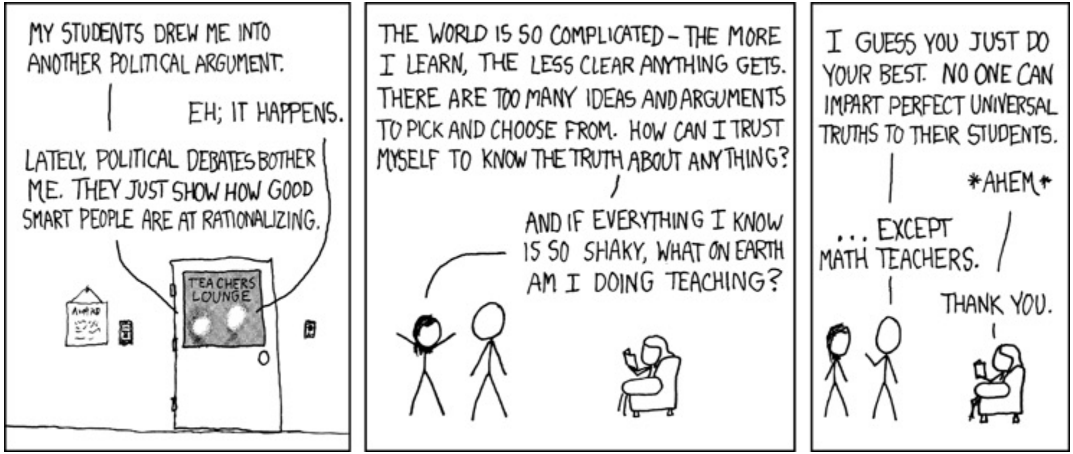


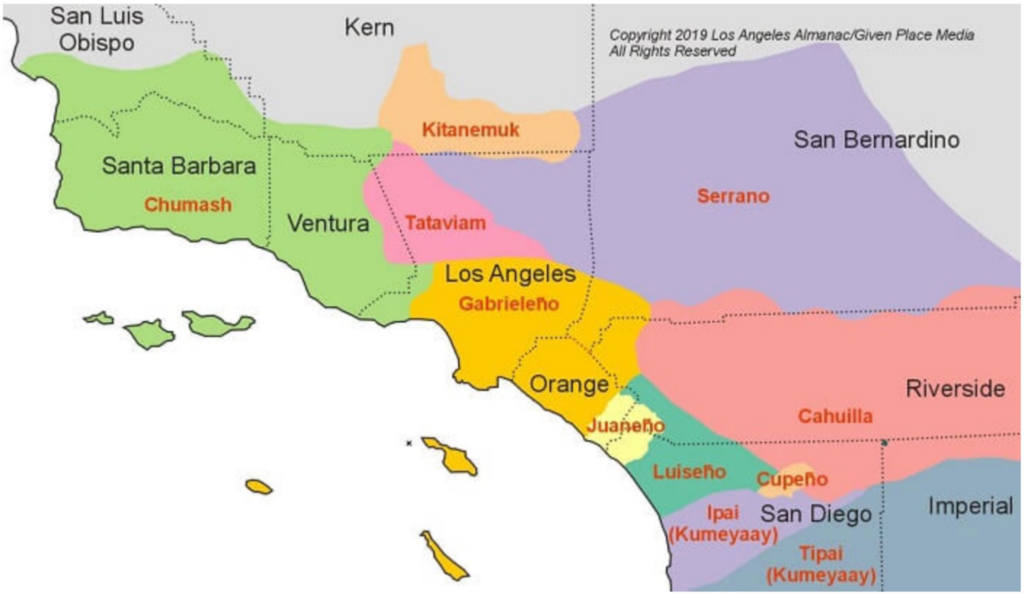
Image credit: xkcd.com “Certainty”

Jo Hardin
Pomona College
October 11, 2022

The Objectivity of My Desire… Data Science & Social Justice

This land acknowledgment recognizes systemic and institutional systems of power that have oppressed Indigenous peoples, with many of the same systems in place today that continue to marginalize those with less power.

Original People of Los Angeles County



Map of territories of Orignal Peoples with county boundaries in Southern California, Los Angeles Almanac, 2019.
Information sources: *Handbook of North American Indians, Vol. 8, California*, William C. Sturtevant (Gen. Editor) & Robert F. Heizer (Vol. Editor), 1978, Smithsonian Institute, and Dr. E. Gary Stickel, Ph.D. (UCLA), Tribal Archeologist, Kizh Nation / Gabrieleno Band of Mission Indians.

Image credit: http://www.laalmanac.com/history/hi05.php

#privilegealert

feel free to call me out publicly or privately

- white
- straight
- cis gendered
- able-bodied
- upper middle class
- educated
- American

Spoilers

1. Computer algorithms ~~don't~~ can't exist without humans making many choices along the way.

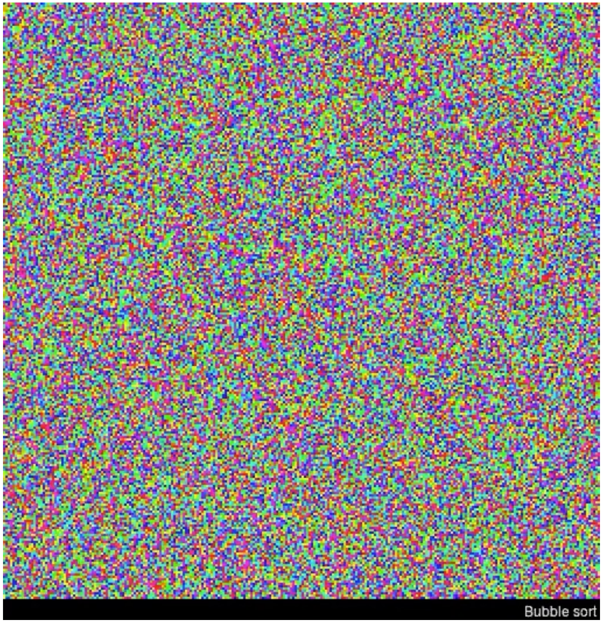
2. Algorithms based on data reproduce patterns in the data.

3. Many datasets represent differences in power dynamics across racial and gender lines.

Definition of Algorithm

a set of instructions for solving a task

Algorithms



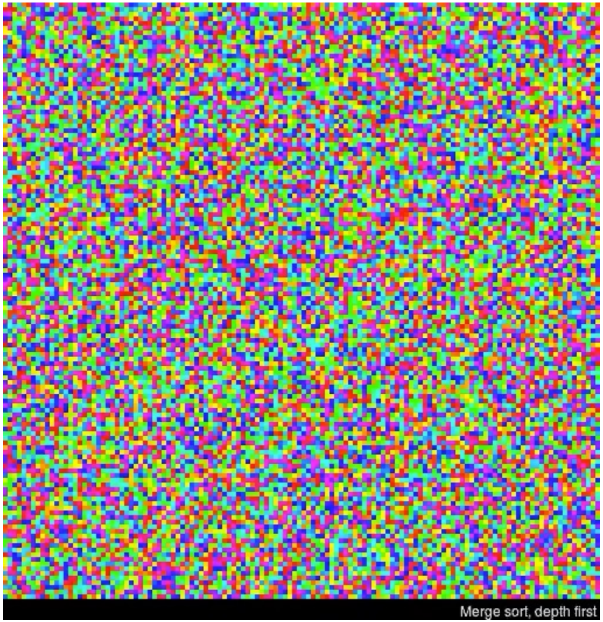


Image credit: https://imgur.com/gallery/voutF

h/t to Amelia McNamara at University of St. Thomas for many of the ideas in this talk

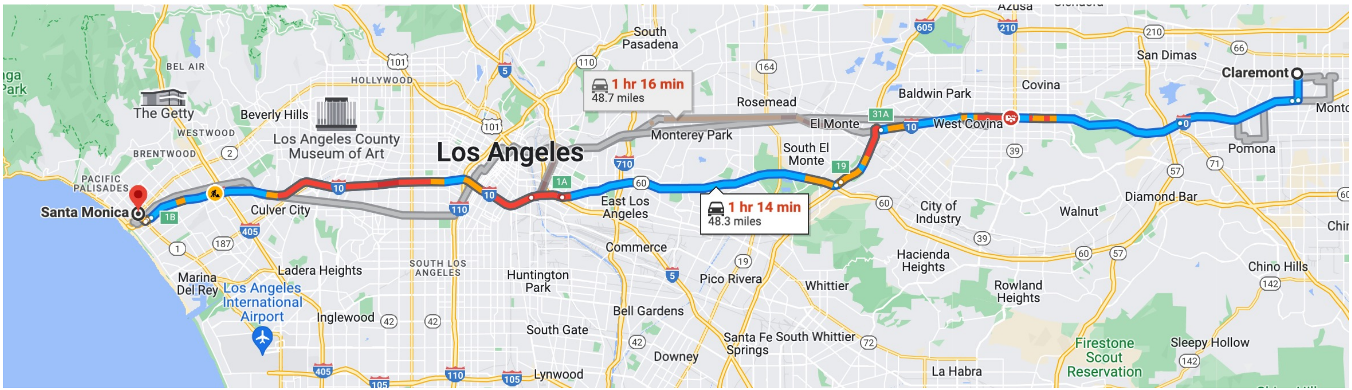
What are data?

DATA

Data exhaust

Image credit: https://www.pexels.com/@khunkorn-laowisit-1359149/

Directions – fastest route



Algorithmic Unfairness without any Bias Baked In

```
talent ~ Normal (100, 15)

grades ~ Normal (talent, 15)

SAT ~ Normal (talent, 15)
```

The example is taken directly (and mostly verbatim) from a blog by Aaron Roth.
http://aaronsadventures.blogspot.com/2019/01/discussion-of-unfairness-in-machine.html

Algorithmic Unfairness without any Bias Baked In

```
talent ~ Normal (100, 15)

grades ~ Normal (talent, 15)

SAT ~ Normal (talent, 15)
```

College wants to admit students with

`talent > 115`

... but the college only has access to **`grades`** and **`SAT`** which are noisy estimates of **`talent`**.

The example is taken directly (and mostly verbatim) from a blog by Aaron Roth.
http://aaronsadventures.blogspot.com/2019/01/discussion-of-unfairness-in-machine.html

Plan for accepting students

- Run a regression on training data (`talent` is known for existing students)
- Find a model which predicts `talent` based on `grades` and `SAT`
- Choose students for whom predicted `talent` is above 115

Flaw in the plan!

- there are two populations of students, the Reds and Blues.
  - Reds are the majority population (99%)
  - Blues are a small minority population (1%)
- the Reds and the Blues are no different when it comes to `talent`: they both have the same `talent` distribution.
- there is no bias baked into the grading or the exams: both the Reds and the Blues also have exactly the same `grade` and `SAT` score distributions

Flaw in the plan!

- there are two populations of students, the Reds and Blues.
    - Reds are the majority population (99%)
    - Blues are a small minority population (1%)
- the Reds and the Blues are no different when it comes to `talent`: they both have the same `talent` distribution.
- there is no bias baked into the grading or the exams: both the Reds and the Blues also have exactly the same `grade` and `SAT` score distributions

But there is one difference: the Blues have more money than the Reds, so they each take the **`SAT`** **twice**, and report only the highest of the two scores to the college.

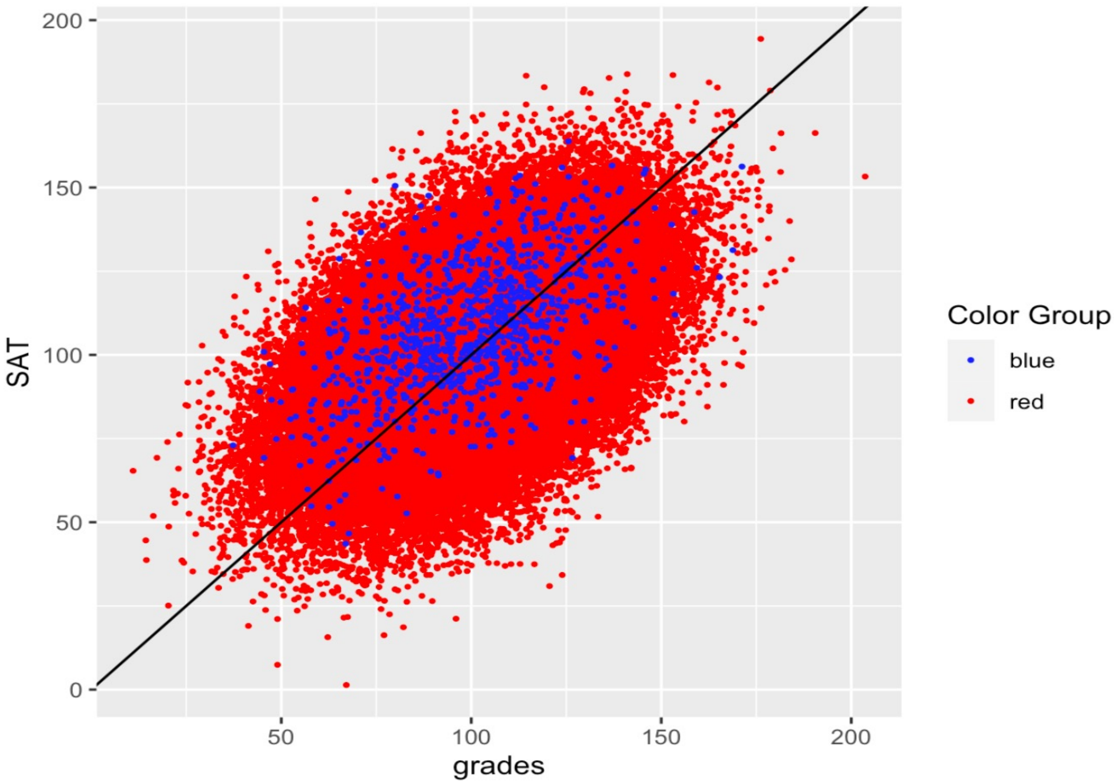
Taking the test twice results in a small but noticeable bump in the average **`SAT`** scores of the Blues, compared to the Reds.

Key Insight:

- The value of `SAT` means something different for the Reds versus the Blues

(They have different feature distributions.)

Let’s see what happens



Two Models:

Red model (SAT taken once):

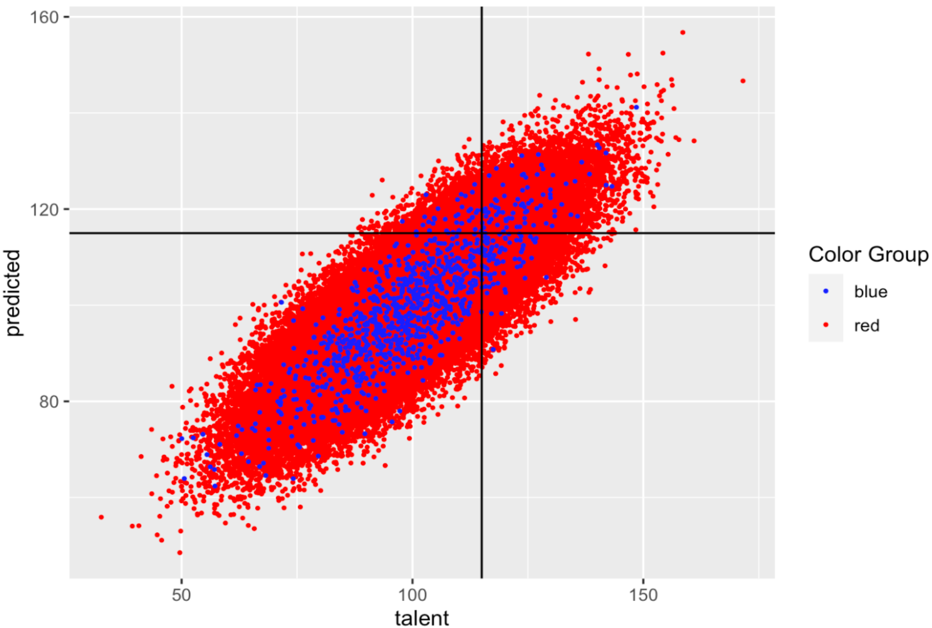
```
## # A tibble: 3 × 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)   33.4     0.152        220.       0
## 2 SAT            0.332   0.00149      223.       0
## 3 grades         0.333   0.00150      223.       0
```

Blue model (SAT is max score of two):

```
## # A tibble: 3 × 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   24.2      1.52        15.9 8.47e- 51
## 2 SAT            0.430    0.0154      27.9 6.53e-127
## 3 grades         0.291    0.0142      20.5 3.15e- 78
```

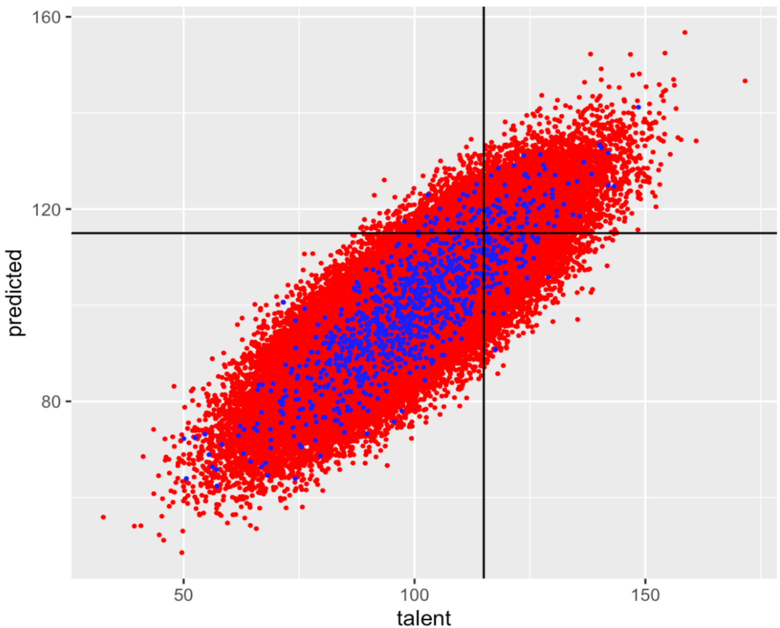
New Data

- Generate new data and use the two models, separately.
- Visually, the prediction of `talent` > 115 seems similar across the two groups.



New Data

- Generate new data and use the two models above.
- Visually, the prediction of `talent` > 115 seems similar across the two groups.



color	tpr	fpr	error
blue	0.510	0.044	0.113
red	0.504	0.037	0.109

$$\texttt{tpr} = \frac{\text{true positives}}{\text{all who should}}$$

$$\texttt{fpr} = \frac{\text{false positives}}{\text{all who should not}}$$

TWO models doesn’t seem right????

- What if we fit only one model to the entire dataset?
- Afterall, there are laws against using protected classes to make decisions (housing, jobs, credit, loans, college, etc.)

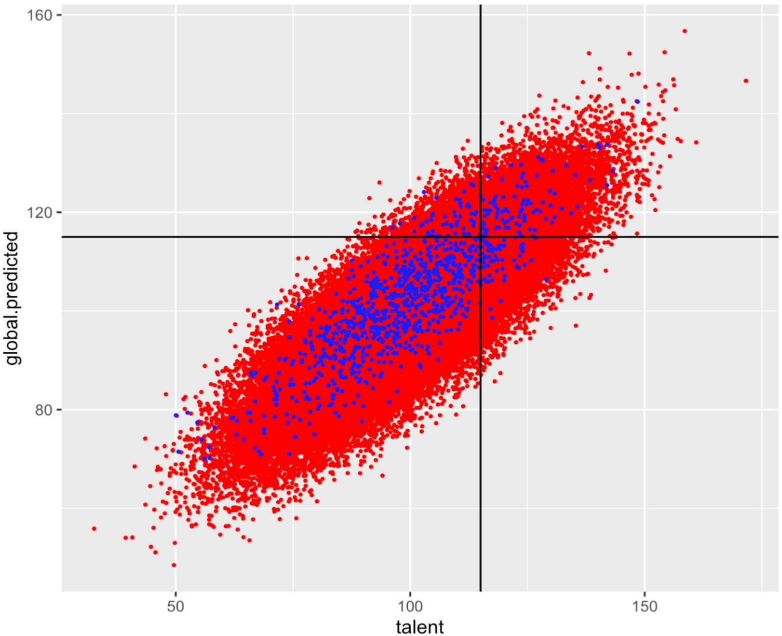
TWO models doesn't seem right????

- What if we fit only one model to the entire dataset?
- Afterall, there are laws against using protected classes to make decisions (housing, jobs, credit, loans, college, etc.)

```
## # A tibble: 3 × 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)   33.4     0.151        221.       0
## 2 SAT            0.332   0.00148      224.       0
## 3 grades         0.334   0.00149      224.       0
```

(The coefficients kind of look like the red model.)

How do the error rates change?



One model:

color	tpr	fpr	error
blue	0.613	0.063	0.113
red	0.502	0.037	0.109

Two separate models:

color	tpr	fpr	error
blue	0.510	0.044	0.113
red	0.504	0.037	0.109

What did we learn?

- with two populations that have different feature distributions, learning a single classifier (that is prohibited from discriminating based on population) will fit the bigger of the two populations

- depending on the nature of the distribution difference, it can be either to the benefit or the detriment of the minority population

- no explicit human bias, either on the part of the algorithm designer or the data gathering process

- the problem is exacerbated if we artificially force the algorithm to be group blind

- well intentioned "fairness" regulations prohibiting decision makers form taking sensitive attributes into account can actually make things less fair and less accurate at the same time

Facial Recognition Software in the Recent News

THE WALL STREET JOURNAL.

MARKETS | JOURNAL REPORTS: WEALTH MANAGEMENT

What Your Face May Tell Lenders About Whether You're Creditworthy

How one of the world's largest insurers uses facial-recognition technology on potential customers to assess risk

Facial-recognition technology has become part of everyday life in China. **PHOTO:** CHINA DAILY/REUTERS

June 10, 2019 10:05 p.m. ET

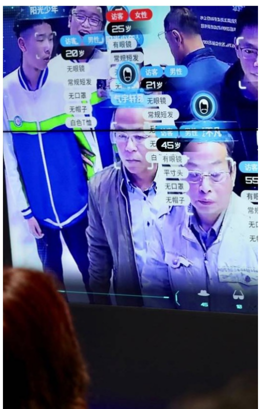
Facial Recognition Software in the Recent News

THE WAI

MARKETS | JOURNAL REPORTS: WEALTH MAN

What Your Face About Whether '

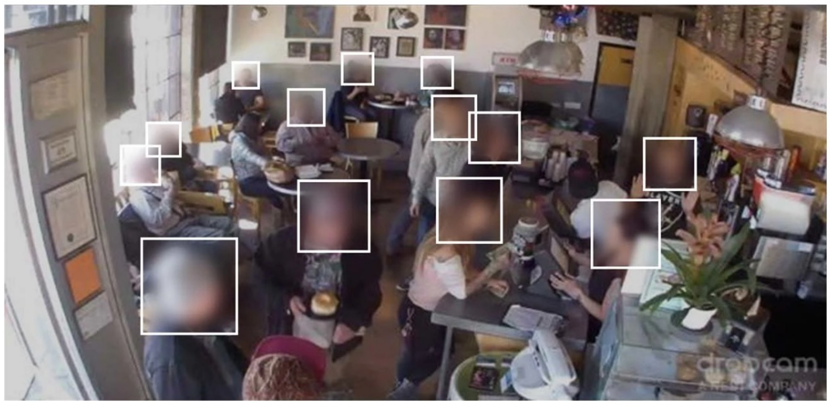
How one of the world's largest on potential customers to asse



Facial-recognition technology has become DAILY/REUTERS

June 10, 2019 10:05 p.m. ET

Facial Recognition Tech Is Growing Stronger, Thanks to Your Face



The Brainwash database, created by Stanford University researchers, contained more than 10,000 images and nearly 82,000 annotated heads.

Open Data Commons Public Domain Dedication and License, via Megapixels

By Cade Metz

July 13, 2019

251

Facial Recognition Software in the Recent News

MARKETS | JOURI

What '
About

How one of t
on potential

Facial-recognitio
DAILY/REUTERS

June 10, 2019 10:(

Police trials of facial recognition backed by home secretary

Sajid Javid supports use of technology despite concern from human rights groups

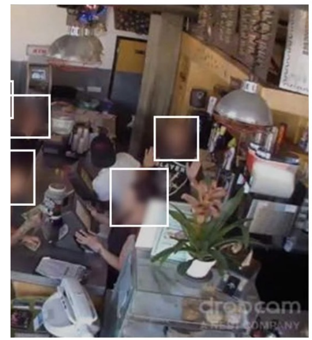
▲ Critics of facial recognition technology have described it as a 'dangerously intrusive and discriminatory'. Photograph: David McNew/AFP/Getty

Jamie Grierson *Home affairs correspondent*

@JamieGrierson

Fri 12 Jul 2019 06.08 EDT

:h Is Growing
›ur Face



ersity researchers, contained more
ads.

ia Megapixels

251

Facial Recognition Software in the Recent News

MARKETS | JOURNA

What Y
About V

How one of th
on potential cu

Facial-recognition t
DAILY/REUTERS

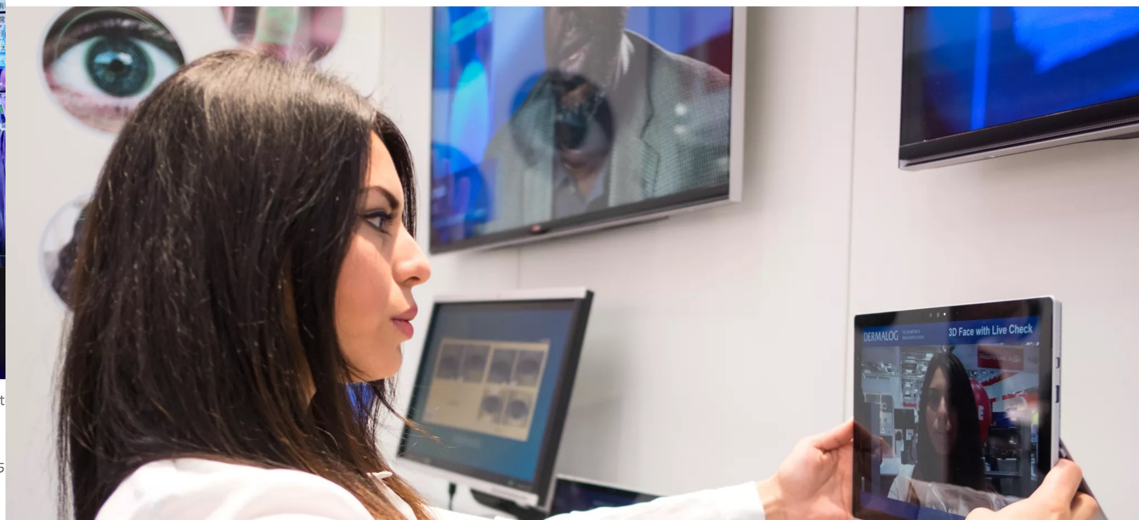
June 10, 2019 10:05

The growing backlash against facial recognition tech

Apple, Amazon, and Microsoft are all mired in controversy over it.

By Sigal Samuel | Apr 27, 2019, 8:00am EDT

SHARE



The New York Times

San Francisco Bans Facial Recognition Technology

Attendees interacting with a facial recognition demonstration at this year's CES in Las Vegas. Joe Buglewicz for The New York Times

By Kate Conger, Richard Fausset and Serge F. Kovaleski

May 14, 2019

367

Facial Re

MARKETS | JOURNAL REPORTS: W

What Your I About Whet

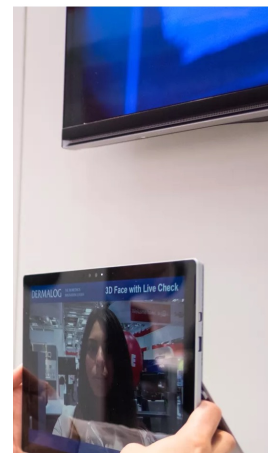
How one of the world's on potential customer

Facial-recognition technology I DAILY/REUTERS

June 10, 2019 10:05 p.m. ET

ed

cial

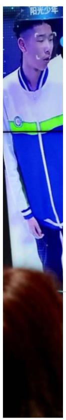


Facial Recognition Software in the Recent News

MARKETS

Wh
Abo

How or
on pote



Facial-rec
DAILY/RE

June 10, 2

The New York Times

The Racist History Behind Facial Recognition

When will we finally learn we cannot predict people's character from their appearance?

Photo illustration by The New York Times; Archive.org (photo plate from "Identification anthropométrique," published 1893); Amazon Rekognition (facial landmarks and boundaries)

By **Sahil Chinoy**

Mr. Chinoy is a graphics editor for The New York Times Opinion section.

July 10, 2019

Facial Recognition Software ir

MARKETS

Wh:
Abo

How or
on pote

Facial-rec
DAILY/RE

June 10, 2

The Racist History Behind Facial Recognition

When will we finally learn we cannot predict people
character from their appearance?

By Sahil Chinoy

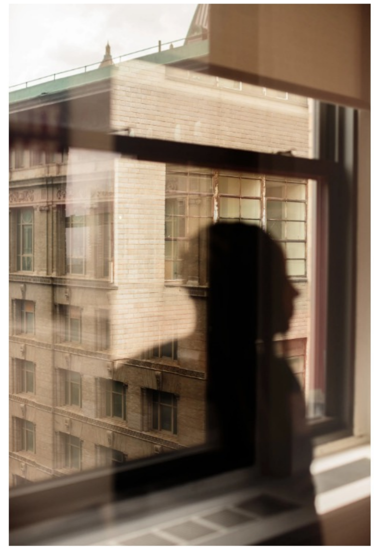
Mr. Chinoy is a graphics editor for The New Y

July 10, 2019

The New York Times

She Was Arrested at 14. Then Her Photo Went to a Facial Recognition Database.

With little oversight, the N.Y.P.D. has been using powerful surveillance technology on photos of children and teenagers.



"It's very disturbing to know that no matter what I'm doing at that moment, someone might be scanning my picture to try to find someone who committed a crime," said Bailey, who pleaded guilty to an assault that occurred when she was 14.
Sarah Blesener for The New York Times

By Joseph Goldstein and Ali Watkins

Aug. 1, 2019

221

late from "Identification
n (facial landmarks and
boundaries)

The New York Times

Facebook, Citing Societal Concerns, Plans to Shut Down Facial Recognition System

Saying it wants “to find the right balance” with the technology, the social network will delete the face scan data of more than one billion users.

Give this article 277

Facebook is shuttering a feature, introduced in December 2010, that automatically identified people who appeared in users’ digital photo albums. Carlos Barria/Reuters

By Kashmir Hill and Ryan Mac

Published Nov. 2, 2021 Updated Nov. 5, 2021

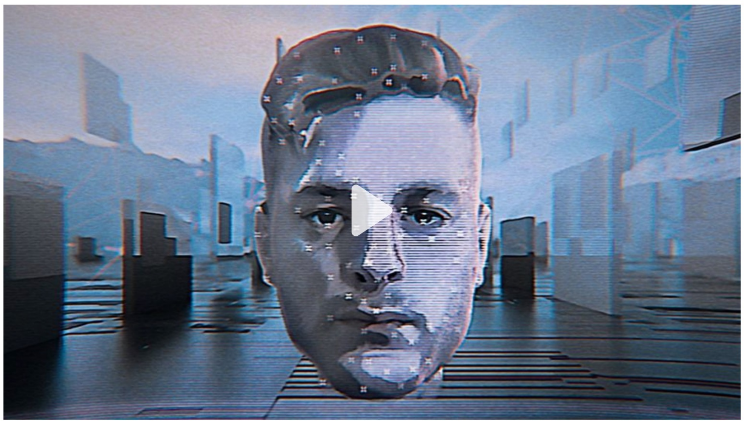
Facial Recognition Software ir

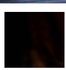
CNN BUSINESS Audio Live TV Log In

Facebook is shutting down its facial recognition software

By Rachel Metz, CNN Business

Updated 11:55 AM EST, Mon November 8, 2021





Facial-rec
DAILY/RE

June 10, 2

By Sahil Chinoy

Mr. Chinoy is a graphics editor for The New Y

July 10, 2019

She
Wer

With li
surveil

picture to try to find someone who committed a crime,” said Bailey, who pleaded guilty to an assault that occurred when she was 14.
Sarah Blesener for The New York Times

boundaries)

By Joseph Goldstein and Ali Watkins

Aug. 1, 2019

221

Facial Recognition Software ir

The New York Times

Facebook, Citing Societal Concerns, Plans to Shut Down Facial Recognition System

Saying it wants "to find the right balance" with the technology, the social network will delete the face scan data of more than one billion users.

Give this article 277

She
Wen
With li
surveil

The Racist History

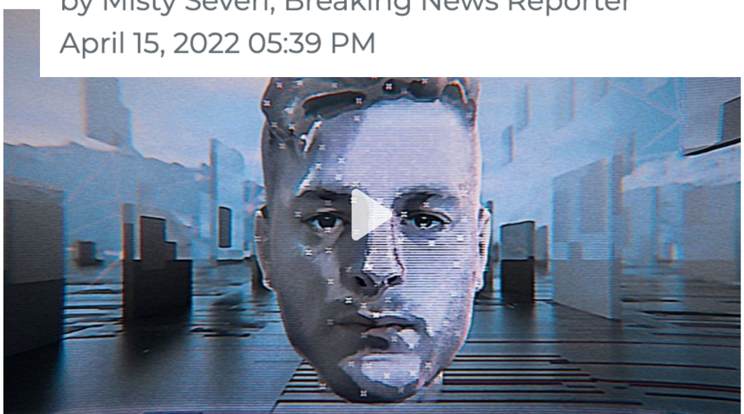
Washington Examiner

NEWS

Ukraine uses facial recognition software to identify dead Russian soldiers

by Misty Severi, Breaking News Reporter
April 15, 2022 05:39 PM

Fac
By Ra
Updat



dentified

people who appeared in users' digital photo albums. Carlos Barria/Reuters

By Kashmir Hill and Ryan Mac
Published Nov. 2, 2021 Updated Nov. 5, 2021

picture to try to find someone who committed a crime," said Bailey, who pleaded guilty to an assault that occurred when she was 14.
Sarah Blesener for The New York Times

boundaries)

By Joseph Goldstein and Ali Watkins

2019

221

Amazon’s algorithm for facial recognition

- Human decisions at different steps
  - What model is used? (e.g., CNN?)
  - What is being optimized (e.g., accuracy? false positives? false negatives?)
  - What data are used to train the model? Where are cameras placed? What data is collected?
  - How is the model tuned (e.g., # layers)? A human decides the cutoff for a “match”.
  - Who buys / uses Amazon’s algorithm?

Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.

Amazon promotes its facial recognition technology on the company's website, saying that the service can track people in a video even when their faces are not visible. Amazon

By **Nick Wingfield**

May 22, 2018

Photo credit: New York Times

NPR
@NPR

Following

Privacy advocates say early forms of facial recognition have been adopted by U.S. police without serious oversight — and the same may happen with newer real-time recognition.

Real-Time Facial Recognition Is Available, But Will U.S. Police Buy It?
Instant facial recognition is ramping up in China and other places, but will U.S. law enforcement follow suit?
npr.org

7:43 PM - 10 May 2018

Photo credit: npr.org

Parliament Benchmark Dataset

Buolamwini & Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, Proceedings of Machine Learning Research 81:1–15, 2018.

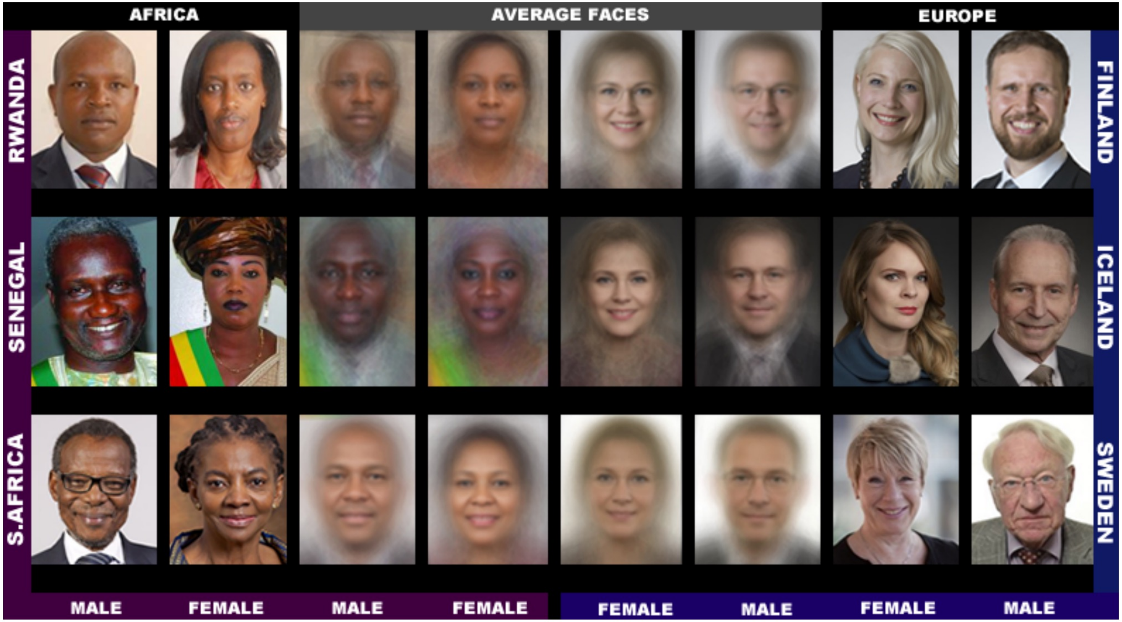


Figure 1: Example images and average faces from the new Pilot Parliaments Benchmark (PPB). As the examples show, the images are constrained with relatively little variation in pose. The subjects are composed of male and female parliamentarians from 6 countries. On average, Senegalese subjects are the darkest skinned while those from Finland and Iceland are the lightest skinned.

$$PPV = \frac{TP}{TP + FP}$$

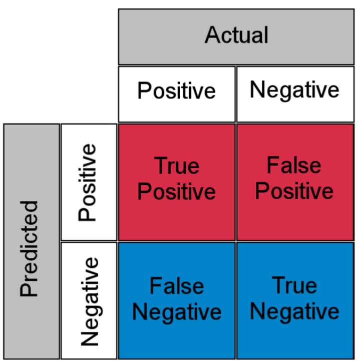
$$Error = \frac{FP}{TP + FP}$$

$$FPR = \frac{FP}{FP + TN}$$

$$TPR = \frac{TP}{TP + FN}$$

Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	PPV(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	TPR (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	FPR (%)	6.3	8.3	3.5	12.9	0.7	**16.3**	7.9	1.3	0.0
Face++	PPV(%)	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	94.0	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	6.0	0.8
	TPR (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	FPR (%)	10.0	14.9	1.1	16.5	4.7	**23.4**	1.2	7.1	1.1
IBM	PPV(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	**99.7**
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	TPR (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	FPR (%)	12.1	14.8	7.9	22.4	3.2	**25.2**	17.7	5.20	0.4

Table 4: Gender classification performance as measured by the positive predictive value (PPV), error rate (1-PPV), true positive rate (TPR), and false positive rate (FPR) of the 3 evaluated commercial classifiers on the PPB dataset. All classifiers have the highest error rates for darker-skinned females (ranging from 20.8% for Microsoft to 34.7% for IBM).



https://commons.wikimedia.org/wiki/File:ConfusionMatrixRedBlue.png

Gender classification

Table credit: Buolamwini & Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, Proceedings of Machine Learning Research 81:1–15, 2018.

$$PPV = \frac{TP}{TP + FP}$$

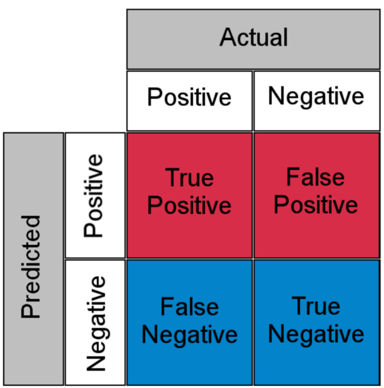
$$Error = \frac{FP}{TP + FP}$$

$$FPR = \frac{FP}{FP + TN}$$

$$TPR = \frac{TP}{TP + FN}$$

Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	PPV(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	TPR (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	FPR (%)	6.3	8.3	3.5	12.9	0.7	**16.3**	7.9	1.3	0.0
Face++	PPV(%)	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	94.0	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	6.0	0.8
	TPR (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	FPR (%)	10.0	14.9	1.1	16.5	4.7	**23.4**	1.2	7.1	1.1
IBM	PPV(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	**99.7**
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	TPR (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	FPR (%)	12.1	14.8	7.9	22.4	3.2	**25.2**	17.7	5.20	0.4

Table 4: Gender classification performance as measured by the positive predictive value (PPV), error rate (1-PPV), true positive rate (TPR), and false positive rate (FPR) of the 3 evaluated commercial classifiers on the PPB dataset. All classifiers have the highest error rates for darker-skinned females (ranging from 20.8% for Microsoft to 34.7% for IBM).



https://commons.wikimedia.org/wiki/File:ConfusionMatrixRedBlue.png

Gender classification

Table credit: Buolamwini & Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, Proceedings of Machine Learning Research 81:1–15, 2018.

Why classify? one-to-many

source: https://www.wired.com/2014/11/algorithms-great-can-also-ruin-lives/

ON APRIL 5, 2011, 41-year-old John Gass received a letter from the Massachusetts Registry of Motor Vehicles informing Gass that his driver’s license had been revoked ... as a conscientious driver who had not received so much as a traffic violation in years, Gass had no idea why it had been sent.

… his image had been automatically flagged by a facial-recognition algorithm designed to scan through a database of millions of state driver’s licenses looking for potential criminal false identities. The algorithm had determined that Gass looked sufficiently like another Massachusetts driver that foul play was likely involved…

The RMV itself was unsympathetic, claiming that it was the accused individual’s “burden” to clear his or her name in the event of any mistakes, and arguing that the pros of protecting the public far outweighed the inconvenience to the wrongly targeted few.

Face Recognition Vendor Test

FP = incorrect match

National Institute of Standards and Technology (2019)
https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

FP results:

- highest in West and East African and East Asian people
- lowest in Eastern European individuals.
- with a number of algorithms developed in China the effect is reversed, with low false positive rates on East Asian faces.
- with domestic law enforcement images, the highest false positives are in American Indians, with elevated rates in African American and Asian populations

- higher in women than men (effect is smaller than that due to race)
- elevated in the elderly and in children

Face Recognition
Vendor Test

National Institute of Standards and Technology (2019)
https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

FN = no match

FN results:

- with domestic mugshots, false negatives are higher in Asian and American Indian individuals, with error rates above those in white and African American faces
- with lower-quality border crossing images, false negatives are generally higher in people born in Africa and the Caribbean, the effect being stronger in older individuals.

Loss functions

most algorithms are trained by optimizing a loss function.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$$

⇒ goal is to maximize the classification of the training data.

Better: use a loss function that maximizes the local maximization. In particular, control with respect to imbalanced data (e.g., “the blues”).

Stephane Gentric, https://nigos.nist.gov/ifpc2018/presentations/44_gentric_Idemia_IFPC.pdf

Thwart the algorithm

- Fawkes "poisons" models that try to learn what you look like, by putting hidden changes into your photos, and using them as Trojan horses to deliver that poison to any facial recognition models of you. Fawkes takes your personal images and makes tiny, pixel-level changes that are invisible to the human eye, in a process we call image cloaking. You can then use these "cloaked" photos as you normally would, sharing them on social media, sending them to friends, printing them or displaying them on digital devices, the same way you would any other photo.

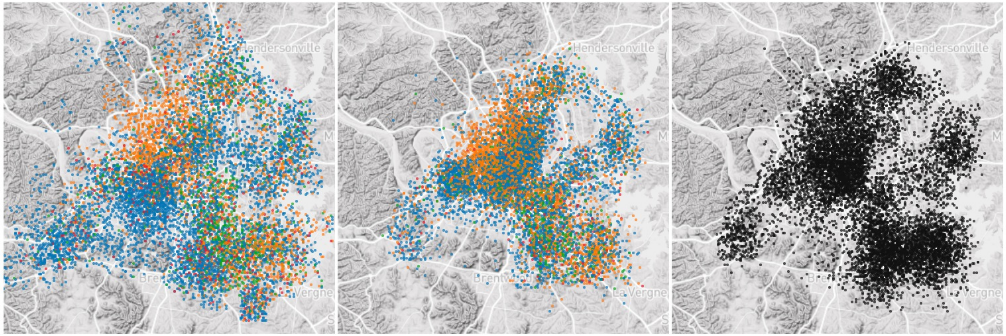
https://sandlab.cs.uchicago.edu/fawkes/

https://policylab.stanford.edu/projects/nashville-traffic-stops.html

Data Science to help marginalized communities

STANFORD COMPUTATIONAL POLICY LAB

Nashville Traffic Stops



The distribution of Nashville's residential driving-age population (left) and locations of non-moving violation stops (middle), colored by race (white, black, Hispanic, and other). Black drivers in Nashville are stopped considerably more often than white drivers, particularly for non-moving violations. These disparities stem, in part, from a strategy that concentrates traffic stops in high-crime areas (right). We find, however, that traffic stops have little impact on curbing serious crime, suggesting the city could safely curtail stops while simultaneously reducing racial disparities.

November 19, 2018

Alex Chohlas-Wood, Sharad Goel, Amy Shoemaker, and Ravi Shroff

Why?

- black drivers are stopped more often than white drivers (even within most zones)
- differences in violation rates (if black drivers are disproportionately more likely to have broken taillights),
- differences in enforcement (implicit bias),
- heterogeneity in population / crime in zone
- some combination of the factors

After the release of the study, MNPD [reduced traffic stops by roughly 75%]() and did not see an associated rise in serious crime.

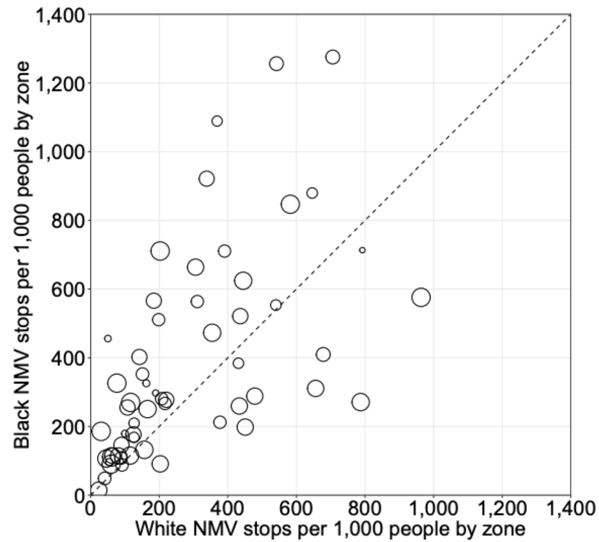


Figure 5: *Black versus white per capita stops for non-moving violations (NMV). Each circle represents a police zone, sized by number of stops (black and white) made in each zone in 2017. More points lie above the reference line than below, indicating that within-location stop rates are higher for black drivers than for white drivers.*

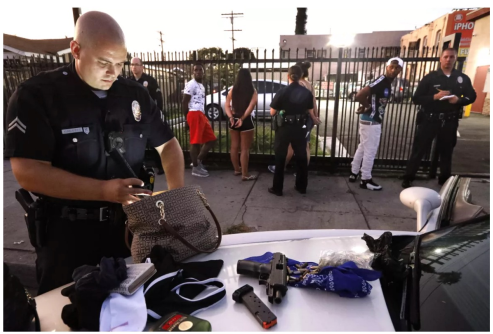
https://policylab.stanford.edu/projects/nashville-traffic-stops.html

Using traffic data

- 100 million traffic stops
- found Black and Hispanic drivers were routinely searched on the basis of less evidence as compared to white drivers

Los Angeles Times

LAPD searches blacks and Latinos more. But they're less likely to have contraband than whites



LAPD Officer Charles Kumlander searches a woman's purse after spotting a gun on the floor of a car he and his partner had pulled over in South Los Angeles. (Genaro Molina / Los Angeles Times)

BY BEN POSTON, CINDY CHANG

OCT. 8, 2019 3:52 PM PT

Los Angeles Times

LAPD will drastically cut back on pulling over random vehicles over racial bias concerns

Community groups in front of LAPD headquarters demanding that the department reform its vehicle stop practices, which disproportionately impact blacks and Latinos. (Al Seib/Los Angeles Times)

BY CINDY CHANG, BEN POSTON

OCT. 13, 2019 7 AM PT

Not all traffic data are created equally

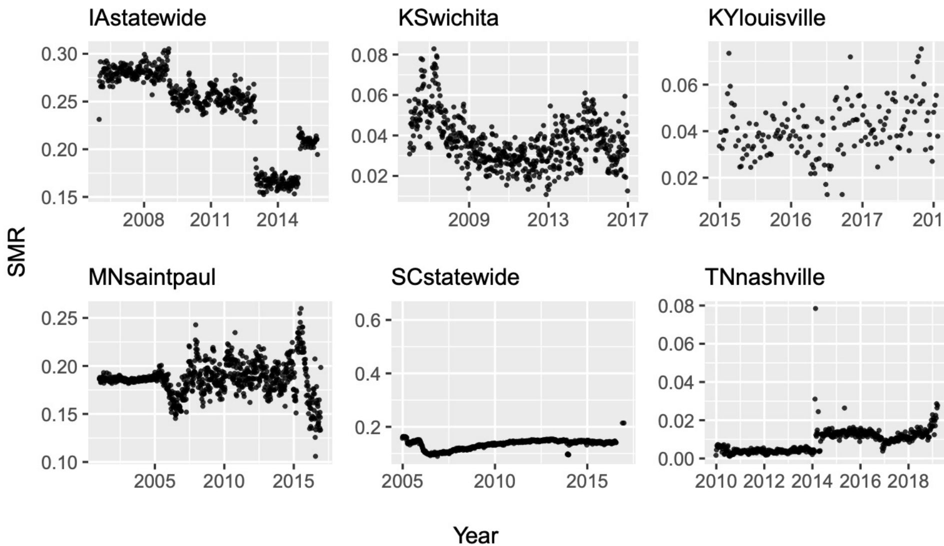
- are there fundamental differences between the traffic stop observations with high and low missingness?
- Does missingness have a certain trend with respect to pertinent variables like race and time?
- Most importantly, are the missingness trends drastic enough to render any results unreliable?

Stop Missingness Rate (SMR)

$$\text{SMR} = \frac{\text{number of unrecorded variables}}{\text{total number of variables}}$$

$$\text{dataset SMR} = \frac{1}{T}\sum^{T} \text{SMR}_i \qquad T = \text{total number of traffic stops}$$

Figure 7: Weekly SMR of Selected Datasets



Scatterplots of the weekly SMR plotted by year. Note that the scaling of the y−axes is different so the details for each trend are clear.

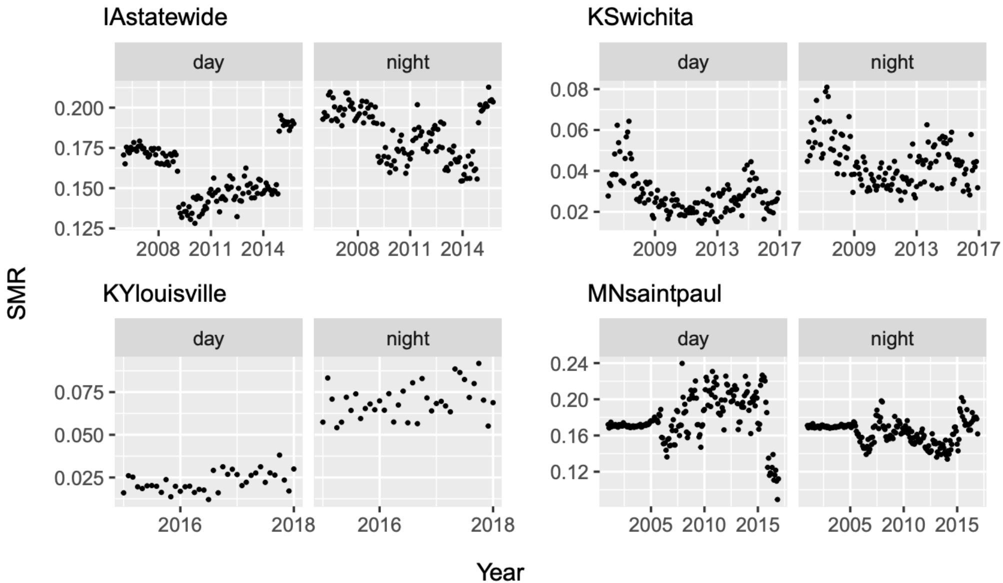
https://www.causeweb.org/usproc/sites/default/files/usresp/2020-2/exploring%20missingness%20and%20its%20implications%20on%20traffic%20stop%20data.pdf

Stop Missingness Rate (SMR)

$$\text{SMR} = \frac{\text{number of unrecorded variables}}{\text{total number of variables}}$$

$$\text{dataset SMR} = \frac{1}{T}\sum_{i=1}^{T}\text{SMR}_i \qquad T = \text{total number of traffic stops}$$

Figure 8: Monthly SMR in Daytime and Nighttime Stops for Selected Datasets



Scatterplots of the monthly SMR for stops occurring during the day and night, plotted by year.
Note that the scaling of the y−axes are different for ease of interpretation.

https://www.causeweb.org/usproc/sites/default/files/usresp/2020-2/exploring%20missingness%20and%20its%20implications%20on%20traffic%20stop%20data.pdf

How does missing race interact with being searched?

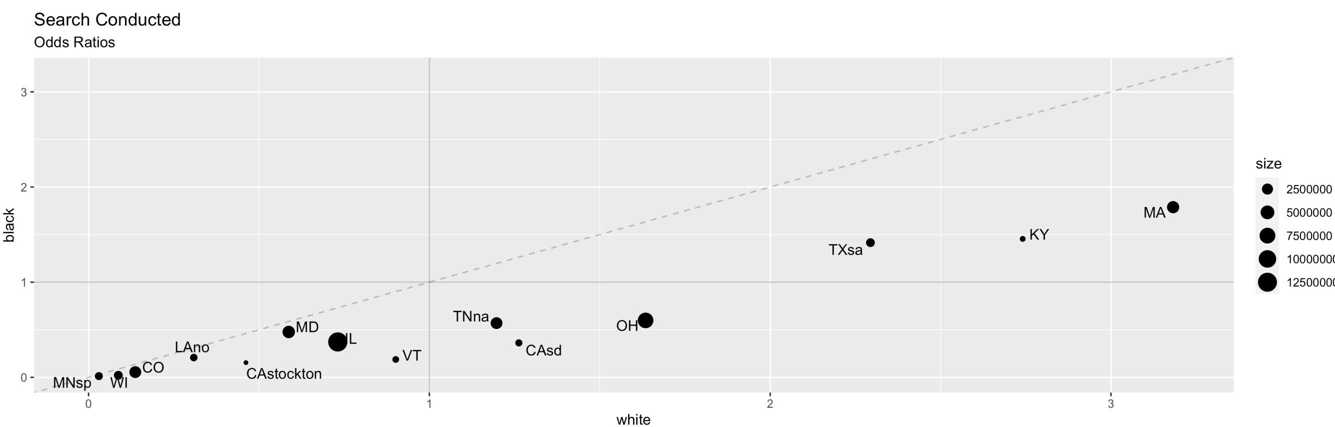
$$\text{odds}_{\text{Black}} = \left.\frac{\text{not searched}}{\text{searched}}\right\} \text{Black population}$$

$$\text{odds}_{\text{white}} = \left.\frac{\text{not searched}}{\text{searched}}\right\} \text{white population}$$

$$\text{odds}_{\text{NA}} = \left.\frac{\text{not searched}}{\text{searched}}\right\} \text{population with missing race}$$

$$\text{OR}_{\text{Black}} = \frac{\text{odds}_{\text{Black}}}{\text{odds}_{\text{NA}}}$$

$$\text{OR}_{\text{white}} = \frac{\text{odds}_{\text{white}}}{\text{odds}_{\text{NA}}}$$



What can we do?

- all types of people at the data science table (especially Black and Brown people)
- data that doesn't replicate existing inequities
- work to find metrics that minimize error rates for marginalized communities (and not just that minimize total error)
- work against pervasive and problematic algorithms by using adversarial methods

The Algorithmic Accountability Act (2022)

THE

NATIONAL LAW REVIEW

Federal Lawmakers in House and Senate Introduce Algorithmic Accountability Act of 2022

Friday, February 11, 2022

Data privacy and AI continues to be a top-of-mind issue in 2022. Consistent with this broader trend, the Algorithmic Accountability Act of 2022 was introduced in the U.S. Senate as S. 3572 and in the U.S. House of Representatives as HR 6580 on February 3, 2022. This bill proposes to direct the Federal Trade Commission ("FTC") to promulgate regulations that require any "covered entity" to perform impact assessments and meet other requirements regarding automated decision-making processes and in particular those that implicate an "augmented critical decision process"— essentially, that result in any legal or other material effects — on a consumer.

https://www.natlawreview.com/article/federal-lawmakers-house-and-senate-introduce-algorithmic-accountability-act-2022

Quote and image credit: Jurassic Park

There is a need for human values and fairness, even at the cost of efficiency.

Thank you

Image credit: Pomona College

jo.hardin@pomona.edu

@jo_hardin47

https://github.com/hardin47

https://hardin47.netlify.app/

Joy Buolamwini – the poet of code

https://www.youtube.com/embed/QxuyfWoVV98

References

- Angwin, Larson, Mattu, and Kirchner, “Machine Bias”, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Buolamwini & Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”, *Proceedings of Machine Learning Research* 81:1–15, 2018.
- Md. Abul Hasnat, Julien Bohne, Jonathan Milgram, St ´ ephane Gentric and Liming Chen, “von Mises-Fisher Mixture Model-based Deep learning: Application to Face Verification”, 2017 https://arxiv.org/abs/1706.04264
- O’Neil, Weapons of Math Destruction, 2016
- Aaron Roth, http://aaronsadventures.blogspot.com/2019/01/discussion-of-unfairness-in-machine.html, 2019
- Shawn Shan, Emily Wenger, Jiayun Zhang, Huiying Li, Haitao Zheng, Ben Y. Zhao, “Fawkes: Protecting Privacy against Unauthorized Deep Learning Models”. *Proc. of USENIX Security Symposium* 2020.