

Brenda Fried – *Brooklyn College*
Harpreet Gaur – *CUNY City Tech*
Adnan Hoq – *St Joseph's College*
Emeka Mbazor – *Lehman College*

Naomi Moreira – *St Joseph's College*
Cindy Muso – *St John's University*
Etta Rapp – *Yeshiva University*
Roymil Terrero – *St Joseph's College*

Replicating "An Empirical Analysis of Racial Differences in Police Use of Force"

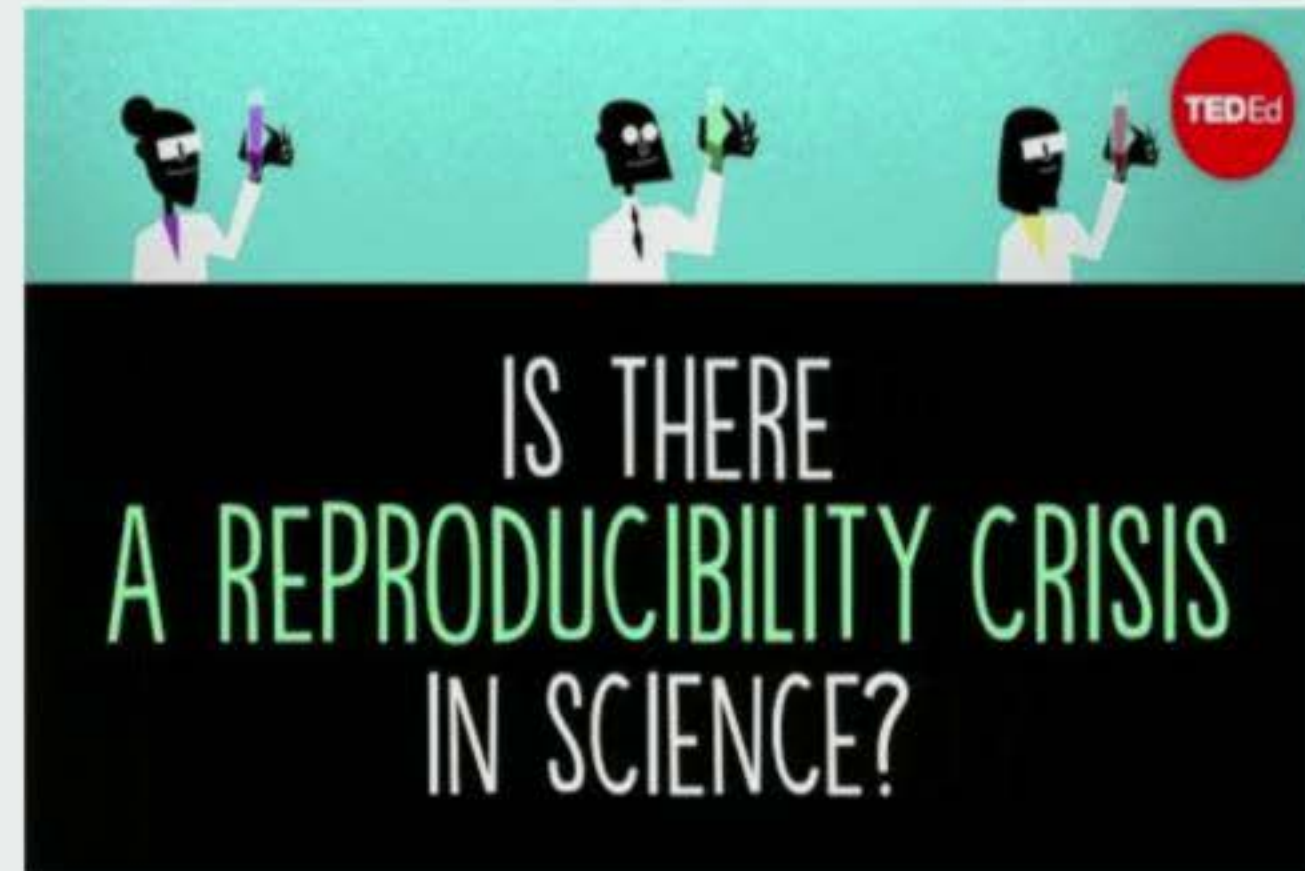
The Importance of Replication

- "Reproducibility is a defining feature of science"



Replication

- A widespread problem in research is that many papers are not reproducible
- Possible reasons for irreproducibility:
 - Code/datasets not provided
 - Methods unclear
 - "p-hacking"
 - And more...



The Paper We Studied



Journal of Political Economy

Volume 127, Number 3 | June 2019

An Empirical Analysis of Racial Differences in Police Use of Force

Roland G. Fryer Jr.

Harvard University and National Bureau of Economic Research

"On nonlethal uses of force, blacks and Hispanics are more than 50 percent more likely to experience some form of force in interactions with police."

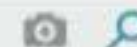
"controls ... cannot fully explain these disparities."

"On the most extreme use of force—officer-involved shootings—we find no racial differences ... "

The Discussion It Generated



bias police use of force fryer



All Images Videos Maps News Shopping | My saves

8,850,000 Results Any time ▾

Surprising New Evidence Shows Bias in Police Use of Force ...

<https://www.nytimes.com/2016/07/12/upshot/surprising-new-evidence-shows-bias-in-police...>

Jul 12, 2016 · Surprising New Evidence Shows **Bias in Police Use of Force** but Not in Shootings Surprising New Evidence Shows **Bias in Police Use of Force** but Not in Shootings. **Lethal uses of force** are exceedingly rare. There were 1.6 million arrests in Houston in the years Mr. Fryer studied. **Officers fired** their weapons 507 times. What is far more common are nonlethal **uses of force**. And in these **uses of force**...

Author: Quoc Trung Bui, Amanda Cox

Paper finding no racial bias in shootings by police criticized

<https://www.usatoday.com/story/news/nation-now/2016/07/21/paper-finding-no-racial-bias...>

Jul 21, 2016 · Paper finding no **racial bias** in shootings by **police** criticized. Fryer examined four datasets to measure racial disparity with regard to **police use of force**: New York's stop-and-frisk program, in which **police** could stop, question and frisk pedestrians; the **Police-Public** ...

Author: Samantha Nelson

Study Finds Racial Bias in Police Use of Force — But Not ...

<nymag.com/intelligencer/2016/07/study-finds-no-bias-in-cops-use-of-lethal-force.html>

Jul 11, 2016 · Study Finds **Racial Bias in Police Use of Force** — But Not in Shootings ... really is going on when it comes to racial differences in **police use of force**." Fryer's study is based off detailed ...

Author: Eric Levitz

No racial bias in police shootings, study by Harvard ...

<https://www.washingtontimes.com/news/2016/jul/11/no-racial-bias-police-shootings-study...>

Jul 11, 2016 · A study by a Harvard professor released this month found no evidence of **racial bias** in **police** shootings even though officers were more likely to interact physically with non-whites than whites.

Author: Valerie Richardson

Our Goal

Can we replicate the main findings of the paper?

An Ideal Scenario (?)



Publicly available datasets

NYPD Stop, Question, and Frisk (SQF)
Police Public Contact Survey (PPCS)



Simple methods

Empirical counts
Standard logistic regression models



Detailed paper

Main text: 50+ pages
Appendix: 60+ pages

Types of Analyses

	Reproduce	Replicate	Extend
Data	Same as original paper	Same as original paper	New
Code	Provided by authors	New	New
Match	Exact	Exact or approximate	Approximate

- Not possible...

- Main focus

- Some of this, too

Types of Analyses

	Reproduce	Replicate	Extend
Data	Same as original paper	Same as original paper	New
Code	Provided by authors	New	New
Match	Exact	Exact or approximate	Approximate

- Not possible...

- Main focus

- Some of this, too

Outline

Outline

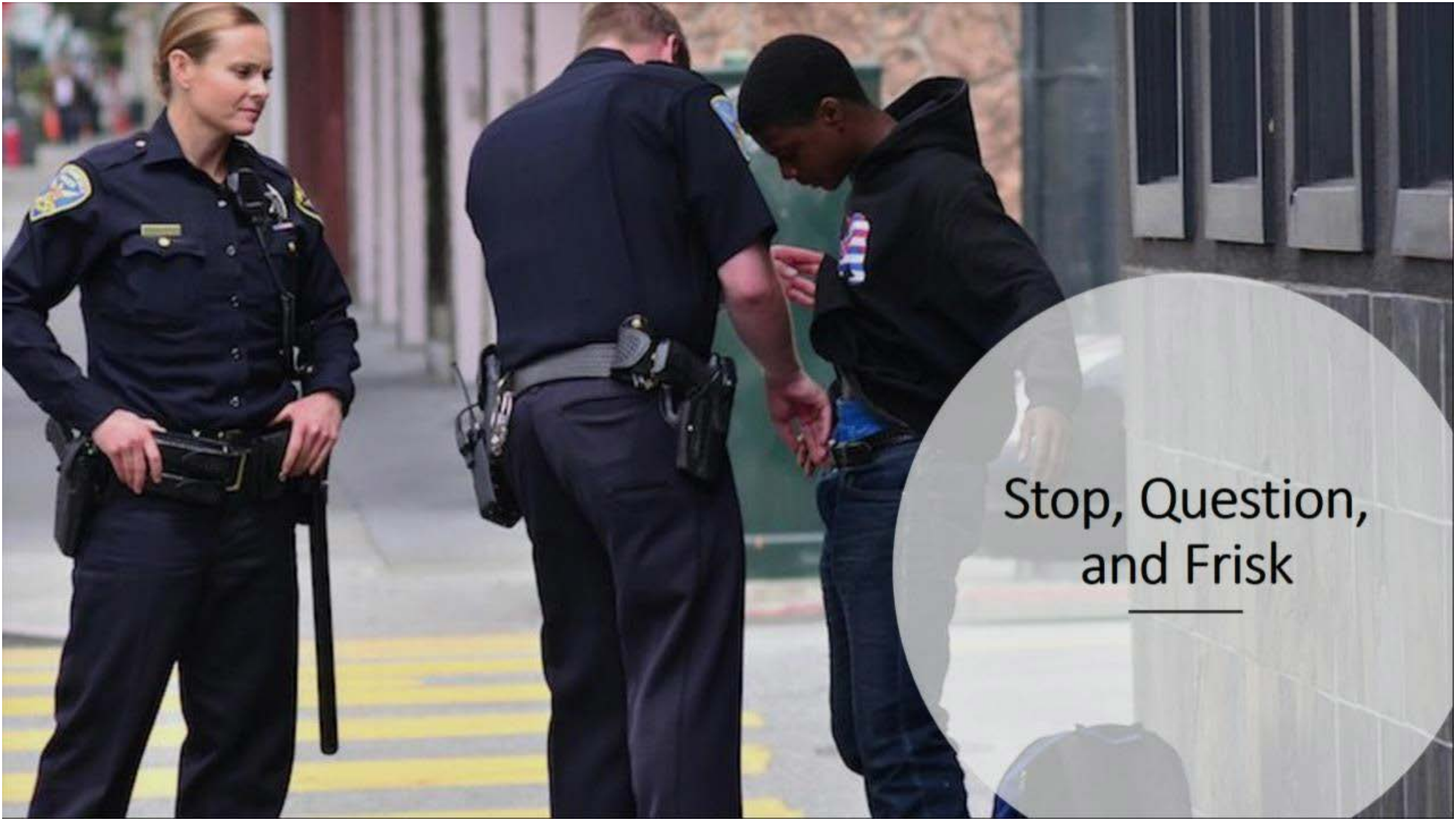
- Replication of Stop, Question, and Frisk analysis

Outline

- Replication of Stop, Question, and Frisk analysis
- Visualization and extension of Stop and Frisk data with maps and census data

Outline

- Replication of Stop, Question, and Frisk analysis
- Visualization and extension of Stop and Frisk data with maps and census data
- Replication of Police Public Contact Survey analysis



Stop, Question,
and Frisk

What is Stop, Question, and Frisk?



NYPD PROGRAM



2003 - PRESENT



**DOCUMENTED AND
REPORTED BY OFFICERS**

Fryer's Paper

- Focuses on use of force
- And how it varies by race

Potential Shortcomings

- Potential misreporting by officers
- Data for stops only

Data Wrangling

- Data collected from NYC.gov
- Bash scripts to download and unzip data files
- Loaded data into R

Challenges

- Mismatching column names
- Different number of columns
- Missing columns
- Nonsensical values
- Inconsistent values across different years

Example

```
Console Terminal x Jobs x
~/
NA
> unique(df17$SUSPECT_RACE_DESCRIPTION)
[1] "(null)" "WHITE"
[3] "BLACK HISPANIC" "BLACK"
[5] "WHITE HISPANIC" "ASIAN/PAC.ISL"
[7] "AMER IND" "MALE"
[9] NA
> |
```

Example

```
Console Terminal x Jobs x  
~/  
NA  
> unique(df17$SUSPECT_RACE_DESCRIPTION)  
[1] "(null)" "WHITE"  
[3] "BLACK HISPANIC" "BLACK"  
[5] "WHITE HISPANIC" "ASIAN/PAC.ISL"  
[7] "AMER IND" "MALE"  
[9] NA  
> |
```

Example

```
Console Terminal x Jobs x
~/
NA
> unique(df17$SUSPECT_RACE_DESCRIPTION)
[1] "(null)" "WHITE"
[3] "BLACK HISPANIC" "BLACK"
[5] "WHITE HISPANIC" "ASIAN/PAC.ISL"
[7] "AMER IND" "MALE"
[9] NA
> |
```

A New Race?

Example

```
Console Termin  
~/  
NA  
> unique(d  
[1] "(null  
[3] "BLACK  
[5] "WHITE  
[7] "AMER  
[9] NA  
> |
```

	year	datestop	race	age	sex
1	2007	07102007	W	453	M
2	2007	07112007	B	388	M
3	2007	07122007	Q	412	M
4	2007	07182007	B	175	M
5	2007	07252007	B	200	M
6	2007	07292007	B	195	M
7	2007	08272007	B	180	M
8	2007	08302007	W	156	M
9	2007	08312007	P	232	F
10	2007	09062007	B	170	M
11	2007	09152007	B	591	M
12	2007	09162007	B	116	M
13	2007	09182007	W	522	M
14	2007	09292007	B	165	M
15	2007	10022007	B	150	M

Example

```
Console Termin  
~/  
NA  
> unique(d  
[1] "(null  
[3] "BLACK  
[5] "WHITE  
[7] "AMER  
[9] NA  
> |
```

	year	datestop	race	age	sex
1	2007	07102007	W	453	M
2	2007	07112007	B	388	M
3	2007	07122007	Q	412	M
4	2007	07182007	B	175	M
5	2007	07252007	B	200	M
6	2007	07292007	B	195	M
7	2007	08272007	B	180	M
8	2007	08302007	W	156	M
9	2007	08312007	P	232	F
10	2007	09062007	B	170	M
11	2007	09152007	B	591	M
12	2007	09162007	B	116	M
13	2007	09182007	W	522	M
14	2007	09292007	B	165	M
15	2007	10022007	B	150	M

Immortal
People?

Replication Differences

- Number of Observations

Number of Observations	
Fryer's	Ours
4,982,925	4,984,393

- Coding Variables Issues
 - Civilian Age
 - Civilian Race

Summary Statistics

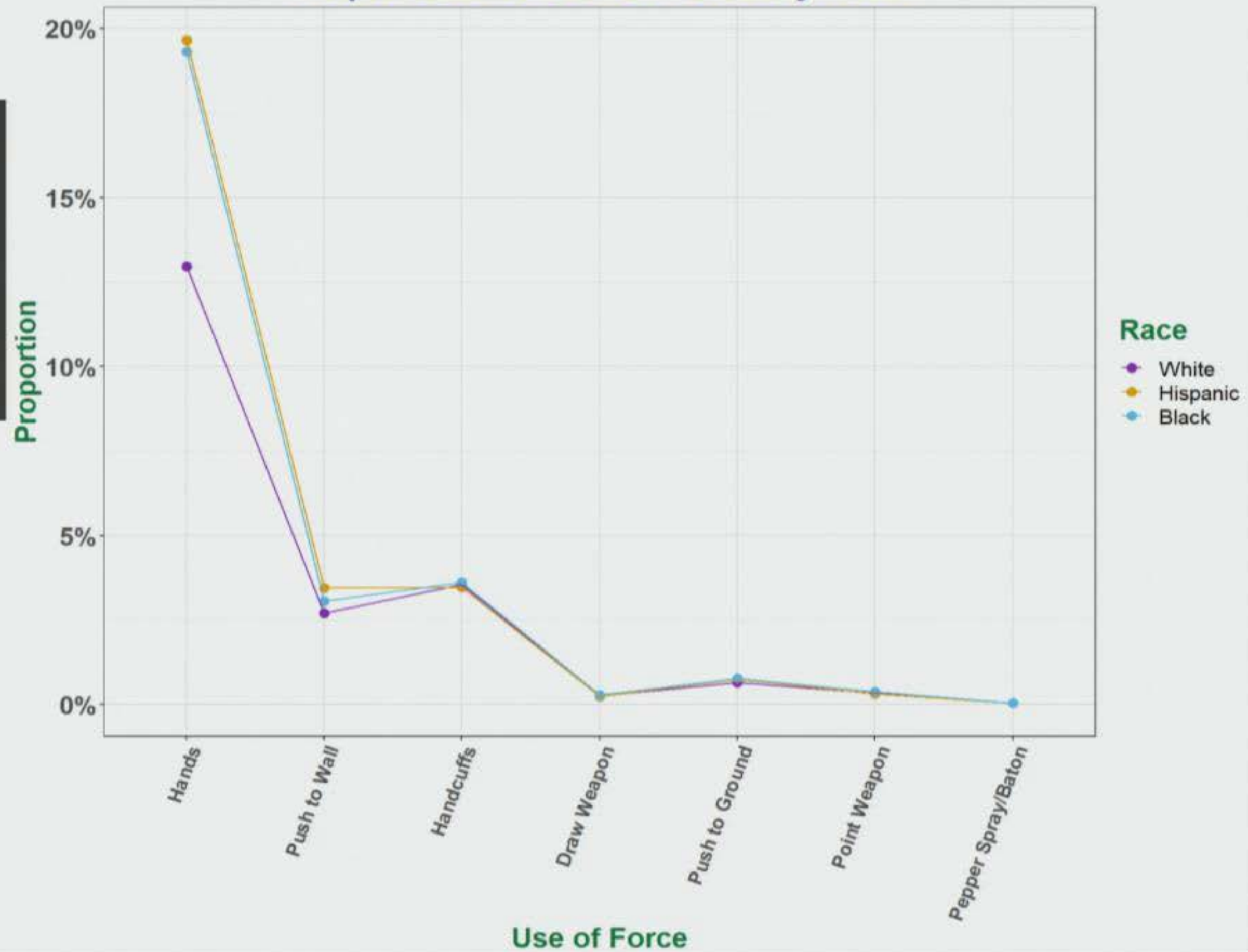
	Fryer's Stats			
	Full Sample	White	Black	Hispanic
Panel A: Baseline Characteristics				
White	0.10	1.00	0.00	0.00
Black	0.58	0.00	1.00	0.00
Hispanic	0.25	0.00	0.00	1.00
Asian	0.03	0.00	0.00	0.00
Other	0.04	0.00	0.00	0.00
Age	28.00	29.25	27.96	27.57
Male	0.93	0.90	0.93	0.93

	Fryer's Stats				Our Stats			
	Full Sample	White	Black	Hispanic	Full Sample	White	Black	Hispanic
Panel A: Baseline Characteristics								
White	0.10	1.00	0.00	0.00	0.10	1.00	0.00	0.00
Black	0.58	0.00	1.00	0.00	0.58	0.00	1.00	0.00
Hispanic	0.25	0.00	0.00	1.00	0.24	0.00	0.00	1.00
Asian	0.03	0.00	0.00	0.00	0.03	0.00	0.00	0.00
Other	0.04	0.00	0.00	0.00	0.04	0.00	0.00	0.00
Age	28.00	29.25	27.96	27.57	27.99	29.24	27.95	27.57
Male	0.93	0.90	0.93	0.93	0.93	0.90	0.93	0.93

Proportions of Use of Force by Race

Summary
Statistics

Disparity decreases
as intensity of force
increases



Modeling Use of Force

- Maybe variables other than race account for differences in use of force
- Ideal situation: controlled experiment
- Fryer uses a model that accounts for as much difference as possible

Modeling Use of Force

Use of Force ~ Race + Civilian Demographics + Encounter Characteristics + Civilian Behavior + Precinct + Year



Main Outcome

Variable of Interest

Controls

- Yes
- No

- White
- Black
- Hispanic
- Asian
- Other

- Sex
- Age

- Place
- Time of Day
- High-crime Area
- High-crime Time
- Police in Uniform
- Civilian ID Type
- Stopped with Others

- Carrying Suspicious Object
- Fit Relevant Description
- Preparing for Crime
- Lookout for Crime
- Dressed in Criminal Attire
- Appearance of Drug Transaction
- Suspicious Movements
- Engaging in Violent Crime

Model Replication

	Fryer Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.153	1.534	1.582	1.044	1.392	4,927,962
(b) +CivilianDemographics		1.480	1.517	1.010	1.346	?
(c) +EncounterCharacteristics		1.655	1.641	1.059	1.452	?
(d) +CivilianBehavior		1.462	1.516	1.051	1.372	?
(e) +PrecinctFE, YearFE		1.178	1.122	0.953	1.060	?

	Our Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.161	1.500	1.548	1.047	1.322	4,984,393
(b) +CivilianDemographics		1.461	1.503	1.021	1.303	4,962,834
(c) +EncounterCharacteristics		1.644	1.611	1.060	1.415	4,398,468
(d) +CivilianBehavior		1.483	1.519	1.073	1.374	4,398,468
(e) +PrecinctFE, YearFE		1.203	1.140	0.975	1.074	4,398,468

Model Replication

	Fryer Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.153	1.534	1.582	1.044	1.392	4,927,962
(b) +CivilianDemographics		1.480	1.517	1.010	1.346	?
(c) +EncounterCharacteristics		1.655	1.641	1.059	1.452	?
(d) +CivilianBehavior		1.462	1.516	1.051	1.372	?
(e) +PrecinctFE, YearFE		1.178	1.122	0.953	1.060	?

	Our Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.161	1.500	1.548	1.047	1.322	4,984,393
(b) +CivilianDemographics		1.461	1.503	1.021	1.303	4,962,834
(c) +EncounterCharacteristics		1.644	1.611	1.060	1.415	4,398,468
(d) +CivilianBehavior		1.483	1.519	1.073	1.374	4,398,468
(e) +PrecinctFE, YearFE		1.203	1.140	0.975	1.074	4,398,468

Model Replication

	Fryer Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.153	1.534	1.582	1.044	1.392	4,927,962
(b) +CivilianDemographics		1.480	1.517	1.010	1.346	?
(c) +EncounterCharacteristics		1.655	1.641	1.059	1.452	?
(d) +CivilianBehavior		1.462	1.516	1.051	1.372	?
(e) +PrecinctFE, YearFE		1.178	1.122	0.953	1.060	?

	Our Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.161	1.500	1.548	1.047	1.322	4,984,393
(b) +CivilianDemographics		1.461	1.503	1.021	1.303	4,962,834
(c) +EncounterCharacteristics		1.644	1.611	1.060	1.415	4,398,468
(d) +CivilianBehavior		1.483	1.519	1.073	1.374	4,398,468
(e) +PrecinctFE, YearFE		1.203	1.140	0.975	1.074	4,398,468

Model Replication

	Fryer Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.153	1.534	1.582	1.044	1.392	4,927,962
(b) +CivilianDemographics		1.480	1.517	1.010	1.346	?
(c) +EncounterCharacteristics		1.655	1.641	1.059	1.452	?
(d) +CivilianBehavior		1.462	1.516	1.051	1.372	?
(e) +PrecinctFE, YearFE		1.178	1.122	0.953	1.060	?

	Our Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.161	1.500	1.548	1.047	1.322	4,984,393
(b) +CivilianDemographics		1.461	1.503	1.021	1.303	4,962,834
(c) +EncounterCharacteristics		1.644	1.611	1.060	1.415	4,398,468
(d) +CivilianBehavior		1.483	1.519	1.073	1.374	4,398,468
(e) +PrecinctFE, YearFE		1.203	1.140	0.975	1.074	4,398,468

Model Replication

	Fryer Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.153	1.534	1.582	1.044	1.392	4,927,962
(b) +CivilianDemographics		1.480	1.517	1.010	1.346	?
(c) +EncounterCharacteristics		1.655	1.641	1.059	1.452	?
(d) +CivilianBehavior		1.462	1.516	1.051	1.372	?
(e) +PrecinctFE, YearFE		1.178	1.122	0.953	1.060	?

	Our Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.161	1.500	1.548	1.047	1.322	4,984,393
(b) +CivilianDemographics		1.461	1.503	1.021	1.303	4,962,834
(c) +EncounterCharacteristics		1.644	1.611	1.060	1.415	4,398,468
(d) +CivilianBehavior		1.483	1.519	1.073	1.374	4,398,468
(e) +PrecinctFE, YearFE		1.203	1.140	0.975	1.074	4,398,468

Model Replication

What does this mean?

Odds ratios are hard to interpret

	Fryer Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.153	1.534	1.582	1.044	1.392	4,927,962
(b) +CivilianDemographics		1.480	1.517	1.010	1.346	?
(c) +EncounterCharacteristics		1.655	1.641	1.059	1.452	?
(d) +CivilianBehavior		1.462	1.516	1.051	1.372	?
(e) +PrecinctFE, YearFE		1.178	1.122	0.953	1.060	?

	Our Stats					
	White Mean	Black	Hispanic	Asian	Other Race	Sample Size
Panel A: NYC Stop, Question & Frisk						
(a) NoControls	0.161	1.500	1.548	1.047	1.322	4,984,393
(b) +CivilianDemographics		1.461	1.503	1.021	1.303	4,962,834
(c) +EncounterCharacteristics		1.644	1.611	1.060	1.415	4,398,468
(d) +CivilianBehavior		1.483	1.519	1.073	1.374	4,398,468
(e) +PrecinctFE, YearFE		1.203	1.140	0.975	1.074	4,398,468

Model Replication

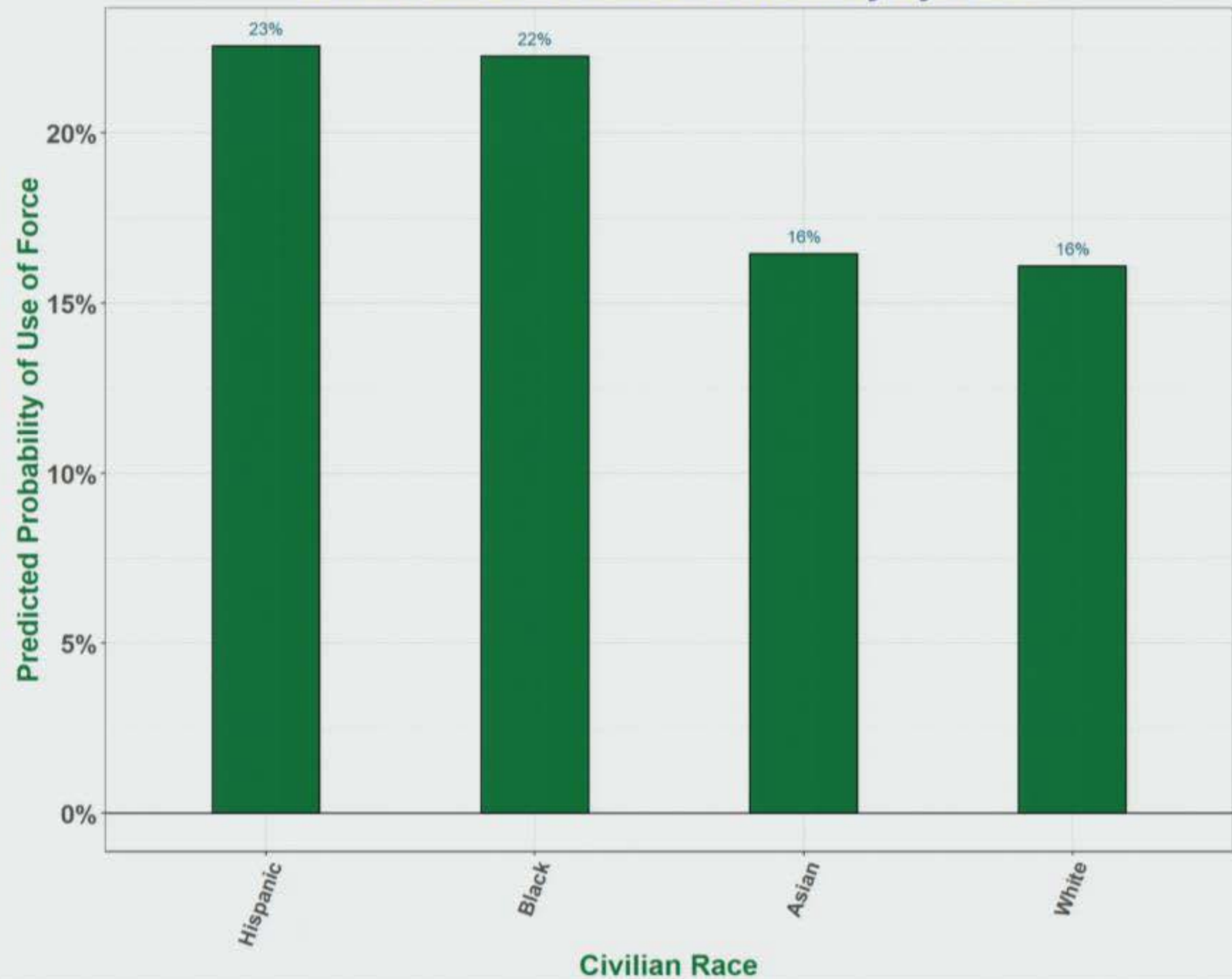
Our Finding:

37.5%



"Blacks are **53%** more likely to experience any use of force relative to a white mean of 15.3%" - Fryer

Difference of Predicted Probability by Race

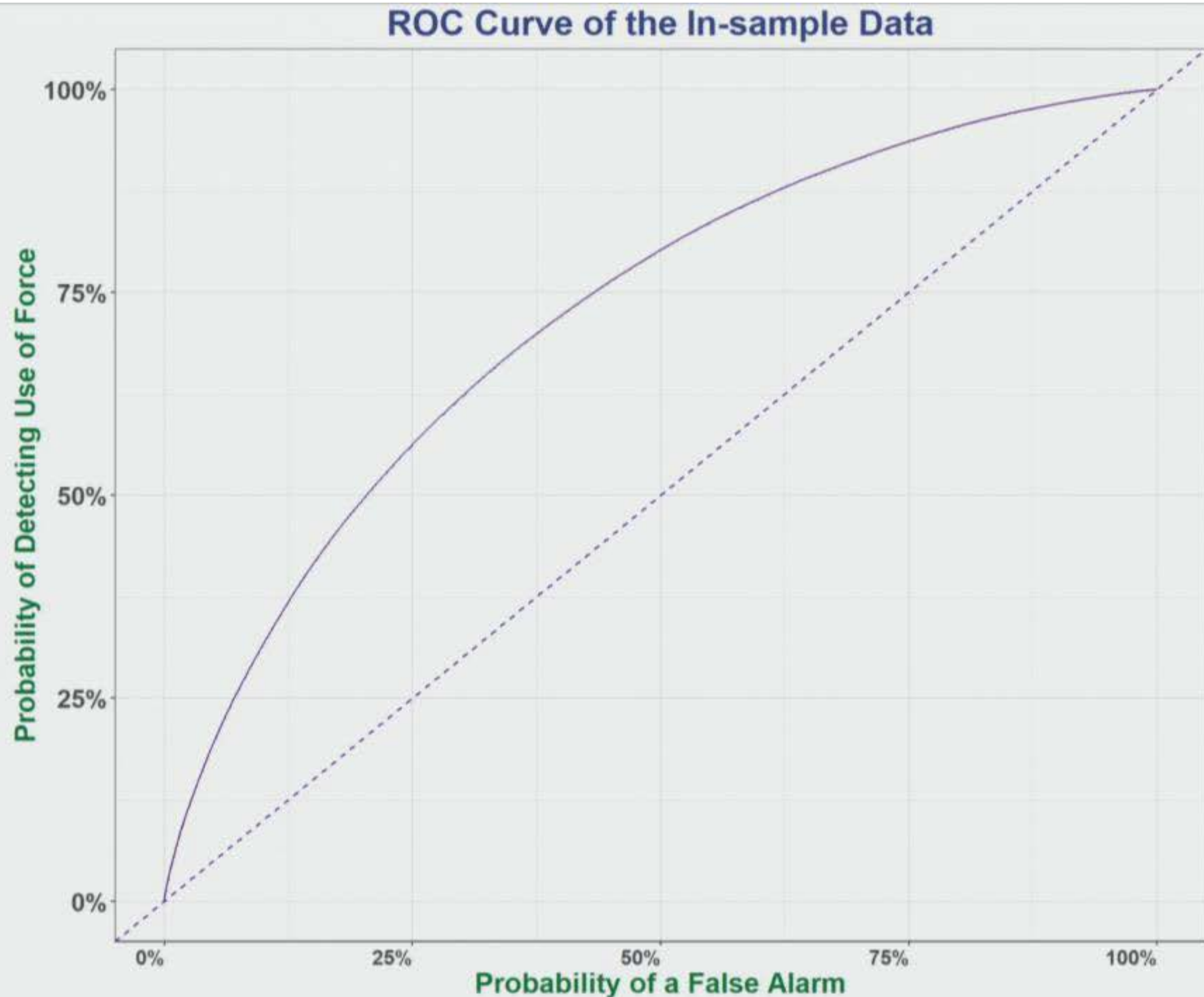


Model Evaluation

AUC (model with race) = 68%

AUC (model without race) = 68%

No significant change!!



Takeaways



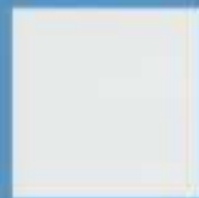
Results did not replicate



But recovered underlying pattern:
controls don't account for racial differences



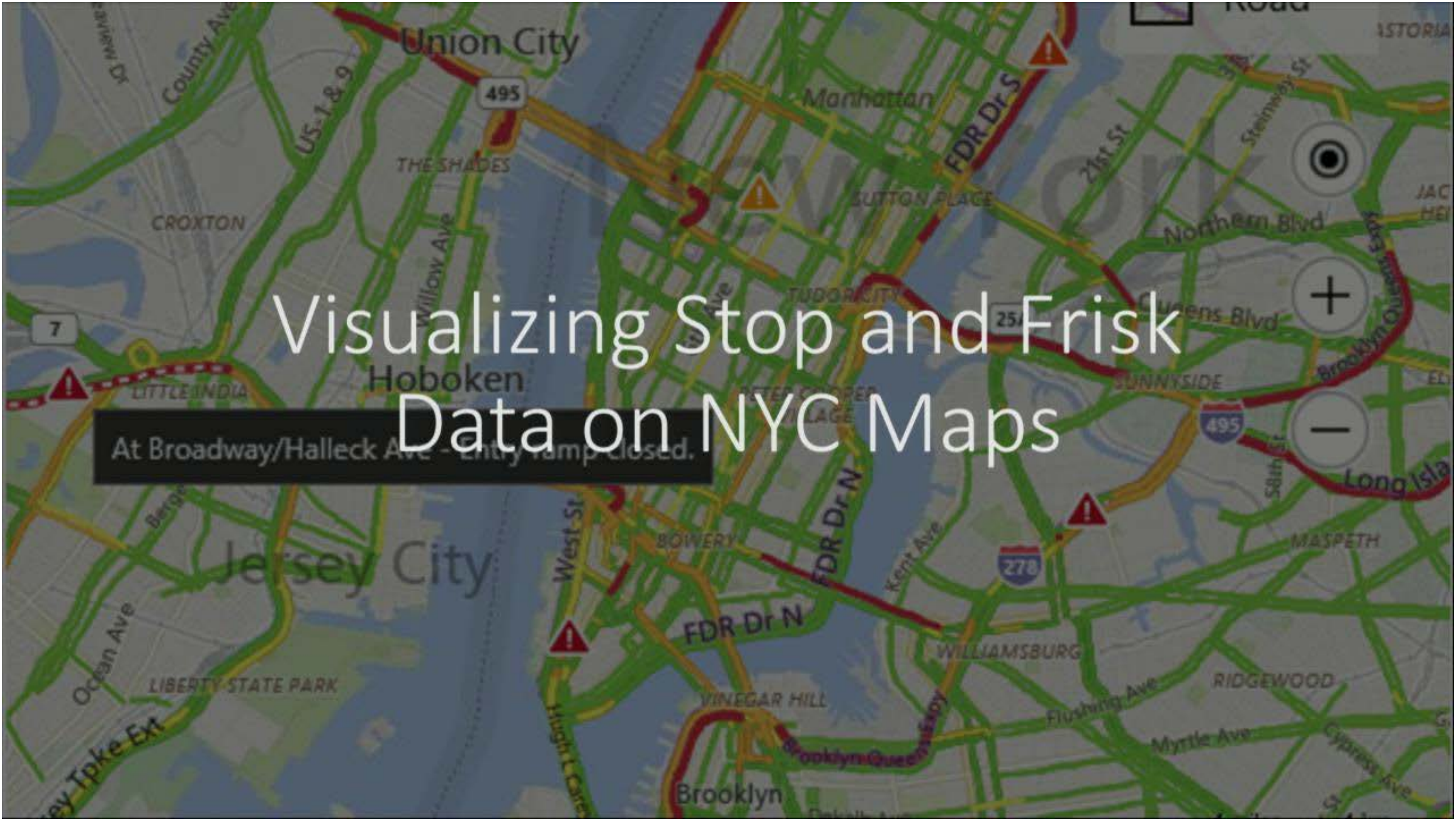
Caveat: model performance not ideal



Removing race does not impact the model
much

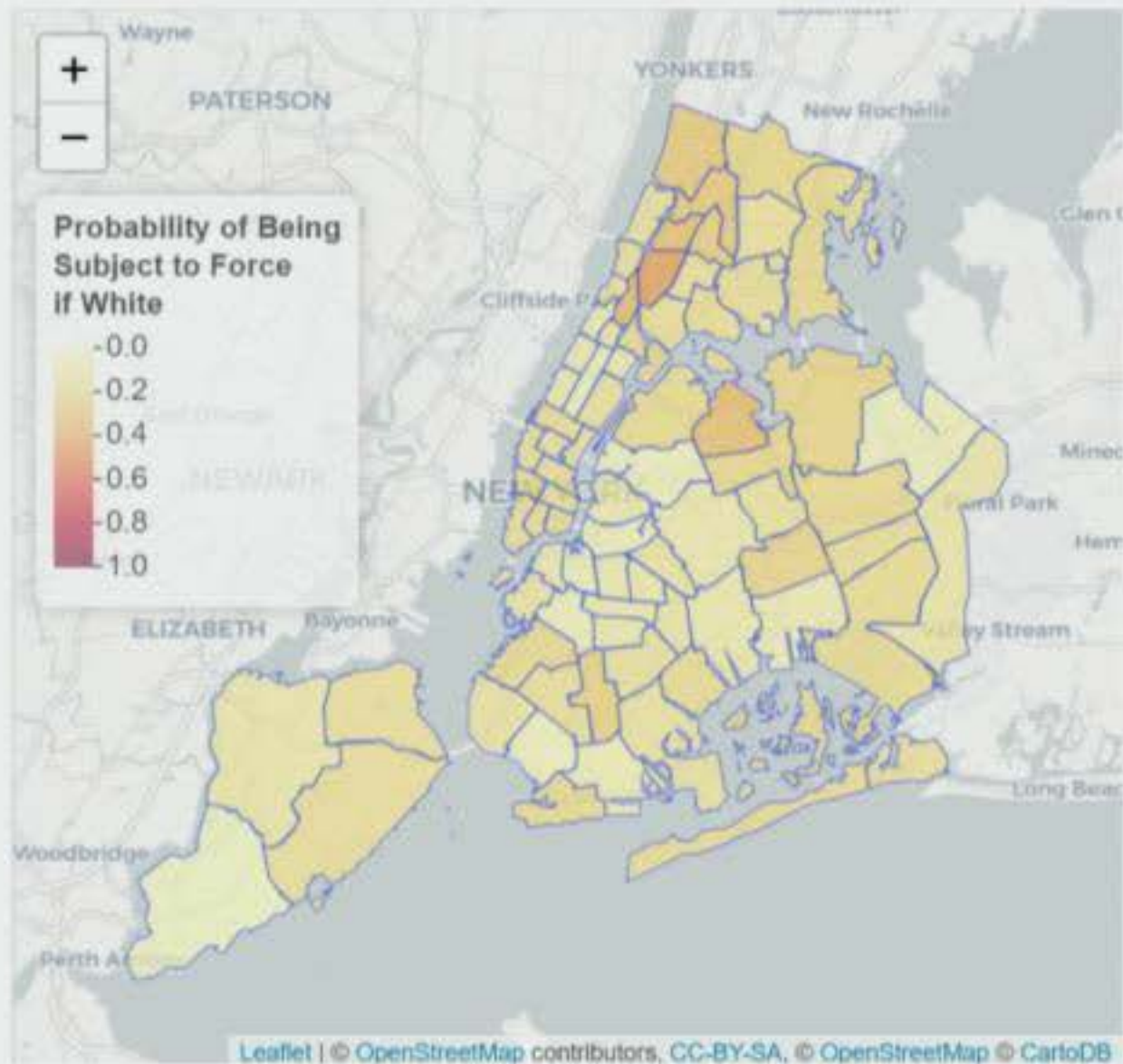
Visualizing Stop and Frisk Data on NYC Maps

At Broadway/Halleck Ave - Entry ramp closed.

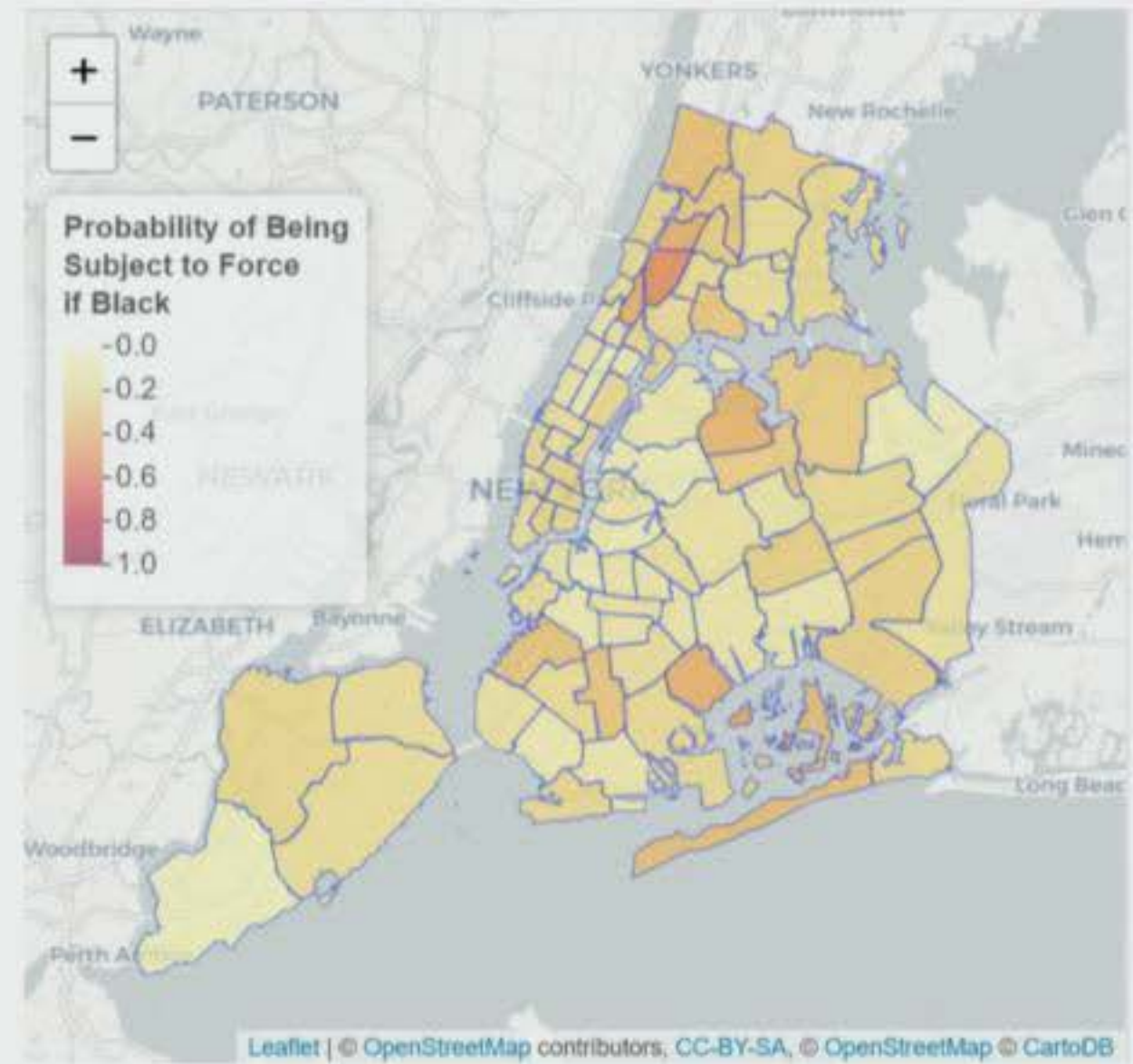


Visualizing the probability of a Stopped Civilian subjected to police use of force, on the basis of race and precinct

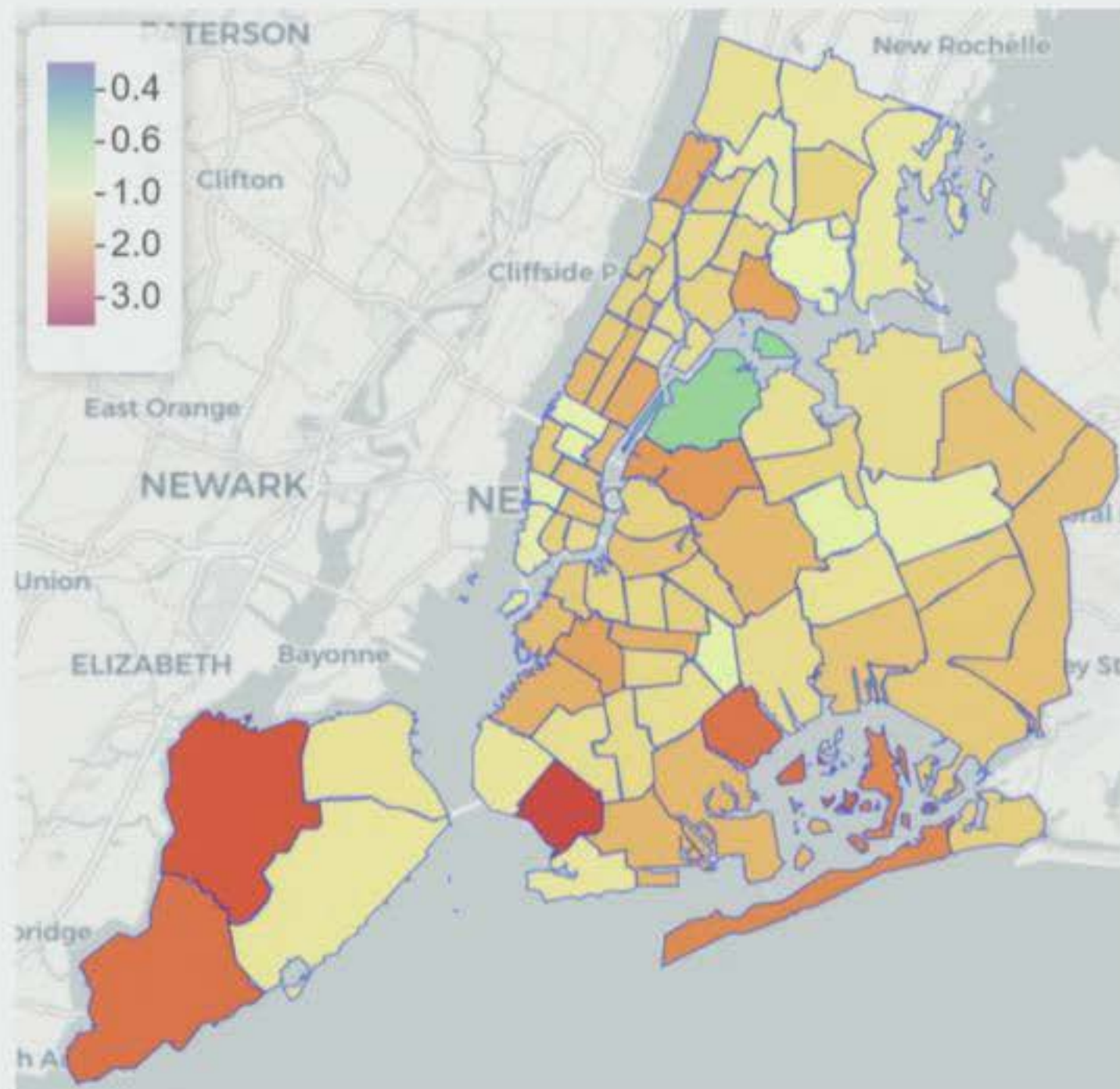
$P(\text{any force} \mid \text{white and stopped in precinct})$



$P(\text{any force} \mid \text{black and stopped in precinct})$



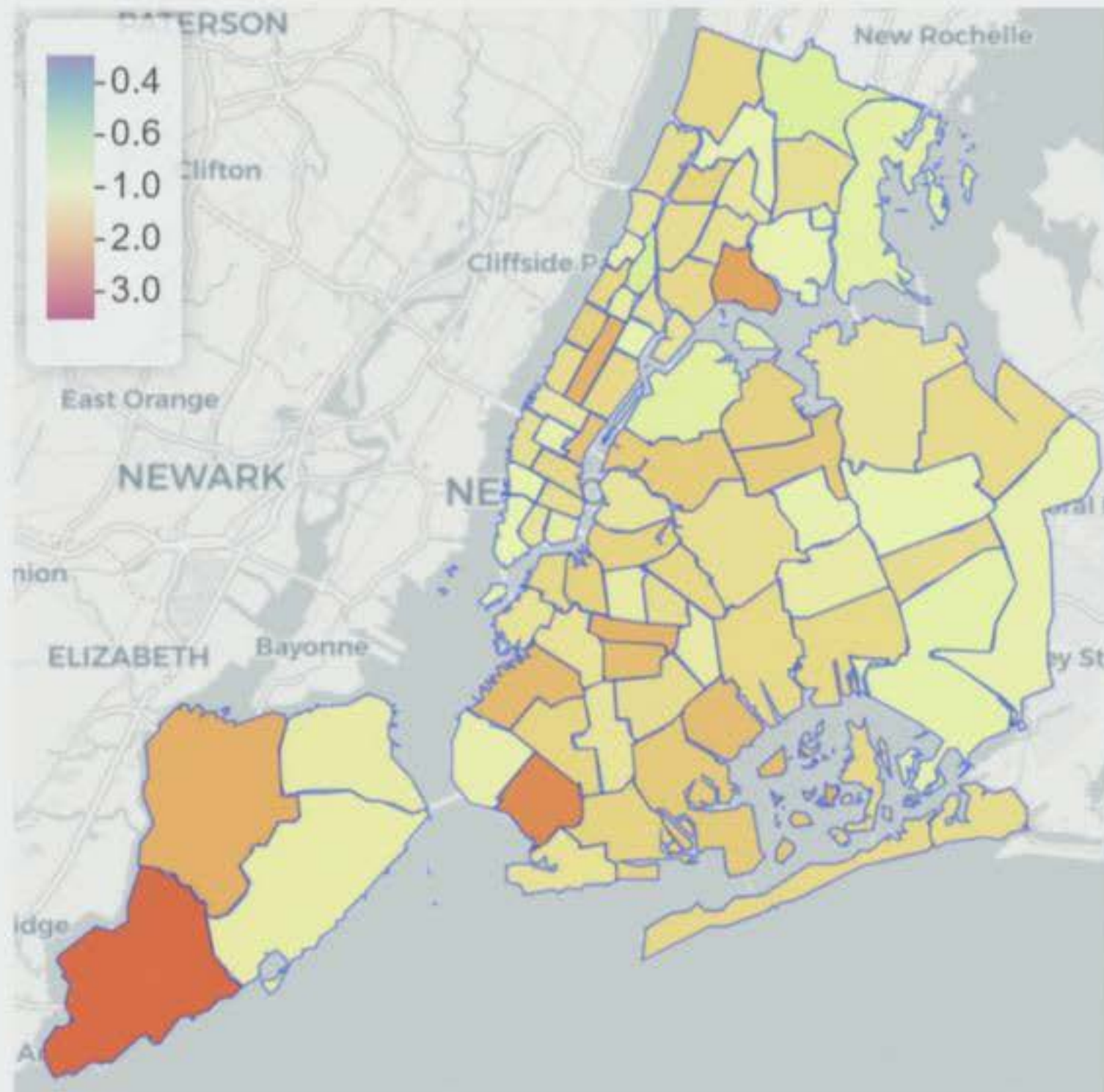
Calculated how much more likely were blacks subjected to any use of force compared to whites



$$\frac{P(\text{any force} \mid \text{black and stopped in precinct})}{P(\text{any force} \mid \text{white and stopped in precinct})}$$

- Blacks are more likely to have force used on them in almost every precinct
- The only exception is Astoria which is highlighted in green

Calculated the disparity between Hispanic and White civilians



$$\frac{P(\text{any force} \mid \text{black and stopped in precinct})}{P(\text{any force} \mid \text{hispanic and stopped in precinct})}$$

- There seems to be less of a disparity between use of force on Hispanic civilians vs White civilians than there is between Black and White
- Nonetheless, the map is more orange than green which still means that there is some discrimination taking place.

MINIMUM
FORCE



How much force was used?



Hands



Handcuffs



Pushed to Wall



Pushed to Floor



Drew Gun



Pointed Gun



Pepper Spray



Baton

Dividing into Two Intensity Levels

Low Intensity



High Intensity



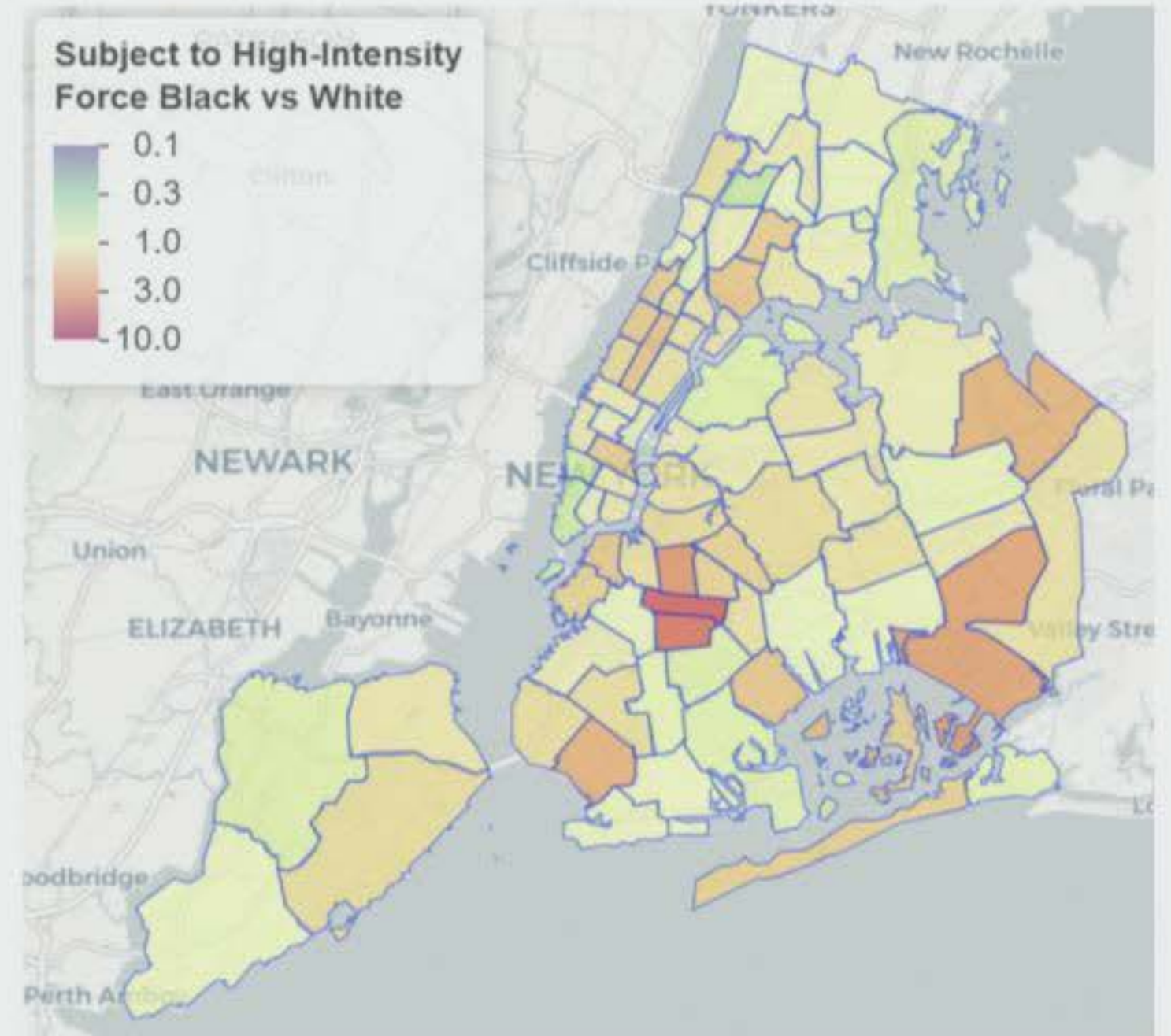
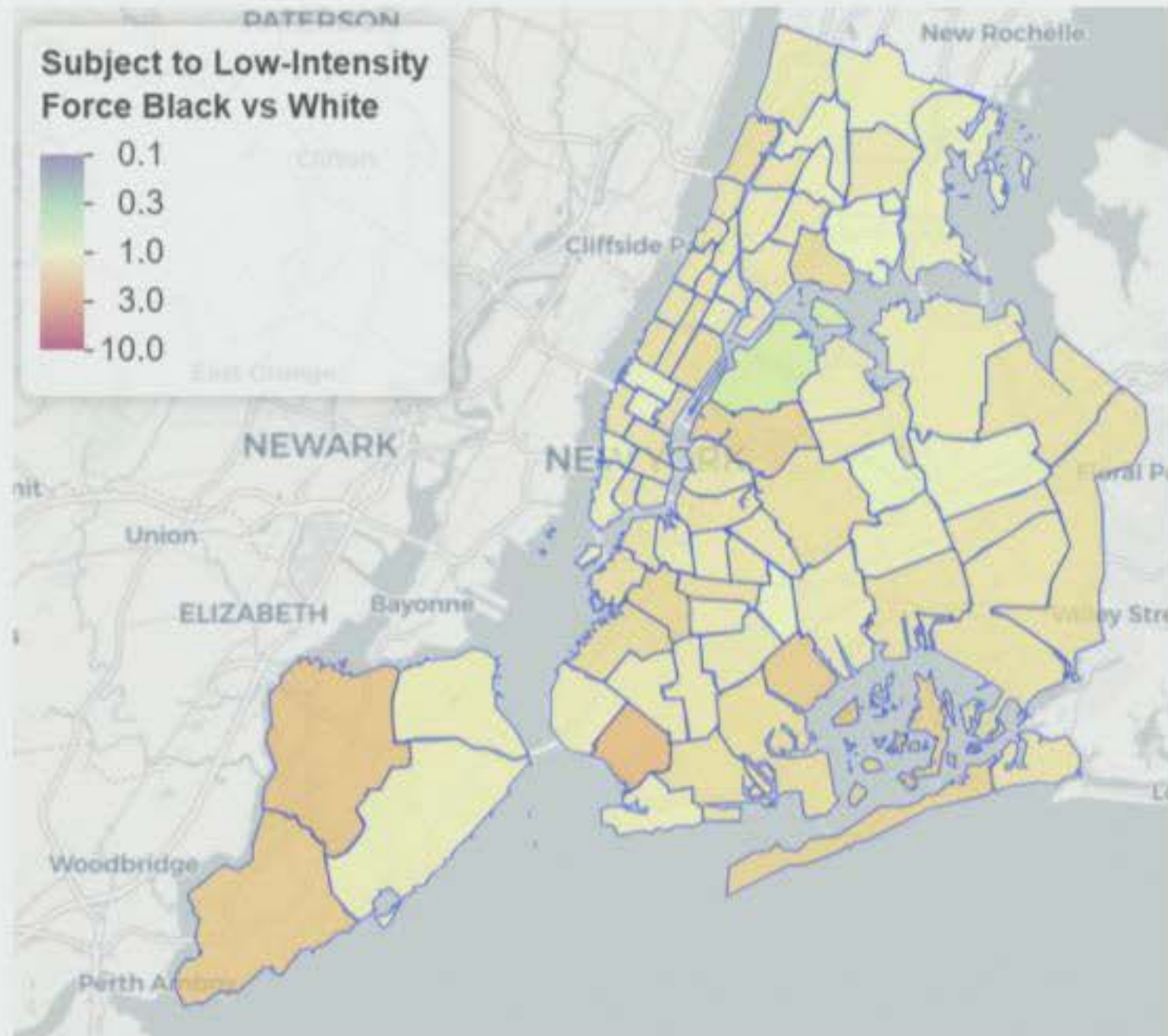
Racial Differences in Different Intensity of Forces

$P(\text{low intensity} | \text{black and stopped in precinct})$

$P(\text{high intensity} | \text{black and stopped in precinct})$

$P(\text{low intensity} | \text{white and stopped in precinct})$

$P(\text{high intensity} | \text{white and stopped in precinct})$

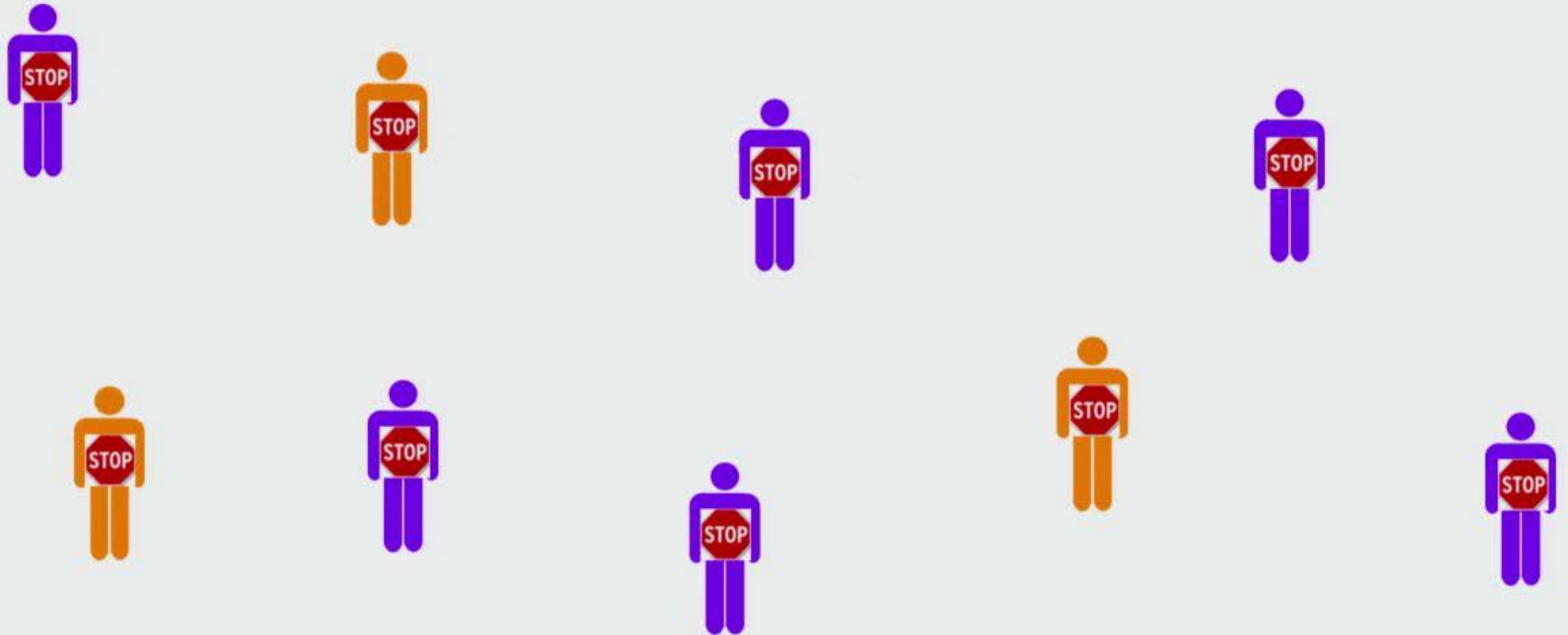


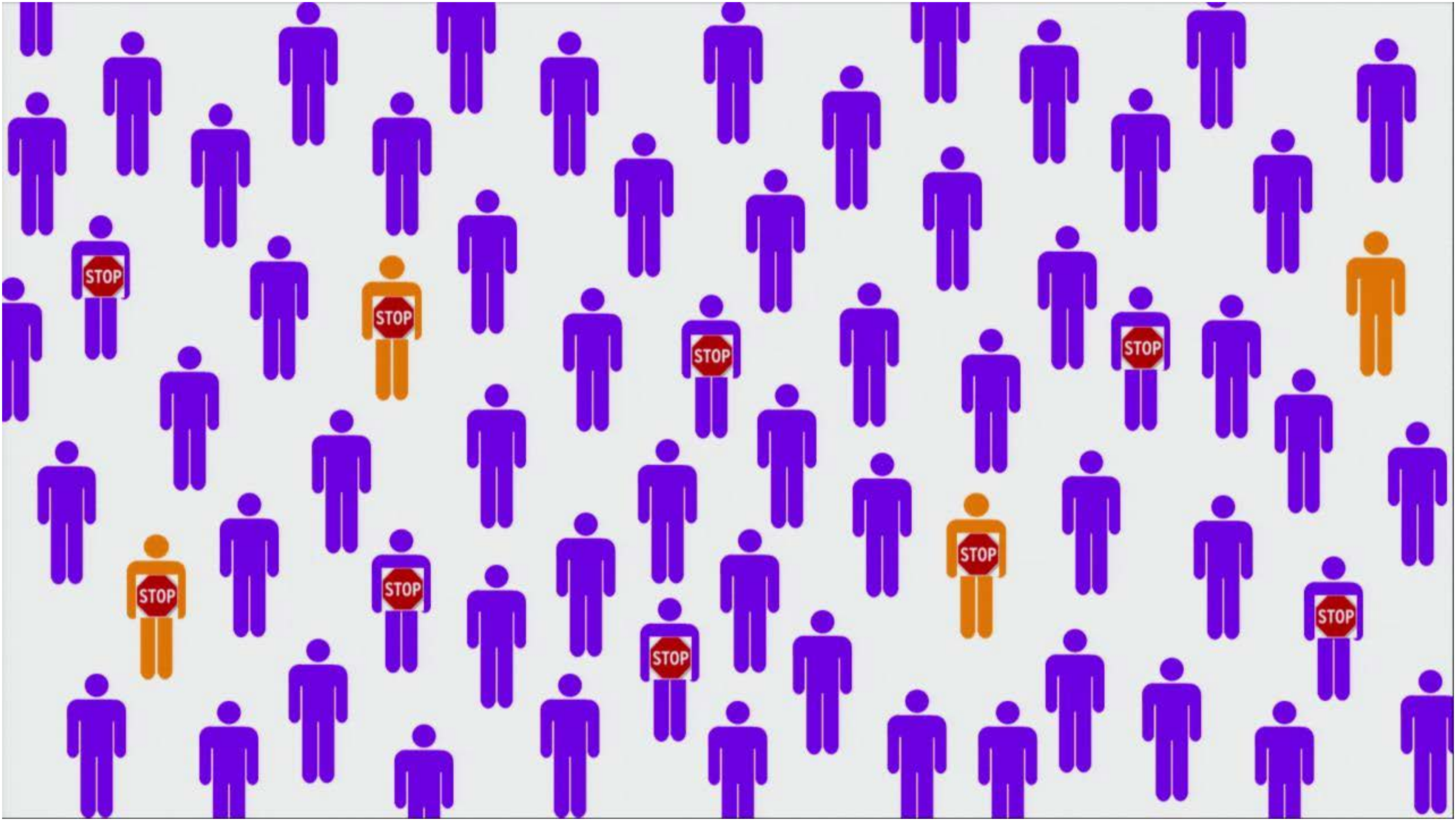
Here we see less of a racial disparity between the races in High intensity forces compared to Low Intensity



WHO IS STOPPED BY POLICE?

A problem in determining who is stopped



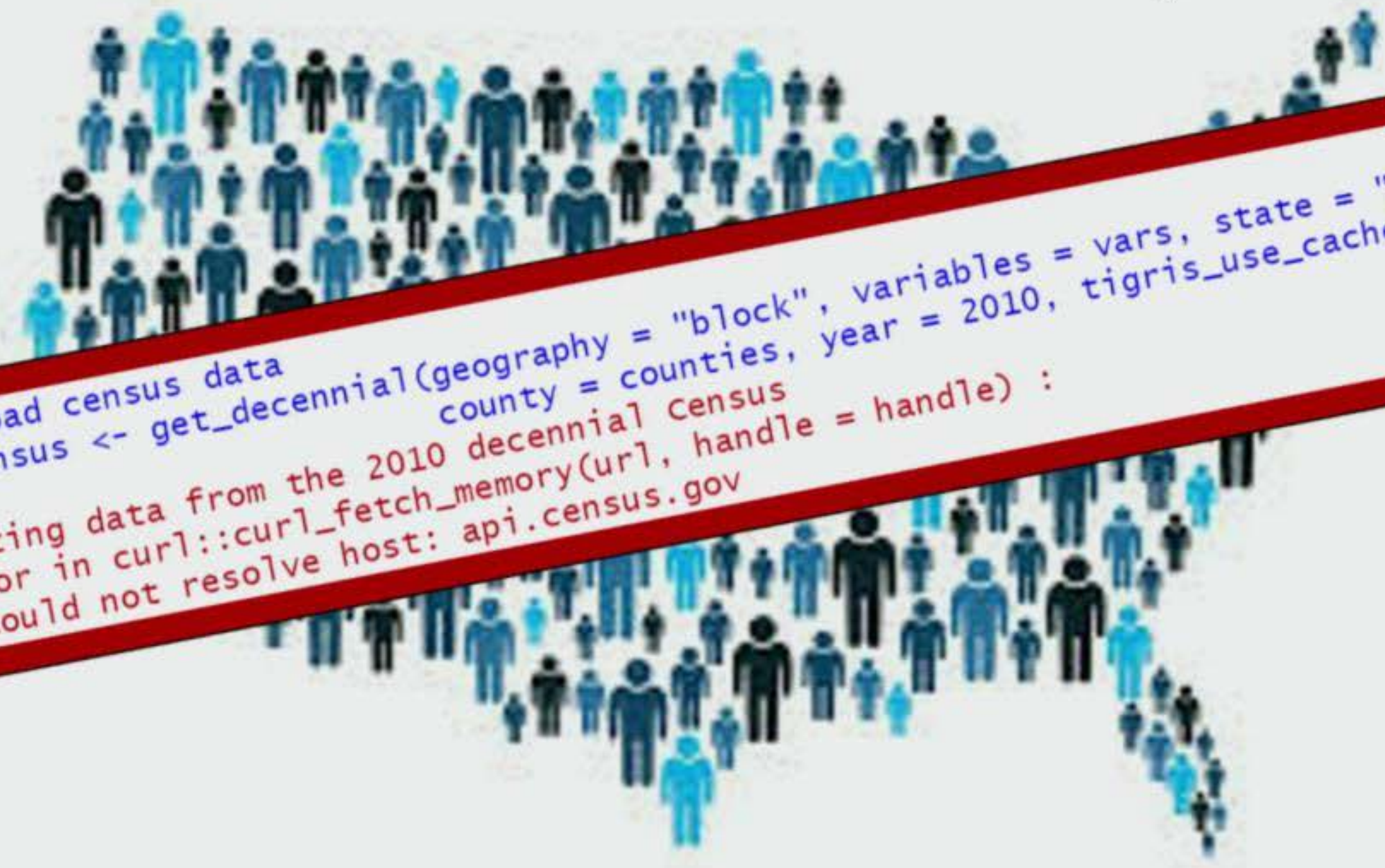


Bringing in Census data

- Census data gives the distribution of civilians that the police are stopping people from



When the Census took a day off...



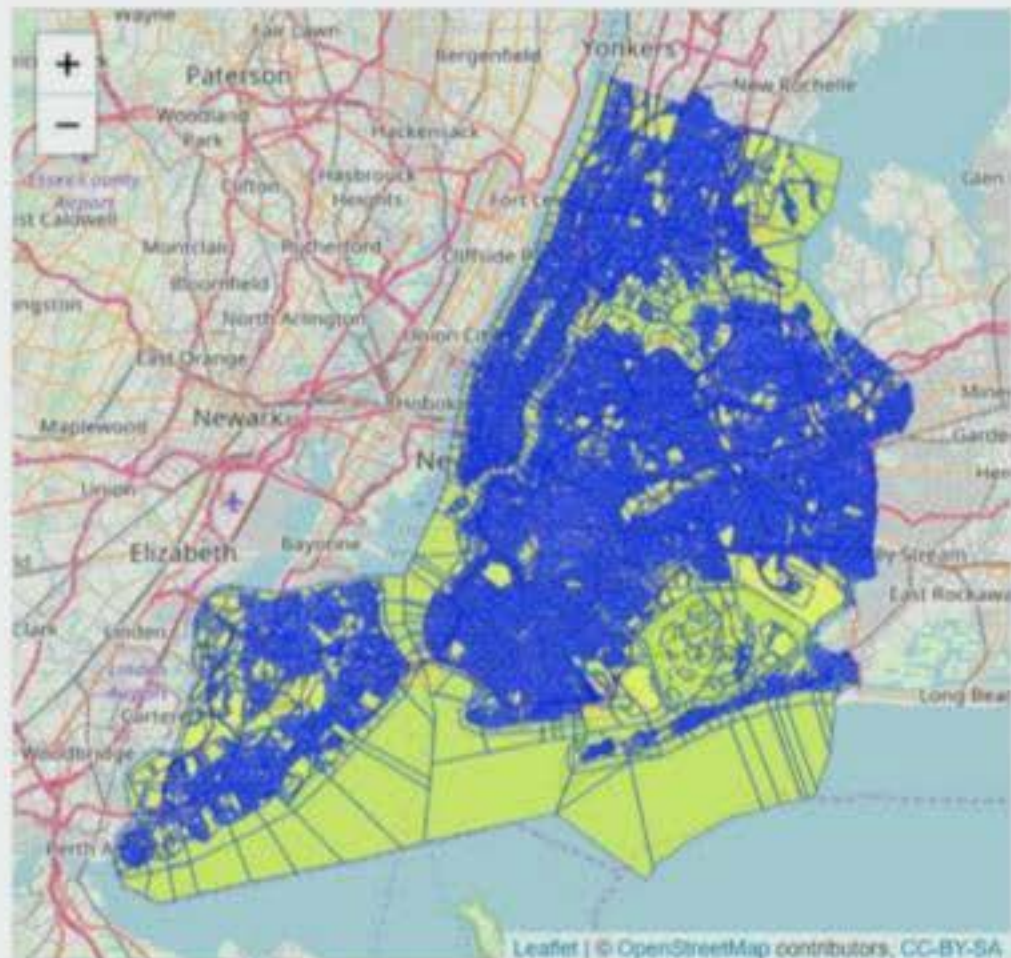
```
> #load census data
> census <- get_decennial(geography = "block", variables = vars, state = "NY",
+                          county = counties, year = 2010, tigris_use_cache = TRUE)
+
+ Getting data from the 2010 decennial census
Error in curl::curl_fetch_memory(url, handle = handle) :
could not resolve host: api.census.gov
```

Harpreet called them up!



Census Data + Geographic Data

Block Level



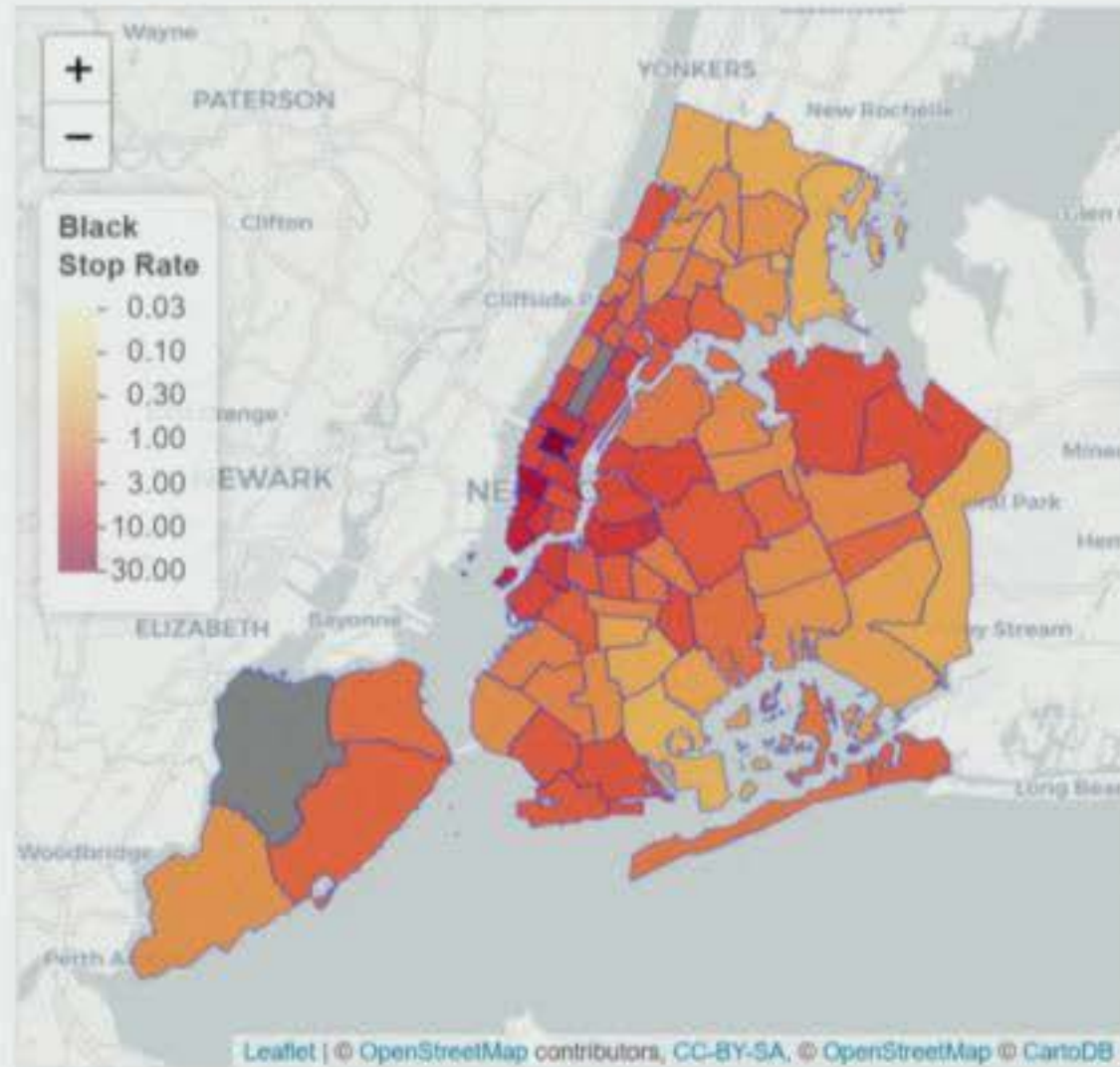
Precinct Level



Thanks to John Keefe, founder of the WNYC Data News Team for his [key](#) that maps blocks to precincts

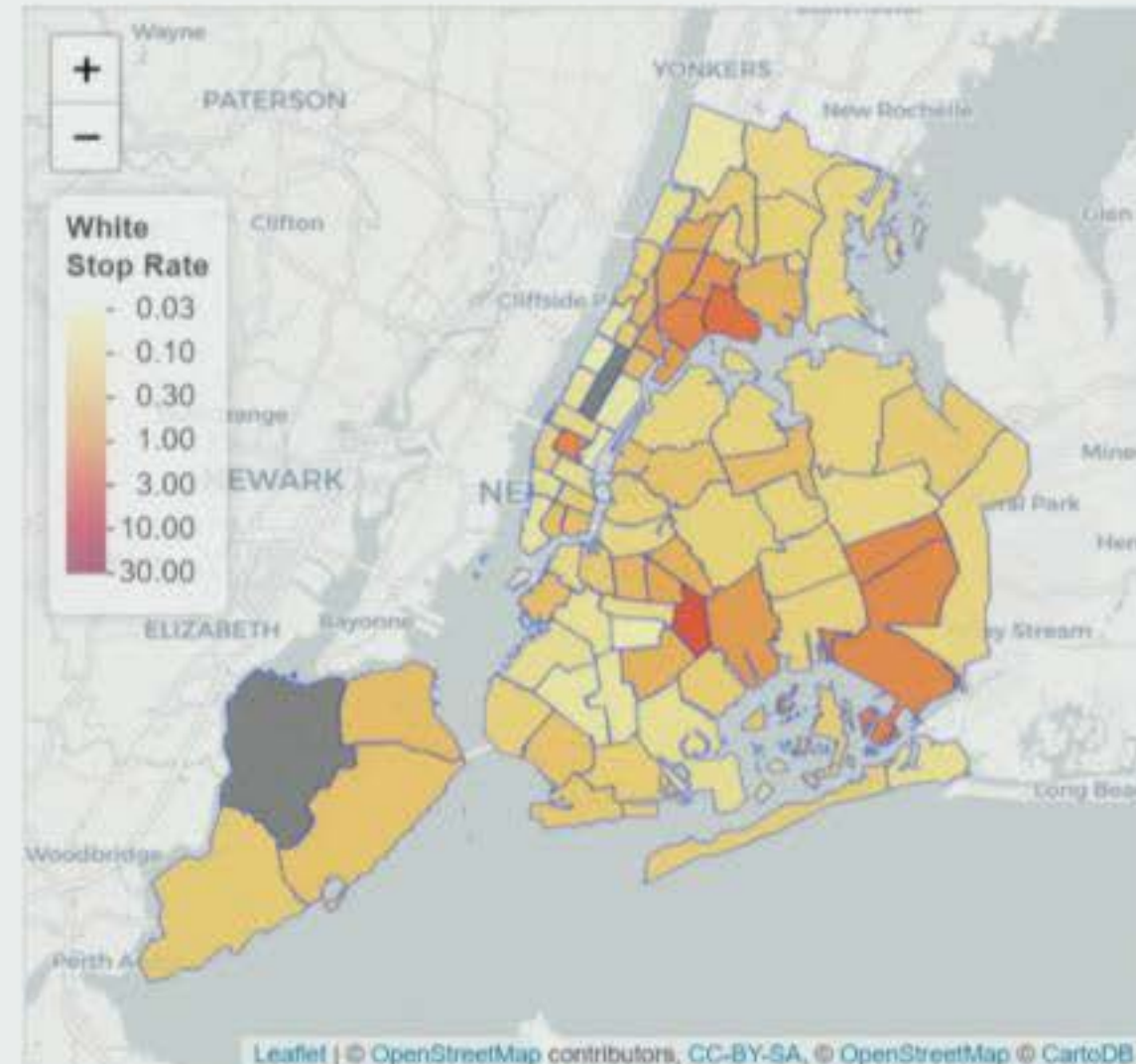
Who is Stopped?

Stop Rates by Race



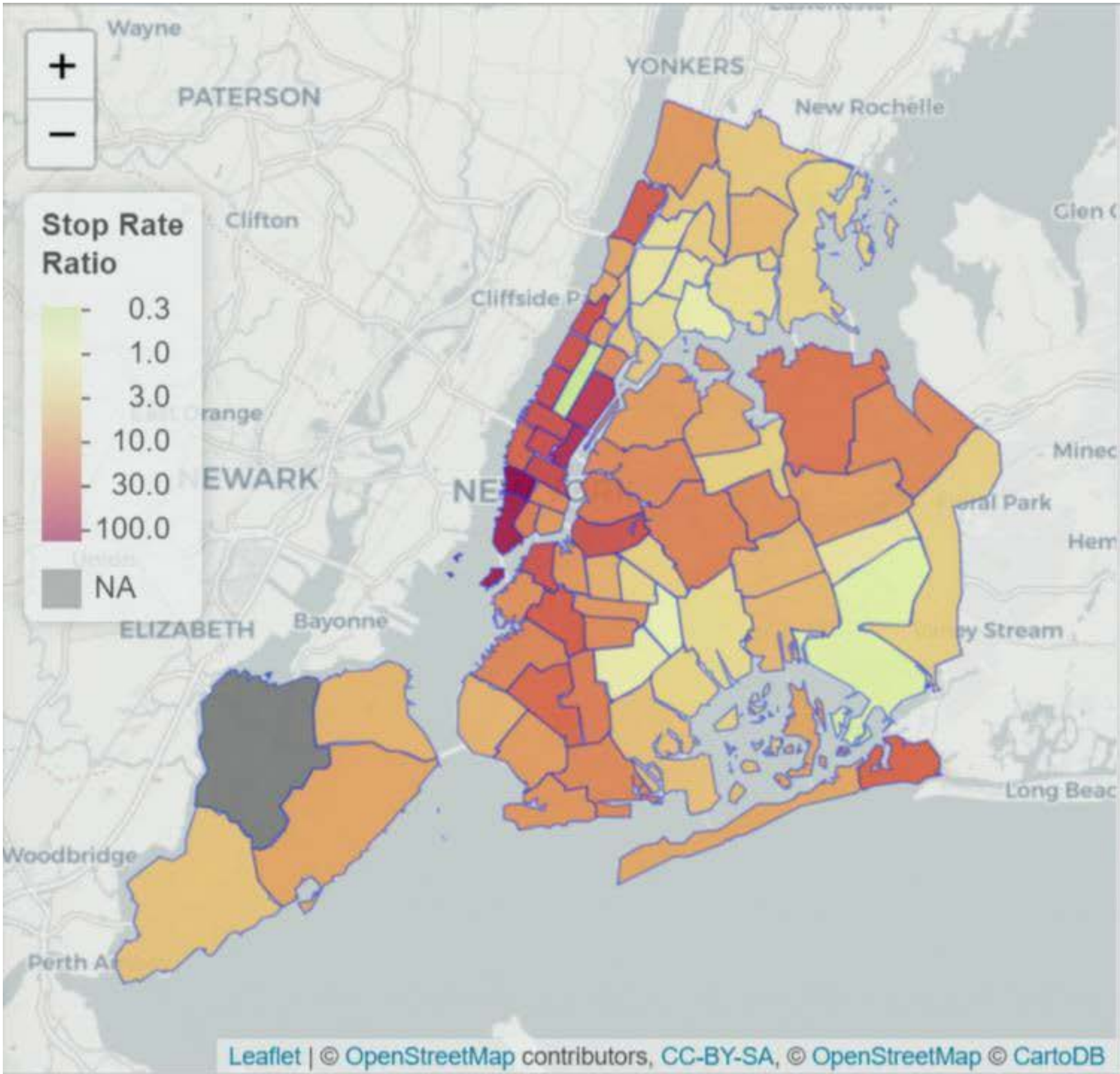
Number of Black People Stopped

Number of Black Residents



Number of White People Stopped

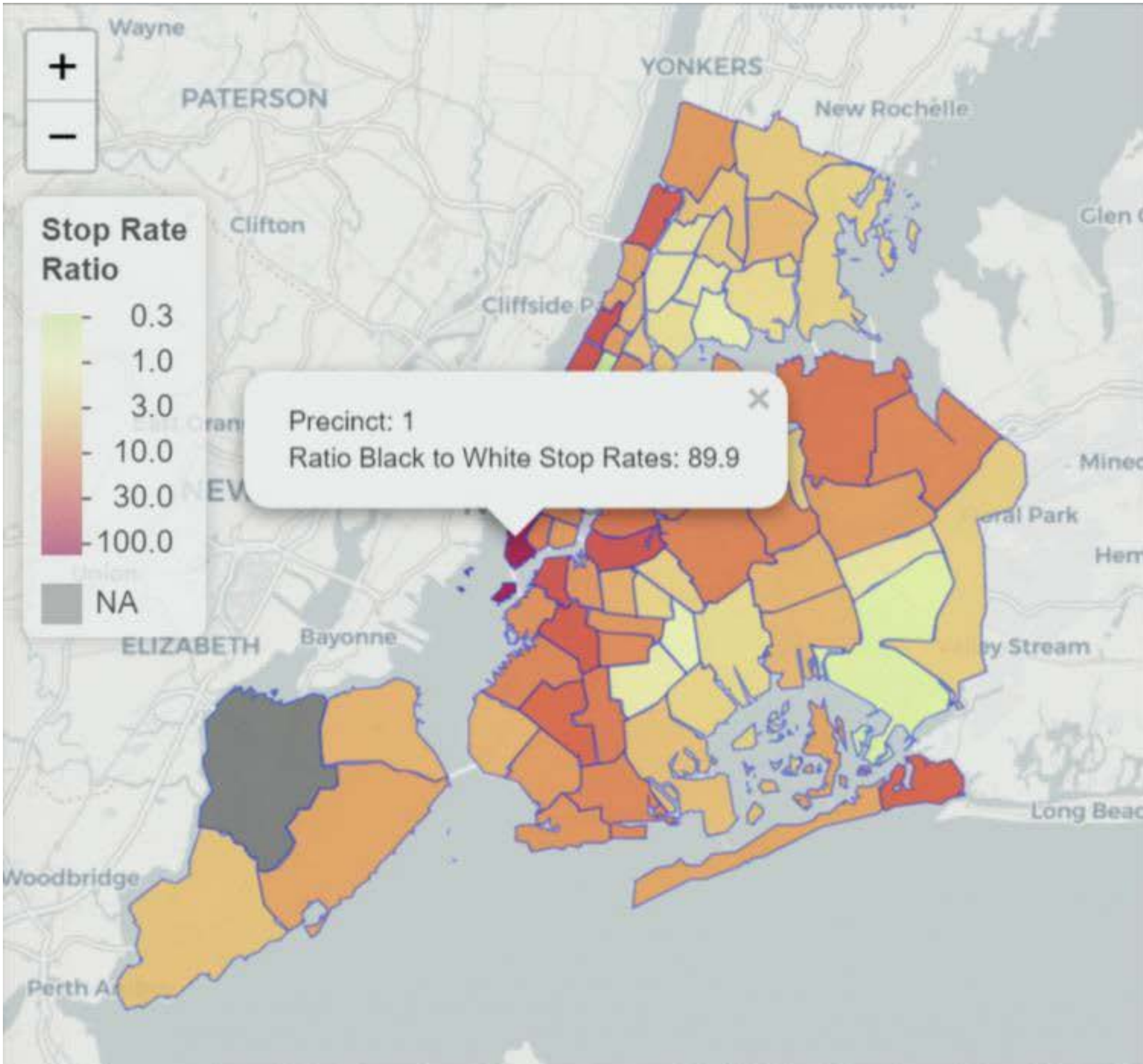
Number of White Residents



Who is Stopped?

Blacks are more likely to be stopped almost everywhere – another reason why Blacks are more often subject to force

$$\frac{\text{Black Stop Rate}}{\text{White Stop Rate}}$$









Who is Stopped?

Blacks are more likely to be stopped almost everywhere – another reason why Blacks are more often subject to force

$$\frac{\text{Black Stop Rate}}{\text{White Stop Rate}}$$

Caveat: Census Population ≠ Daytime Population

Commuter Flows (Journey to Work)		
Topic/ Year(s)	NYC & Boroughs	Census Tracts
NYC Residence to NYC Workplace		
2006-07-08-09-10		
Metro Counties to NYC Boroughs (1990-2010)		
2006-07-08-09-10	 	
NYC Boroughs to Metro Counties (1990-2010)		
2006-07-08-09-10	 	

- Census tells us where people live, not where they spend their time
- We found [data](#) describing commuting patterns, but it was on a census tract level and didn't include racial data



PPCS Data Set

Police Public Contact Survey



A Civilian's Point of View



- Provides us with a **civilian's perspective** of their interactions with the police



- Could reveal differences masked by police misreportings

What is Police Public Contact Survey (PPCS)?

- Nationally representative survey of civilians aged 16+
- Provides information of a civilian's perspective of their encounter(s) with the police
- Collected by the Bureau of Justice Statistics 1996 - 2015
- The data is publicly available on *NACJD's* website



Problems with Survey Data



People fill out nonsensical answers



People are less likely to admit to their own shortcomings



People might not remember things accurately



People fill out nonsensical answers

year	any_force	grab_push	hit_kick	point_gun	handcuffed	pepper_stun	contact	num_face_to_face	face_to_face
2011	0	0	0	0	0	0	NA	600	(1) Yes
2011	0	0	0	0	0	0	NA	365	(1) Yes
2011	0	0	0	0	0	0	NA	365	(2) No
2011	0	0	0	0	0	0	NA	365	(1) Yes
2011	0	0	0	0	0	0	NA	350	(1) Yes
2011	0	0	0	0	0	0	NA	300	(1) Yes
2011	0	0	0	0	0	0	NA	300	(2) No
2011	0	0	0	0	0	0	NA	300	(1) Yes
2011	0	0	0	0	0	0	NA	300	(2) No
2011	0	0	0	0	0	0	NA	261	(1) Yes
2011	0	0	0	0	0	0	NA	200	(1) Yes
2011	0	0	0	0	0	0	NA	129	(1) Yes
2011	0	0	0	0	0	0	NA	100	(2) No
2011	0	0	0	0	0	0	NA	100	(2) No
2011	0	0	0	0	0	0	NA	100	(1) Yes

Example: How many face to face contacts did you have with a police officer in the past 12 months?
600? 365? 350?



People are less likely to admit to their own shortcomings

8. Were you injured as a result of this contact?	1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No - <i>Skip</i> to Item 9
8a. What were the injuries you suffered? Mark all that apply. What other injuries did you suffer?	1 <input type="checkbox"/> Gun shot, bullet wounds 2 <input type="checkbox"/> Broken bones or teeth knocked out 3 <input type="checkbox"/> Internal injuries 4 <input type="checkbox"/> Knocked unconscious 5 <input type="checkbox"/> Bruises, black eye, cuts, scratches, swelling, chipped teeth 6 <input type="checkbox"/> Emotionally traumatized 7 <input type="checkbox"/> Some other type of injury- specify
9. During this contact were you arrested?	1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No 3 <input type="checkbox"/> Don't know
10. During this contact were you handcuffed?	1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No 3 <input type="checkbox"/> Don't know
11. At any time during this contact, did you argue with, curse at, insult, or verbally threaten the police?	1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No 3 <input type="checkbox"/> Don't know
12. At any time during this contact, did you ... a. Disobey or interfere with the officer(s)? b. Try to get away? c. Push, grab, or hit the police officer(s)? d. Resist being handcuffed, arrested, or searched? e. Physically do anything else? What else did you physically do?	YES NO 1 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> - If "Yes" specify below <i>Please specify</i> - _____
G. REASON FOR CONTACT	
13. Did this contact occur because you were involved in or witnessed a traffic ACCIDENT?	1 <input type="checkbox"/> Yes - <i>SKIP</i> to item 45 2 <input type="checkbox"/> No

Survey Question: "Did you hit the officer?"

Civilian:



A Few Challenges with Data Cleaning

- PPCS data columns were spread out for the variables that Fryer used
- We had to make subjective decisions when categorizing our columns
- Some of the variables used in the paper were missing



Differences in Officer Race Columns?

Differences in Officer Race Columns?

2005

Differences in Officer Race Columns?

2005

2 columns

V46 RACE OF OFFICER																			
Location:	361-362 (width: 2; decimal: 0)																		
Variable Type:	numeric (ISO)																		
Interval:	discrete																		
	<table border="1"><thead><tr><th>Value</th><th>Label</th></tr></thead><tbody><tr><td>0 (M)</td><td>Missing/NA</td></tr><tr><td>1</td><td>White</td></tr><tr><td>2</td><td>Black</td></tr><tr><td>3</td><td>Other</td></tr><tr><td>4</td><td>Do not know</td></tr></tbody></table>	Value	Label	0 (M)	Missing/NA	1	White	2	Black	3	Other	4	Do not know						
Value	Label																		
0 (M)	Missing/NA																		
1	White																		
2	Black																		
3	Other																		
4	Do not know																		
Summary statistics not available for this variable.																			
V47 RACE OF OFFICERS																			
Location:	363-364 (width: 2; decimal: 0)																		
Variable Type:	numeric (ISO)																		
Interval:	discrete																		
	<table border="1"><thead><tr><th>Value</th><th>Label</th></tr></thead><tbody><tr><td>0 (M)</td><td>Missing/NA</td></tr><tr><td>1</td><td>All White</td></tr><tr><td>2</td><td>All Black</td></tr><tr><td>3</td><td>All some other race</td></tr><tr><td>4</td><td>Mostly White</td></tr><tr><td>5</td><td>Mostly Black</td></tr><tr><td>6</td><td>Mostly some other race</td></tr><tr><td>7</td><td>Equally mixed</td></tr></tbody></table>	Value	Label	0 (M)	Missing/NA	1	All White	2	All Black	3	All some other race	4	Mostly White	5	Mostly Black	6	Mostly some other race	7	Equally mixed
Value	Label																		
0 (M)	Missing/NA																		
1	All White																		
2	All Black																		
3	All some other race																		
4	Mostly White																		
5	Mostly Black																		
6	Mostly some other race																		
7	Equally mixed																		

2011

12 columns

Differences in Officer Race Columns?

2005

2 columns

V46 RACE OF OFFICER	
Location:	381-382 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Value	Label
0 (M)	Missing/NA
1	White
2	Black
3	Other
4	Do not know
Summary statistics not available for this variable.	
V47 RACE OF OFFICERS	
Location:	383-384 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Value	Label
0 (M)	Missing/NA
1	All White
2	All Black
3	All some other race
4	Mostly White
5	Mostly Black
6	Mostly some other race
7	Equally mixed

2011

12 columns

V67 - WAS THE POLICE OFFICER OF HISPANIC OR LATINO ORIGIN	
Location:	225-226 (width: 2; decimal: 0)
Variable Type:	numeric
Question:	
Was the police officer of Hispanic or Latino origin?	
Value	Label
1	Yes
2	No
3	Don't know
8	Refused
-9 (M)	Out of universe/missing
V68 - OFFICER 1 RACE	
Location:	227-228 (width: 2; decimal: 0)
Variable Type:	numeric
Question:	
Was the race of the police officer White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander? You may select one or more.	

- 48 -

- Study 34276 -

Value	Label
0	Not selected
1	White

Differences in Officer Race Columns?

2005

2 columns

V46 RACE OF OFFICER	
Location:	381-382 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Value	Label
0 (M)	Missing/NA
1	White
2	Black
3	Other
4	Do not know
Summary statistics not available for this variable.	
V47 RACE OF OFFICERS	
Location:	383-384 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Value	Label
0 (M)	Missing/NA
1	All White
2	All Black
3	All some other race
4	Mostly White
5	Mostly Black
6	Mostly some other race
7	Equally mixed

2011

12 columns

V6 V6 V72 V77 - OFFICER 3 RACE		
Loc	Location: 245-246 (width: 2; decimal: 0)	
Var	Variable Type: numeric	
Q	Question:	
W	Were one or more of the police officers White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander? You may select one or more for each officer.	
Value	Label	
0	0	Not selected
1	1	American Indian or Alaska Native
2	8	Refused
3	-9 (M)	Out of universe/missing
V73 V78 - OFFICER 4 RACE		
Loc	Location: 247-248 (width: 2; decimal: 0)	
Var	Variable Type: numeric	
Q	Question:	
W	Were one or more of the police officers White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander? You may select one or more for each officer.	
Value	Label	
0	0	Not selected
1	8	Asian
2	-9 (M)	Out of universe/missing

- 51 -

- Study 34276 -

Differences in Officer Race Columns?

2005

2 columns

V46 RACE OF OFFICER	
Location:	381-382 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Value	Label
0 (M)	Missing/NA
1	White
2	Black
3	Other
4	Do not know
Summary statistics not available for this variable.	
V47 RACE OF OFFICERS	
Location:	383-384 (width: 2; decimal: 0)
Variable Type:	numeric (ISO)
Interval:	discrete
Value	Label
0 (M)	Missing/NA
1	All White
2	All Black
3	All some other race
4	Mostly White
5	Mostly Black
6	Mostly some other race
7	Equally mixed

2011

12 columns

V79 - OFFICER 5 RACE	
Location:	249-250 (width: 2; decimal: 0)
Variable Type:	numeric
Question:	
Were one or more of the police officers White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander? You may select one or more for each officer.	
Value	Label
0	Not selected
1	Native Hawaiian or other Pacific Islander
8	Refused
-9 (M)	Out of universe/missing
V80 - OFFICER 6 RACE	
Location:	251-252 (width: 2; decimal: 0)
Variable Type:	numeric
Question:	
Were one or more of the police officers White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander? You may select one or more for each officer.	
Value	Label
0	Not selected
1	Don't know
8	Refused
-9 (M)	Out of universe/missing
V81 - ABOUT WHAT TIME OF DAY DID THIS CONTACT OCCUR	
Location:	253-254 (width: 2; decimal: 0)
Variable Type:	numeric
Question:	
About what time of day did this contact occur?	
Value	Label
01	After 6 a.m. - 12 noon
02	After 12 noon - 6 p.m.
03	Don't know what time of day
04	After 6 p.m. - 12 midnight
05	After 12 midnight - 6 a.m.
06	Don't know what time of night

Subjective Decisions:

TRAFFIC STOP? STREET STOP? OTHER?

We spent quite some time figuring out the labels for each reason, for each year:

Subjective Decisions:

TRAFFIC **STOP?** STREET **STOP?** OTHER?

We spent quite some time figuring out the labels for each reason, for each year:

1996

- REASON_REPORT_A_CRIME,
- REASON_ASK_FOR_ASSISTANCE
- REASON_PROBLEM_IN_THE_NEIGHBORHOOD
- REASON_TRAFFIC_PARKING_VIOLATION
- REASON_INVOLVED_IN_TRAFFIC_INCIDENT
- REASON_WITNESS_TO_TRAFFIC_ACCIDENT
- REASON_VICTIM_OF_A_CRIME
- REASON_WITNESS_TO_A_CRIME
- REASON_CRIME_THOUGHT_INVOLVED_IN
- REASON_WHAT_YOU_WERE_DOING_IN_THE_AREA
- REASON_HAD_WARRANT_FOR_YOUR_ARREST
- REASON_CASUAL_ENCOUNTERS
- REASON_COMMUNITY_MEETINGS
- REASON_SOME_OTHER_REASON
- REASON_NO_SPECIFIC_REASON

2011

- REASON FOR STOP: SOMEONE WITH YOU MATCH DESCRIPTION
- REASON FOR STOP: PROVIDE ASSISTANCE TO YOU
- REASON FOR STOP: INVESTIGATING A CRIME
- REASON FOR STOP: SEEKING INFORMATION
- REASON FOR STOP: SUSPECT YOU OF SOMETHING
- REASON FOR STOP – JAYWALKING
- REASON FOR STOP - CREDENTIAL VALIDATION
- REASON FOR STOP - NOISE VIOLATION
- REASON - OUT AFTER CURFEW
- REASON FOR STOP - DRINKING RELATED REASON
- REASON FOR STOP - PARKING VIOLATION
- REASON FOR STOP - SECURITY CHECK
- REASON FOR STOP - NOISE VIOLATION
- REASON FOR STOP - CREDENTIAL VALIDATION
- REASON FOR STOP - BIKE VIOLATION

Missing Data

Fryer was **not specific** about how the data was aggregated

Some years were missing variables that were used for the regression

For example:

- 1996 & 2008 are missing income and population size
- 1999 is missing the time of encounter
- 2005 is missing civilian behavior

As a result, the missing variables affected the regression!



Data Cleaning

After merging all the years together, and finding our case counts, we ran into a case count discrepancy:

Fryer's total sample: 426,000 observations

Our total sample: 409,678 observations

And after many hours and days of trying to fix this discrepancy, we came up with a new solution.....

We decided to email the Data Manager from
the *NATIONAL ARCHIVE OF CRIMINAL JUSTICE DATA* :

"I am unsure why there is a case
count discrepancy. I took a look at all
the data files, and none of them
have been updated since 2014
(when PPCS 2011 was released).

I do not recall the exact number, but
I know it was approximately 409,000"

- Data Manager
NACJD

The mystery of the extra cases still remains...

Summary Statistics

	Fryer Stats			
	Full Sample	White	Black	Hispanic
Panel A: Civilian Demographics				
White	0.77	1.00	0.00	0.00
Black	0.10	0.00	1.00	0.00
Other	0.04	0.00	0.00	0.00
Hispanic	0.09	0.00	0.00	1.00
Age	40.92	42.06	38.98	34.99
Male	0.50	0.50	0.45	0.54
Female	0.50	0.50	0.55	0.46
Employed last week	0.72	0.72	0.69	0.74
Income	2.09	2.15	1.75	1.89
Population Size	1.41	1.31	1.79	1.77

Summary Statistics

	Fryer Stats				Our Stats			
	Full Sample	White	Black	Hispanic	Full Sample	White	Black	Hispanic
Panel A: Civilian Demographics								
White	0.77	1.00	0.00	0.00	0.77	1.00	0.00	0.00
Black	0.10	0.00	1.00	0.00	0.09	0.00	1.00	0.00
Other	0.04	0.00	0.00	0.00	0.04	0.00	0.00	0.00
Hispanic	0.09	0.00	0.00	1.00	0.09	0.00	0.00	1.00
Age	40.92	42.06	38.98	34.99	40.89	41.99	39.32	34.99
Male	0.50	0.50	0.45	0.54	0.50	0.50	0.45	0.54
Female	0.50	0.50	0.55	0.46	0.50	0.50	0.55	0.46
Employed last week	0.72	0.72	0.69	0.74	0.70	0.70	0.68	0.73
Income	2.09	2.15	1.75	1.89	2.09	2.15	1.74	1.88
Population Size	1.41	1.31	1.79	1.77	1.41	1.31	1.80	1.77

Summary Statistics

	Fryer Stats				Our Stats			
	Full Sample	White	Black	Hispanic	Full Sample	White	Black	Hispanic
Panel A: Civilian Demographics								
White	0.77	1.00	0.00	0.00	0.77	1.00	0.00	0.00
Black	0.10	0.00	1.00	0.00	0.09	0.00	1.00	0.00
Other	0.04	0.00	0.00	0.00	0.04	0.00	0.00	0.00
Hispanic	0.09	0.00	0.00	1.00	0.09	0.00	0.00	1.00
Age	40.92	42.06	38.98	34.99	40.89	41.99	39.32	34.99
Male	0.50	0.50	0.45	0.54	0.50	0.50	0.45	0.54
Female	0.50	0.50	0.55	0.46	0.50	0.50	0.55	0.46
Employed last week	0.72	0.72	0.69	0.74	0.70	0.70	0.68	0.73
Income	2.09	2.15	1.75	1.89	2.09	2.15	1.74	1.88
Population Size	1.41	1.31	1.79	1.77	1.41	1.31	1.80	1.77
Panel B: Civilian Behavior								
Disobeyed	0.00	0.00	0.01	0.01				
Tried to get away	0.00	0.00	0.00	0.00				
Resisted	0.00	0.00	0.00	0.00				
Complained	0.06	0.05	0.07	0.06				
Argued	0.02	0.02	0.02	0.02				
Threatened Officer	0.01	0.01	0.02	0.01				
Used Physical Force	0.00	0.00	0.00	0.00				
					0.01	0.01	0.02	0.01
Panel C: Contact and Officer Characteristics								
Street Stop	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Traffic Stop	0.55	0.54	0.56	0.57	0.48	0.47	0.49	0.53
Other Stop	0.45	0.45	0.44	0.42	0.51	0.52	0.51	0.46
Daytime	0.68	0.69	0.62	0.66	0.71	0.71	0.66	0.70
Night	0.32	0.31	0.38	0.34	0.29	0.29	0.34	0.30
Off majority Hispanic	0.02	0.02	0.04	0.07	0.00	0.00	0.00	0.00
Off majority White	0.88	0.90	0.76	0.84	0.88	0.90	0.75	0.85
Off majority Black	0.06	0.05	0.16	0.05	0.07	0.06	0.18	0.06
Off majority Other	0.03	0.02	0.03	0.07	0.03	0.02	0.03	0.07
Off split race	0.02	0.02	0.04	0.02	0.02	0.02	0.04	0.02
Panel D: Alternative Outcomes								
Injured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Perceived excess force	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
Searched	0.05	0.04	0.11	0.09	0.02	0.01	0.03	0.02
Arrested	0.03	0.02	0.05	0.04	0.02	0.02	0.04	0.03
Guilty of carrying ille	0.14	0.16	0.10	0.13	0.00	0.00	0.00	0.01
Panel E: Use of Force								
Any use of force	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
Grab or push	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01
Hit or kick	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Point Gun	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Handcuffed	0.02	0.02	0.04	0.03	0.02	0.02	0.04	0.03
Pepper spray/stungu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Type of Incident

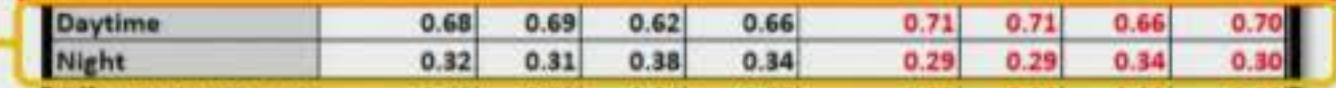


Summary Statistics

	Fryer Stats				Our Stats			
	Full Sample	White	Black	Hispanic	Full Sample	White	Black	Hispanic
Panel A: Civilian Demographics								
White	0.77	1.00	0.00	0.00	0.77	1.00	0.00	0.00
Black	0.10	0.00	1.00	0.00	0.09	0.00	1.00	0.00
Other	0.04	0.00	0.00	0.00	0.04	0.00	0.00	0.00
Hispanic	0.09	0.00	0.00	1.00	0.09	0.00	0.00	1.00
Age	40.92	42.06	38.98	34.99	40.89	41.99	39.32	34.99
Male	0.50	0.50	0.45	0.54	0.50	0.50	0.45	0.54
Female	0.50	0.50	0.55	0.46	0.50	0.50	0.55	0.46
Employed last week	0.72	0.72	0.69	0.74	0.70	0.70	0.68	0.73
Income	2.09	2.15	1.75	1.89	2.09	2.15	1.74	1.88
Population Size	1.41	1.31	1.79	1.77	1.41	1.31	1.80	1.77
Panel B: Civilian Behavior								
Disobeyed	0.00	0.00	0.01	0.01				
Tried to get away	0.00	0.00	0.00	0.00				
Resisted	0.00	0.00	0.00	0.00				
Complained	0.06	0.05	0.07	0.06				
Argued	0.02	0.02	0.02	0.02				
Threatened Officer	0.01	0.01	0.02	0.01				
Used Physical Force	0.00	0.00	0.00	0.00				
					0.01	0.01	0.02	0.01
Panel C: Contact and Officer Characteristics								
Street Stop	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Traffic Stop	0.55	0.54	0.56	0.57	0.48	0.47	0.49	0.53
Other Stop	0.45	0.45	0.44	0.42	0.51	0.52	0.51	0.46
Daytime	0.68	0.69	0.62	0.66	0.71	0.71	0.66	0.70
Night	0.32	0.31	0.38	0.34	0.29	0.29	0.34	0.30
Off majority Hispanic	0.02	0.02	0.04	0.07	0.00	0.00	0.00	0.00
Off majority White	0.88	0.90	0.76	0.84	0.88	0.90	0.75	0.85
Off majority Black	0.06	0.05	0.16	0.05	0.07	0.06	0.18	0.06
Off majority Other	0.03	0.02	0.03	0.07	0.03	0.02	0.03	0.07
Off split race	0.02	0.02	0.04	0.02	0.02	0.02	0.04	0.02
Panel D: Alternative Outcomes								
Injured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Perceived excess force	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
Searched	0.05	0.04	0.11	0.09	0.02	0.01	0.03	0.02
Arrested	0.03	0.02	0.05	0.04	0.02	0.02	0.04	0.03
Guilty of carrying ille	0.14	0.16	0.10	0.13	0.00	0.00	0.00	0.01
Panel E: Use of Force								
Any use of force	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
Grab or push	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01
Hit or kick	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Point Gun	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Handcuffed	0.02	0.02	0.04	0.03	0.02	0.02	0.04	0.03
Pepper spray/stungu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Type of Incident

Time of Encounter



Summary Statistics

	Fryer Stats				Our Stats			
	Full Sample	White	Black	Hispanic	Full Sample	White	Black	Hispanic
Panel A: Civilian Demographics								
White	0.77	1.00	0.00	0.00	0.77	1.00	0.00	0.00
Black	0.10	0.00	1.00	0.00	0.09	0.00	1.00	0.00
Other	0.04	0.00	0.00	0.00	0.04	0.00	0.00	0.00
Hispanic	0.09	0.00	0.00	1.00	0.09	0.00	0.00	1.00
Age	40.92	42.06	38.98	34.99	40.89	41.99	39.32	34.99
Male	0.50	0.50	0.45	0.54	0.50	0.50	0.45	0.54
Female	0.50	0.50	0.55	0.46	0.50	0.50	0.55	0.46
Employed last week	0.72	0.72	0.69	0.74	0.70	0.70	0.68	0.73
Income	2.09	2.15	1.75	1.89	2.09	2.15	1.74	1.88
Population Size	1.41	1.31	1.79	1.77	1.41	1.31	1.80	1.77
Panel B: Civilian Behavior								
Disobeyed	0.00	0.00	0.01	0.01				
Tried to get away	0.00	0.00	0.00	0.00				
Resisted	0.00	0.00	0.00	0.00				
Complained	0.06	0.05	0.07	0.06				
Argued	0.02	0.02	0.02	0.02				
Threatened Officer	0.01	0.01	0.02	0.01				
Used Physical Force	0.00	0.00	0.00	0.00				
					0.01	0.01	0.02	0.01
Panel C: Contact and Officer Characteristics								
Street Stop	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Traffic Stop	0.55	0.54	0.56	0.57	0.48	0.47	0.49	0.53
Other Stop	0.45	0.45	0.44	0.42	0.51	0.52	0.51	0.46
Daytime	0.68	0.69	0.62	0.66	0.71	0.71	0.66	0.70
Night	0.32	0.31	0.38	0.34	0.29	0.29	0.34	0.30
Off majority Hispanic	0.02	0.02	0.04	0.07	0.00	0.00	0.00	0.00
Off majority White	0.88	0.90	0.76	0.84	0.88	0.90	0.75	0.85
Off majority Black	0.06	0.05	0.16	0.05	0.07	0.06	0.18	0.06
Off majority Other	0.03	0.02	0.03	0.07	0.03	0.02	0.03	0.07
Off split race	0.02	0.02	0.04	0.02	0.02	0.02	0.04	0.02
Panel D: Alternative Outcomes								
Injured	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Perceived excess force	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
Searched	0.05	0.04	0.11	0.09	0.02	0.01	0.03	0.02
Arrested	0.03	0.02	0.05	0.04	0.02	0.02	0.04	0.03
Guilty of carrying ille	0.14	0.16	0.10	0.13	0.00	0.00	0.00	0.01
Panel E: Use of Force								
Any use of force	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
Grab or push	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01
Hit or kick	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Point Gun	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Handcuffed	0.02	0.02	0.04	0.03	0.02	0.02	0.04	0.03
Pepper spray/stungu	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

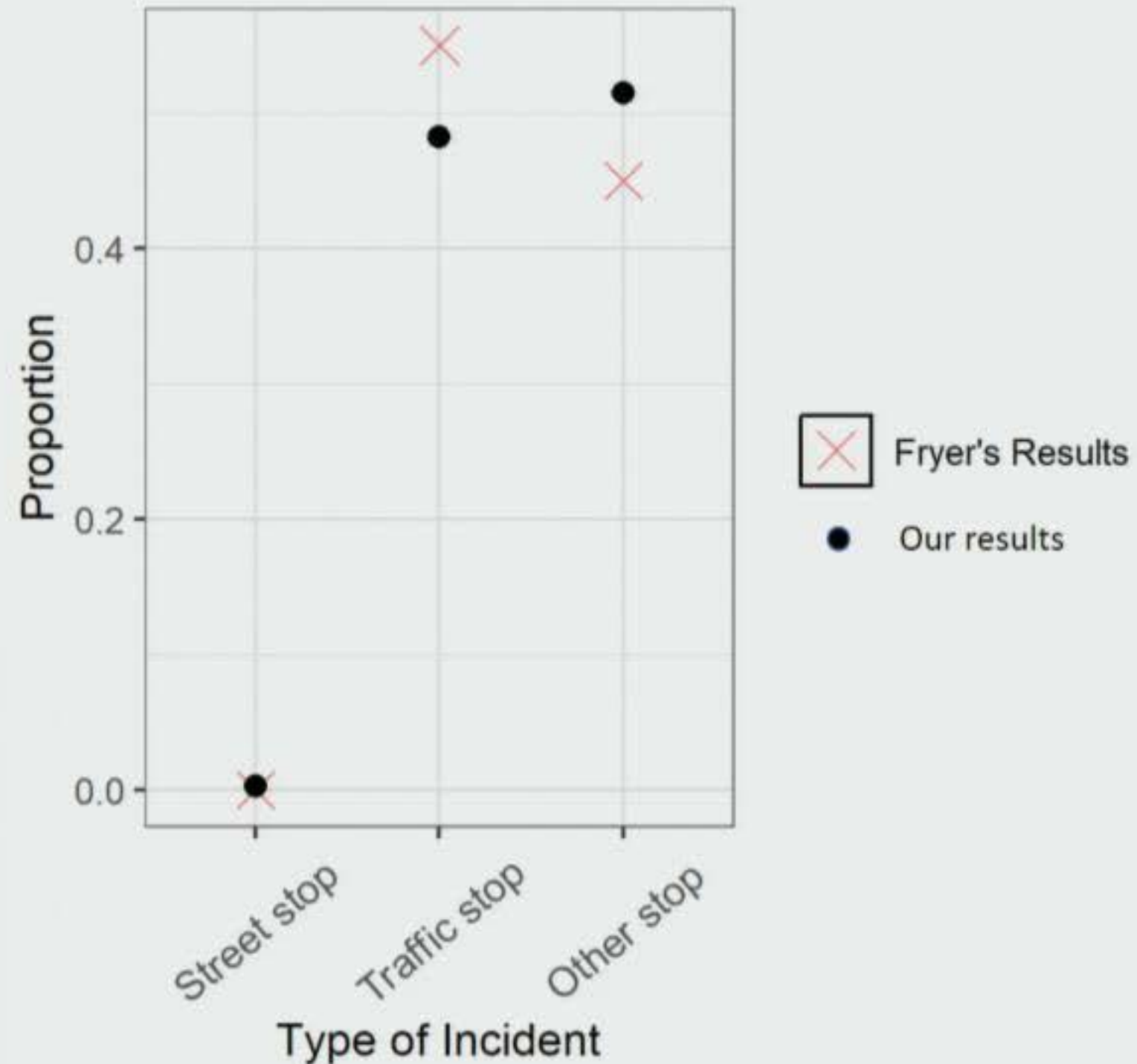
Type of Incident

Time of Encounter

Alternative Outcome

Summary Statistics

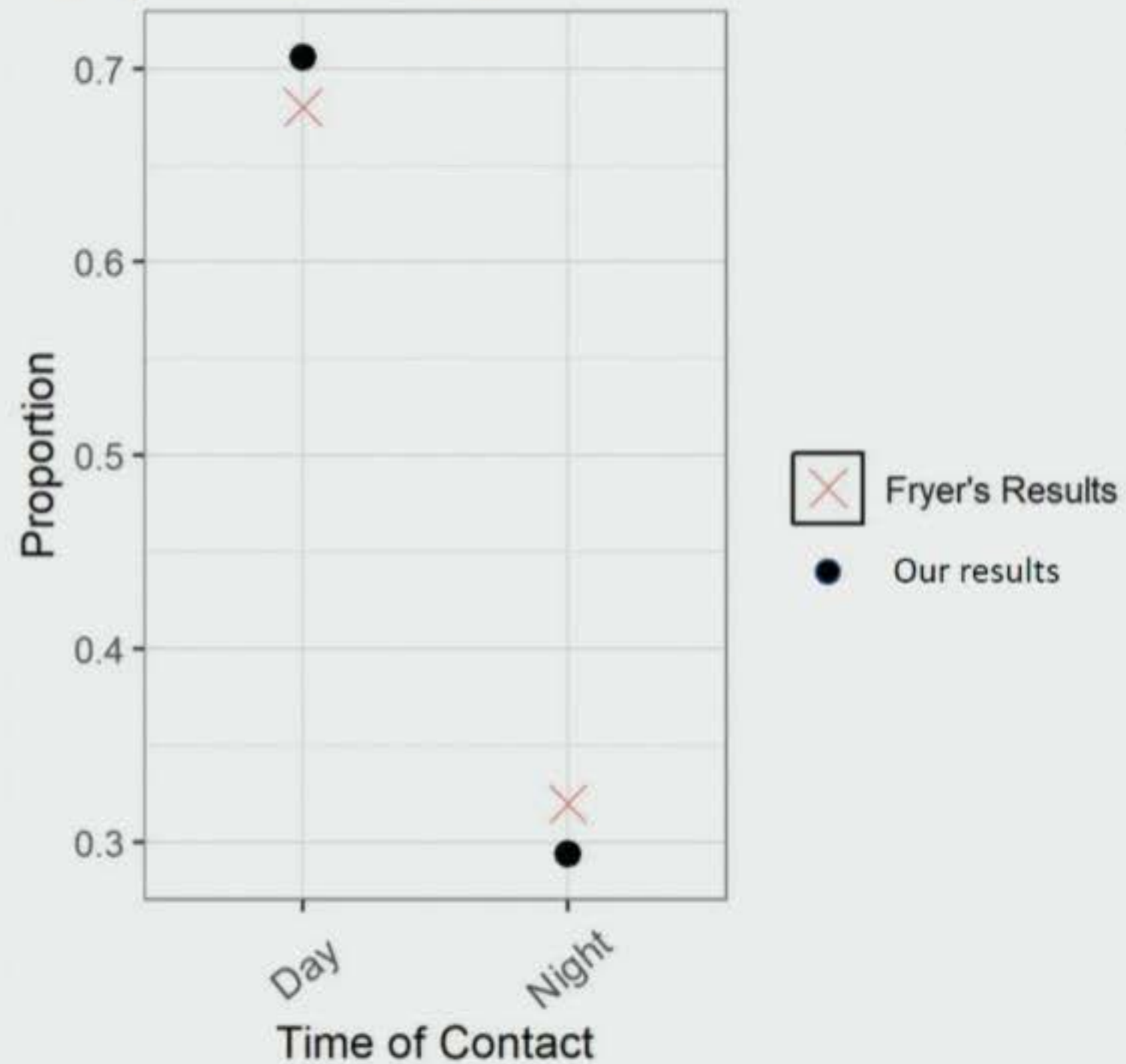
Proportions of Type of Incident



- Despite following Fryer's coding schemes we were not able to match his statistics for what isn't a street stop

Summary Statistics

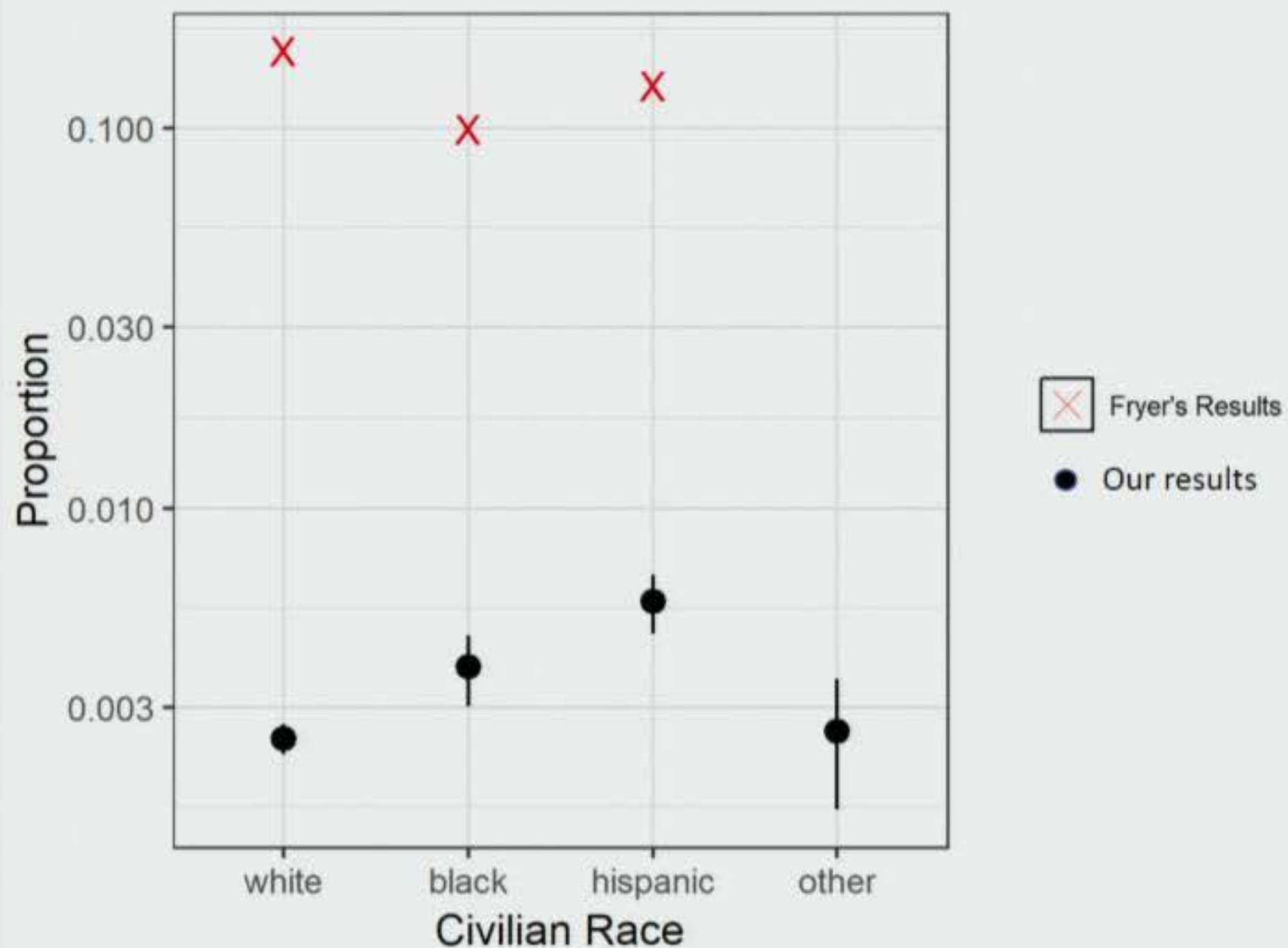
Proportions of Time of Contact



- Similar to the SQF data, there are discrepancies in what counts as a daytime stop

Summary Statistics

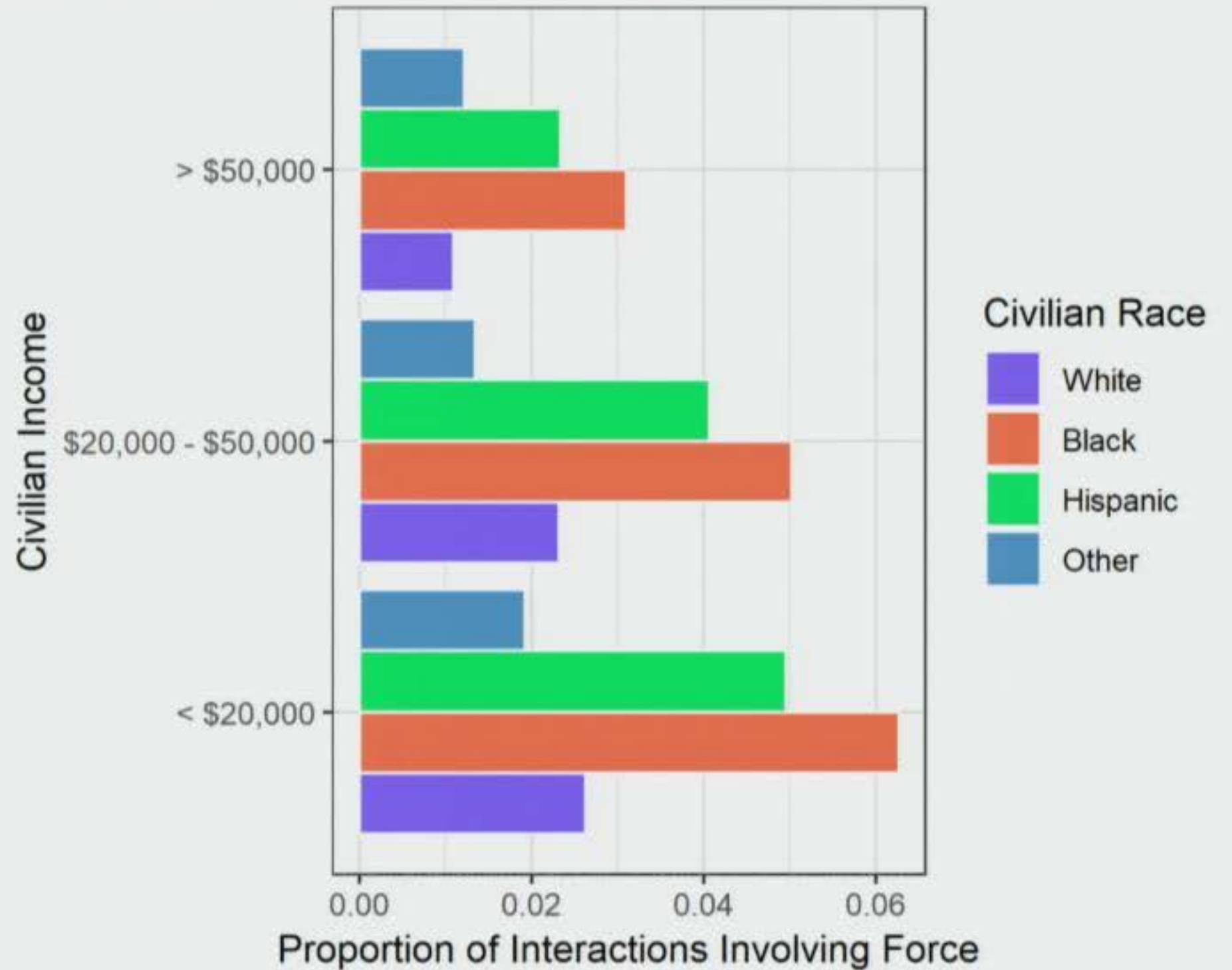
Proportions of Civilians Guilty of Carrying Illegal Items by Race



- Very large differences in proportion carrying illegal items by race

Force-Related Incidents Extension

- Within each income group, blacks are more likely to experience use of force than whites



Modeling Use of Force by Race

Main Outcome

Variable of
interest

Civilian reports ~ Civilian Race
use of force

- Yes
- no
- White
- Black
- Hispanic
- Other

Restricted to people who had a face – to –face contact with a police officer
61,499 out of 409,678

Modeling Use of Force by Race

Main Outcome

Variable of interest

Controls

Civilian reports use of force

Civilian Race

Civilian Demographics

Civilian Behavior

Contact and Officer Characteristics

Year

- Yes
- no

- White
- Black
- Hispanic
- Other

- Income
- Population size
- Age
- Employed in last 12 months
- Gender

- Disobeyed
- Tried to get away
- Resisted
- Complained
- Argued
- Threatened officer
- Used physical force

- Incident Type: Street Stop
- Incident Type: Traffic Stop
- Incident Type: Other
- Time of contact was day
- Time of contact was night
- Officers majority Hispanic
- Officers majority White
- Officers majority Black
- Officers majority other race
- Officers split race

- 1996
- 1999
- 2002
- 2005
- 2008
- 2011

Restricted to people who had a face – to –face contact with a police officer

61,499 out of 409,678

Logistic Regression Model

	Eryer Stats				
	White Mean	Black	Hispanic	Other Race	Sample Size
Panel B: PPCS Data					
(a) No Controls	0.007	3.50	2.70	1.13	59,668
(b) + Civilian Demographics		2.75	1.72	0.79	???
(c) + Encounter Characteristics		2.66	1.70	0.81	???
(d) + Civilian Behavior		2.78	1.82	0.76	???
(e) + Year		2.77	1.82	0.76	???

	Our Stats				
	White Mean	Black	Hispanic	Other Race	Sample Size
Panel B: PPCS Data					
(a) No Controls	0.02	2.80	2.06	0.97	61,499
(b) + Civilian Demographics		2.40	1.42	0.60	51,372
(c) + Encounter Characteristics		2.25	1.75	0.75	12,421
(d) + Civilian Behavior		2.23	1.76	0.76	12,421
(e) +Year		2.27	1.76	0.79	12,421

- Our numbers don't match well
- Same trend

Logistic Regression Model

	Fryer Stats				
	White Mean	Black	Hispanic	Other Race	Sample Size
Panel B: PPCS Data					
(a) No Controls	0.007	3.50	2.70	1.13	59,668
(b) + Civilian Demographics		2.75	1.72	0.79	???
(c) + Encounter Characteristics		2.66	1.70	0.81	???
(d) + Civilian Behavior		2.78	1.82	0.76	???
(e) + Year		2.77	1.82	0.76	???

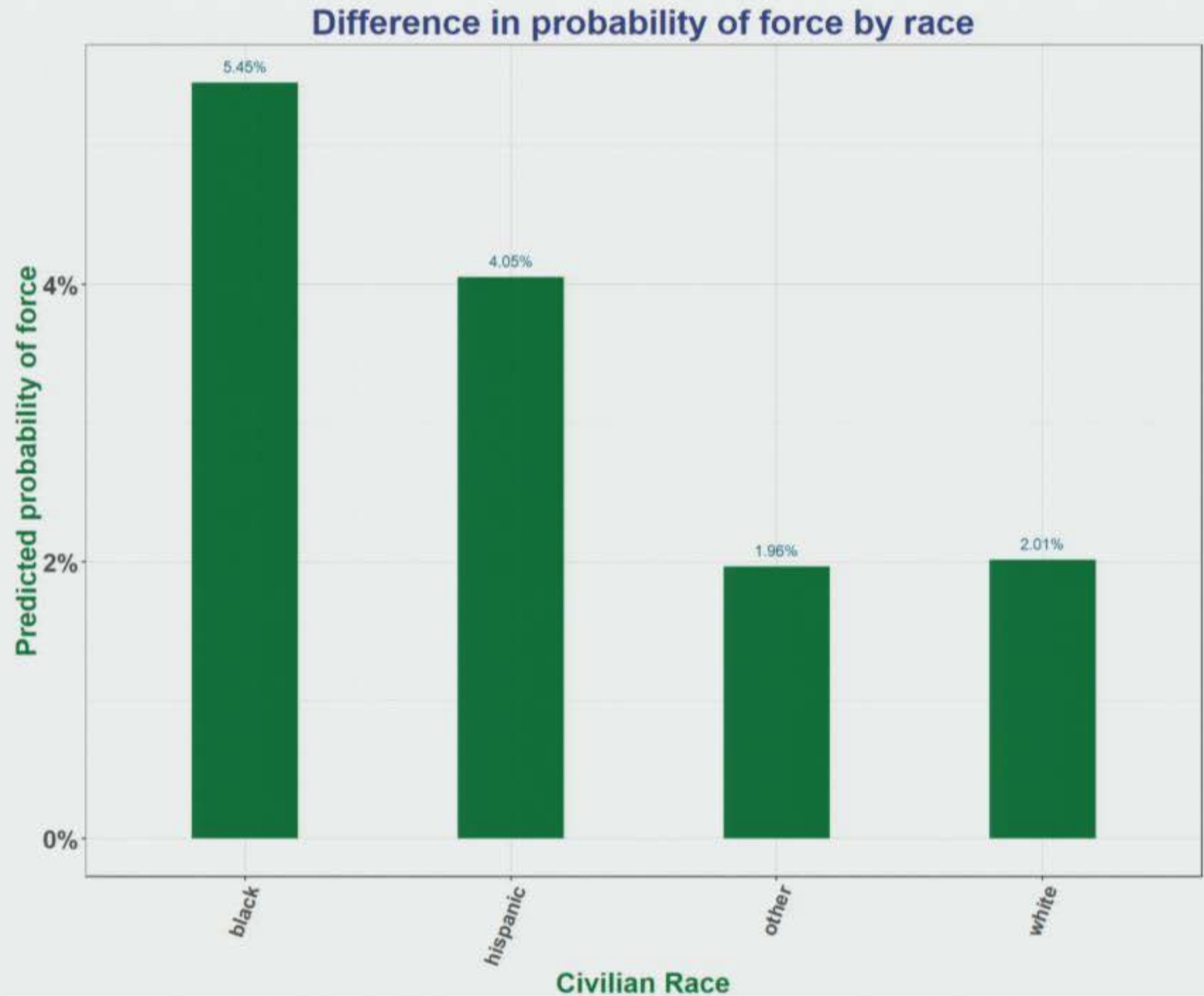
	Our Stats				
	White Mean	Black	Hispanic	Other Race	Sample Size
Panel B: PPCS Data					
(a) No Controls	0.02	2.80	2.06	0.97	61,499
(b) + Civilian Demographics		2.40	1.42	0.60	51,372
(c) + Encounter Characteristics		2.25	1.75	0.75	12,421
(d) + Civilian Behavior		2.23	1.76	0.76	12,421
(e) +Year		2.27	1.76	0.79	12,421

What do these numbers mean?

Odds ratios are hard to interpret!

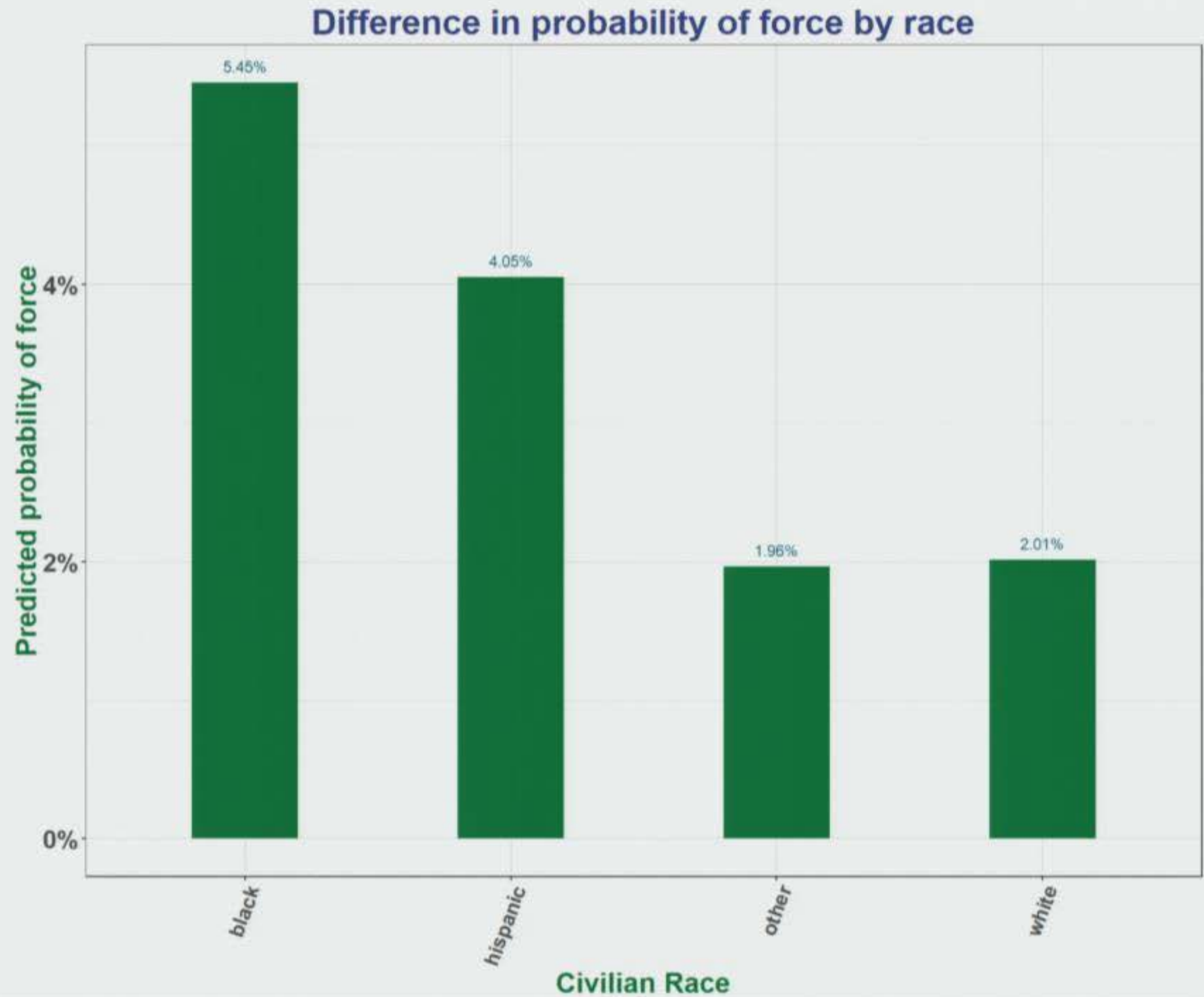
Model Replication

"Blacks are 3.5 times more likely to report use of force by police in an interaction"



Model Replication

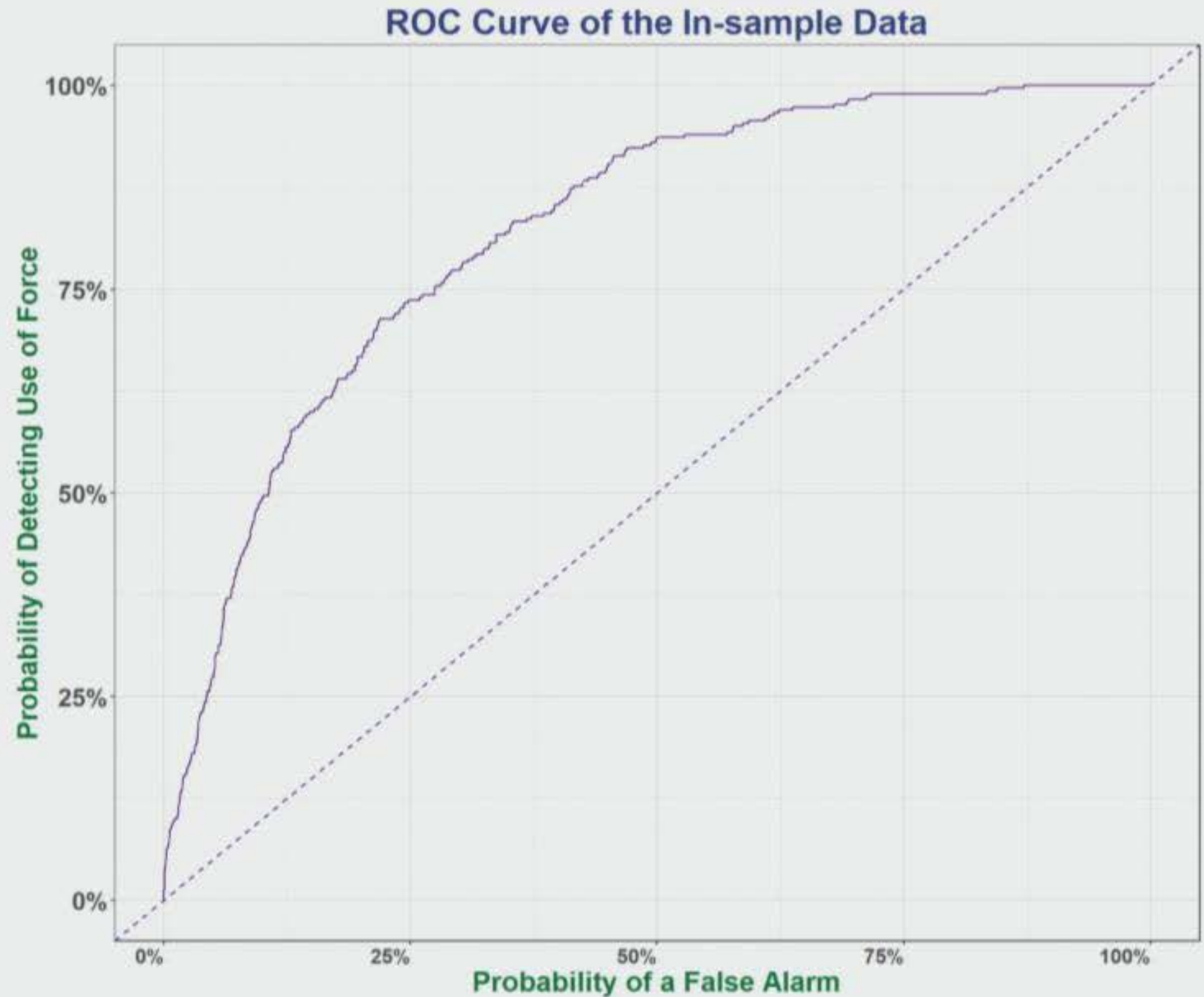
"Blacks are ~~3.X~~ ^{2.7} times more likely to report use of force by police in an interaction"



Model Evaluation

Area under the curve = 78%

Area under the curve without
race as a predictor = 74%



PPCS Takeaways

- Fryer's results did not replicate exactly
- But recover underlying pattern: Controls don't account for racial differences
- Caveats: Missing data eliminated many observations, model doesn't capture everything going on

Conclusion

Takeaways on Research Replication

- Replication is hard, even in a nearly ideal scenario
 - Public data, simple methods, extensive documentation
- Papers often leave out important details
 - Subjective coding, missing data
- Releasing data and code are key for reproducibility
 - Otherwise takes lots of effort to recreate
- Important to think critically about easily misinterpreted results
 - "Blacks are 53% more likely to experience use of force"?

Our Contribution

Clean data and code for anyone to use

The screenshot shows the GitHub interface for the repository 'msr-ds3 / stop-question-frisk'. At the top, there is a navigation bar with links for 'Why GitHub?', 'Enterprise', 'Explore', 'Marketplace', and 'Pricing', along with a search bar and 'Sign in'/'Sign up' buttons. Below the repository name, there are statistics: 'Watch 7', 'Star 0', and 'Fork 0'. A navigation menu includes 'Code', 'Issues 0', 'Pull requests 0', 'Projects 0', 'Security', and 'Insights'. A message states 'No description, website, or topics provided.' Below this, a summary bar shows '559 commits', '1 branch', '0 releases', and '8 contributors'. There are buttons for 'Branch master', 'New pull request', 'Find file', and 'Clone or download'. A file list is shown with columns for file name, description, and commit time. The files listed are: 'Naomoreira alt outcomes plot' (latest commit 870d8f6, 1 hour ago), 'clean_data' (directory structure with some notes, 19 days ago), 'figures' (Merge branch 'master' of https://github.com/msr-ds3/stop-question-frisk, 1 hour ago), 'notes' (renaming files, 6 days ago), 'raw_data' (alt outcomes plot, 1 hour ago), '.gitignore' (script updated to download all the data, 16 days ago), and 'README.md' (directory structure with some notes, 19 days ago). Below the file list, there is a section for 'README.md' with the title 'Stop, Question, Frisk' and a description: 'A replication and extension project from the Microsoft Research Data Science Summer School class of 2019'.

A best faith effort of 8 hard-working individuals

Thank You!

Questions?