour format	DateTime
Area Name	Police division	String
Reporting District	Sub-area within division	Integer

Age	Age of arrestee	Integer
Sex Code	Sex of arrestee	String
Descent Code	Race of arrestee	String
Charge Group Code	Crime type code	String
Charge Group Description	Crime type description	String
Arrest Type Code	A code to indicate the type of charge the individual was arrested for. D - Dependent F - Felony I - Infraction M - Misdemeanor O - Other	String
Charge	Normalized charge description	String
Charge Description	Charge the individual was arrested for	String
Year	Year	Integer
PredPol Deployed	Whether it occurred when PredPol was active	Boolean
Total	Location of incident	Shapely Point

Exploratory Data Analysis

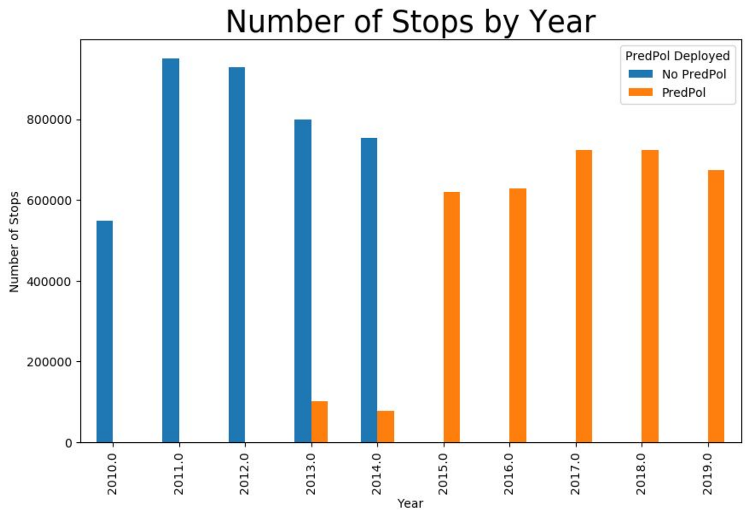
Our study used three datasets from the LAPD to analyze the effect of PredPol: pedestrian and traffic stops, crimes, and arrests. In our setup, the treatment was defined as the reassignment by PredPol. Two cases were considered when classifying an observation into the treatment and control group. The first case was when a mismatch occurred between the officer division number and the geographical division number. The second case was if an observation occurred when PredPol was deployed in its area. More specifically, an observation was tagged as treatment if: 1. it was between 2013-2015 and in either Foothill, Southeast, or North Hollywood, or 2. It occurred after 2015 in all 21 divisions. Thus, there was some uncertainty when looking at instances from the second case as it was not guaranteed that they were observed due to PredPol. A map of the 21 LAPD divisions are displayed below. (Note: The division without a label is North Hollywood.)



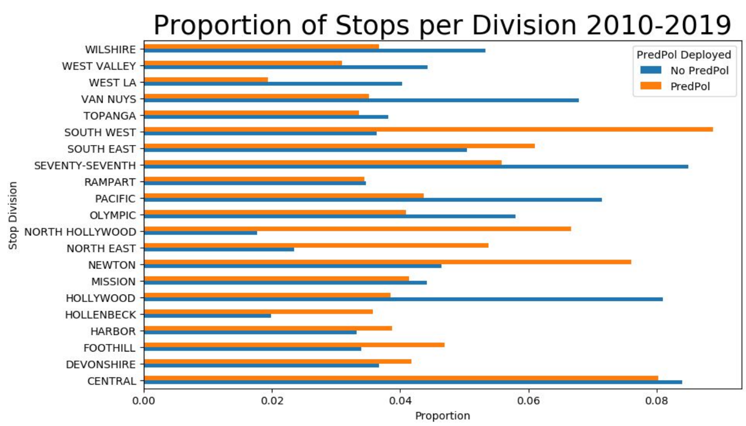
The following subsections will walk through the exploratory data analyses performed on each dataset and how they motivated us to perform further statistical testing. (Note: Observations made in 2020 were excluded in all figures and analyses due to the relatively low amount of data.)

Stops

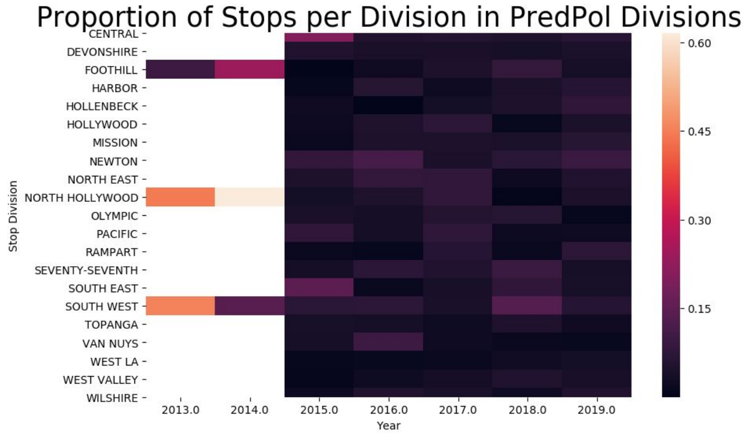
The *Stops* data originally contained features describing both the officer and the stopped individual, as well as the time and reporting district of the stop. This allowed us to identify the stops from the treatment group with the first case. The figure below shows the number of stops made from 2010 to 2019. We can see how the overall numbers generally followed a negative slope while numbers from the treatment group remained fairly constant. Stops made by reassigned officers began to emerge in 2013 which aligns with the year when LAPD began to deploy PredPol in the North Hollywood, Foothill, and Southwest divisions. It can also be observed how the number of stops from the control group experienced a dip in 2015 which coincides with the year when PredPol was fully deployed in all 21 divisions.

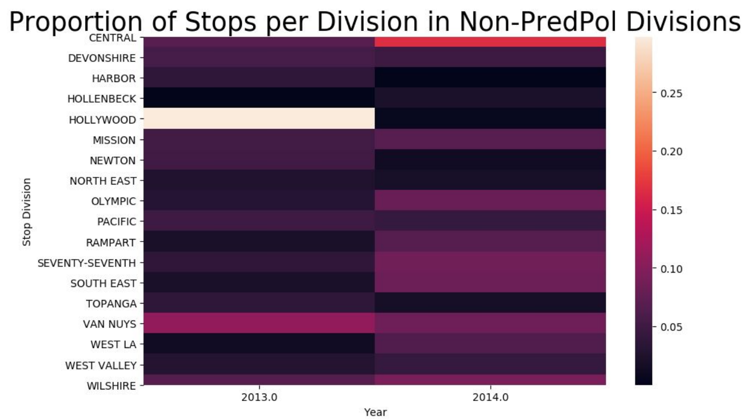


We can also look at how the stops were distributed among the divisions. The following figure shows how the overall pattern remained unchanged between the control and treatment groups. However, there has been a widening of the gap between the low and high proportions. This raised the question of whether PredPol was predicting or reinforcing past, observed trends.

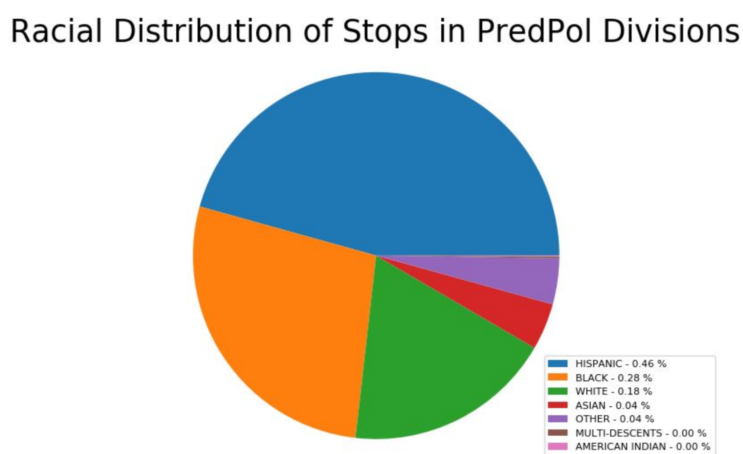


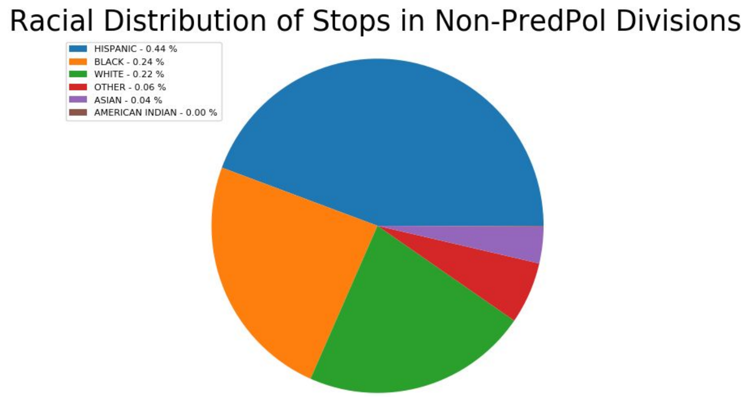
The proportion of stops per division was also broken down into a year-by-year comparison. Similar observations were made where annual trends remained the same, but the treatment group occurred at a larger scale as seen in the two figures below.





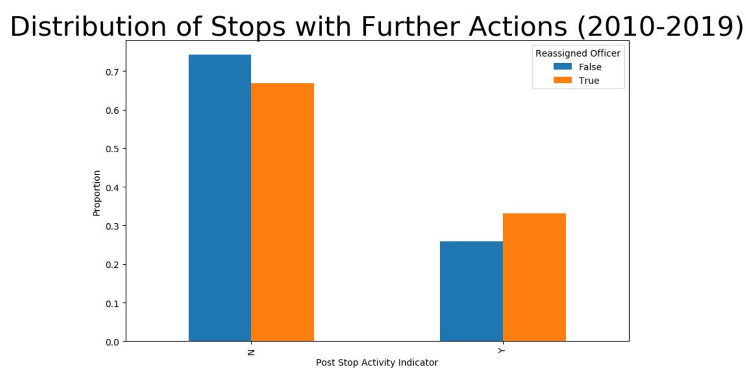
Ever since PredPol was deployed, there has been arguments of how it targets racial minorities (15). The two pie charts below compare the racial distribution of the stops made in the two groups.





It can be seen that there were minor differences in most races except for Black and White individuals. Black drivers and pedestrians saw a 0.07% increase in proportion while White drivers and pedestrians decreased by 0.03%.

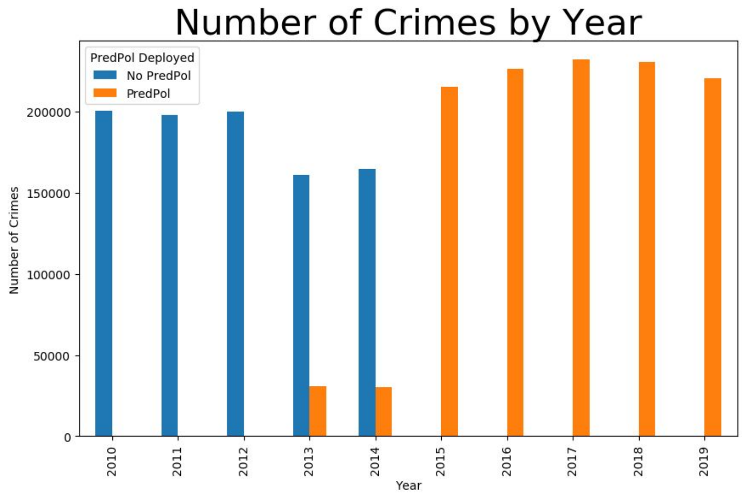
Lastly, we also wanted to see if there were any disparities in how reassigned and non-reassigned officers approached stopped individuals. The figure below shows how reassigned officers tend to take further post-stop actions more than officers patrolling in their original divisions. One possible explanation could be that an officer becomes more meticulous due to knowing that the area is supposedly a “hotspot”.



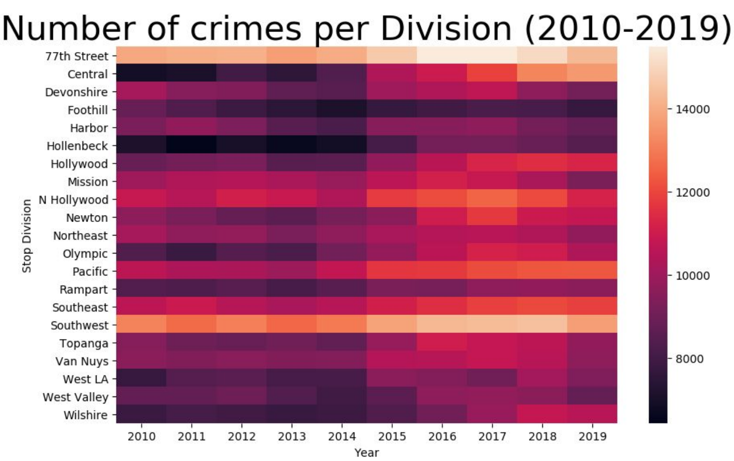
The *Stops* data also required some non-trivial processing. Please refer to Appendix A for a description on the data cleaning process.

Crimes

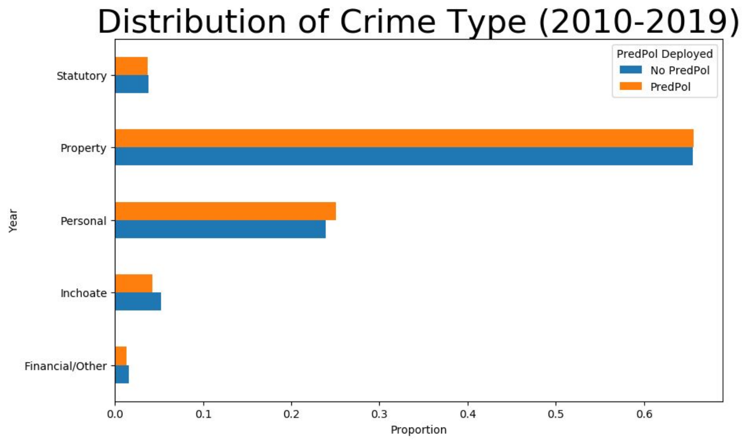
Unlike the *Stops* data, the *Crimes* data did not contain features on the officer. Thus, the treatment group was identified using the second case signifying crimes that occurred during the years and areas where PredPol was active. The graph below shows how the number of crimes varied from 2010 to 2019. It can be seen how the total number of crimes slightly decreased in 2013 when PredPol was introduced, but gradually increased in 2015. A possible explanation for the observation is that increasing police resources in “hotspots” led to more crimes being reported there and it outweighed the decrease in crimes in other areas.



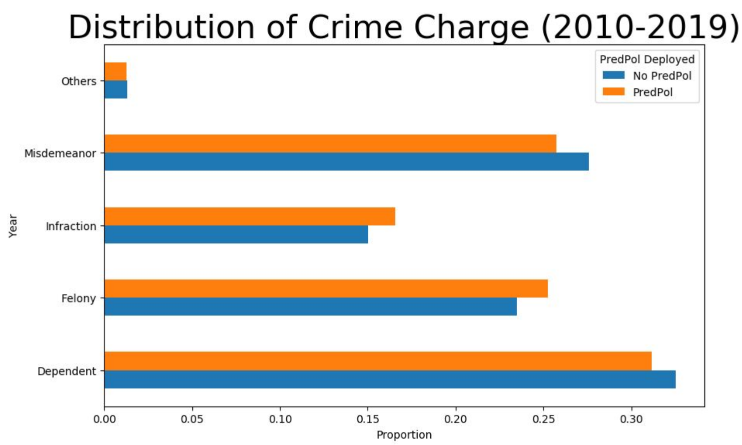
This motivated us to analyze the crime numbers by division as shown in the heatmap below. We expected to see similar trends as with the *Stops* data where "hotspots" saw even more increase in observations while other areas had a decrease. However, that was not the case with crimes. Starting from 2015, all divisions saw an increase in crime numbers. This trend continued until 2019 when some divisions such as Foothill slightly decreased.



Our speculation was that perhaps not all types of crimes were impacted by PredPol equally. Thus, we compared the distribution of crimes by crime type in the figure below.



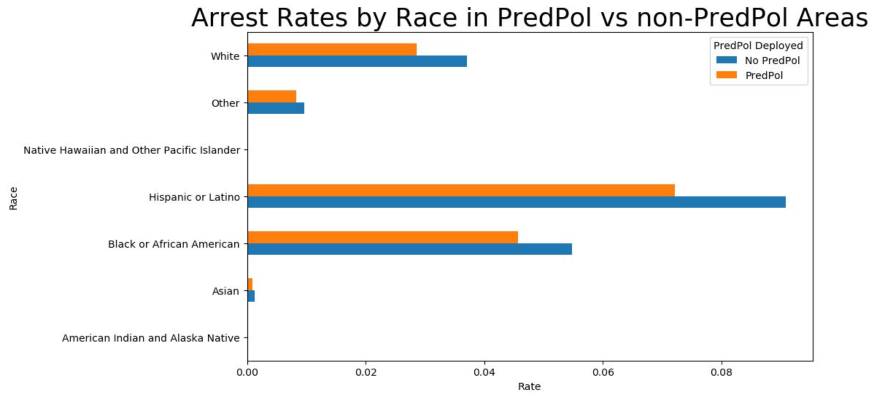
Although property crime was the most common type of crime, there was no visible difference between the two groups. Meanwhile, personal crimes increased ever since PredPol was deployed and inchoate crimes decreased. However, statistical testing is required to draw any conclusion.



Another distribution we looked at was by the crime charge as shown in the above figure. Similar to the distribution of crime types, not all charges were affected the same way by PredPol. The definitions of crime types and crime charges as well as the classification process are defined in Appendix B.

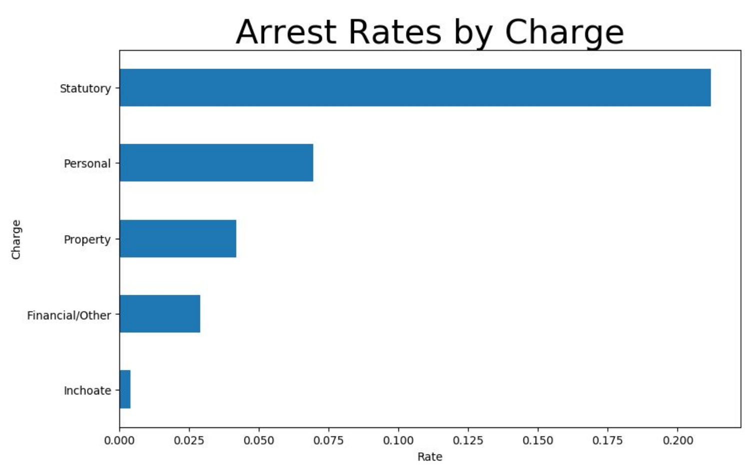
Arrests

The *Arrests* data was used to see how PredPol has affected arrest rate. The denominator used to compute the rates was obtained from the 2010 census data. Similar to the *Crimes* data, it did not contain officer features. Thus, the treatment group identification was based on year and division.

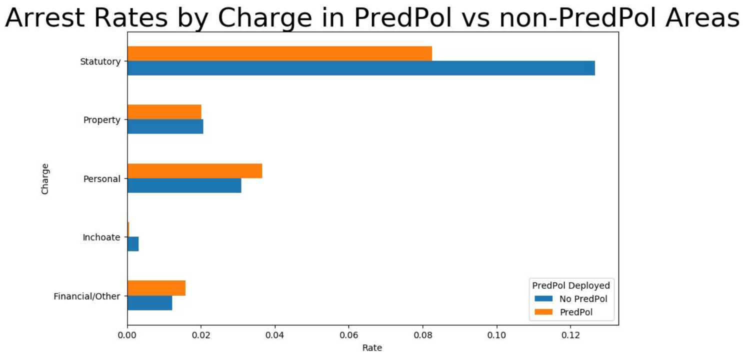


The figure above compares the arrest rates of different races between the control and treatment group. It can be seen that all races had lower arrest rates in areas that PredPol was active which led to an overall decrease in arrest rates.

We also looked at the arrest rates based on different charges. PredPol aims to reduce crimes by increasing police presence in "hotspots". However, not all crimes are equal. For example, we would expect to see a decrease in arrest rates for crimes that are often committed in public and easily spotted. On the other hand, crimes such as fraud can occur through other means that are more subtle. Thus, increasing the presence of police officers might not affect such crimes as much as the former.



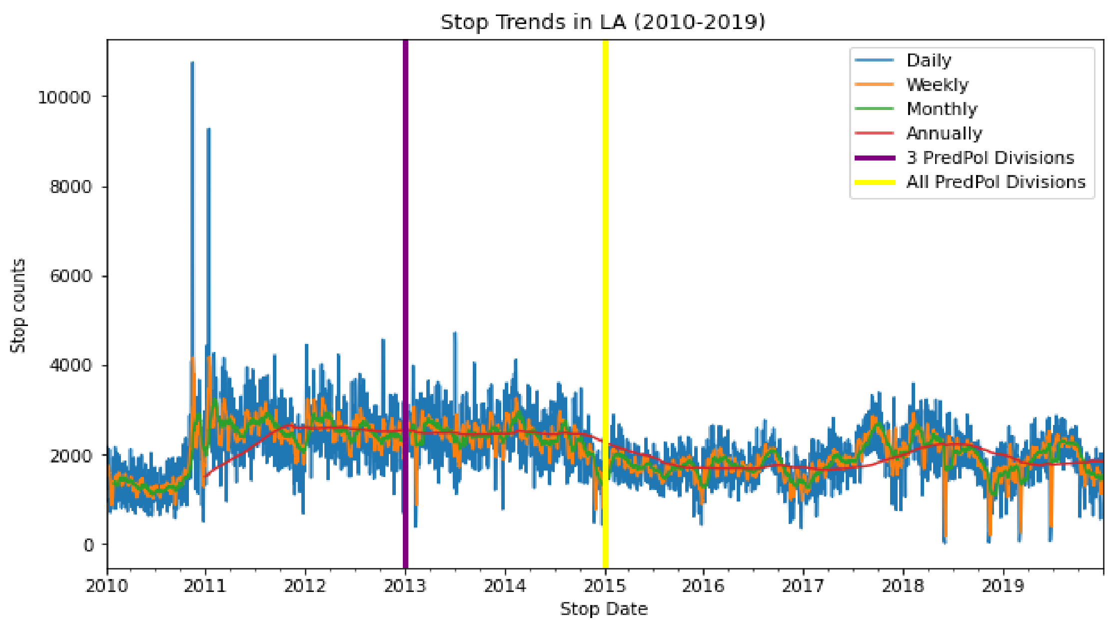
The graph below shows the differences between the arrest rates by charge in PredPol areas and non-PredPol areas. It can be seen that there are certain crimes that saw higher arrest rates in deployed areas such as personal crimes. On the other hand, statutory crimes, which had high arrest rates in general from the graph above had lower rates in deployed areas. Statutory crimes include DUI and several traffic violations which are often easily spotted.



Analysis / Results

Stops

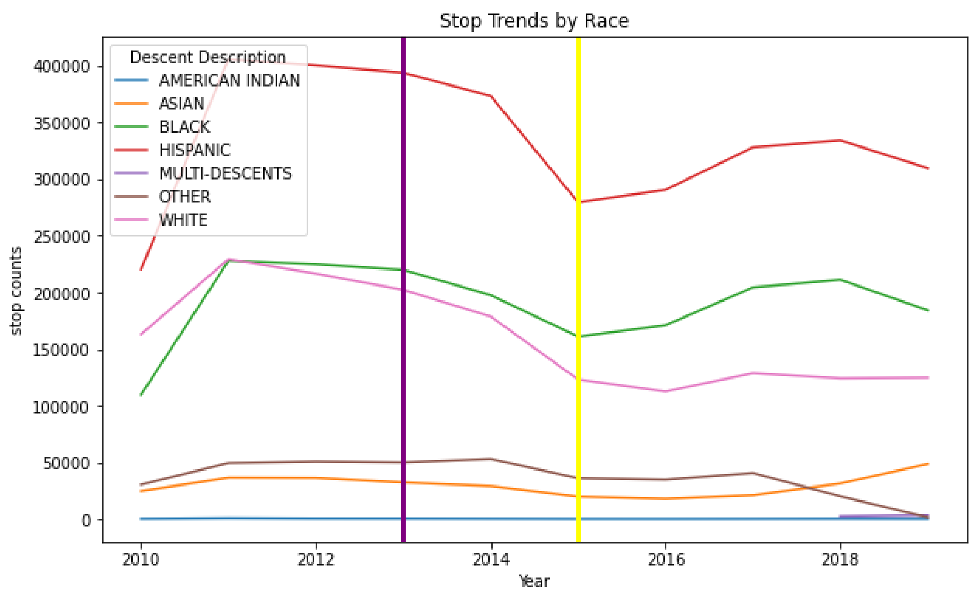
Analyzing the traffic stops data we look at how stop trends changed over the years. First we look at the overall number of traffic stops and observe the trends.



In the chart above, we can observe that during the period that PredPol was only deployed in three divisions the number of traffic stops remained the same following the years before its deployment. After it was deployed in all 21 divisions we see that the number of stops goes down for a bit before spikes continue. To observe how the traffic stops changed during these different time periods we perform the Dickey Fuller test to show if they are stationary.

Time Period	Pre-2013	2013-2014	Post-2014
Test Statistic	-2.6369	-3.5723	-3.646771
p-value	0.0856	0.0063	0.004927
Critical Value (1%)	-3.4365	-3.4395	-3.433988
Critical Value (5%)	-2.8642	-2.8656	-2.863147
Critical Value (10%)	-2.5682	-2.5689	-2.567626

Thus we can see that stops were stationary and did not change during any of the periods where PredPol was deployed but only was non-statioary in the pre-PredPol time period. We then proceed to study how PredPol affected stop rates of specific racial groups.



Looking at the number of stops by race we see that the raw number of stops during the 2013-2014 period goes down for about all drivers regardless of race. Following PredPols deployment to all divisions, we see that stops increase for minority drivers while going down for white drivers. Breaking this down even further we look at the stop rates by race over the total number of stops and stop rates over the population of the area.

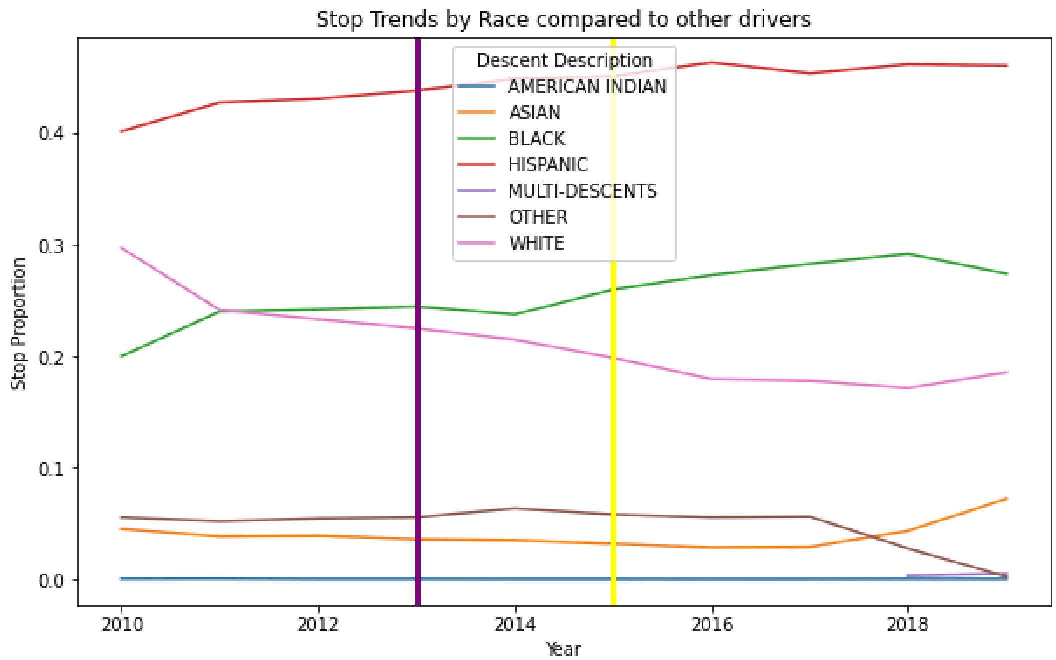


Figure 1: Racial stop rates by overall stops

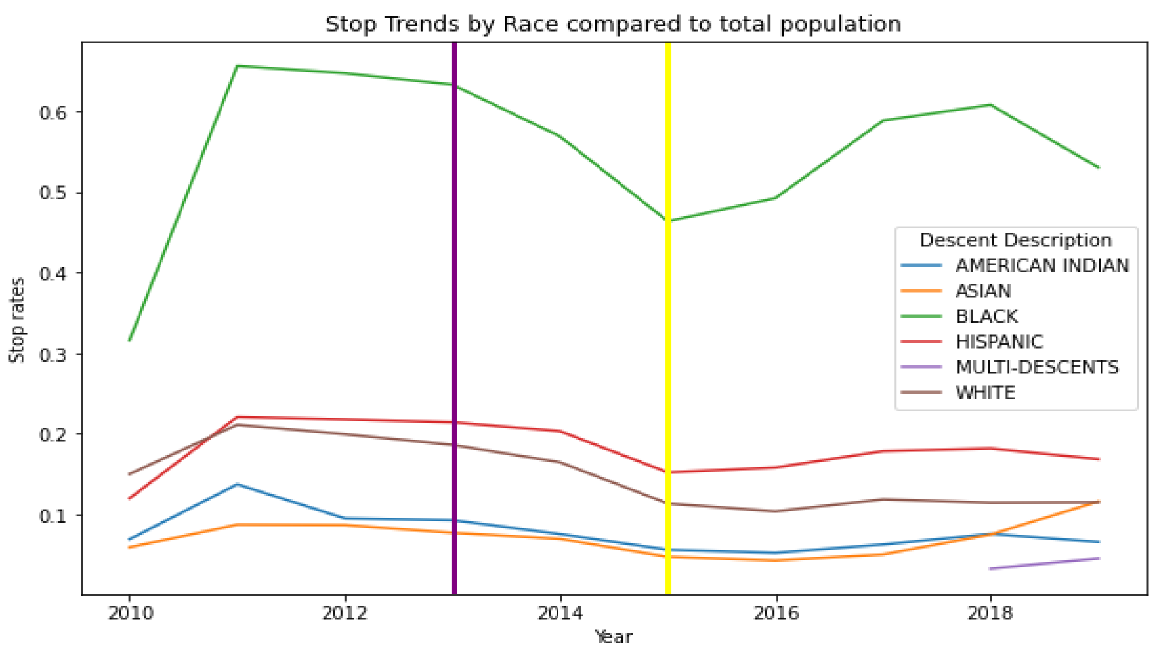


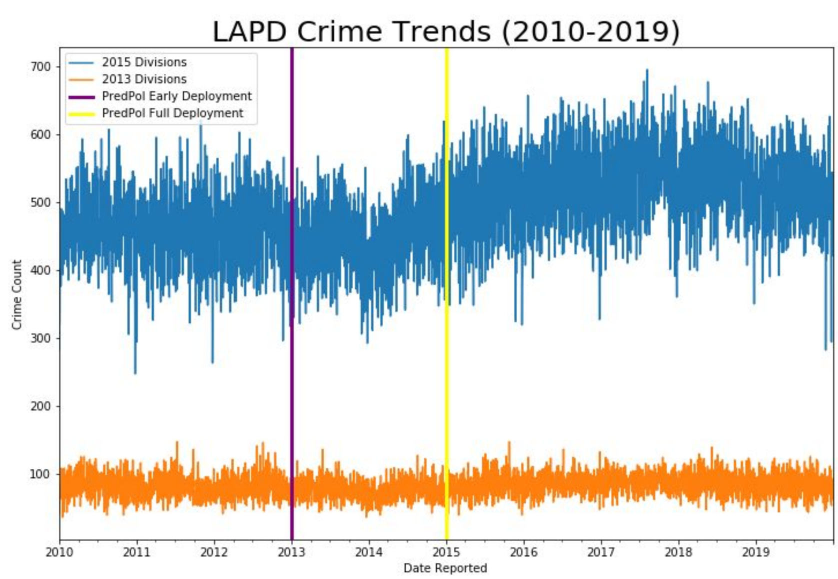
Figure 2: Racial stop rates by populations

From Figure 1 we can see that the stop rates for minority drivers continue to increase during PredPol's deployment and decrease for white drivers. Looking at stop rates in Figure 2 we can see that stop rates for minority drivers decrease during PredPol's early deployment but increase again during PredPol's full deployment. But we cannot attribute this changein stop rates from PredPol alone so we run t-tests to see that PredPol has on stops. There we see that PredPol resulted in an increase in Stops for black drivers while a decrease in stops for white drivers. The results are displayed below with significant values highlighted.

Race	**Statistic**	**p-value**
American Indian	0.9377	0.34835
Asian	-0.4746	0.6350
Black	3.4581	0.0005
Hispanic	0.5324	0.5944
White	-0.416	0.0007
Other	-3.364	0.6773

Crimes & Arrests

We first separate the crimes into those that occurred in the three divisions with PredPol from 2013 and those that occurred in the remaining 18 divisions.



From the plot above, we can see that daily crimes were daily constant in the divisions with early deployment. On the other hand, the other divisions had varying trends especially during the 2013 to 2014 period when PredPol first started out. We then performed the Augmented Dickey Fuller Test on each series so see if they are stationary.

Time Series	Early Deployment	Regular Deployment
Statistic	-4.808845	-2.712778
P-Value	0.000052	0.071853
Critical Value (1%)	-3.432	-3.432
Critical Value (5%)	-2.862	-2.862
Critical Value (10%)	-2.567	-2.567

With a confidence level of 0.95, the outcome tells us that data from the three divisions with early division is actually stationary. This implies that the mean and variance remained constant over time. Meanwhile, divisions with regular deployment are not. We then took into account the fact that PredPol was deployed in certain divisions at different times by first splitting the time series into three segments (pre-2013, 2013-14, post-2014) and performed the Dickey Fuller test on each for the regular deployment series. The results are displayed in the table below.

Table 4. Dickey Fuller Test Outcomes on Regular Deployment Divisions

Time Period	Pre-2013	2013-2014	Post-2014
Test Statistic	-5.793692	-2.633557	-4.878484
p-value	0.000000	**0.086258**	0.000038
Critical Value (1%)	-3.436	-3.440	-3.434
Critical Value (5%)	-2.864	-2.866	-2.863
Critical Value (10%)	-2.568	-2.569	-2.568

It can be seen that daily crimes in non-PredPol divisions between 2013 and 2014 were not stationary and actually had a slight decrease. Thus, we narrowed the scope of the analysis to crimes that occurred during this timeframe which coincides with the time when PredPol was deployed in only three divisions. We then proceeded to analyze how criminal behavior may have been affected by PredPol by looking at how the distribution of crime types differed between areas with PredPol and areas without. The results are shown in **Table 5**.

Table 5. Distribution of Crime Types in PredPol and non-PredPol Divisions

Crime Type	No PredPol	PredPol
Financial/Other	0.014483	0.013499
Inchoate	0.048344	0.049947
Personal	0.241325	0.232723
Property	0.658628	0.666976
Statutory	0.037221	0.036855

From the table we can see that in areas with PredPol deployed, there was a higher proportion in inchoate and property crimes. We performed t-tests on the difference of the proportions with respect to PredPol areas to see if any were significant.

Table 6. T-Test Outcomes on Proportion of Crime Types Committed in Different Areas

Crime Type	Statistic	p-value
Financial/Other	-2.395509252354022	0.016599122469128813
Inchoate	1.3708963002320382	0.17041036162649842
Personal	-2.0534995699630065	0.040026746299820865
Property	3.4441761860529043	0.0005730369896344116
Statutory	0.6697855444149966	0.5029960564855552

The outcomes of the t-tests are displayed in the table above with significant values highlighted. We then performed similar tests on the proportion of *arrests* made on different crime types to see if there were similar patterns. The results are displayed below.

Table 6. T-Test Outcomes on Proportion of Crime Types Arrested in Different Areas

Arrest Type	Statistic	p-value
Financial/Other	-21.10464427159605	1.1847715929607052e-98
Inchoate	1.7713431545803482	0.07650670051305225
Personal	9.386460525033932	6.3293390912986235e-21
Property	17.826200411274847	5.701650536736053e-71
Statutory	-7.916222910535208	2.4732603700415266e-15

Discussion / Conclusion:

Findings: Traffic Stops

From the analysis on the stops data we see that the overall number of stops did not change much during PredPol's deployment and remained stationary throughout the entire period. Looking at how drivers of different races are affected, during the 2013-2014 we saw drivers of every race either saw a decrease in stops or remained the same. But when we looked at the stop rates over the total number of drivers and the populations, we saw that stop rates were much different with minority drivers still being impacted, Overall we see that PredPol had a negative effect on black drivers, with stops increasing, and positive effect on white drivers where the stops decreased.

Findings: Crimes and Arrests

The analysis on crimes and arrests tells us that there was a minor reduction in daily crimes when PredPol was deployed in just three out of the 21 divisions. However, the decrease occurred in the divisions without PredPol, so this raises the question of whether crimes were reduced because people are discouraged from committing them, or because police resources are allocated elsewhere.

During that period, there was also an increase in the proportion of property crimes being reported in PredPol divisions which also led to more arrests of that particular type of crime.

Ultimately, daily crimes increased once PredPol was deployed in all divisions while the arrests kept seeing a gradual decrease over the years.

<u>Confounding Factors:</u>

To answer our question, “Is predictive policing preventative” using PredPol, we must also acknowledge the assumptions and possible confounders of our project. When aggregated by division level, the results of crime analysis were mixed. This could be due to there being an actual reduction of crime in some places, or just a reduction in police presence in some areas (due to reassignment). Additionally, the location, or whether the area is commercial or residential may have an impact on if PredPol is more effective.

When the crime analysis was aggregated by crime type, the results were also mixed. This may imply that PredPol targets certain crime types more than others. For example, we saw a major decrease in financial crimes but this could be due to underreporting; we would not necessarily expect financial crime to be impacted by sending more officers to an area.

One major confounder that prevents us from drawing causal inferences between the increase in crimes to PredPol is another predictive algorithm called LASER which was also deployed around the same time as PredPol. To generalize to the effect of all predictive policing, it would be helpful to look at other predictive policing algorithms (like LASER) and additionally, other cities or states that use PredPol specifically for comparison.

Overall, although the increase or decrease in crimes cannot be fully attributed to PredPol, we can at least say that there was no reduction in crimes as PredPol has claimed to do.

References

1. Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased? (2018, March 5). Retrieved April 5, 2020, from https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/
2. Haskins, C. (2019, February 14). Academics Confirm Major Predictive Policing Algorithm is Fundamentally Flawed. Retrieved April 5, 2020, from https://www.vice.com/en_us/article/xwbag4/academics-confirm-major-predictive-policing-algorithm-is-fundamentally-flawed
3. Ensign, D. (2017, June 29). Runaway Feedback Loops in Predictive Policing. Retrieved April 5, 2020, from https://arxiv.org/abs/1706.09847
4. Wolpert, S. (2015, October 7). Predictive policing substantially reduces crime in Los Angeles during months-long test. Retrieved April 5, 2020, from https://newsroom.ucla.edu/releases/predictive-policing-substantially-reduces-crime-in-los-angeles-during-months-long-test
5. https://data.lacity.org/
6. Smith, M. P. (2019, March 8). PDF. Los Angeles, http://www.lapdpolicecom.lacity.org/031219/BPC_19-0072.pdf
7. E. S. Levine, Jessica Tisch, Anthony Tasso, Michael Joy (2017) The New York City Police Department's Domain Awareness System. Interfaces Published online in Articles in Advance 18 Jan 2017 . http://dx.doi.org/10.1287/inte.2016.0860
8. *Predictive Policing: Using Technology to Reduce Crime | FBI: Law Enforcement Bulletin*. (2020). *FBI: Law Enforcement Bulletin*. Retrieved 12 April 2020, from https://leb.fbi.gov/articles/featured-articles/predictive-policing-using-technology-to-reduce-crime
9. @randyrieland, F. (2018). Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?. Smithsonian Magazine. Retrieved 12 April 2020, from https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/
10. Smith, M. P. (2019, March 8). PDF. Los Angeles, http://www.lapdpolicecom.lacity.org/031219/BPC_19-0072.pdf

11. *LAPD pioneered predicting crime with data. Many police don't think it works*. (2019). *Los Angeles Times*. Retrieved 12 April 2020, from https://www.latimes.com/local/lanow/la-me-lapd-precision-policing-data-20190703-story.html
12. Lum, K. and Isaac, W. (2016), To predict and serve?. Significance, 13: 14-19. doi:10.1111/j.1740-9713.2016.00960.x
13. Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M. E., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (1970, January 1). Randomized Controlled Field Trials of Predictive Policing. Retrieved from https://www.semanticscholar.org/paper/Randomized-Controlled-Field-Trials-of-Predictive-Mohler-Short/e6287d4c4ed33707843321995246fdbe42ff82e9
14. Creative Commons License Deed. (n.d.). Retrieved from https://creativecommons.org/publicdomain/zero/1.0/
15. Reynolds, M. (2020). *Biased policing is made worse by errors in pre-crime algorithms*. *New Scientist*. Retrieved 2 May 2020, from https://www.newscientist.com/article/mg23631464-300-biased-policing-is-made-worse-by-errors-in-pre-crime-algorithms/
16. *Types of Criminal Offenses*. (2018). *Justia*. Retrieved 2 May 2020, from https://www.justia.com/criminal/offenses/

Appendices

Appendix A: Processing Stops Data

Some of the columns in the *Stops* data were irrelevant or can be inferred from others. Such columns were dropped. For example, most of the missing data were from columns describing the second officer. The assumption was that officers always patrolled in pairs. Thus, knowing just the division number of the first officer should suffice for identifying whether a stop belongs in the control or treatment group.

For *Officer Serial Number*, instances where that value was null were dropped since there was no relationship between the missing value and values from the other columns.

For *Officer 1 Division Number* and *Division Description 1*, there were no instances where both fields were empty. Thus, the method of imputation was to infer from the other column that was not null. Both columns indicated the same information, but were both kept because the division number was later used to compare with the division number that the reporting district falls in. Different division numbers implied that the officers did not usually patrol the area where the stop occurred. For such stops to be classified as the treatment group, the stopped area must also be an area that PredPol was deployed and the year must also align.

For *Post Stop Activity Indicator*, there was a clear relationship between the column and *Officer 1 Serial Number*. The first step was filtering the stops data to show only stops made by officers who had null values in the indicator column. It was observed that all the non-null instances have a single unique value, N. Thus, the decision was to impute null values with N which meant that no further activities were taken after a stop was made.

One column that was added during the process was the *Stop Division* column - the division where the *Reporting District* is located. This was done by merging the district column with a spatial DataFrame containing census data, districts, divisions, and blocks in Los Angeles. The column will later be used to compare with the officer division number to identify whether the officer was reassigned or not.

Appendix B: Processing Crimes Data

Most of the missing values in the *Crimes* data were either pertained to the victim's information or could be inferred from other columns with no null values. Thus, only trivial cleaning procedures were required after dropping such columns.

There were two additional columns, namely, Crime Type and Crime Severity that were derived from the *Crm Cd Desc* column since there were 141 unique values in the original column. Among the 141 values, some were the same, but with minor typos. Some others were just the same type of crime, but with different charges.

The *Crime Type* column classified all the values into five types of criminal offenses. These offenses are defined as the following (16):

- Personal - Crimes related to either physically or mentally harming another person.
- Property - Crimes that involve the interference with property that belongs to another.
- Inchoate - Crimes that were initiated but incomplete, as well as contribution to crimes.
- Statutory - Crimes that are proscribed by law such as alcohol related crimes and traffic violations.
- Financial/Other - These are white-collar crimes where one uses deception or fraud for financial reasons.

On the other hand, the *Crime Severity* column indicated the charge assigned to the crime. With severity increasing from left to right, the possible values are: Infraction, Wobblette, Misdemeanor, Wobbler, and Felony. A fifth value, Others, was designated for extreme and rare crimes. An example is illustrated below of how Theft is classified into different severities:

- Felony - Thefts of property over $950.01.
- Misdemeanor - Thefts of property under $950.
- Wobbler - Thefts without indication of value that can be either felony and misdemeanor.
- Wobblettes - crimes that require more information to determine whether or not they belong in infraction or misdemeanor.