Policing Lab Research

Mengxuan Yan and Niki Vasan

Big Picture Goal:

Merge Census data with the pre-existing officer-involved shooting dataset (OIS_master)

More Specifically...

Find the smallest geographical unit (Census block) where each shooting incident occurred and gather relevant demographic features of the units from Census Bureau data

OIS_master dataset

4086 officer-involved shootings in 37 cities in the US

- Incident ID
- City
- Date
- Time
- *Location (Address)
- *Coordinates (lat/long)
- Offender ID
- *Offender Race
- Offender Fatality (Fatal / Struck / Not-Struck)
- *Officer Race
- *Officer Fatality (Fatal / Struck / Not-Struck)

* many missing values/incorrectly formatted observations

Basic Process

1

Block-level GEOIDs

Use the Census and FCC APIs to get GEOIDs for every shooting location based off of address and coordinates respectively

2

Census Demographic Information

Use the get_decennial() function in the *tidycensus* package to get block-level demographic info for every block (identified by a 15-digit GEOID) in every county/state combination that had a shooting incident

3

Join the two datasets

Join the OIS_master data with the block-level Census data so that each shooting incident has relevant demographic features

***Details will be explained in later slides**

Methods

Stage 1

Gathering Census Information

- Used the get_decennial() function from the package *tidycensus* (API call to the decennial Census) on each county-state combination individually for a total of the 37 cities and obtained block-level data from the Decennial Census: Summary File 1 (2010), sf1.
- Used variables beginning with “P” which give us population info down to the block level
- Each observation in the result is a block that has a 15-digit GEOID as the identifier and variables we want

```
# San Francisco CA
San_Francisco <- get_decennial(geography = "block", county = "San Francisco", state = "ca",
                               variables = c(totPop = "P001001",
                                             race = "P003001",
                                             totBlack  = "P003003",
                                             totalHispanic = "P004001",
                                             raceAlloc = "P046001",
                                             male = "P012002",
                                             maleBlack = "P012B003"),
                               year = 2010)
San_Francisco<- San_Francisco  %>% spread("variable", "value")
write.csv(San_Francisco, "San_Francisco.csv")
```

	GEOID	NAME	birac	blackhisp	blackmale	female	hispmale	male
1	060750101001000	Block 1000, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	0
2	060750101001001	Block 1001, Block Group 1, Census Tract 101, San Fra...	0	0	0	13	1	31
3	060750101001002	Block 1002, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	0
4	060750101001003	Block 1003, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	0
5	060750101001004	Block 1004, Block Group 1, Census Tract 101, San Fra...	0	0	0	1	0	0
6	060750101001005	Block 1005, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	0
7	060750101001006	Block 1006, Block Group 1, Census Tract 101, San Fra...	0	0	0	1	0	1
8	060750101001007	Block 1007, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	0
9	060750101001008	Block 1008, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	0
10	060750101001009	Block 1009, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	1
11	060750101001010	Block 1010, Block Group 1, Census Tract 101, San Fra...	0	0	0	1	0	0
12	060750101001011	Block 1011, Block Group 1, Census Tract 101, San Fra...	0	0	0	1	0	0
13	060750101001012	Block 1012, Block Group 1, Census Tract 101, San Fra...	0	0	0	0	0	0

Ex: San Francisco

Stage 1 (continued)

Gathering Census Information

- Repeating this process individually for each of the 37 cities was tedious
- Created an efficient and replicable version of the code where we looped over vectors containing the counties and respective states and assigned the result to the city names
- Saved all the cities to a list to later join with the OIS_master

```
# retrieve data from sf1
for(i in 1:length(county)){
  assign(cityname[i], get_decennial(geography = "block", county = county[i], state = state[i],
                                    variables = c(totpop = "P001001",
                                                  male = "P012002",
                                                  female = "P012026",
                                                  blackmale = "P012B003",
                                                  hispmale = "P012H002",
                                                  totblack = "P003003",
                                                  tothisp = "P004003",
                                                  totwhite = "P008003",
                                                  totasian = "P003005",
                                                  totnative = "P008005",
                                                  birac = "P008011",
                                                  blackhisp = "P005012",
                                                  toturban = "P002002",
                                                  singleparent_f = "P018006",
                                                  singleparent_m = "P018005"),
                                    year = 2010, geometry = TRUE, summary_var = "P001001")
         %>% spread(key = "variable", value = "value"))
}
```

Ex of more efficient get_decennial() call

Stage 2

Finding the GEOID

- First tried the US Census Geocoder which takes an address batch as an input and returns the GEOID of each address. However, many addresses were formatted incorrectly so we had around 1600 NA values for the GEOID column (40%)

Correctly Formatted Addresses

3653	13581	180014	12/29/10	15:40	626 Bacon St
3654	13583	180015	1/4/11	10:33	1380 Howard St
3655	13585	180016	6/7/11	17:23	65 Buena Vista East
3656	13586	180016	6/7/11	17:23	65 Buena Vista East
3657	13587	180017	6/29/11	7:55	1155 Ellis St
3658	13588	180017	6/29/11	7:55	1155 Ellis St

Incorrectly Formatted Addresses

13653	180047	1/4/15	17:21	Mission Station
13654	180048	2/26/15	21:47	2800 block of Folsom St
13655	180048	2/26/15	21:47	2800 block of Folsom St
13656	180049	3/17/15	19:07	1500 block of Van Ness Ave at Pine St
13657	180049	3/17/15	19:07	1500 block of Van Ness Ave at Pine St
13658	180050	7/26/15	7:00	Collector Ramp, San Bruno Ave/US-101 NB
13659	180051	9/18/15	8:00	San Mateo, California

Ex: San Francisco

Stage 2 (continued)

Finding the GEOID

- Then used the Federal Communications Commission (FCC) geocoding API which takes coordinates as an input and
- This cut the number of NAs in the GEOID column down to 559

Incorrect address with coordinates

City	Date	Time	Location	Lat	Long	GEOID
San Frar	All	All	All	All	All	All
San Francisco	4/12/10	17:00	1100 block of Folsom	37.76309	-122.4219	6.08e+13
San Francisco	7/3/10	1:31	Columbus and Grant St	37.79818	-122.4071	6.08e+13
San Francisco	7/3/10	1:31	Columbus and Grant St	37.79818	-122.4071	6.08e+13
San Francisco	7/24/10	5:30	700 block of Arkansas	37.75722	-122.3982	6.08e+13
San Francisco	8/28/10	21:43	Reardon and Kiska Rd	37.73000	-122.3772	6.08e+13

Incorrect address with no coordinates (559)

City	Date	Time	Location	Lat	Long	GEOID
San Frar	All	All	All	All	All	All
San Francisco	7/26/15	7:00	Collector Ramp, San Bruno Ave/US-101 NB	NA	NA	NA
San Francisco	9/18/15	8:00	San Mateo, California	NA	NA	NA
San Francisco	10/24/15	8:00	Main Gate, Ave of the Palms, Treasure Island	NA	NA	NA
San Francisco	10/24/15	8:00	Main Gate, Ave of the Palms, Treasure Island	NA	NA	NA

Ex: San Francisco

Stage 2 (continued)

Finding the GEOID

- Finally used the *ggmaps* package with the Google Maps API to find coordinates of remaining data points that had incorrectly formatted addresses and no coordinates
- Then used FCC API to find GEOID; this resulted in only 3 missing GEOIDs
  - one in Philadelphia and two cities that were unspecified

City	Date	Time	Location	Lat	Long	GEOID
mento	All	All	All	All	All	All
Sacramento	3/15/2010	20:35	2000 Block of Oneil Way	NA	NA	x060670042011002
Sacramento	5/17/2010	15:20	3100 Block of Martin Luther King Blvd	NA	NA	x060670027004000
Sacramento	6/26/2010	3:38	3400 Block of Lumley Ln	NA	NA	x060670070151092
Sacramento	7/30/2011	16:24	3800 Block of Florin Rd	NA	NA	x060670049031002
Sacramento	8/26/2011	12:38	2900 Block of Silk Ct	NA	NA	x060670042031013
Sacramento	9/6/2011	15:30	2500 Block of 24th St	NA	NA	x060670026001009

Ex: Sacramento

(SF had no missing coordinates)

City	Date	Time	Location	Lat	Long	GEOID
Philadelphia	5/9/2013	20:56	7000 block of Pascual Ave	NA	NA	xNA
No information				NA	NA	xNA
No information				NA	NA	xNA

Final Result - only 3 NAs

Dataset Functionality

Variables (flexible)

- total population
- total male
- total female
- total black male
- total hispanic male
- total black population
- total hispanic population
- total white population
- total asian population
- total native population
- total bi-racial (black/white) population
- total black hispanic
- total urban population
- single parent household (mother present)
- single parent household (father present)

Logistical Information

- Incident ID
- City
- Date
- Time
- *Location
- *Coordinates (lat/long)
- GEOID (15 digit block-level)
- Offender ID
- *Offender Race
- Offender Fatality
- *Officer Race
- *Officer Fatality

https://api.census.gov/data/2010/dec/sf1/variables.html

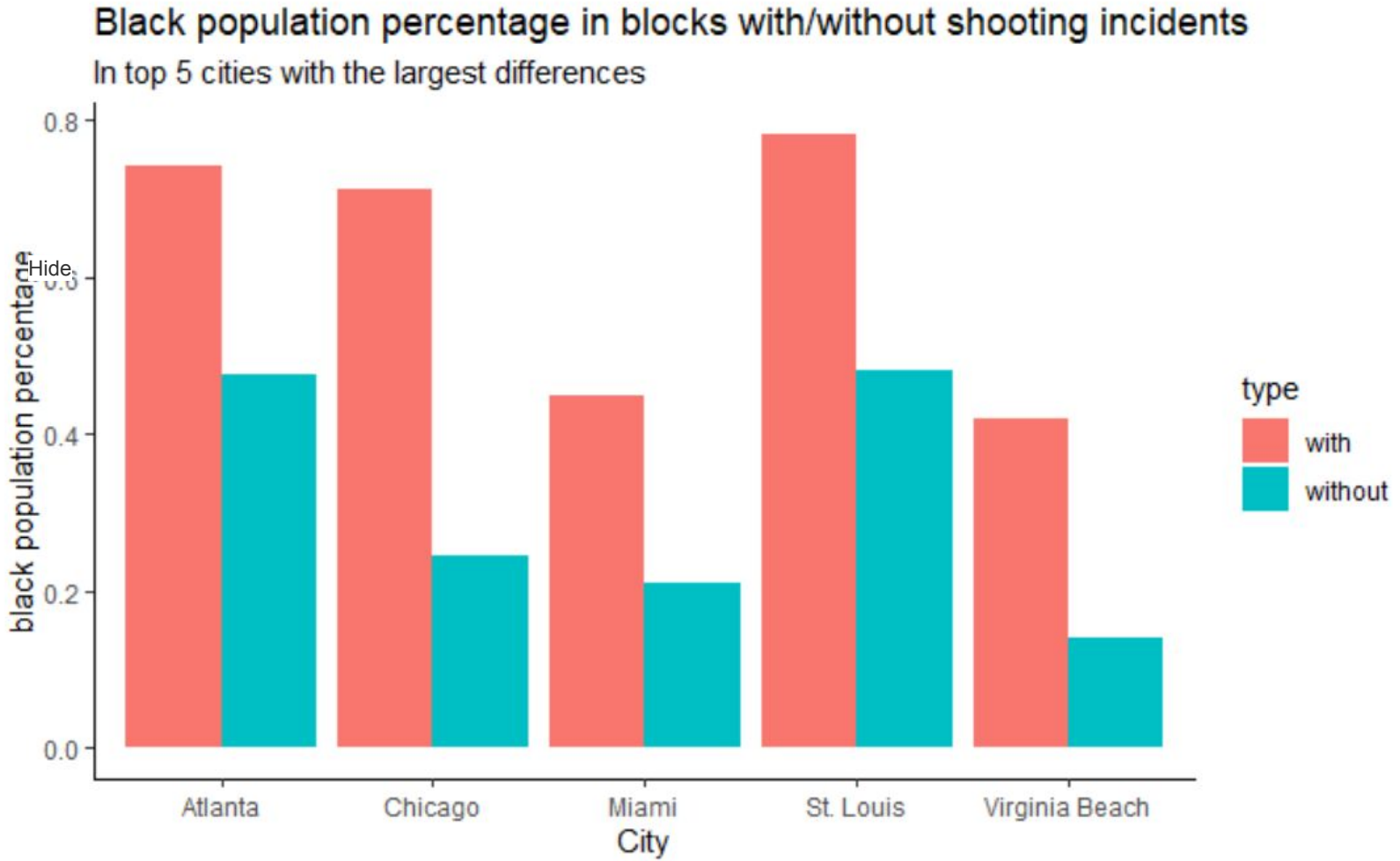
* many missing values/incorrectly formatted observations

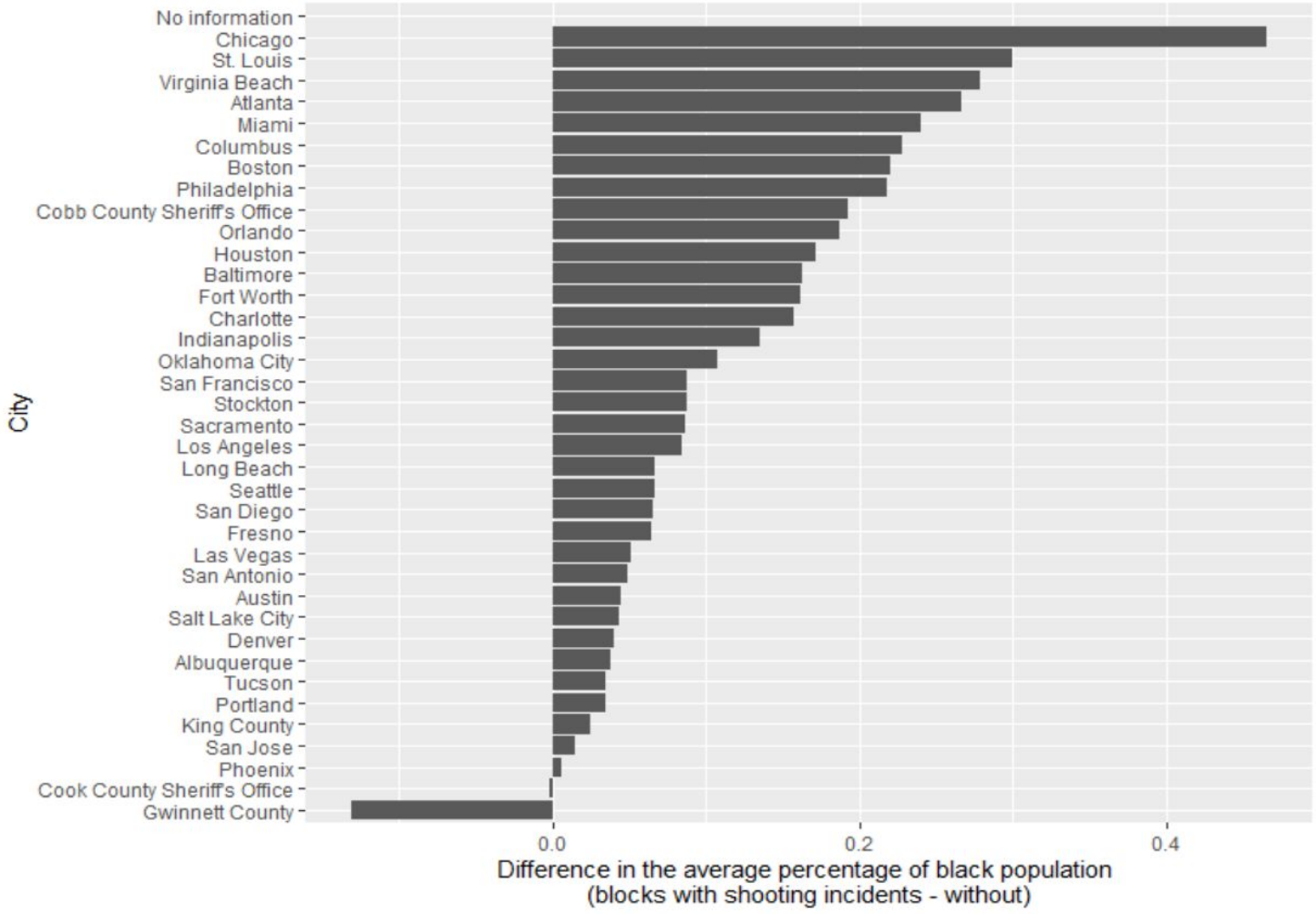
Findings

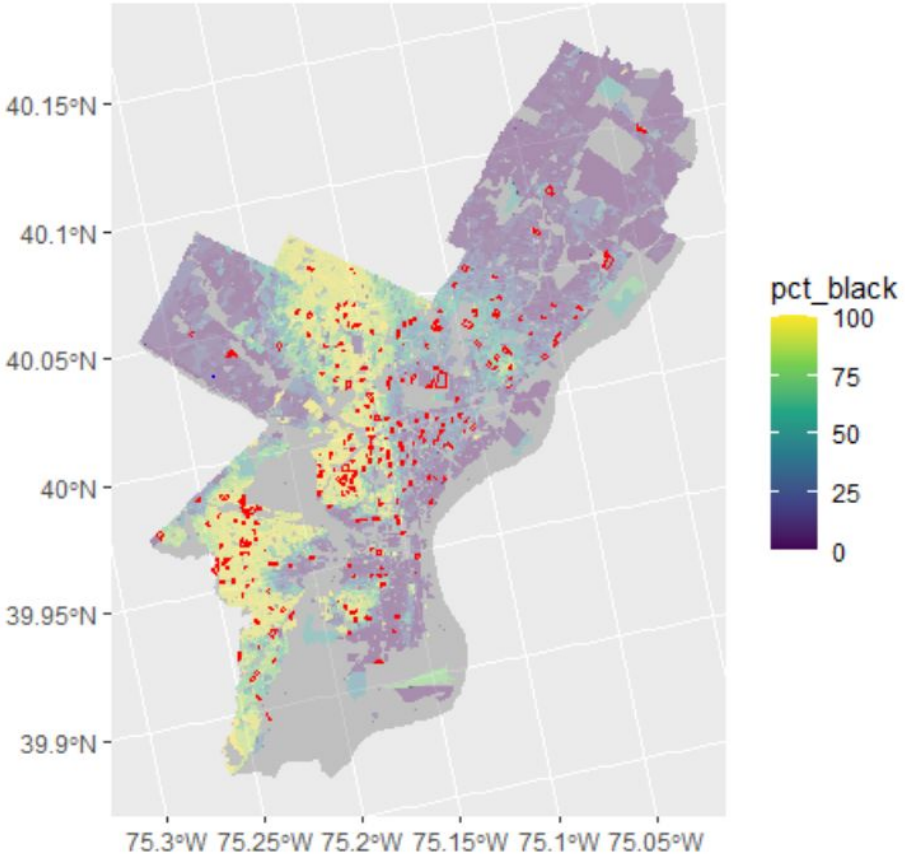
Top 5 Cities with the Largest Differences in Black Population Proportion

	Black Population Proportion		
City	**Blocks with shooting**	**Blocks without shooting**	**Difference**
Chicago	0.7120	0.2463	0.4657
St. Louis	0.7817	0.4816	0.3001
Virginia Beach	0.4192	0.1400	0.2792
Atlanta	0.7420	0.4759	0.2661
Miami	0.4502	0.2103	0.2398

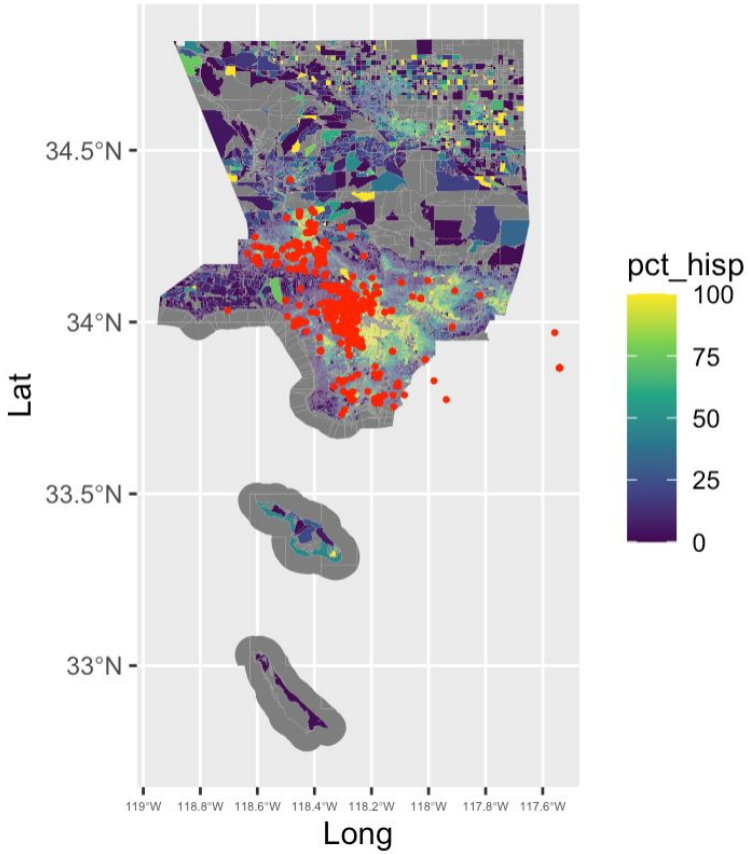
Note: Difference is calculated by blocks with shooting incidents - blocks without


Black population percentage in blocks with/without shooting incidents
In top 5 cities with the largest differences
Hide
black population percentage
0.8
0.4
0.2
0.0
Atlanta
Chicago
Miami
St. Louis
Virginia Beach
City
type
with
without

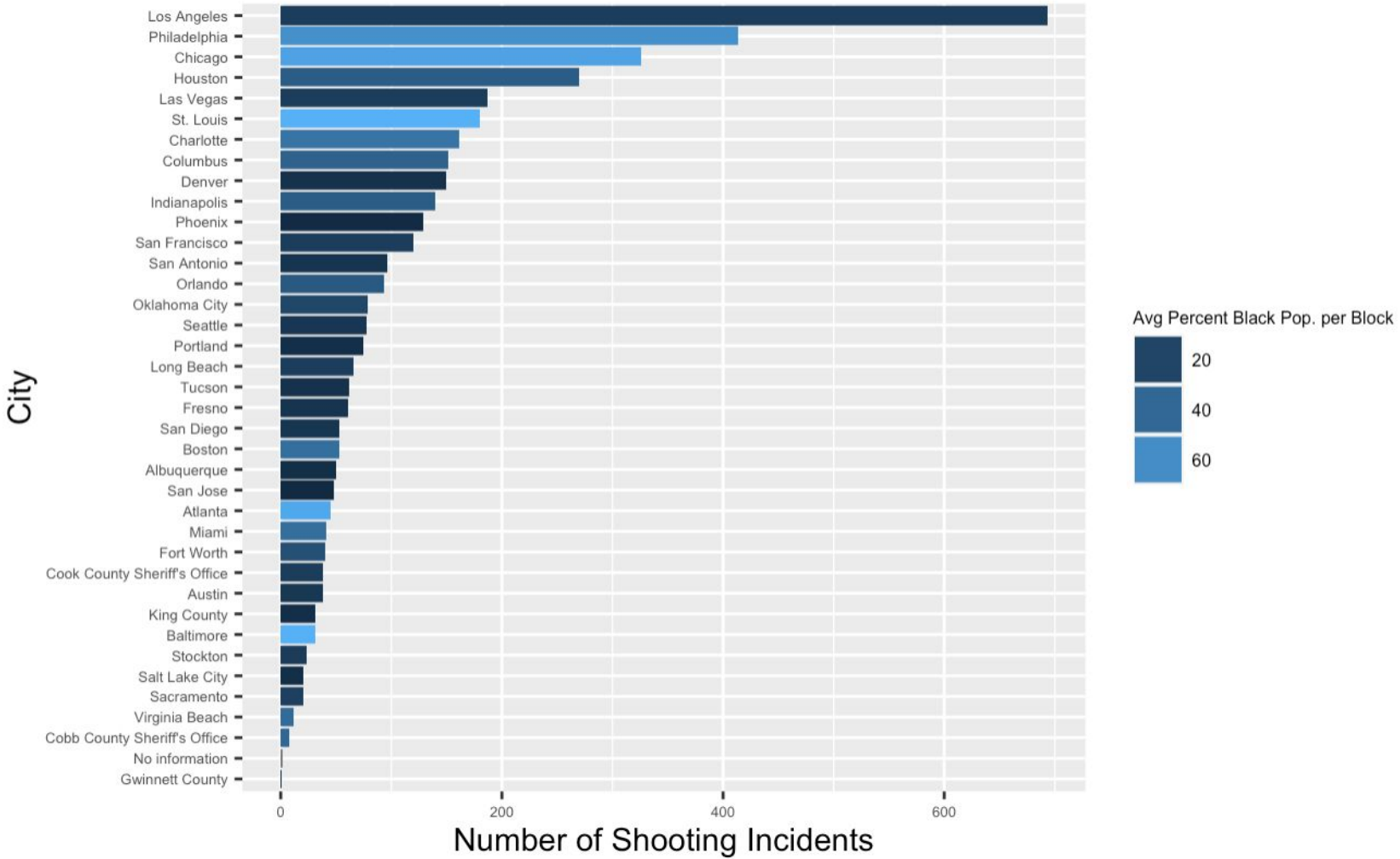

No information
Chicago
St. Louis
Virginia Beach
Atlanta
Miami
Columbus
Boston
Philadelphia
Cobb County Sheriff's Office
Orlando
Houston
Baltimore
Fort Worth
Charlotte
Indianapolis
Oklahoma City
San Francisco
Stockton
Sacramento
Los Angeles
Long Beach
Seattle
San Diego
Fresno
Las Vegas
San Antonio
Austin
Salt Lake City
Denver
Albuquerque
Tucson
Portland
King County
San Jose
Phoenix
Cook County Sheriff's Office
Gwinnett County
City
0.0
0.2
0.4
Difference in the average percentage of black population
(blocks with shooting incidents - without)

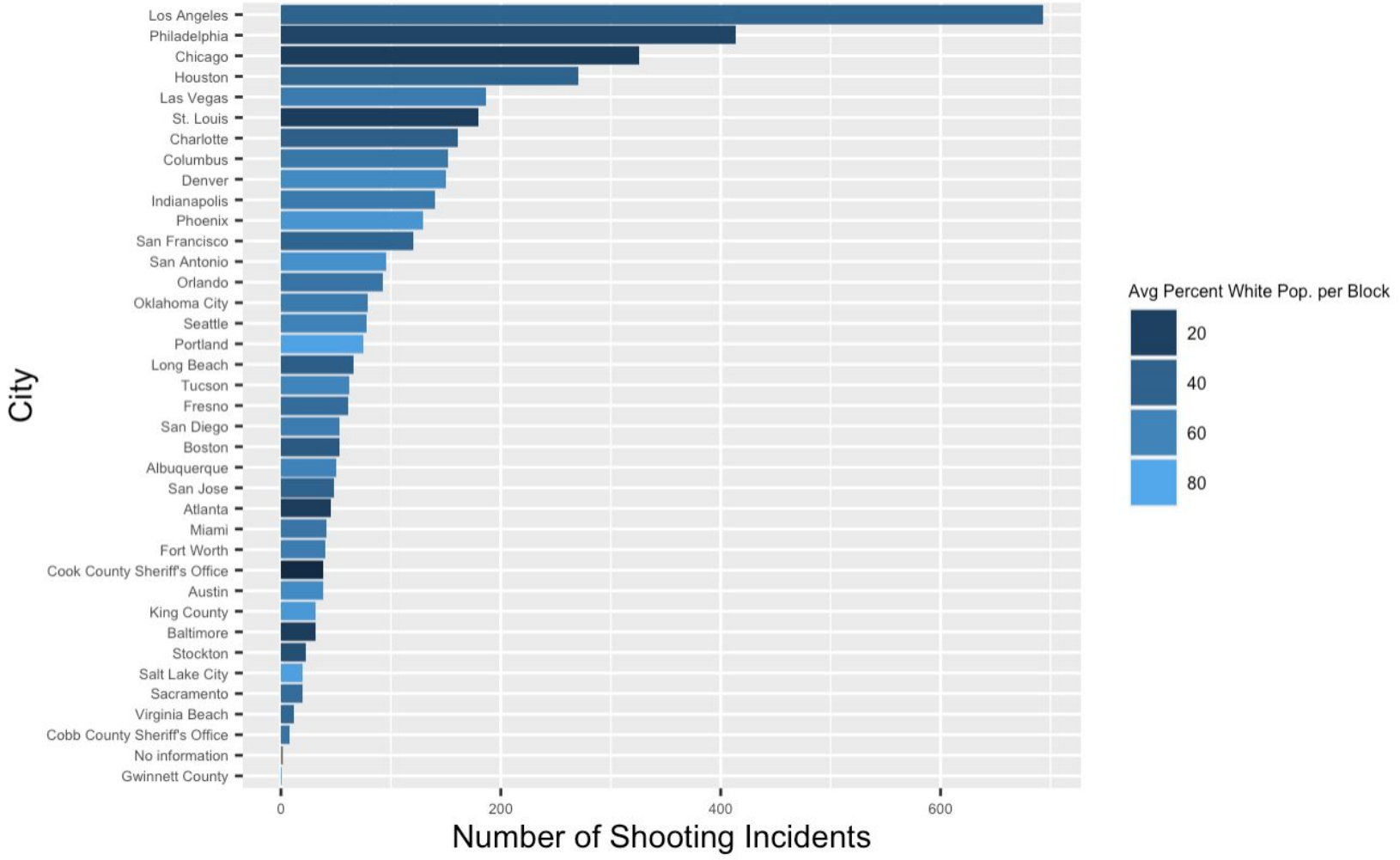


Map of Philadelphia officer-involved shootings colored according to percent black population per block



Map of Los Angeles officer-involved shootings colored according to percent Hispanic population per block


Los Angeles
Philadelphia
Chicago
Houston
Las Vegas
St. Louis
Charlotte
Columbus
Denver
Indianapolis
Phoenix
San Francisco
San Antonio
Orlando
Oklahoma City
Seattle
Portland
Long Beach
Tucson
Fresno
San Diego
Boston
Albuquerque
San Jose
Atlanta
Miami
Fort Worth
Cook County Sheriff's Office
Austin
King County
Baltimore
Stockton
Salt Lake City
Sacramento
Virginia Beach
Cobb County Sheriff's Office
No information
Gwinnett County
City
0
200
400
600
Number of Shooting Incidents
Avg Percent Black Pop. per Block
20
40
60


Los Angeles
Philadelphia
Chicago
Houston
Las Vegas
St. Louis
Charlotte
Columbus
Denver
Indianapolis
Phoenix
San Francisco
San Antonio
Orlando
Oklahoma City
Seattle
Portland
Long Beach
Tucson
Fresno
San Diego
Boston
Albuquerque
San Jose
Atlanta
Miami
Fort Worth
Cook County Sheriff's Office
Austin
King County
Baltimore
Stockton
Salt Lake City
Sacramento
Virginia Beach
Cobb County Sheriff's Office
No information
Gwinnett County
City
0
200
400
600
Number of Shooting Incidents
Avg Percent White Pop. per Block
20
40
60
80