



To Predict and Serve? Predictive Policing with Biased Training Data

William Isaac - PhD Candidate, Department of Political Science, MSU

Kristian Lum - Lead Statistician, HRDAG

October 2016

Algorithms = Models+ Data

Algorithms = Models+ **Data**

Why Biased Data Matters

Predictive policing uses police records to learn patterns in the occurrence of crime.

Why Biased Data Matters

Predictive policing uses police records to learn patterns in the ~~occurrence of crime.~~

**police
records**

Why Biased Data Matters

Predictive policing uses police records to learn patterns in the ~~occurrence of crime.~~

**police
records**

Using these patterns, the computer then predicts the most likely locations of future crimes.

Why Biased Data Matters

Predictive policing uses police records to learn patterns in the ~~occurrence of crime.~~

**police
records**

Using these patterns, the computer then predicts the most likely locations ~~of future crimes.~~

future recorded crimes

Why Biased Data Matters

Predictive policing uses police records to learn patterns in the ~~occurrence of crime.~~

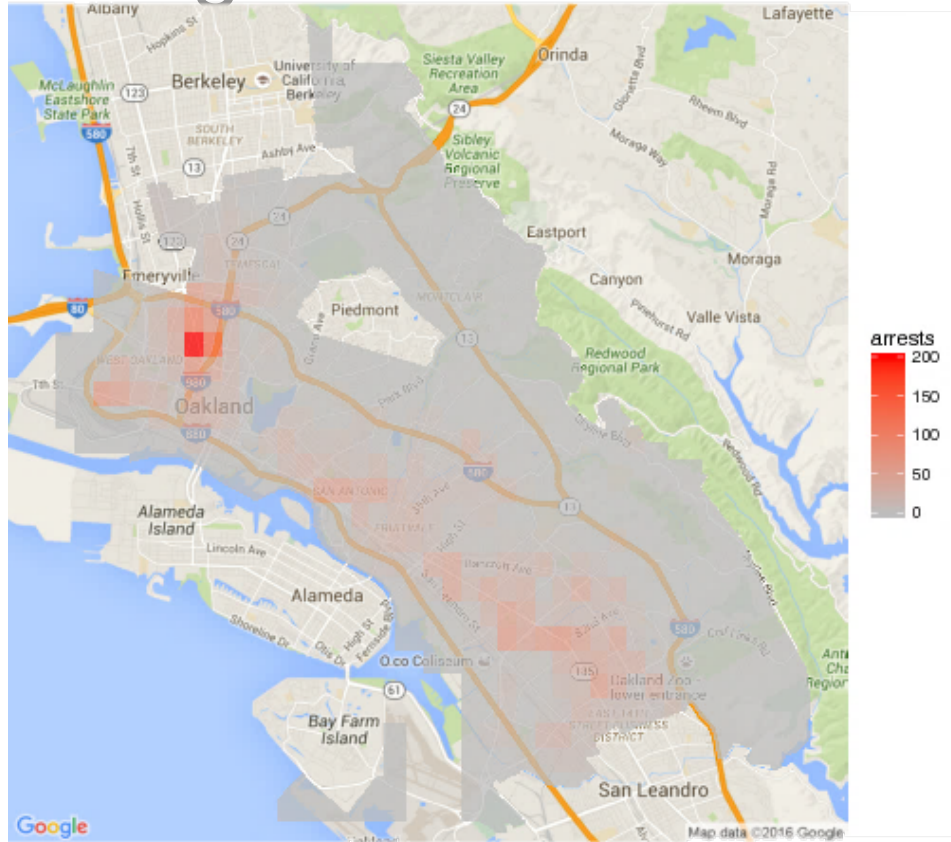
**police
records**

Using these patterns, the computer then predicts the most likely locations ~~of future crimes.~~

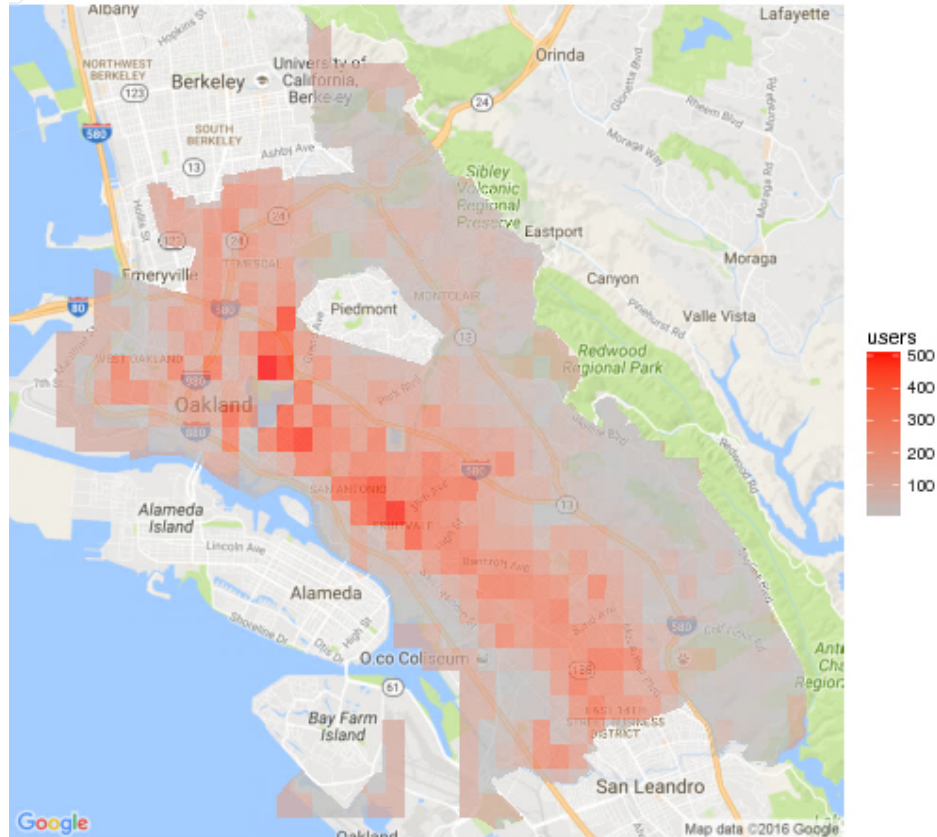
future recorded crimes

As police are then dispatched to the locations with the highest predicted rate of crime, new recorded crimes **validate the initial biased predictions and can amplify their influence in subsequent forecasting.**

2010 Drug Crimes in Oakland, CA



Estimated Number of Drug Users



What did we do?

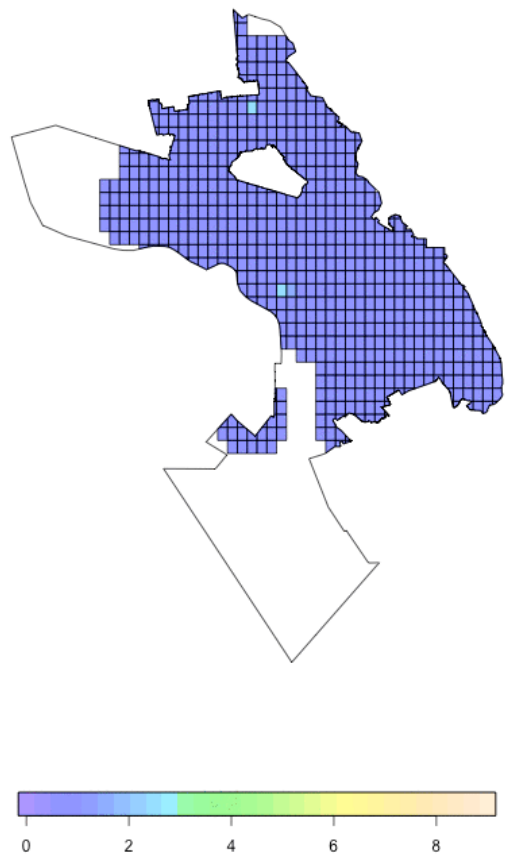
- We apply the predictive policing algorithm given in [3] to drug crimes documented by the Oakland PD in 2011.
 - We begin by observing the locations of drug crimes— this is our initial training data.
 - Once the predictive policing begins, each day the algorithm predicts which locations are likely to contain the most drug crime based on the previous six months' police records.
 - These locations are targeted for further police monitoring and are shown in red.
- In following example, we use the unaltered Oakland PD data, i.e. we investigate what would happen under the assumption that additional policing does not lead to increased rate of arrests in the targeted locations.

Targeted

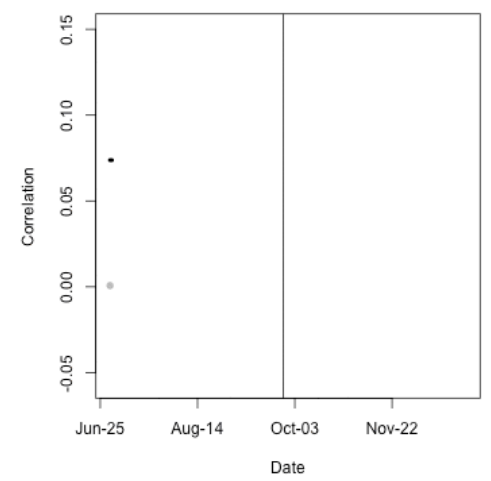
2011-Jun-30

Applying predictive policing

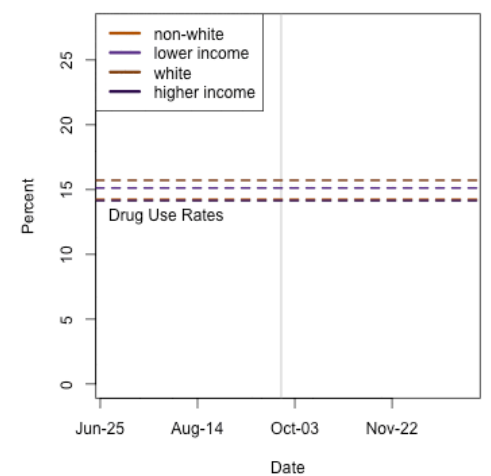
- The algorithm directs additional policing to the over-policed areas.
- Biases that previously existed in the police database are confirmed.
- Minority and low-income communities receive a disproportionate amount of targeted policing.



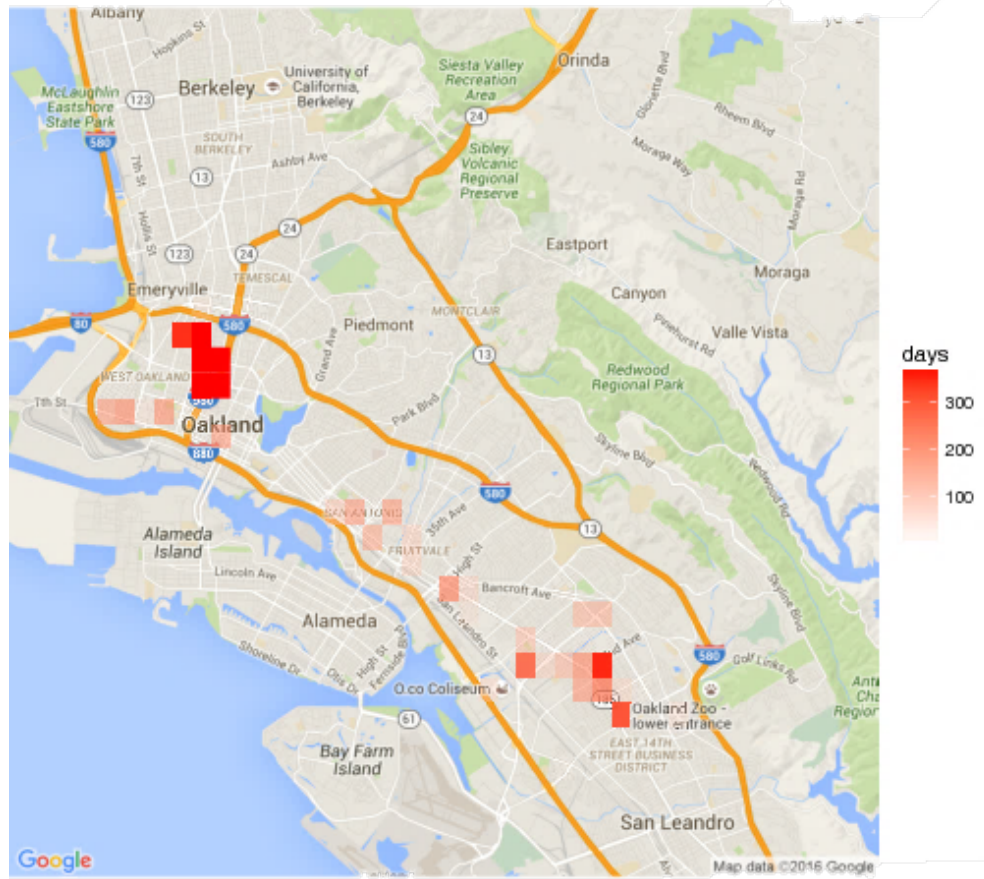
Correlation Between % Low Income & Arrest Rate



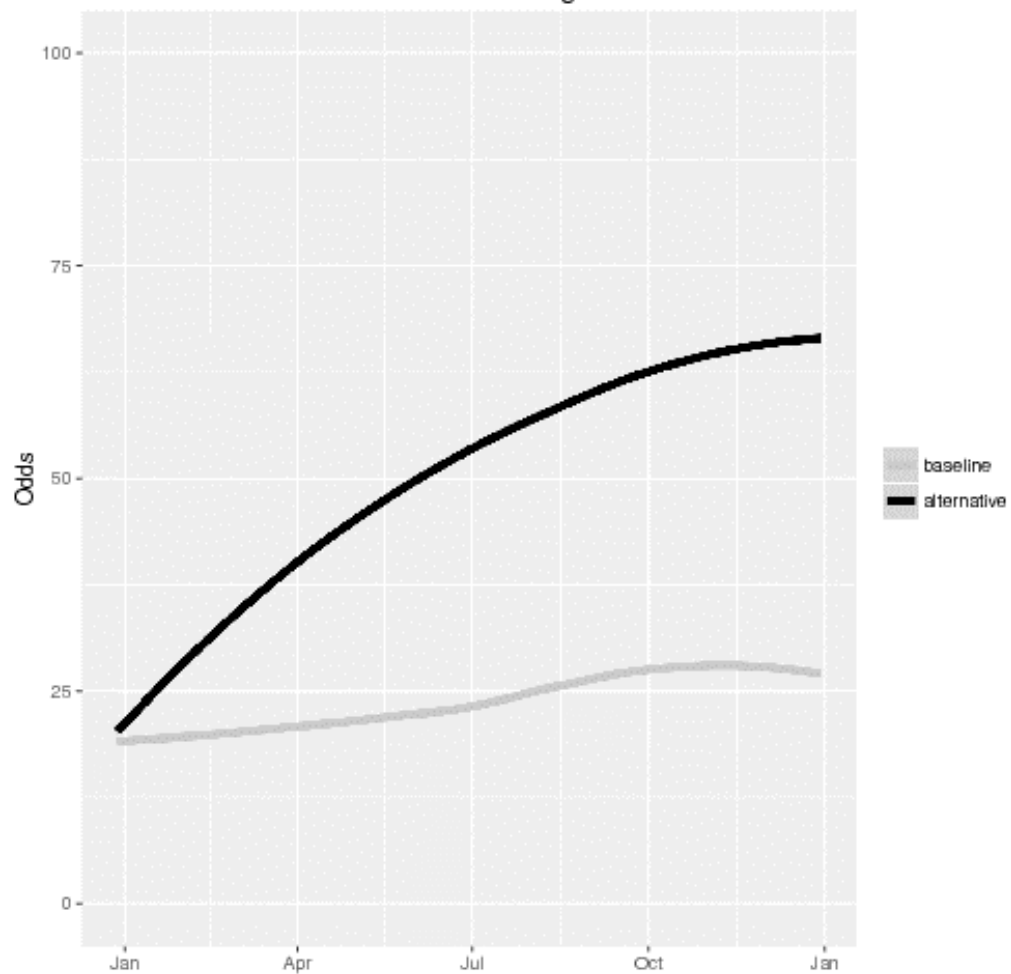
% Experiencing Targeted Policing



Number of Days with Targeted Policing



Predicted Odds of Crime in Targeted Locations



Summary

- Crime is everywhere, and police data on crime is not representative of all crime.
- Biased data leads to biased predictions in any machine learning algorithm— predictive policing included.
- Predictive policing reproduces the biases that already exist in the police data in a much more precise and targeted way.
- If increased targeted policing leads to an increase in the number of arrests, targeted policing will magnify pre-existing biases.
- Because of the biases in the existing police data, predictive policing will disproportionately affect affect historically over-policed communities.



Thank You!

From me and **from the team at HRDAG.**

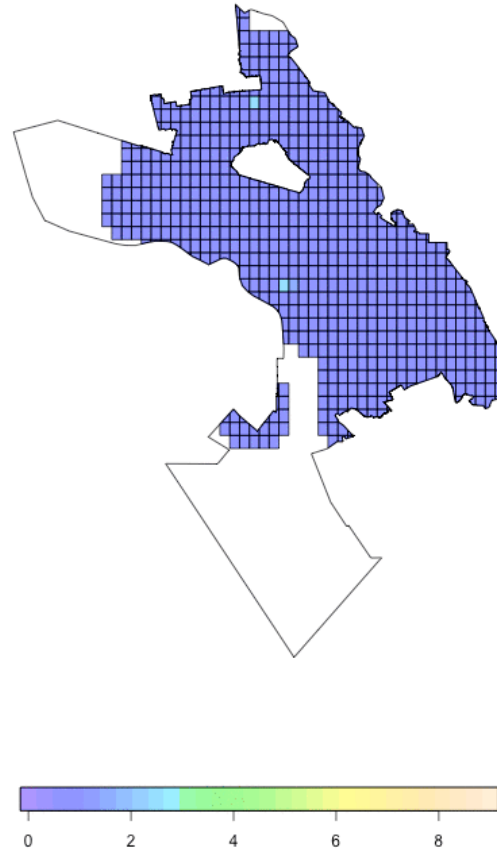
contact: isaacwil@msu.edu

Targeted

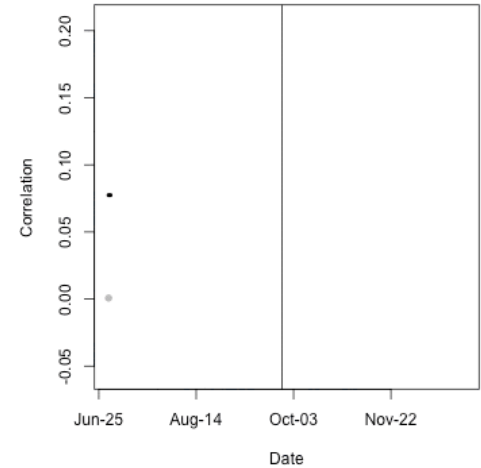
2011-Jun-30

Simulating an alternative scenario

- We assume that sending targeted policing to an area results in a slight increase in the number of drug crimes discovered by police.
- Biases in the data are *amplified*.
- Minority and low-income communities receive a disproportionate amount of targeted policing.



Correlation Between % Low Income & Arrest Rate



% Experiencing Targeted Policing

