CSE 6242 Final Report – Team 4

J. Carbonell, C. O'Brien, N. Ramallo, A. Sharma, I. Tumey, W. Wirono

I. INTRODUCTION

By identifying where crimes are likely to occur and restructuring patrol routes to increase police presence in these areas, place-based predictive policing technologies offer the promise of deterring rather than responding to crimes. They also promise more "objective" direction of police activity – pushing police to where crime occurs rather than being guided racist notions of which neighborhoods harbor criminals [1]. Proponents of these technologies often overlook the extent to which crime data are themselves an artefact of racist historical patterns of policing [2].

II. PROBLEM STATEMENT

We aim to develop granular maps of crime predictions and neighborhood demographics to evaluate the extent to which place-based crime prediction has disparate racial impacts. We approximate predictive policing techniques by using time-series forecasting and crimes and calls for service data from Baltimore, MD, and the year 2018, to predict where crime is most likely to occur – crime "hot spots". To investigate the influence of data inputs on which neighborhoods (and racial groups) are identified as hot spots, we compare predictions based on calls for service and crime data.

Understanding the difference between calls for service and crime is key to appreciating the value of the proposed approach. A call for service is generated when a Baltimore resident calls 911 or police internally request the assistance of additional officers for public safety issues, representing a broad range of potential police activity. Not every call for service results in a crime being reported or an arrest, nor does every crime reported correspond to a call for service. The production processes for each type of data are therefore distinct. We investigate whether these distinct data types generate different predictions of where crime will occur, as well as the implications for what racial groups are policed more heavily.

To understand how these different types of data might over- or under-represent different parts of the city, we look at the number of arrests, crimes, calls for service, and Type 1 calls for service (a subset of calls for service that relate to Type 1 or felony crimes, the same crimes that are included in the 'Crime' dataset) by census tract. We then calculate the extent of over-/under-representation of each data input for all census tracts, relative to the population of each tract. One clear finding is that predominantly non-African American census tracts are underrepresented, regardless of the data input. The other, more subtle finding is that the most over-/under-represented tracts in each data set have different African American population shares. This suggests that crime predictions trained on historical data might predict crime in locations that differ systematically in terms of neighborhood demographics. We evaluate this possibility in the Experiments and Evaluation section below.

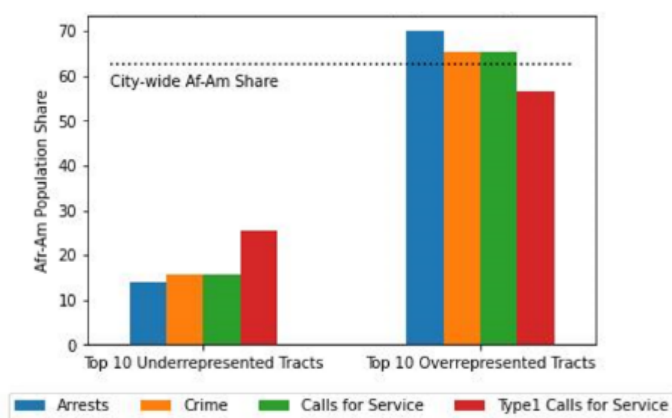

Figure 1: Demographic comparison of top over- and under-represented census tracts by data type

III. LITERATURE SURVEY

We explored the roots of big data's usage in police departments, starting with Ridgeway [4] and his analysis of the President's Commission on Law Enforcement and Administration of Justice in 1967. The technologies envisioned by the commission are now feasible and can be implemented throughout the country. Benbouzid [5] explores how crime heat map visualizations have been helpful by allowing officers to understand trends. To build upon our understanding of visualizations, we consulted Stoffel et al. [6], where the authors took a deep dive into a method utilizing geospatial and time-based features to derive insights into crime patterns. Furthermore, we consulted Ferguson [7], who outlined some of the most common methods being employed, discussed the pros and cons of each model, and discussed limitations to their implementation. To further illustrate these limitations, Isaac and Lum [8] tried predicting drug crime against arrest data, exposing the bias in arrests. However, drug arrests exhibit greater racial bias and represent a small proportion of crime relative other crimes, limiting the applicability of these findings.

Next, we needed to explore any potential issues in the methods discussed above. The article from the Brennan Center for Justice [9] provides an overview of the issues plaguing predictive policing systems, namely prejudicial data being used in models, the lack of accountability, and Civil Rights violations. We consulted Egge [10] and Chainey [3], who both explored the accuracy of these models and how the data had collection issues and were manipulated. While these papers did not address the issues of bias still present within the data, Richardson et al. [2] examines how these practices in policing, data collection, and manipulation work in concert to create a positive feedback loop for the bias. In addition, Brayne [11] studied the LAPD, to show how bias and unjustified surveillance by the police created a positive feedback loop, entrenching the biases already present. The authors do not offer solutions, but we believe that issues in the current processes of police departments can be solved by increasing the granularity and accountability of the methods. In addition, we picked up on issues regarding the erosion of Fourth Amendment rights. We consulted Andrew Ferguson's book The Rise of Big Data Policing [12], where he discusses technologies, like aerial cameras, facial recognition, and automated license plate readers, being employed with predictive policing models. Ferguson [12] explores not only the potential errors resulting from the above tools, but also how these methods can have a disproportionate impact on minority communities. In addition, he addresses the privacy issues that arise from the use of these technologies, leading to an erosion of our Fourth Amendment rights. While Ferguson [12] does suggest some policy changes which could stem these issues, we would like to take a deeper dive into how we can balance the need for good data and people's privacy when constructing predictive policing models.

Through further research, we found Levitt [13] investigates datasets that exhibit reporting and recording bias due to police force size policy changes. Although this policy does not pertain to our research, the study's insights inspire us to build an algorithm based on calls of service. Reasoning for this stems from a paper by Perry [14], which states how many police departments continue to employ predictive policing models without assessing their predictions accuracy. The author also discusses the Supreme Court ruling that relaxed the reasonable suspicion requirement for "high crimes" areas, which impact citizens' civil and privacy rights. Similarly, Ferguson [15] reviews issues with "black data" that is currently being used in predictive policing methods. He argues that the predictions could lead to new bias, where police are more likely to engage individuals, leading to a roll-back of our Fourth Amendment rights. Mohler [16] offers one of the only empirical tests of this possibility to date. He evaluates the impact of algorithmically-directed hot spot patrols relative to patrols directed by traditional crime forecasts, where crime forecasts were based on calls for service. He finds higher rates of arrests in algorithmically-identified hot spots but with no significant demographic differences.

To dive deeper, we explored findings from the Chicago Police Department. Reports from their OIG office [17] detail discontinuation of one of their policing programs due to unreliable scoring and the lack of internal controls. Another report discusses the CPD's implementation of Strategic Decision Support Centers [18]. This report will aid our research as it evaluates the implementation of the policing technologies we are researching even though underlying data used to generate these models are not discussed. Furthermore, we gained a deeper insight from Jefferson [19] regarding predictive crime mapping that highlights racialized approaches to policing. Geographic information systems provide more precision with which locations to target but these mapping systems are built by "nonscientists" working in an organization of racialized policing. The author addresses precision issues with mapping that our team hopes to improve by using hourly data points.

Ultimately Perry [20] reminds us that predictive policing cannot foretell crimes, motivating the need for a probabilistic approach in visualizing the geospatial data. One shortcoming is the article's untenable approach in consolidating multiple crime data from different states. Hence, this prompts our team to provide a clear scope (i.e. by zip code, city, or state-level) in obtaining crime databases while also addressing the need for a risk assessment framework [21]. The author does not discuss data preprocessing in meaningful detail, so we will identify and further examine schemes dealing with incomplete data (e.g. imputation, avoiding selection bias, etc.) in our project.

IV. PROPOSED METHOD

A. *Contribution to the State of the Art*

Our main contribution is to predict criminal activity as a descriptive exercise. Our approach asks where police might spend their time if directed by predictions of criminal activity "hot spots". Furthermore, it asks whether the location of these hot spots differs significantly based on the data inputs used to generate the predictions. Where these locations differ, we compare their demographic compositions. There is a rich methodological literature on approaches to predicting crime. There is comparably deep theoretical work on potential bias or harms from such predictions, but relatively little work to date has empirically evaluated different hot spot prediction model inputs in terms of spatial or racial bias. To our knowledge, it is the first effort to directly compare hot spot predictions separately generated using calls for service and crime data based on neighborhood demographics

B. *Description of Approach*

To build geographic crime hot spots, we treat the crime and calls for service data as time series data. We used Google's Geocoding API to convert calls for service addresses to latitude and longitude coordinates. Due to the large number of observations in this dataset and geocoding cost constraints, we extracted a subset of roughly 450,000 address points. This subset of observations is a random sample, stratified by census tract and month to ensure the sample is geographically and temporally representative. Using these data, we produce localized event predictions using an ARIMA time-series forecasting model. We chose an ARIMA model as opposed to a more complicated model because we wanted a simple, interpretable methodology for generating crime predictions. This addresses the issues of opaqueness that are seen in many models currently employed within police departments (see Furguson [15]). Further, it has been shown that crime in a specific area tends to signal both future re-victimization and the spread to areas nearby (see Ferguson [21]). ARIMA models can incorporate this attribute using its autoregressive element. Additionally, an ARIMA model can handle cyclicality or seasonality trends in the data. We believe that with our localized crime predictions we can inform police departments about where different types of criminal activity are likely to occur and the disparate racial impacts of different model inputs.

After specifying our forecasting model, we use it to predict crimes and calls for service for different geographic regions within Baltimore. We draw on our sample calls for service dataset of 473,153 observations and 19 features and a crime dataset with 48,464 observations and 13 features to generate our predictions. These data are drawn from the City of Baltimore's Open Data portal [22]. We divide the city of Baltimore into 478 equal sized geographic regions, each approximately 0.22 square miles. We link each geographic region to the census tract it resides in, allowing us to approximate the demographic composition of each region.

Our visualization (Figure 2) displays the forecasted crimes and calls for service for geographic regions in the city of Baltimore, coloring each region according to the proportion of the total amount of predicted crimes or calls for service for single day. We create the visualization in Python using Folium, a data visualization tool built on *leaflet.js.* We also convert the tool into a web framework with Flask and deploy it using the cloud platform Heroku so users can experiment with the visualization online [23].

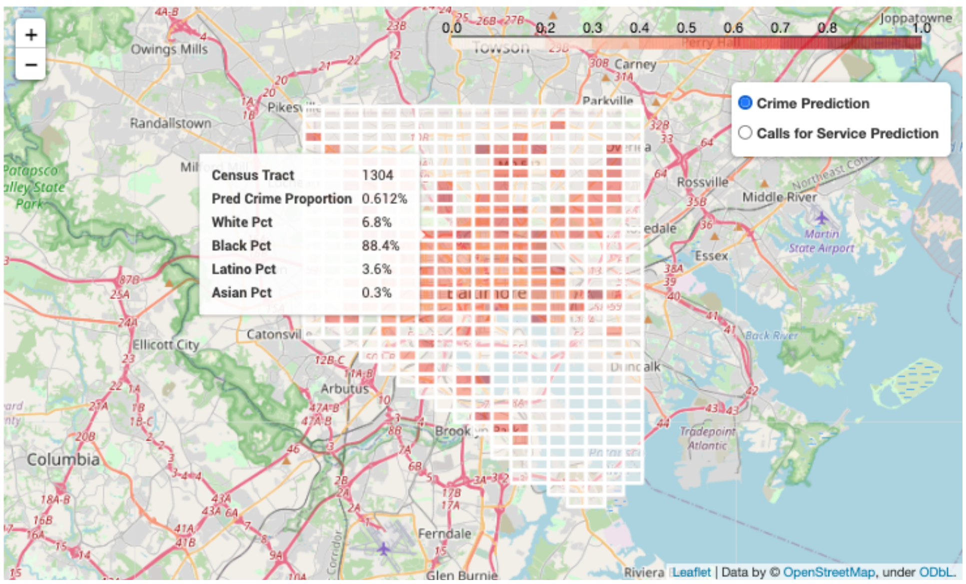

Figure 2: Visualization of forecasted crimes in Baltimore.

Users can select which prediction input (crimes or calls for service) they would like visualized on the map. This allows users to compare the forecasted values between the crime and calls for service maps, as well the forecasts for different geographic regions. When hovering over a geographic region, users are presented with a tooltip that displays its corresponding census tract, predicted proportion of crimes or calls for service, and demographic composition. This allows users to explore how different data inputs for predictive policing models may over or under-represent certain neighborhoods, and what the demographic composition of those neighborhoods are.

V. EXPERIMENTS AND EVALUATION

In determining what models to use to predict future rates of crime and calls for service, we attempt to balance two concerns. We want a model that would serve us well to predict the crime, but we also need to keep the model transparent and accountable, so we do not end up with the same issues seen in the black-box methods. Following the first concern, treating crimes and calls for service as time-series addresses an issue identified in our literature survey (Ferguson [21]): crime tends to lead to future crimes in that location. An ARIMA model also addresses the second concern because it is a well-understood model that can be transparently assessed to determine its efficacy.

To optimize our model parameters, we used a Python library titled *pmdarima* that tests different ARIMA specifications and optimizes their parameters while minimizing the Akaike information criterion (AIC). AIC estimates a model's prediction error, so we believe it to be a good measure to find the model with the best fit. We found that an ARIMA(3,0,1) model was the best fit for both the crime and calls for service data.

	Parameters	Crime Model AIC	CFS Model AIC
	ARIMA(1,0,1)	3095.037	4582.802
	ARIMA(0,0,0)	4625.196	6274.308
	ARIMA(1,0,0)	3260.674	inf
	ARIMA(0,0,1)	4209.413	inf
	ARIMA(2,0,1)	3091.008	4553.091
	ARIMA(2,0,0)	inf	inf
	ARIMA(3,0,1)	3089.782	4547.913
	ARIMA(3,0,0)	inf	inf
	ARIMA(4,0,1)	3091.081	4548.769
	ARIMA(3,0,2)	3094.048	4626.752
	ARIMA(2,0,2)	3089.944	inf
	ARIMA(4,0,0)	inf	inf
	ARIMA(4,0,2)	3093.586	inf
	ARIMA(3,0,1)	3107.905	4571.641
Best Fit Model:	**ARIMA(3,0,1)**	**3089.782**	**4547.913**

Figure 3: Crime and Calls for Service ARIMA Model Evaluation

To further assess the quality of the models, we calculate the MSE for the final 30 days of data. The model built on crime data had a Mean Squared Error (MSE) of 562.993, and the model built on calls for service data had a MSE of 19,041.247. In the interest of employing a readily interpretable model, we didn't include seasonality components as they didn't provide a dramatic improvement over the models without them.

Next we evaluate the demographic composition of the neighborhood in which hot spots reside. To assess the impact of different data inputs on which neighborhoods are considered hot spots, we consider the administrative context in which patrol decisions are made. Police district commands use crime intelligence and their knowledge of the community to inform patrol patterns. Crime predictions are an input to patrol pattern decisions made at the district level. Therefore, we evaluate the predicted crimes and calls for service for all geohashes that reside within a single police district boundary and classify the geohash with the highest predicted criminal activity as that police district's hot spot. Using this approach, we identify a two crime hot spots for each police district, one based on predictions using crime data, the other using calls for service data. In the table below, we compare the neighborhood demographics of these two hotspots by district:

Police District 'Hot Spot'	Pct. White		Pct. People of Color		Pct. Below Poverty Level	
	CFS	***Crime***	***CFS***	***Crime***	***CFS***	***Crime***
Central	49.2	72.3	50.8	27.7	19.8	8.1
Eastern	3.0	36.0	97.0	64.0	40.7	25.9
Northeastern	18.3	30.1	81.7	69.9	7.9	7.7
Northern	36.7	79.3	63.3	20.7	26.1	5.0
Northwestern	0.6	4.2	99.4	95.8	36.2	20.0
Southeastern	63.7	61.6	36.3	38.4	29.9	11.9
Sourthern	24.5	57.9	75.5	42.1	41.9	25.9
Southwestern	57.4	6.1	42.6	93.9	20.4	16.1
Western	0.7	1.4	99.3	98.6	21.4	47.2
Average	**28.2**	**38.8**	**71.8**	**61.2**	**27.1**	**18.6**

Figure 4: Demographic comparison of district hot spots by data type

Although there is considerable variation in hot spot demographics across police districts, consistent with patterns of residential segregation, there is a striking consistency in the difference between hot spots predicted using crime data vs. calls for service. Hot spots predicted based on calls for service data are consistently less white (10.6 pp on average) and poorer (8.5 pp on average) than hot spots predicted based on crime data. In 7 out of 9 police districts call for service hot spots are less white, and in 8 out of 9 police districts they are poorer.

VI. CONCLUSION AND DISCUSSION

Our investigation makes three important contributions to the literature on hot spot policing. First, we create a public-facing, interactive visualization of crime predictions. Typically, predictions at this level of geographic granularity are not available to the public, nor is the ability to visualize how hot spot locations change based on data inputs. Second, we outline a transparent methodology by which predictions of criminal activity are produced and provide detail on their degree of accuracy. Finally, we evaluate the demographic characteristics of hot spot neighborhoods and find that hot spots predicted using calls for service reside in significantly blacker/browner and poorer Baltimore neighborhoods.

It is important to emphasize that the crime and calls for service data that underpin this analysis attempt to make different real-world phenomena legible (likely with different degrees of fidelity) and carry with them different implications about the function of police in preserving public safety. Calls for service are more representative of where police presently spend their time. Traffic-related calls occupy a quarter of the top 20 most frequently occurring calls for service, and just 1 of the top 20 relates to violent crime ("common assault"). In contrast, the crime data exclusively capture serious property crime and violent offenses, like murder, rape, and aggravated assault, on which patrol police spend comparably little time. It is beyond the scope of this analysis to comment on which type of data should inform what spaces police occupy and what function they perform in the community. Instead, we hope that our work illustrates how the choice of data input reflects not only a subjective - and therefore, political - judgement of which real-world phenomena demand police attention, but also the implications of this choice in terms of which communities are disproportionately surveilled by police.

References:

1. Muhammad, Khalil Gibran. The condemnation of Blackness: Race, crime, and the making of modern urban America, with a new preface. Harvard University Press, 2019.
2. Richardson, Rashida and Schultz, Jason and Crawford, Kate, Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice (February 13, 2019). 94 N.Y.U. L. REV. ONLINE 192 (2019), Available at SSRN: https://ssrn.com/abstract=3333423
3. Chainey, Spencer et al. "The Utility Of Hotspot Mapping For Predicting Spatial Patterns Of Crime". Security Journal, vol 21, no. 1-2, 2008, pp. 4-28. Springer Science And Business Media LLC, doi:10.1057/palgrave.sj.8350066
4. Ridgeway, Greg. "Policing in the era of big data." *Annual review of criminology* 1 (2018): 401-419.
5. Benbouzid, Bilel. "To predict and to manage. Predictive policing in the United States." Big Data & Society 6.1 (2019): 2053951719861703.
6. Stoffel, Floarian, Hanna Post, Marcus Stewen, and Daniel A. Keim. "Polimaps: Supporting Predictive Policing with Visual Analytics." EuroVis Visual Analytics (2018). http://dx.doi.org/10.2312/eurova.20181111
7. Ferguson, Andrew Guthrie, Predictive Policing and Reasonable Suspicion (May 2, 2012). 62 Emory Law Journal 259 (2012), Available at SSRN: https://ssrn.com/abstract=2050001
8. Lum, Kristian, and William Isaac. "To predict and serve?." Significance 13.5 (2016): 14-19.
9. Lau, Tim. "Predictive Policing Explained." Brennan Center for Justice, 1 Apr. 2020, www.brennancenter.org/our-work/research-reports/predictive-policing-explained.
10. Egge, Jeff. "Experimenting with Future-Oriented Analysis at Crime Hot Spots in Minneapolis." NCJ, Mar. 2011.
11. Brayne, Sarah. "Big Data Surveillance: The Case of Policing." American Sociological Review, vol. 82, no. 5, Oct. 2017, pp. 977–1008, doi:10.1177/0003122417725865.
12. Ferguson, Andrew Guthrie. The Rise of Big Data Policing, New York, USA: New York University Press, 2017. https://doi.org/10.18574/9781479854608
13. Ridgeway, Greg. "Policing in the Era of Big Data." Annual Review of Criminology, vol. 1, Jan. 2018, pp. 401–419., www.annualreviews.org/doi/full/10.1146/annurev-criminol-062217-114209#_i24.
14. Perry, Walter L., Brian McInnis, Carter C. Price, Susan Smith, and John S. Hollywood. "Predictive Policing: Forecasting Crime for Law Enforcement." Santa Monica, CA: RAND Corporation, 2013.
15. Ferguson, Andrew Guthrie, Illuminating Black Data Policing (March 16, 2018). 15 Ohio State Journal of Criminal Law 503 (2018), Available at SSRN: https://ssrn.com/abstract=3142263
16. Mohler, George O., et al. "Randomized controlled field trials of predictive policing." Journal of the American statistical association 110.512 (2015): 1399-1411.
17. Hollywood, John S., et al. Real-time Crime Centers in Chicago: Evaluation of the Chicago Police Department's Strategic Decision Support Centers. RAND Corporation, 2019.
18. Brian Jordan Jefferson. "Predictable Policing: Predictive Crime Mapping and Geographies of Policing and Race", Annals of the American Association of Geographers, 2016, 108:1, 1-16, DOI: 10.1080/24694452.2017.1293500
19. City of Chicago, Office of the Inspector General, The Public Safety Section of the Office of the inspector General, Advisory Concerning the Chicago Police Department's Predictive Risk Models. Chicago: Office of the Inspector General, 2020. Web. https://igchicago.org/wp-content/uploads/2020/01/OIG-Advisory-Concerning-CPDs-Predictive-Risk-Models-.pdf
20. Perry, Walter L., and John Gordon IV. "Analytic Support to Intelligence in Counterinsurgencies." Santa Monica, CA: RAND Corporation, 2013.
21. Andrew G. Ferguson. "Policing Predictive Policing." 94 Wash. U. L. Rev. 1109 (2017). https://openscholarship.wustl.edu/law_lawreview/vol94/iss5/5
22. https://data.baltimorecity.gov/
23. https://predictive-policing-analysis.herokuapp.com/