Chapel Hill Police District:

Is a racial biased justice observable between 2010 and 2019?

Pr. Thibault Vatter, Data Science

Maeva Marin
Eugénie Mathieu

UNIL | Université de Lausanne
HEC Lausanne

Tables of contents

1 **Introduction**

Motivation

Data

Goals

2 **Exploratory analysis**

Arrestees

Victims

Employees

3 **Modeling**

The process

4 **Conclusion**

Results

Learning points

UNIL | Université de Lausanne
HEC Lausanne

The **cliché rift between black/white people** is said observable through a difference in judicial treatment between races. But what are the characteristics explaining this denounced racial bias in the American judicial system?

Data from Chapel Hill Open Data

	Observations	Variables
Arrestees	82'525	12
Victims	16'040	18
Police Employees	174	7

#Introduction - Data

Glimpse on the arrestees dataset

Date of Arrest	Arrestee's Age	Arrestee's Race	Arrestee's Gender	Type of Arrest	Drugs or Alcohol Present	Weapon Present
2010-01-01	18	White	Male	SUMMONED/CITED	Yes	UNARMED
2010-01-01	18	White	Male	TAKEN INTO CUSTODY	No	UNARMED
2010-01-01	27	White	Male	ON VIEW	Unknown	UNARMED
2010-01-05	55	White	Male	SUMMONED/CITED	Yes	UNARMED
2010-01-06	23	White	Male	ON VIEW	Yes	UNARMED

#Introduction - Data

Glimpse on the victims dataset

Date of Report	Date of Occurrence	Reported As	Victim Age	Victim Race	Victim Gender
2019-06-29	2019-06-29	THEFT/LARCENY	21	White	Female
2019-06-29	2019-06-29	ADMINISTRATIVE	73	White	Male
2019-06-29	2019-06-29	DISTURBANCE/NUI	29	Black	Female
2019-06-29	2019-06-29	DAMAGE/VANDALIS	25	White	Male
2019-06-28	2019-06-28	TRAFFIC/TRANSPO	78	White	Male

#Introduction - Data

Glimpse on the police employees dataset

Job Class Description	Age	Service Years	Gender	Race
POLICE OFFICER IV	50	10	Male	White
POLICE OFFICER IV	46	13	Male	White
POLICE OFFICER IV	41	15	Male	White
RELIEF PARKING LOT ATT	22	2	Female	Black
RELIEF PARKING LOT ATT	38	1	Female	Black
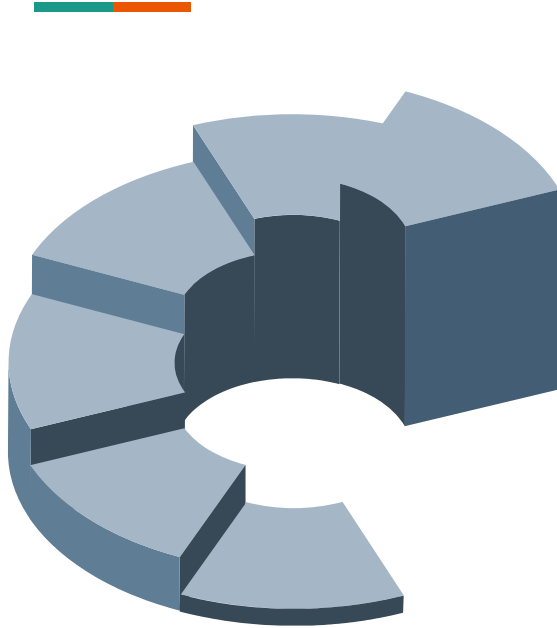
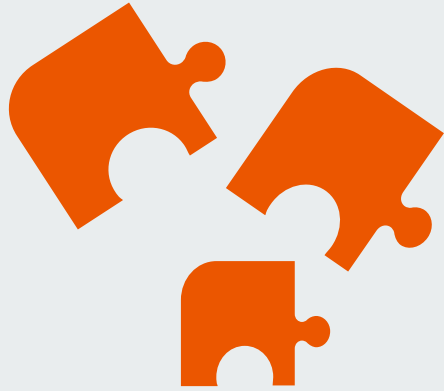Analyse the distribution of victims and arrestees profiles to check the representativeness of our samples

Sketch the typical profile of a victim and of an arrestee to check if race has an impact

Our samples are
not representative
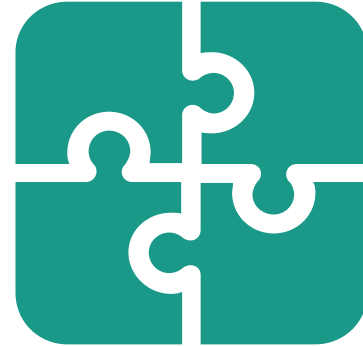but why...?

Samples population
70-30 of white / black victims
50-50 of white / black arrestees

Population of Chapel Hill
~ 90% of white citizens
~ 10% of black citizens
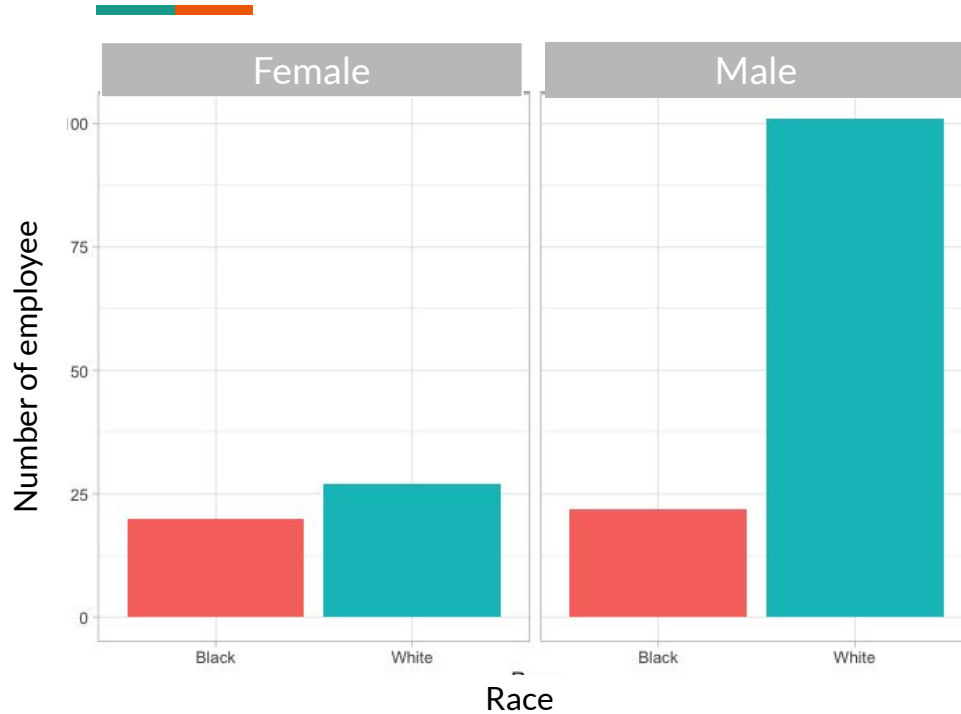
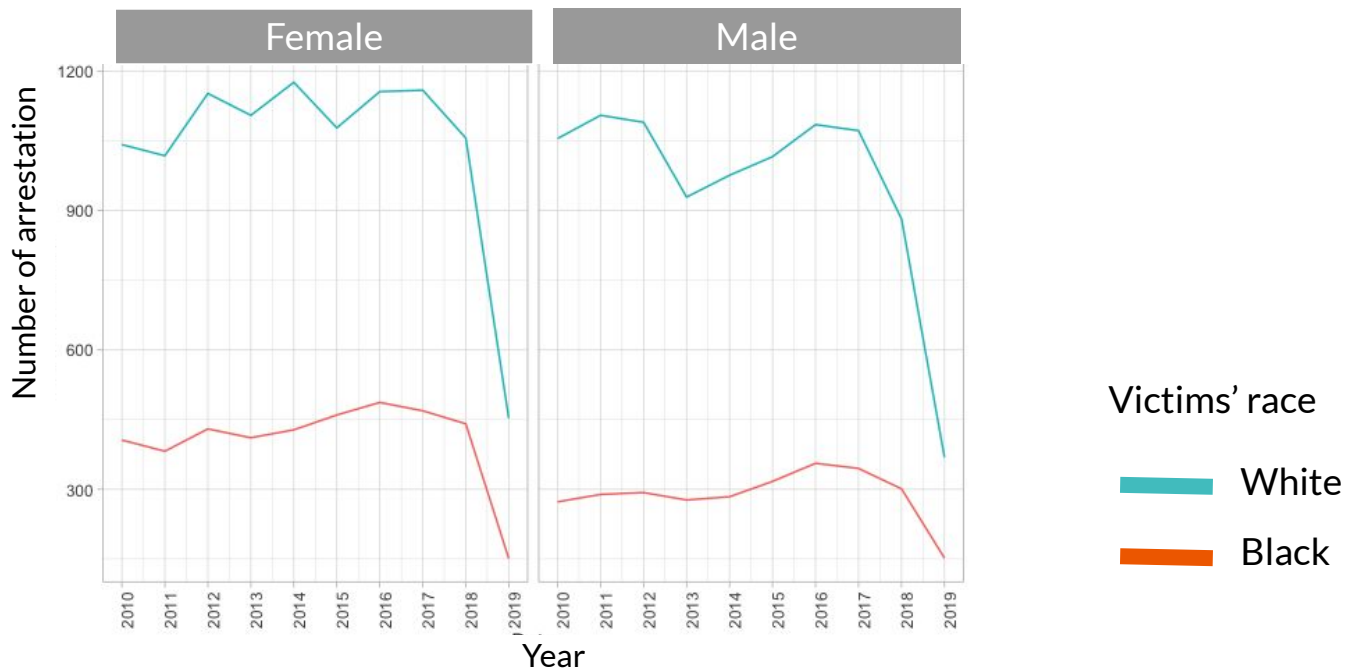Clear majority of males among positions in the Chapel Hill Police District
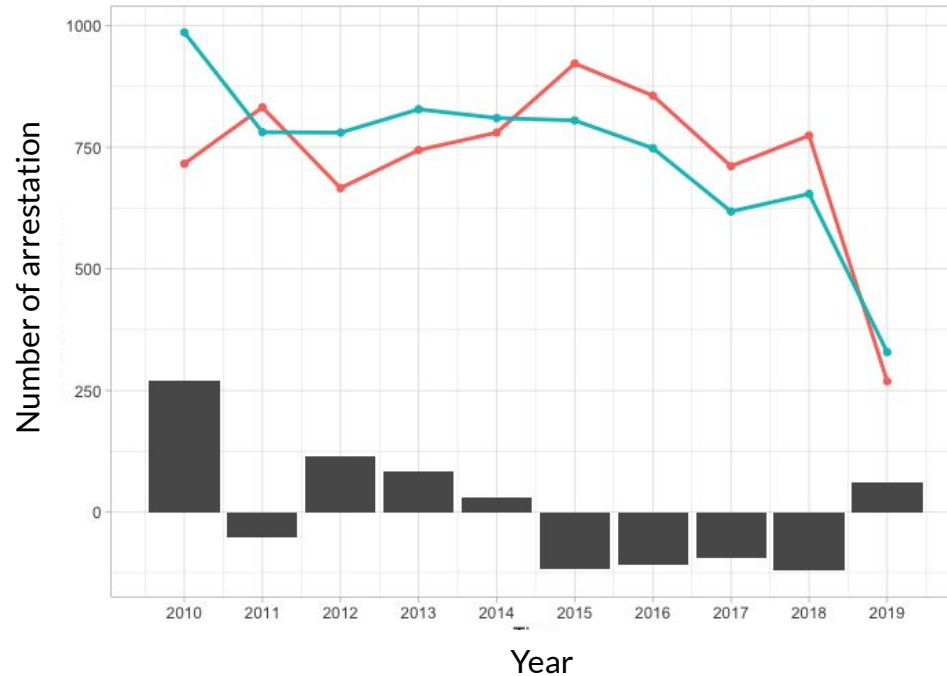
Employees' race

White

Black

White victims outnumber black ones.
This dataset better represents the real Chapel Hill's race distribution

11

#Exploration analysis - Arrestees

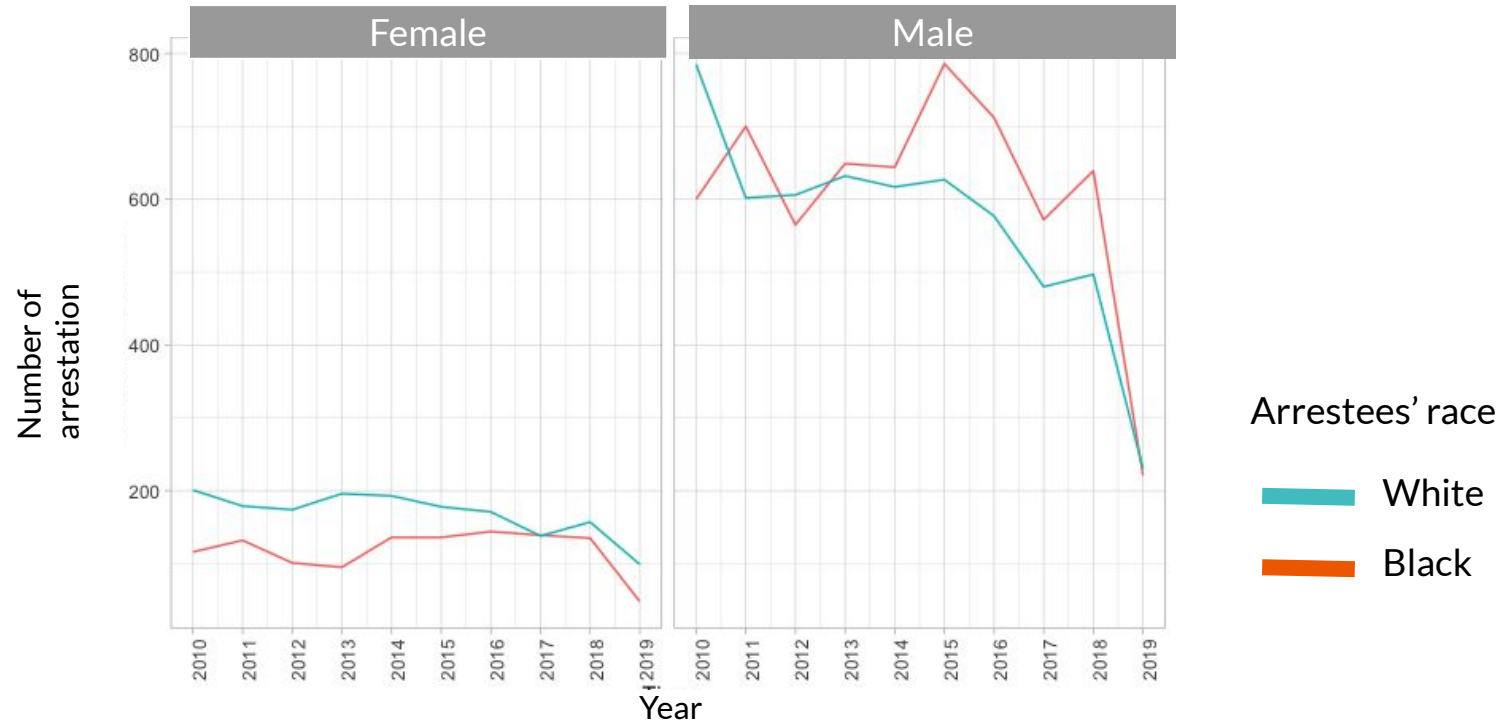Higher number of arrestation for black people since 2015

Arrestees' race

White

Black

Men outnumber women in the arrestees sample

#Exploration analysis - Arrestees

Black people are systematically more taken into custody than white people.
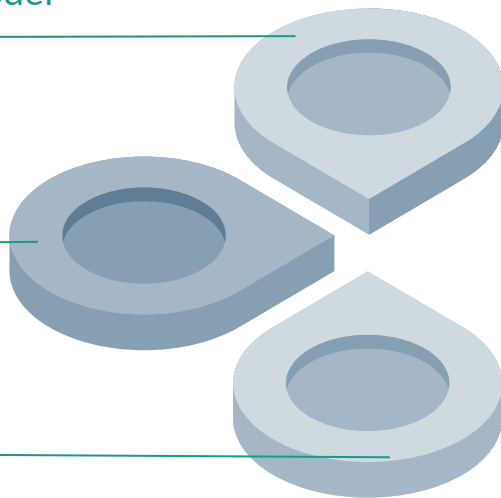
Arrestees' race

White

Black

To quantitatively assess our exploratory analysis...

Gaussian linear model

Poisson
linear model

Random Forest

Linear models to check the representativeness of our samples

Classification method to check the racial bias in arrestee profile

#Conclusion

Results

Non significance for race in linear models due to the biased dataset (black overrepresented)

A satisfying classification, even if fixing the race distribution bias would have steer better generalizing results

Learning points

Always contextualize datasets to confirm its representativeness

Manage the different time of data collection when working with many datasets

Let the figures do the talking, not extrapolate on unverified assumptions