Redlining and Predictive Policing: Using historic crime data to prove racial bias

Rojin Bakhti, Leon Durrenberger, Alice Gusev

Outline

Background & Motivation

Goals

Related Work

Data Exploration

Procedure

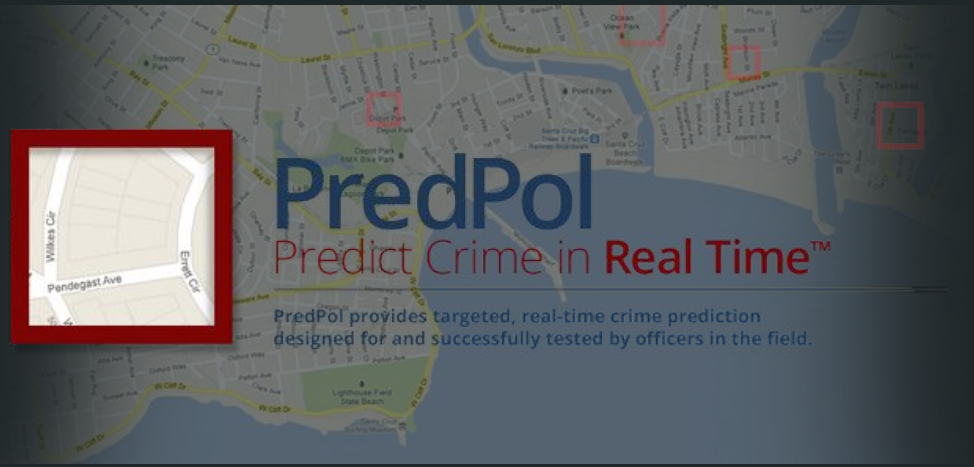
Analysis

Future Work

Background and Motivation

Predictive Policing

- Over-policing is a product of systematic and historic racism
- Stanford Open Policing Project 2020 showed disproportionate traffic stops among POC communities
- The LAPD has used a predictive policing tool (PredPol)



Redlining

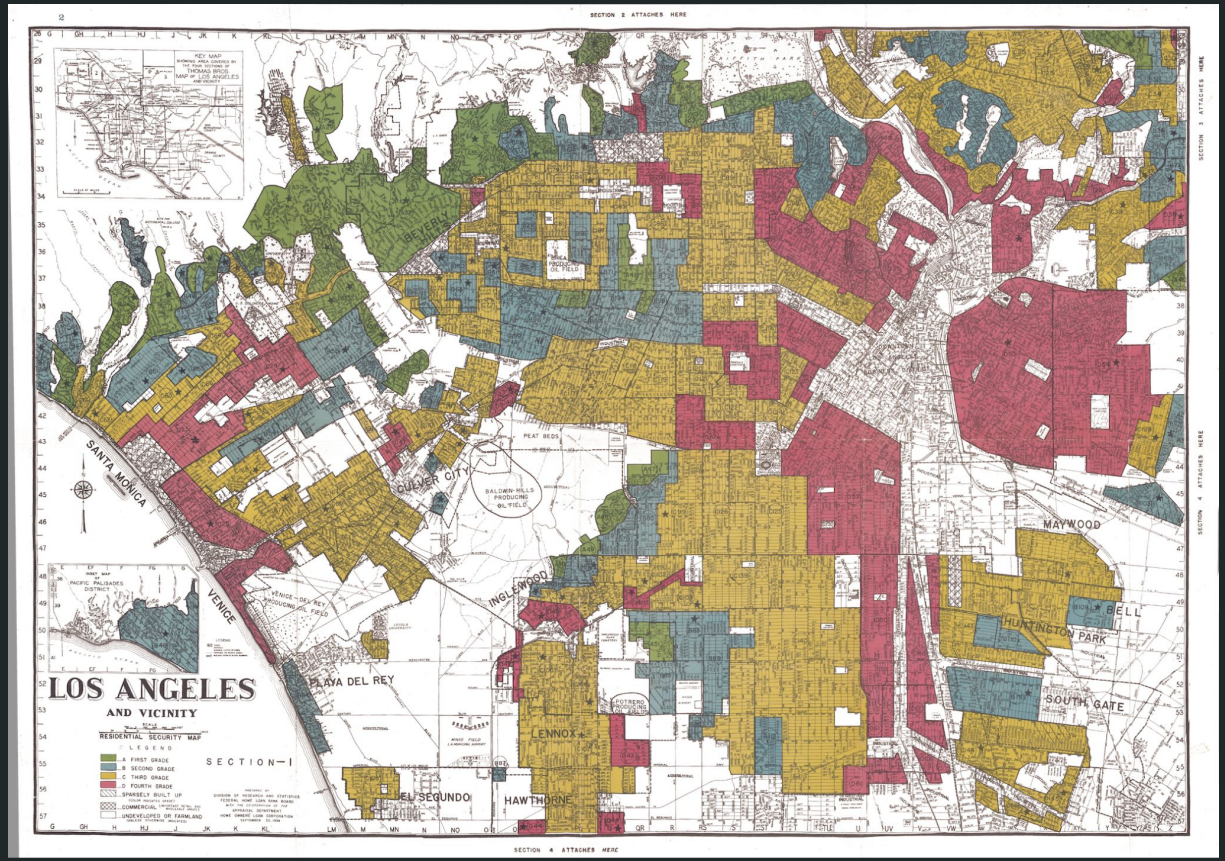
“The presumed practice of mortgage lenders of drawing red lines around portions of a map to indicate areas or neighborhoods in which they do not want to make loans.”

Home Owner’s Loan Corp’ (HOLC) mapped areas as Green(“best”), Blue(“still desirable”), Yellow(“definitely declining”), and Red for (“hazardous.”)

Based On Race, Income, Religion

Repercussions

Home Ownership = Generational Wealth and Long Term Efffects


SECTION 2 ATTACHES HERE
SECTION 3 ATTACHES HERE
SECTION 4 ATTACHES HERE
SECTION 4 ATTACHES HERE
KEY MAP
SHOWING AREA COVERED BY THIS SECTION OF THOMAS BROS. MAP OF LOS ANGELES AND VICINITY
LOS ANGELES
AND VICINITY
RESIDENTIAL SECURITY MAP
SECTION—I
LEGEND
A FIRST GRADE
B SECOND GRADE
C THIRD GRADE
D FOURTH GRADE
SPARSELY BUILT UP
COMMERCIAL
UNDEVELOPED OR FARMLAND
SANTA MONICA
VENICE
PACIFIC PALISADES DISTRICT
BEVERLY HILLS
CULVER CITY
PEAT BEDS
BALDWIN HILLS PRODUCING OIL FIELD
INGLEWOOD
VENICE-DEL REY PRODUCING OIL FIELD
PLAYA DEL REY
EL SEGUNDO
HAWTHORNE
LENNOX
MAYWOOD
BELL
HUNTINGTON PARK
SOUTH GATE

Goals

Goals

1) To implement and analyze ETAS (algorithm used by Predpol)
2) ARIMA for crime rate prediction in 10 Los Angeles regions (5 red-lined vs. 5 non-redlined)
3) Visualize the parities found among these models, and overlay results with historically red-lined zones

Related Work

Related Work

- “Randomized Controlled Field Trials of Predictive Policing” by Mohler and Malinowski (2016).

- K. Lum and W. Isaac, "To predict and serve?", *Royal Statistical Society* (2018).

Data Exploration

Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	LURN_SAK	INCIDENT_DATE	INCIDENT_REPORTED_DATE	CATEGORY	STAT	STAT_DESC	ADDRESS	STREET	CITY	ZIP	INCIDENT_ID	REPORTING_D	SEQ	GANG_RELATE	UNIT_ID	UNIT_NAME	LONGITUDE	LATITUDE	PART_CATEGORY
2	18692378	09/09/2018 12:00:00 AM	09/09/2018	VEHICLE / BOA	250	VEHICLE AND E	44800 SIERRA I	44800 SIERRA I	LANCASTER		918-18370-1122	1122	18370	NO	CA0190024	LANCASTER	-118.136	34.696	2
3	18691063	09/07/2018 3:17:00 AM	09/07/2018	BURGLARY	62	BURGLARY, RE	26000 BATES P	26000 BATES P	STEVENSON R	91381	918-12897-0661	661	12897	NO	CA0190006	SANTA CLARIT	-118.593	34.394	1
4	18577911	05/13/2018 6:15:00 PM	05/13/2018	DRUNK / ALCOI	201	DRUNK: Alcohol	800 W RAYMON	800 W RAYMON	COMPTON	90220	918-06659-2833	2833	6659	NO	CA0190042	COMPTON	-118.238	33.887	2
5	18714783	09/30/2018 4:00:00 AM	10/02/2018	GRAND THEFT	94	GRAND THEFT	1000 LAS LOMA	1000 LAS LOMA	DUARTE	91010	918-80174-0542	542	80174	NO	CA0190005	TEMPLE	-117.946	34.14	1
6	18578181	05/14/2018 3:38:00 AM	05/14/2018	GRAND THEFT	91	GRAND THEFT	500 SUNSET AV	500 SUNSET AV	LA PUENTE	91744	918-06368-1430	1430	6368	NO	CA0190014	INDUSTRY	-117.968	34.04	1
7	18639785	07/15/2018 8:00:00 PM	07/15/2018	NON-AGGRAVA	146	ASSAULT, NON	1300 CORDON	1300 CORDON	LOS ANGELES	90023	018-10585-0273	273	10585	NO	CA0190002	EAST LOS ANG	-118.174	34.054	2
8	18449603	01/07/2018 4:00:00 PM	01/07/2018	NARCOTICS	185	Misdemeanor Pc	17200 LAKEWO	17200 LAKEWO	BELLFLOWER	90706	918-00484-1335	1335	484	NO	CA0190013	LAKEWOOD	-118.14	33.875	2
9	18594227	05/26/2018 3:00:00 AM	05/29/2018	AGGRAVATED /	50	ASSAULT, AGG	SIERRA HWY A	SIERRA HWY A	PALMDALE		018-07879-2649	2649	7879	NO	CA01900W9	PALMDALE	-118.114	34.557	1
10	18539834	03/17/2018 11:17:30 AM	04/06/2018	FRAUD AND NS	113	FRAUD: Fraud E	10900 IMPERIA	10900 IMPERIA	NORWALK	90650	918-05385-0451	451	5385	NO	CA0190004	NORWALK	-118.101	33.918	2
11	18574822	05/10/2018 12:50:00 AM	05/10/2018	VANDALISM	261	VANDALISM MI	38700 20TH ST	38700 20TH ST	PALMDALE	93550	918-06772-2610	2610	6772	NO	CA01900W9	PALMDALE	-118.094	34.583	2
12	18620758	06/25/2018 12:00:00 AM	06/25/2018	VANDALISM	261	VANDALISM MI	800 AVENIDA L	800 AVENIDA L	SAN DIMAS	91773	918-03331-0817	817	3331	NO	CA0190008	SAN DIMAS	-117.821	34.088	2
13	18607270	06/11/2018 1:52:30 PM	06/11/2018	LARCENY THEI	340	VEHICLE BURG	1301 AVENIDA	1301 AVENIDA	MONTEREY PA	91754	918-00365-8724	8724	365	NO	CA01900R5	COMMUNITY C	-118.15	34.04	1
14	18618873	06/19/2018 8:30:00 PM	06/22/2018	LARCENY THEI	389	THEFT, PETTY:	4700 COMMON	4700 COMMON	CALABASAS		918-03736-2245	2245	3736	NO	CA0190022	MALIBU/LOST H	-118.644	34.154	1
15	18573509	05/08/2018 9:30:00 PM	05/08/2018	MISDEMEANOF	399	MISDEMEANOF	1800 E AVENUE	1800 E AVENUE	PALMDALE	93550	918-06696-2608	2608	6696	NO	CA01900W9	PALMDALE	-118.095	34.568	2

Data Exploration

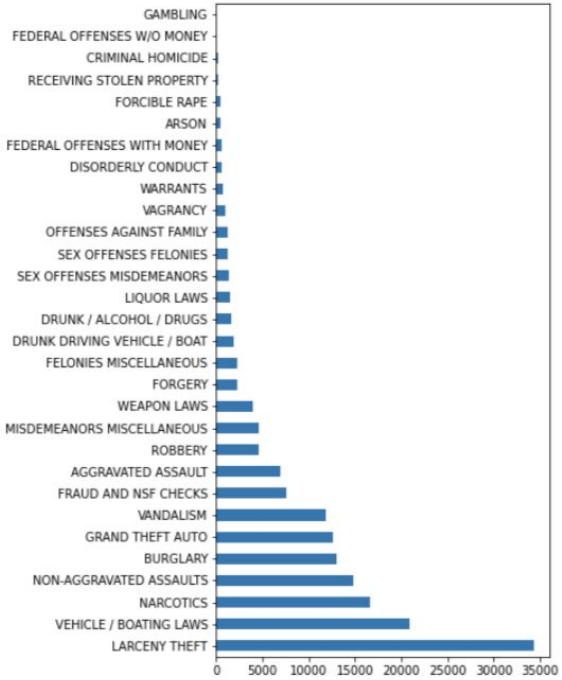


Figure 1 - Categories of crime and relative frequency in 2017

Data Exploration

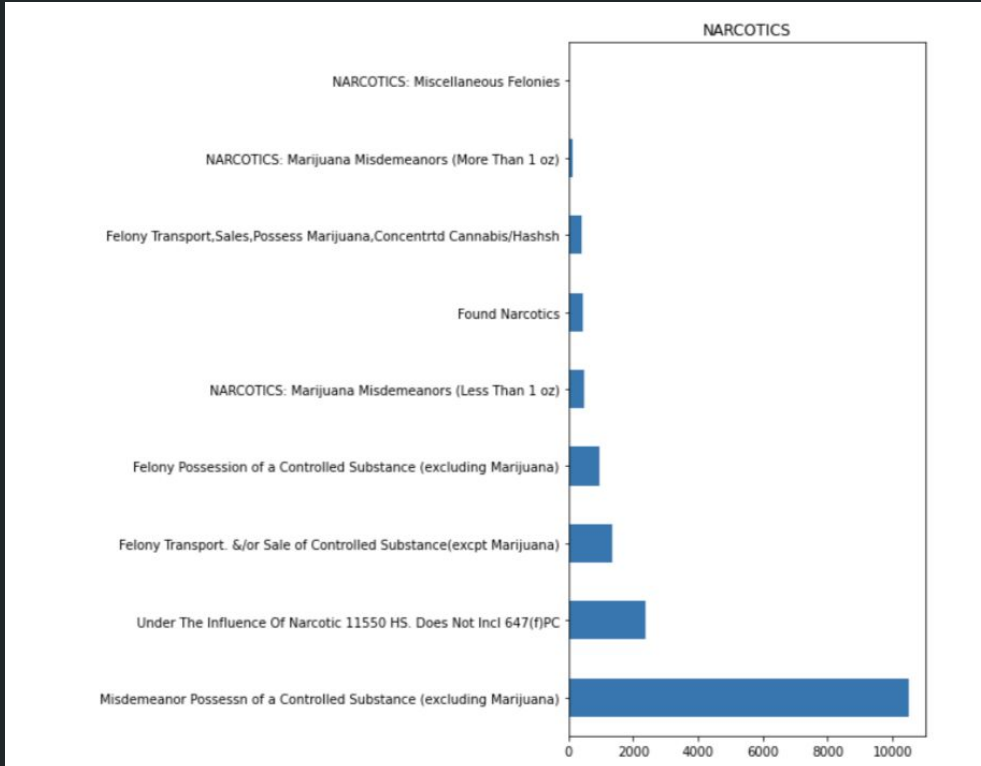


Figure 2 - Narcotics crime details

Data Exploration

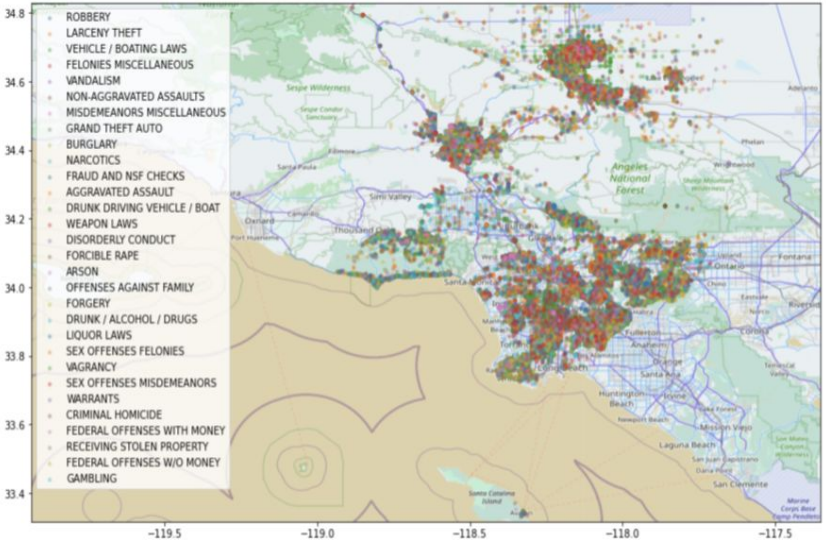


Figure 3 - Mapping of all crime in 2017

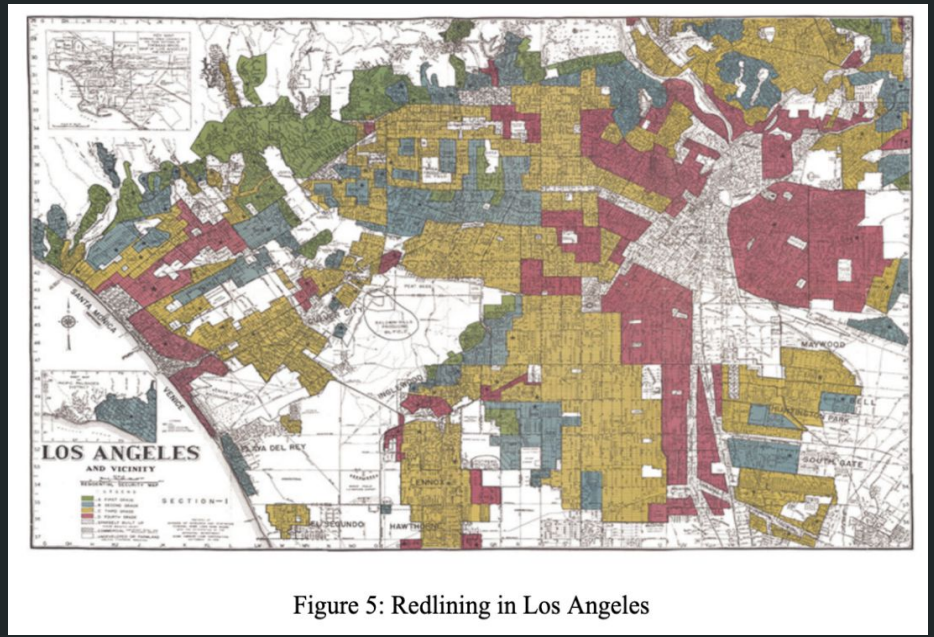
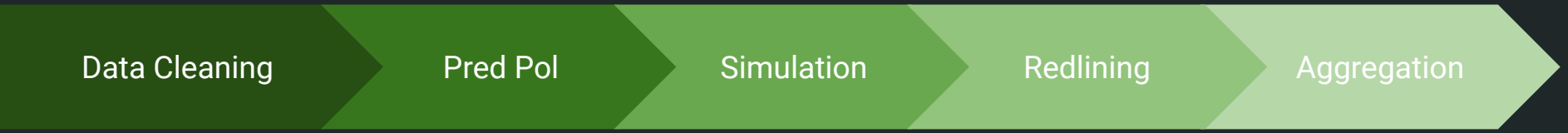


Figure 5: Redlining in Los Angeles

Procedure

Procedure



Data Cleaning	Pred Pol	Simulation	Redlining	Aggregation
Parse, clean, and edit dataset for use	Run the predictive policing algorithm	Use results to simulate LAPD responses	Interpret simulation in context of redlining	Combine data to produce visualization

Data Cleaning

- Need to turn two years of LAPD crime records into clean dataset
- Divided Los Angeles into 150x150m bins
- Translated each longitude/latitude pair into specified bin
- Isolated only Narcotics category
- Final dataset:
  - ID
  - Bin
  - Date
  - Lag

import > input > drug_crimes_with_bins.csv

```
id,bin,OCCURRED,LAG
12353,6294,01/01/17,0
4191,2879,01/01/17,0
13814,5797,01/01/17,0
4820,3239,01/01/17,0
2642,1964,01/01/17,0
10465,5644,01/01/17,0
9024,5066,01/01/17,0
13791,392,01/01/17,0
10473,1271,01/01/17,0
5408,631,01/01/17,0
812,153,01/01/17,1
9258,463,01/01/17,0
71,71,01/01/17,1
4034,918,01/01/17,0
12043,1385,01/01/17,0
8064,434,01/01/17,0
2396,305,01/01/17,0
15122,7149,01/01/17,0
11845,392,01/01/17,0
7883,4633,01/02/17,0
660,208,01/02/17,0
7317,4405,01/02/17,0
2784,2040,01/02/17,0
```

Pred Pol

- Ran dataset through Pred Pol ETAS algorithm
- Poisson distribution - epidemic model
- Sliding window technique
  - Use past 365 days of data to predict where crimes would occur
  - Typically done once per day
- Output odds of crime for every day

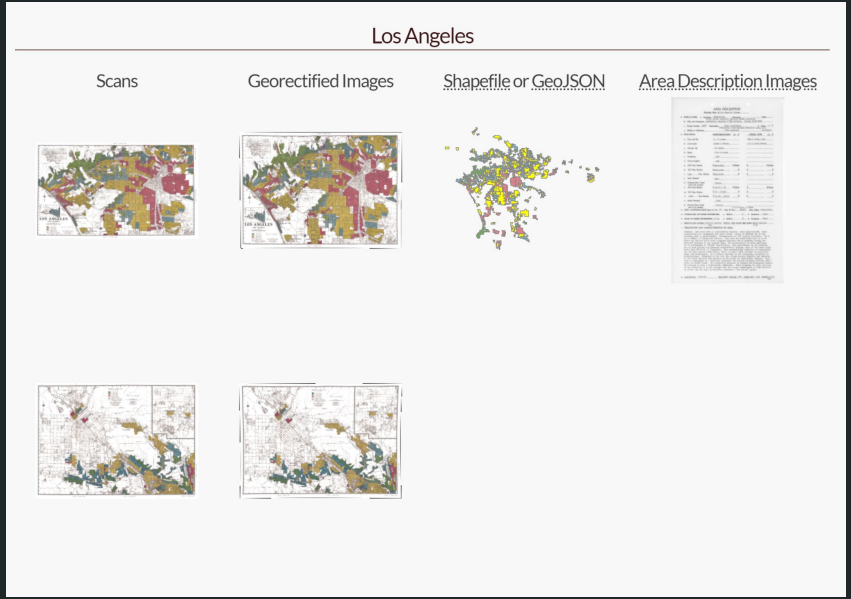
Simulation

- Generate predictions of where LAPD would go based on odds
- Ambiguous threshold for odds of crime, send police if over threshold
- Repeat for every day of 2018
- Generate how often police would be sent for every bin over course of year

	key	freq	bin
0	3	230	648349
1	53	95	615325
2	83	60	731488
3	152	338	714245
4	170	37	833721

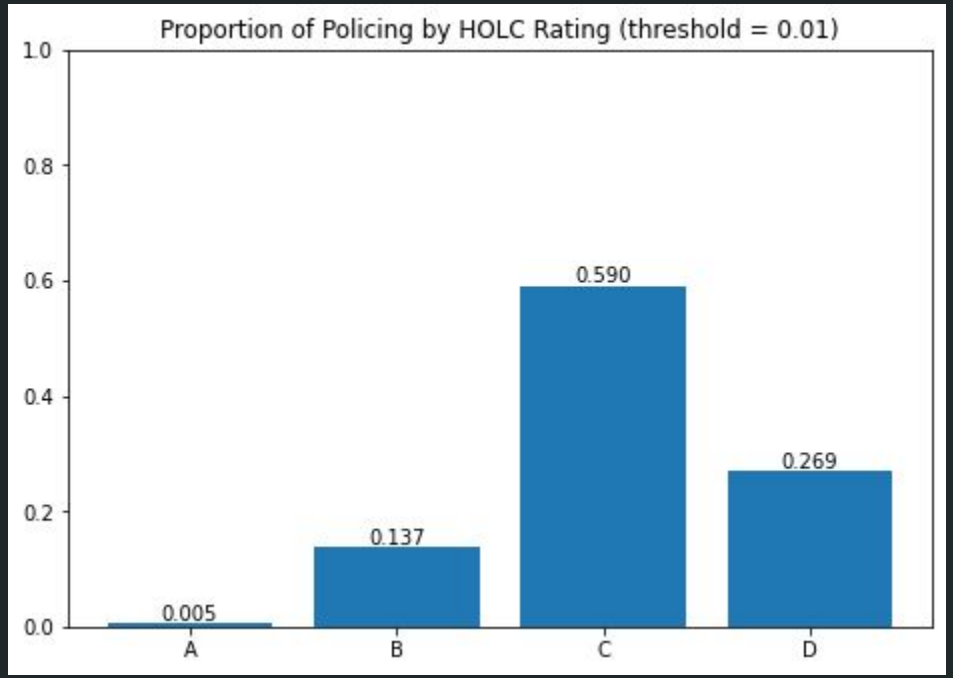
Redline Contextualization

- Use University of Richmond “Mapping Inequality” GeoJSON dataset to see what HOLC ratings certain neighborhoods of LA received
- Convert bins back into latitude/longitude and determine what rating each bin belongs to
- Aggregate frequency of police simulation by HOLC rating and see results!

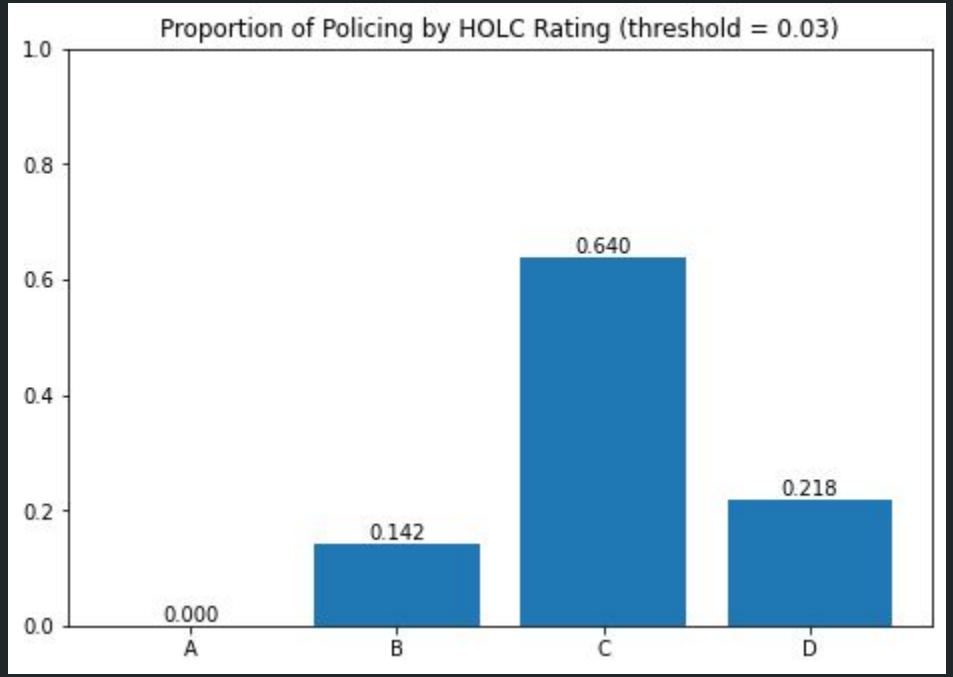


Analysis

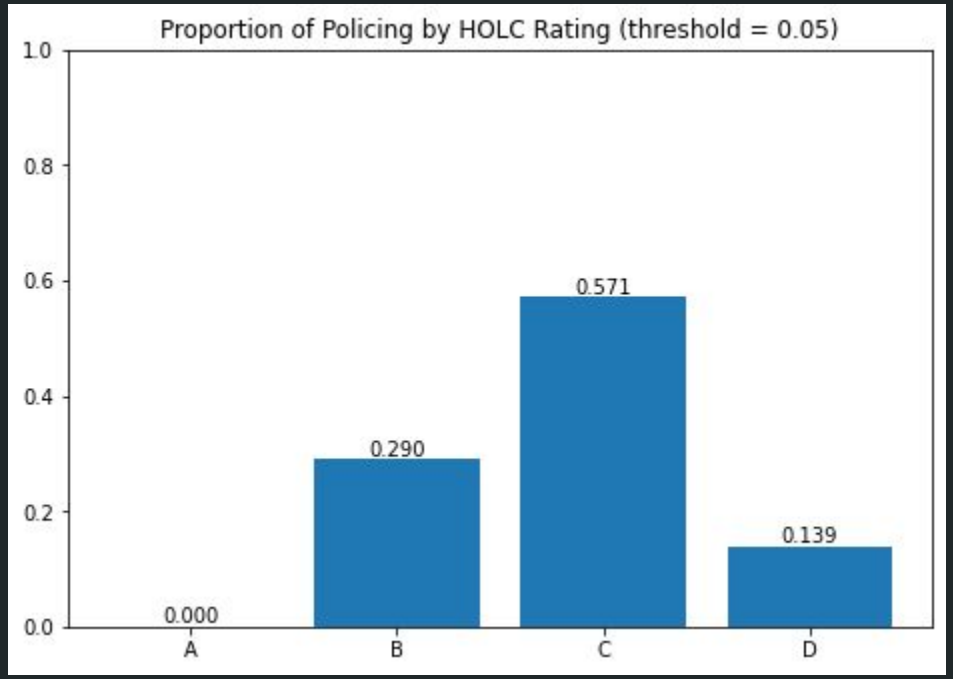
Impact of Redlining on Policing



Impact of Redlining on Policing



Impact of Redlining on Policing



map options

Search for city

+ -

Los Angeles, CA

Areas by Grade

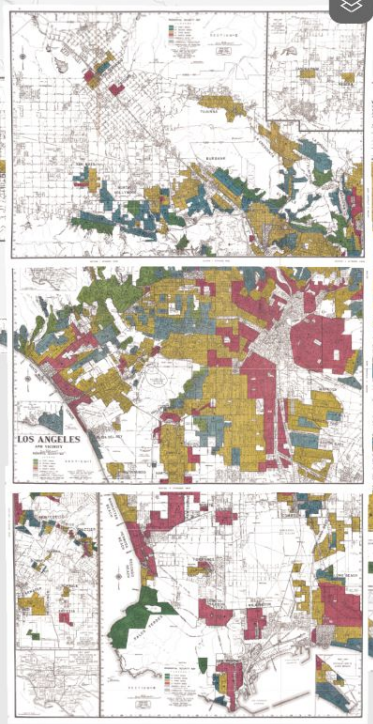
Area	Grade
12%	A "Best"
20%	B "Still Desirable"
46%	C "Definitely Declining"
22%	D "Hazardous"

Demographics

1,504,277	Total Population (1940)
14.3%	Foreign-born white
0.6%	Foreign-born Japanese

Area Descriptions

click to select



Los Angeles

Further Analysis

Question: Using HOLC Ratings generated by our Predpol results, are some historically red-lined areas over-policed unfairly?

Predpol Analysis: Generated Frequency of Police Deployment over time

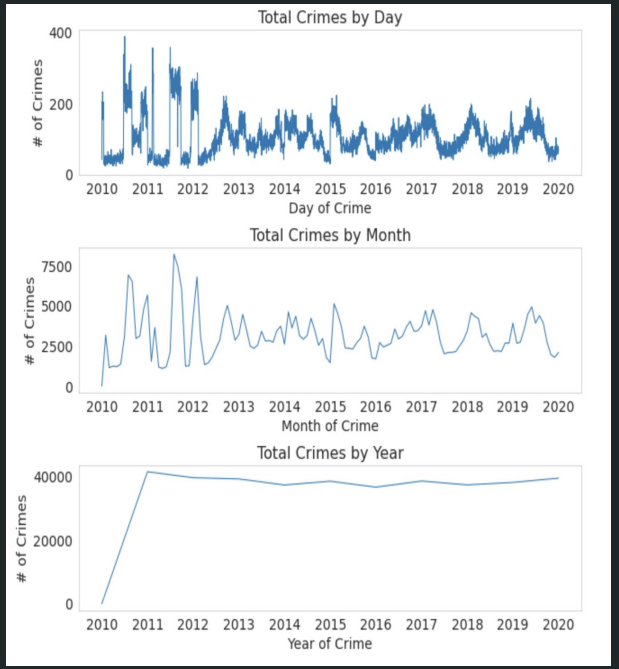
Using Bias Tests from COMPASS and Traffic Stop Papers

Goal:

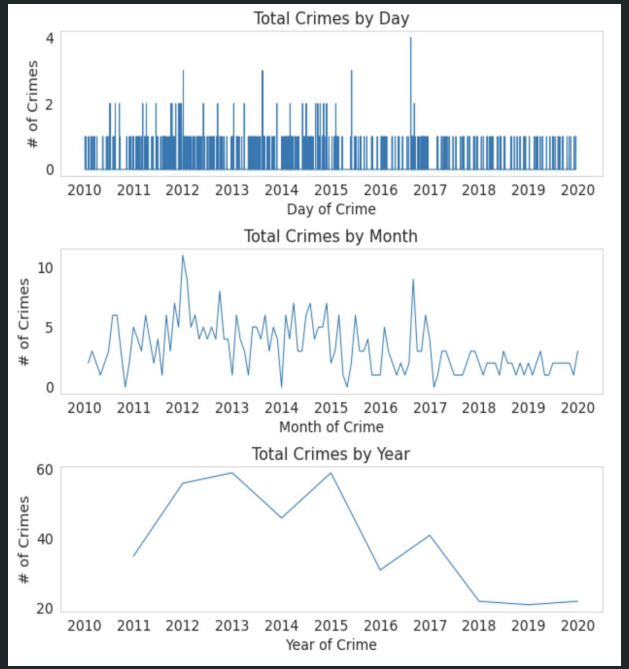
(1) Compare policing deployment frequency with actual crime statistics
(2) Find Discrepancies among different red-lined districts
(3) Analyze disparate impact among demographics

Crime Statistics

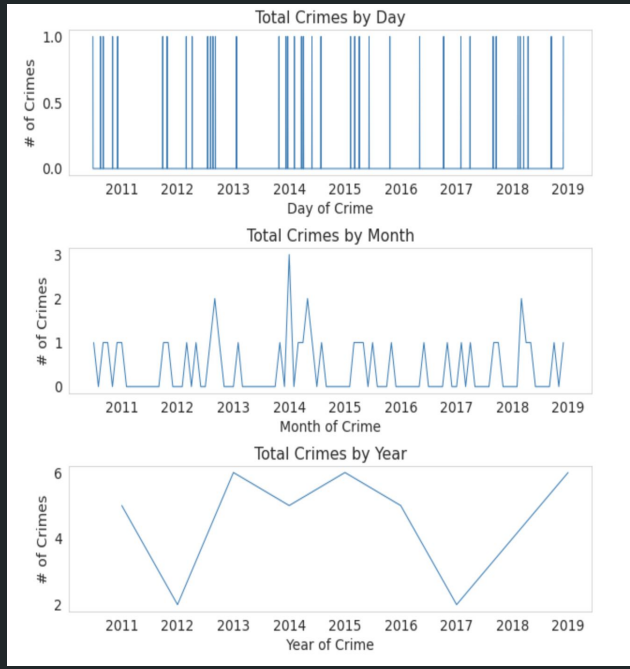
Aggregate

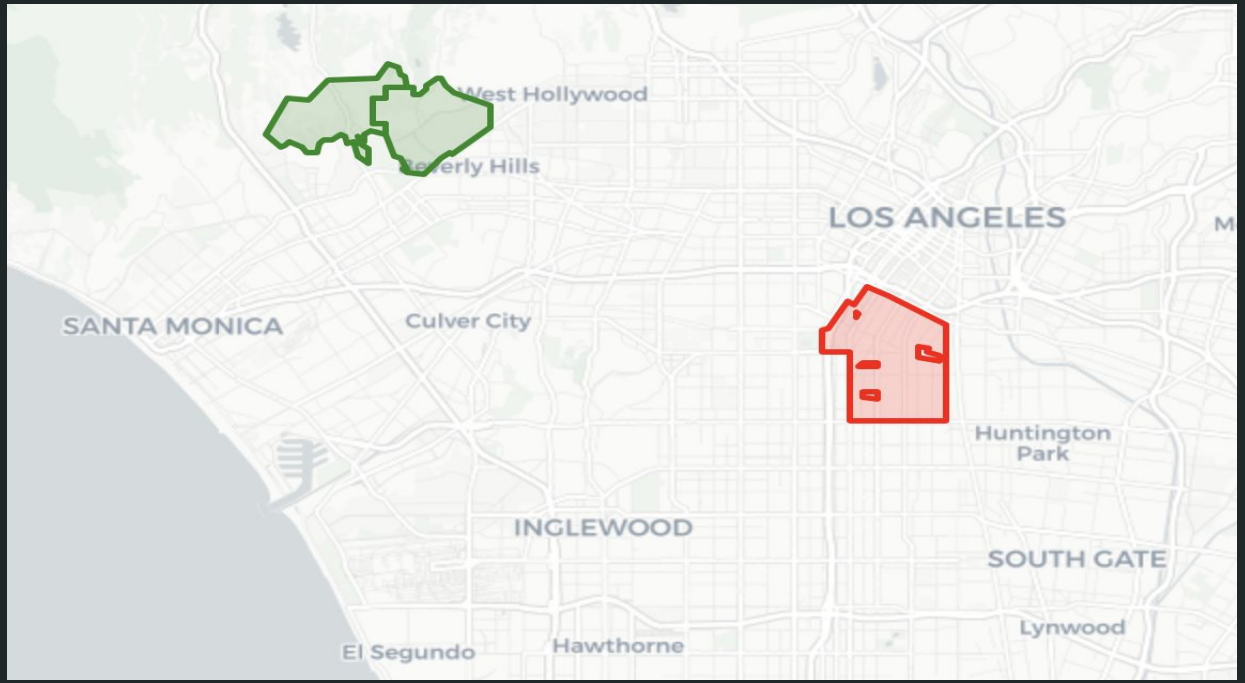


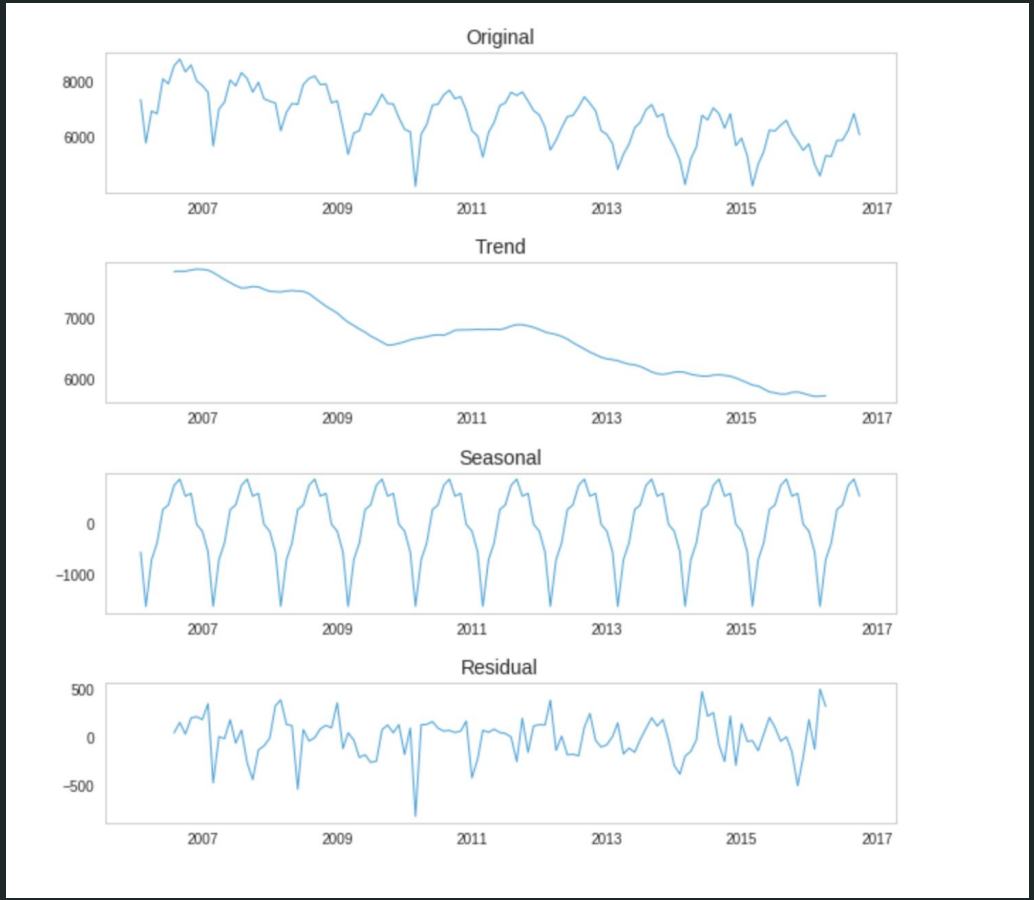
Red Zone



Green Zone




West Hollywood
Beverly Hills
LOS ANGELES
SANTA MONICA
Culver City
Huntington Park
INGLEWOOD
SOUTH GATE
Lynwood
El Segundo
Hawthorne


Original
8000
6000
2007
2009
2011
2013
2015
2017
Trend
7000
6000
2007
2009
2011
2013
2015
2017
Seasonal
0
−1000
2007
2009
2011
2013
2015
2017
Residual
500
0
−500
2007
2009
2011
2013
2015
2017

Future Work

What’s left?

- Better visualize results with mapping libraries
- Incorporate census data to investigate relationships between predictive policing and race
- ARIMA model for additional model comparison
- Fairness metrics and comparison with baseline drug distribution along race/class

Extensions

- Epidemic modeling
  - information, pathogens,
- Extensible to other cities
- Public pressure to stop predictive policing
- Understanding historical factors on today’s society