Algorithmic accountability

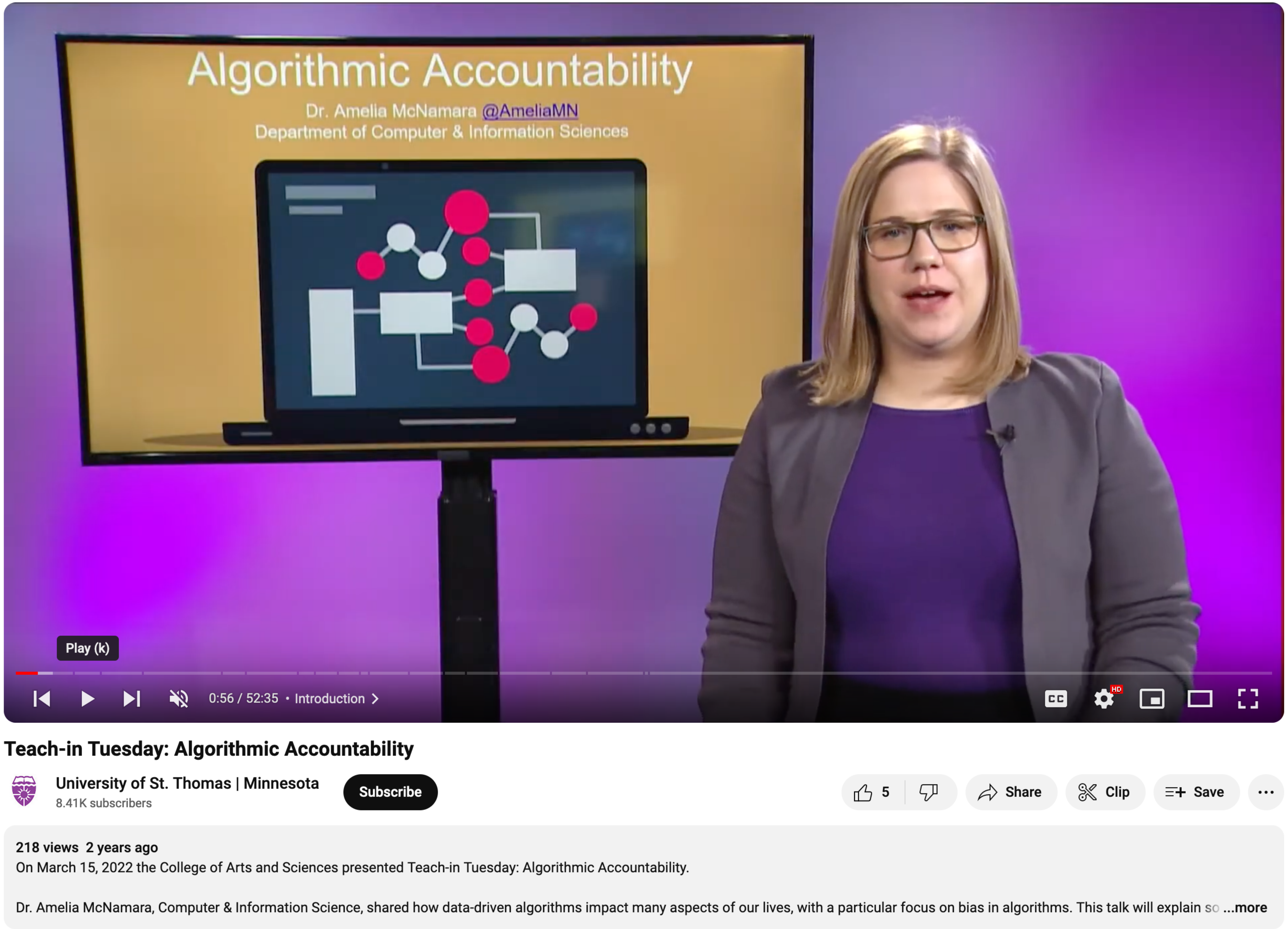
Amelia McNamara

flickr: Monika Hoinkis

Main tasks for algorithms

- Prediction (“what will the value be?”)
- Classification (“is this A or B?”)

All of these rely on training data, so all of them will be limited by what has happened in the past, and what they are trained on.


Algorithmic Accountability
Dr. Amelia McNamara @AmeliaMN
Department of Computer & Information Sciences
Play (k)
0:56 / 52:35 • Introduction
University of St. Thomas | Minnesota
8.41K subscribers
Subscribe
5
Share
Clip
Save

Teach-in Tuesday: Algorithmic Accountability

218 views 2 years ago

On March 15, 2022 the College of Arts and Sciences presented Teach-in Tuesday: Algorithmic Accountability.

Dr. Amelia McNamara, Computer & Information Science, shared how data-driven algorithms impact many aspects of our lives, with a particular focus on bias in algorithms. This talk will explain so ...more

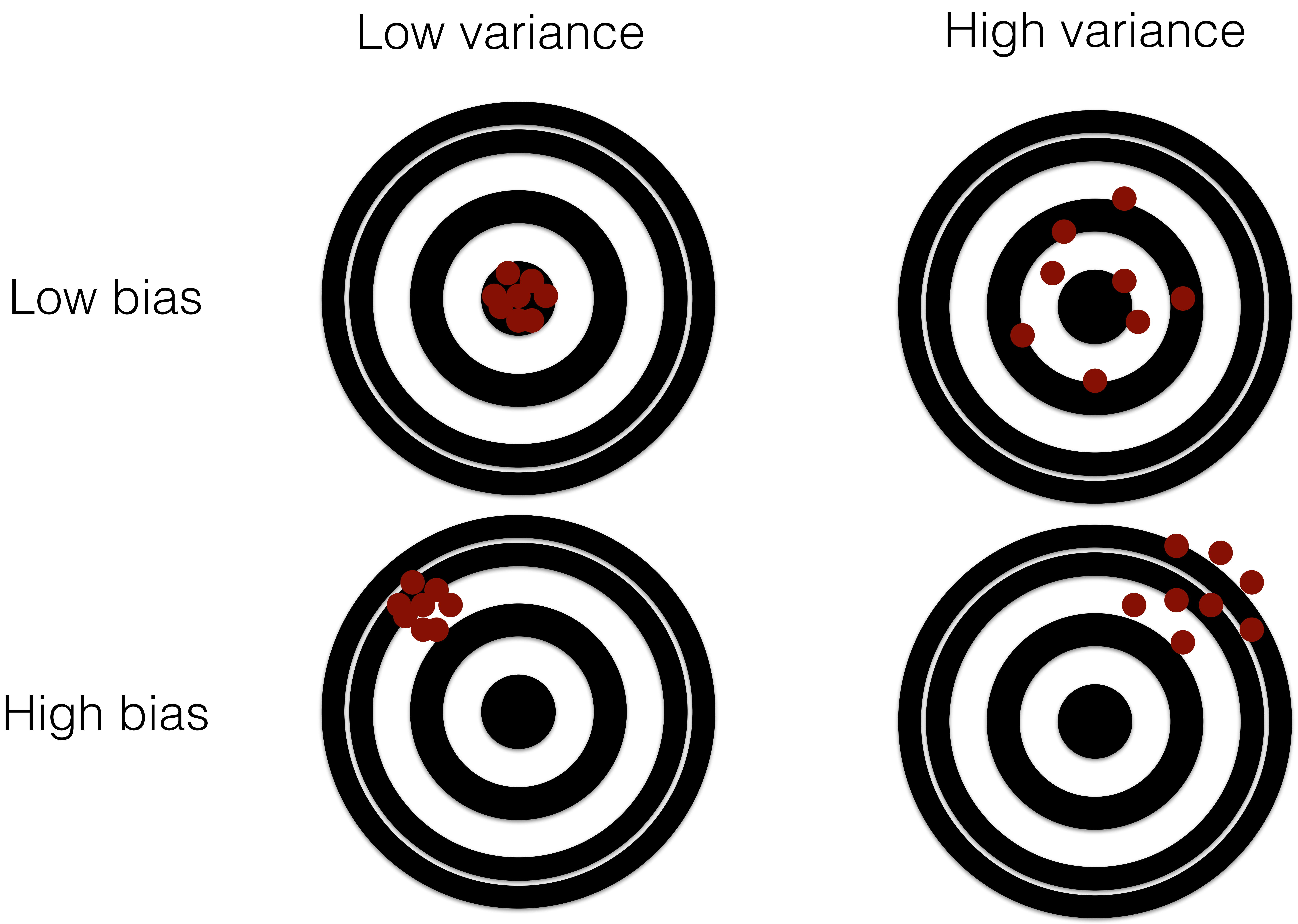
Teach-in Tuesday on Algorithmic Accountability, March 2022. https://www.youtube.com/watch?v=jeG3RgOO2c8

flickr: jcolman

Canoga Park
Reseda
Van Nuys
North Hollywood
La Cañada Flintridge
Altadena
Upper Las Virgenes Canyon Open Space Preserve
Woodland Hills
Burbank
Sierra Madre
Encino
Sherman Oaks
Glendale
Pasadena
Arcadia
Duarte
Calabasas
Universal City
Creek
ark
Topanga State Park
Los Feliz
Northeast Los Angeles
Alhambra
Hollywood
Rosemead
Baldwin Park
Monte Nido
Bel Air
Beverly Hills
Central LA
Westwood
Chinatown
El Monte
West Covina
Monterey Park
South El Monte
Brentwood
Los Angeles
Mid-City
Pacific Palisades
East Los Angeles
Montebello
Santa Monica
Culver City
Industry
Commerce
Hacienda Heights
South Los Angeles
Huntington Park
Pico Rivera
Whittier
Marina Del Rey
Ladera Heights
Los Angeles International Airport
Inglewood
Bell Gardens
La Habra Heights
Santa Fe Springs
South Whittier
South Gate
Downey
La Habra
Lynwood
El Segundo
Hawthorne
Willowbrook
La Mirada
Norwalk
Compton
Paramount
Manhattan Beach
Gardena
Buena Park
Cerritos
Lakewood
Beach
Azu
Row
Hei
Ful
27
101
170
5
210
134
2
1
164
10
710
60
405
110
187
605
19
72
39
105
91
Live traffic
Fast
Slow

We want to ensure algorithms are fair

pixabay: mohamed_hassan



Low variance

High variance

Low bias

High bias

bias noun

Definition of *bias*

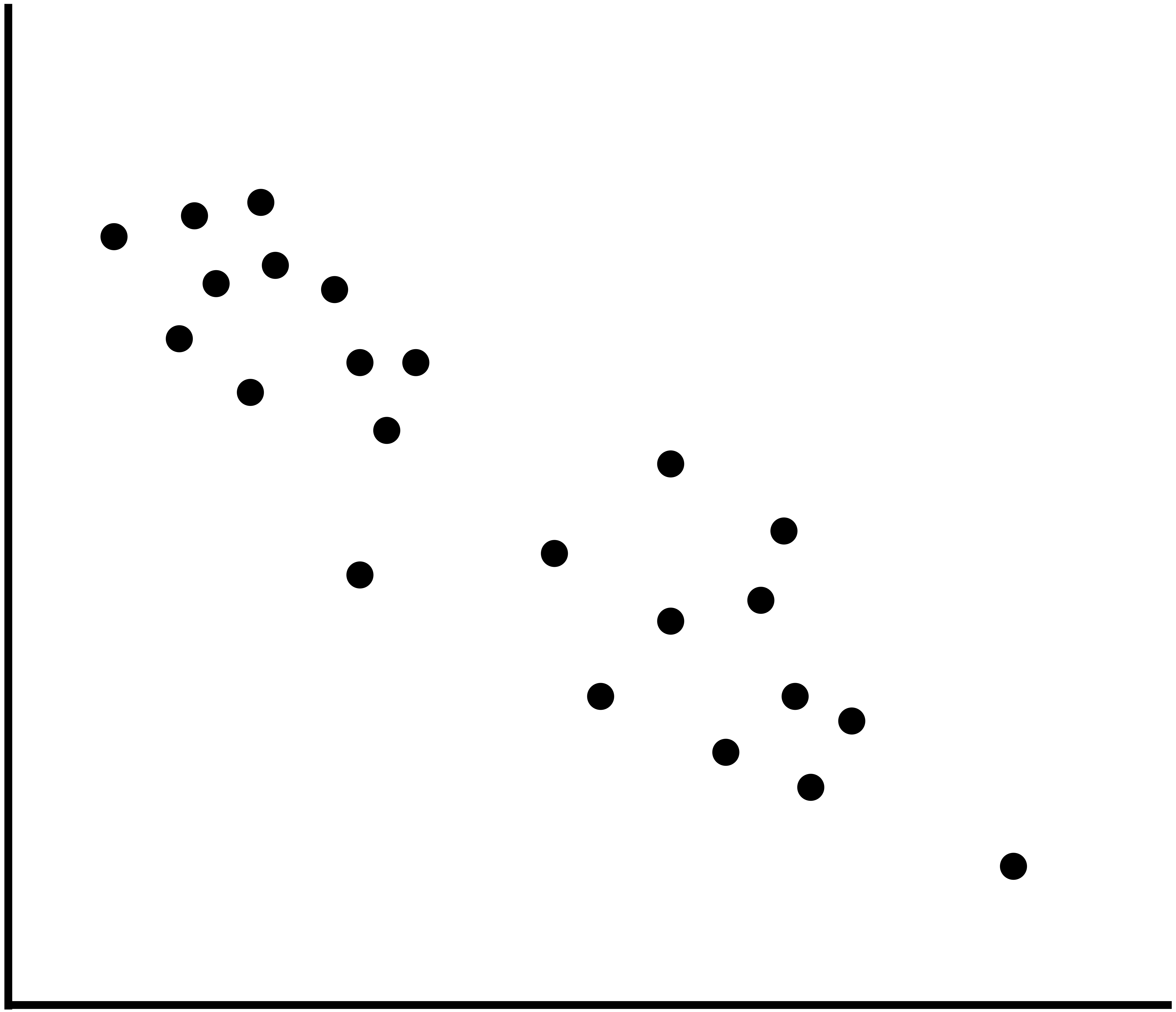
1.

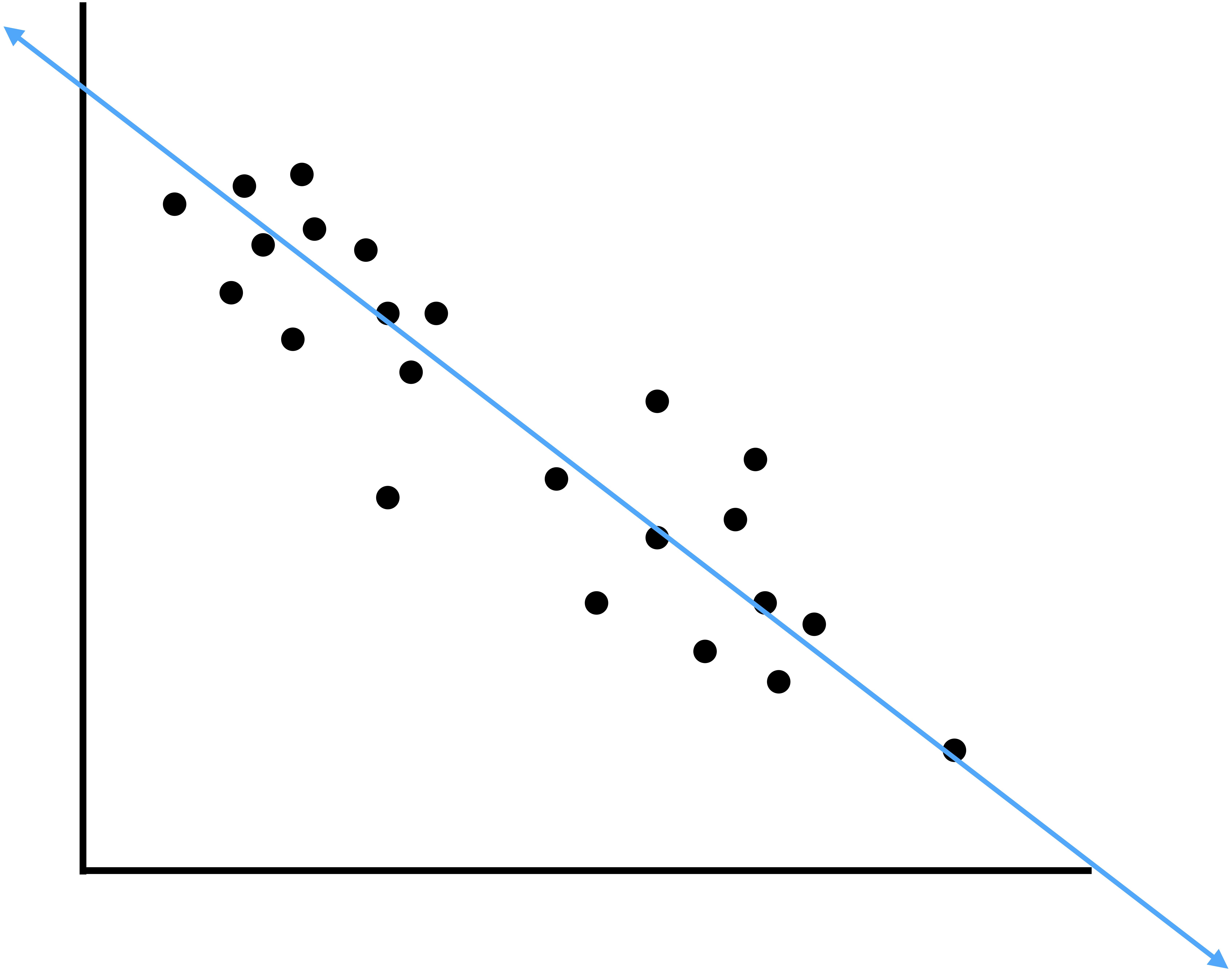
a.

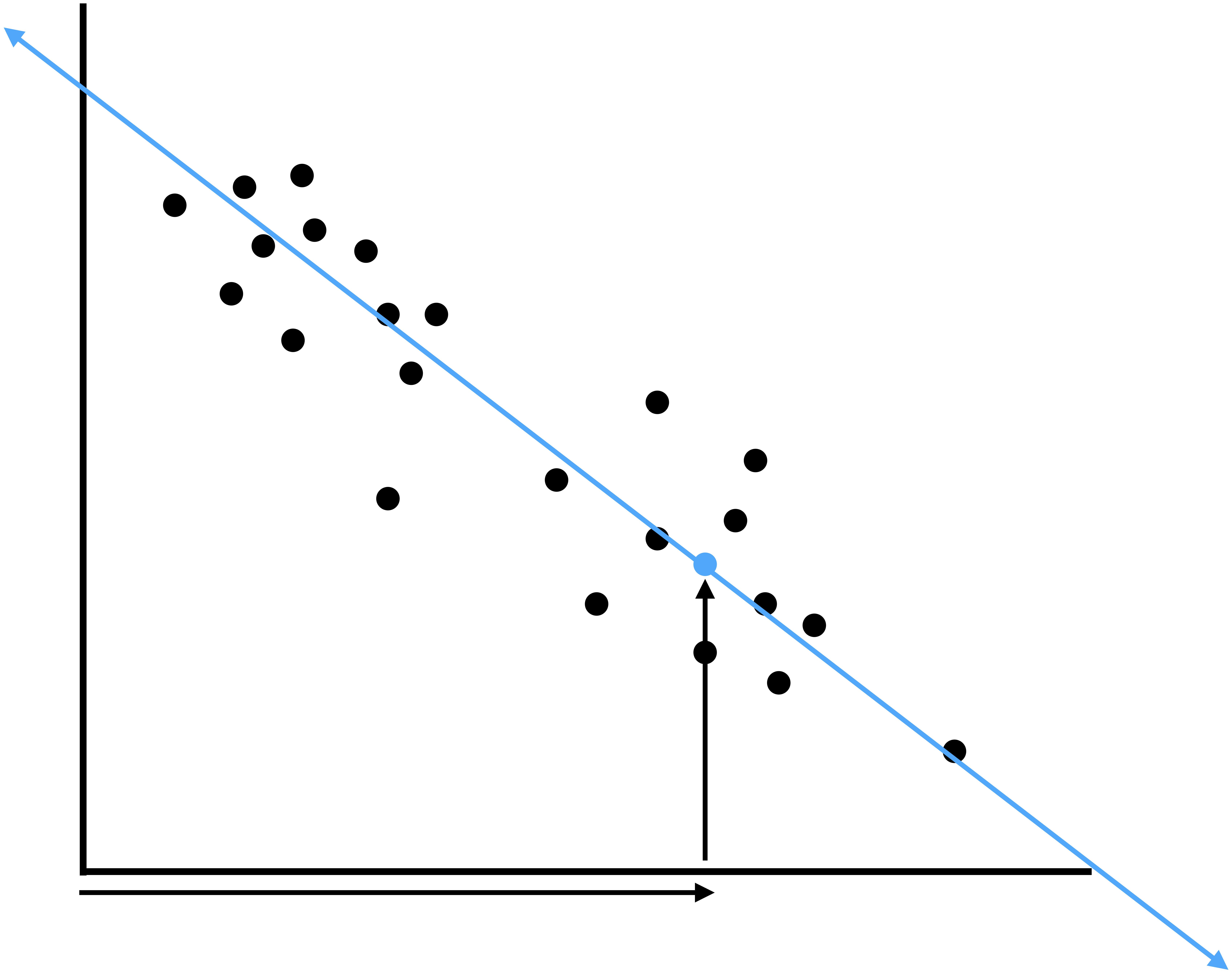
b.

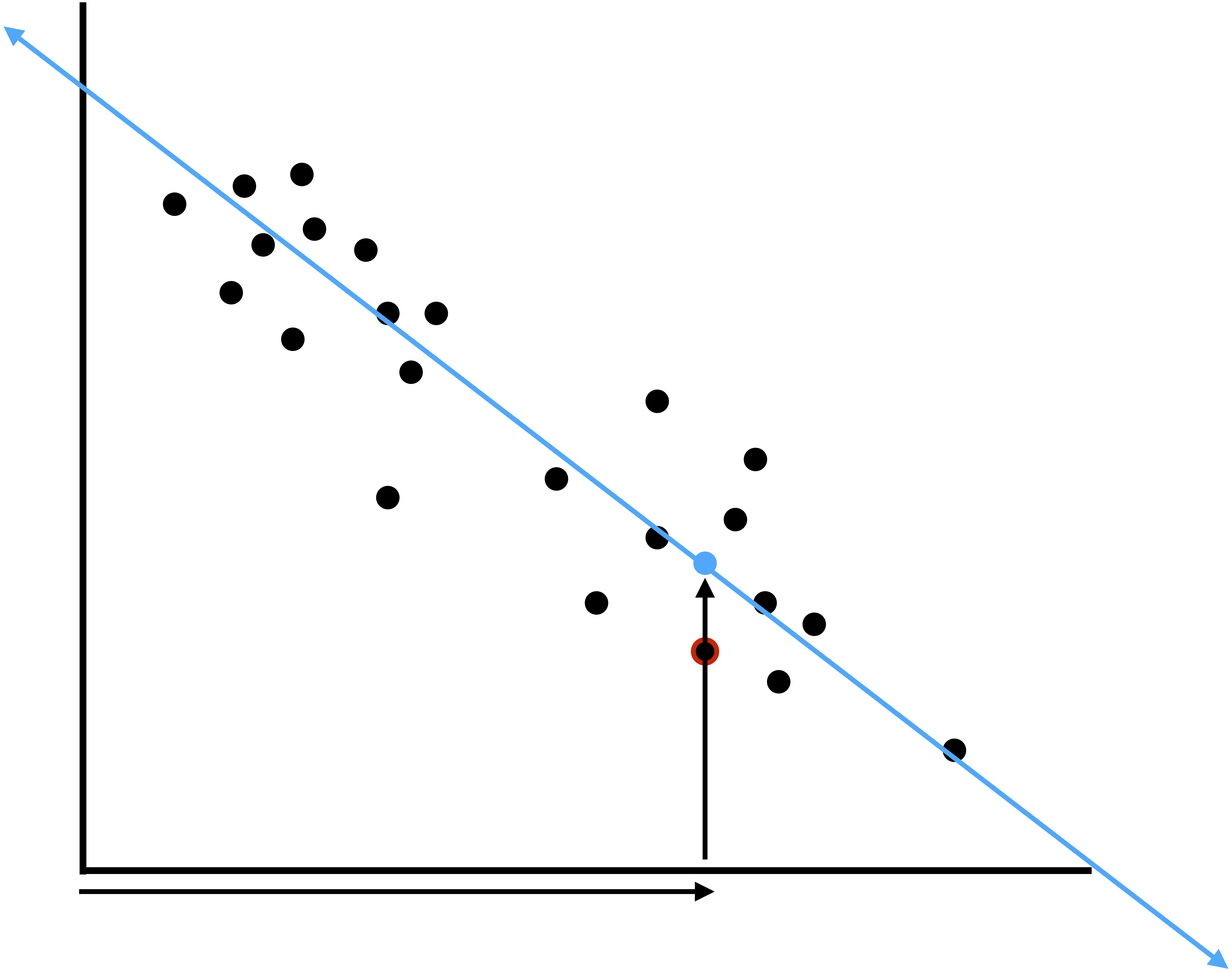
c.

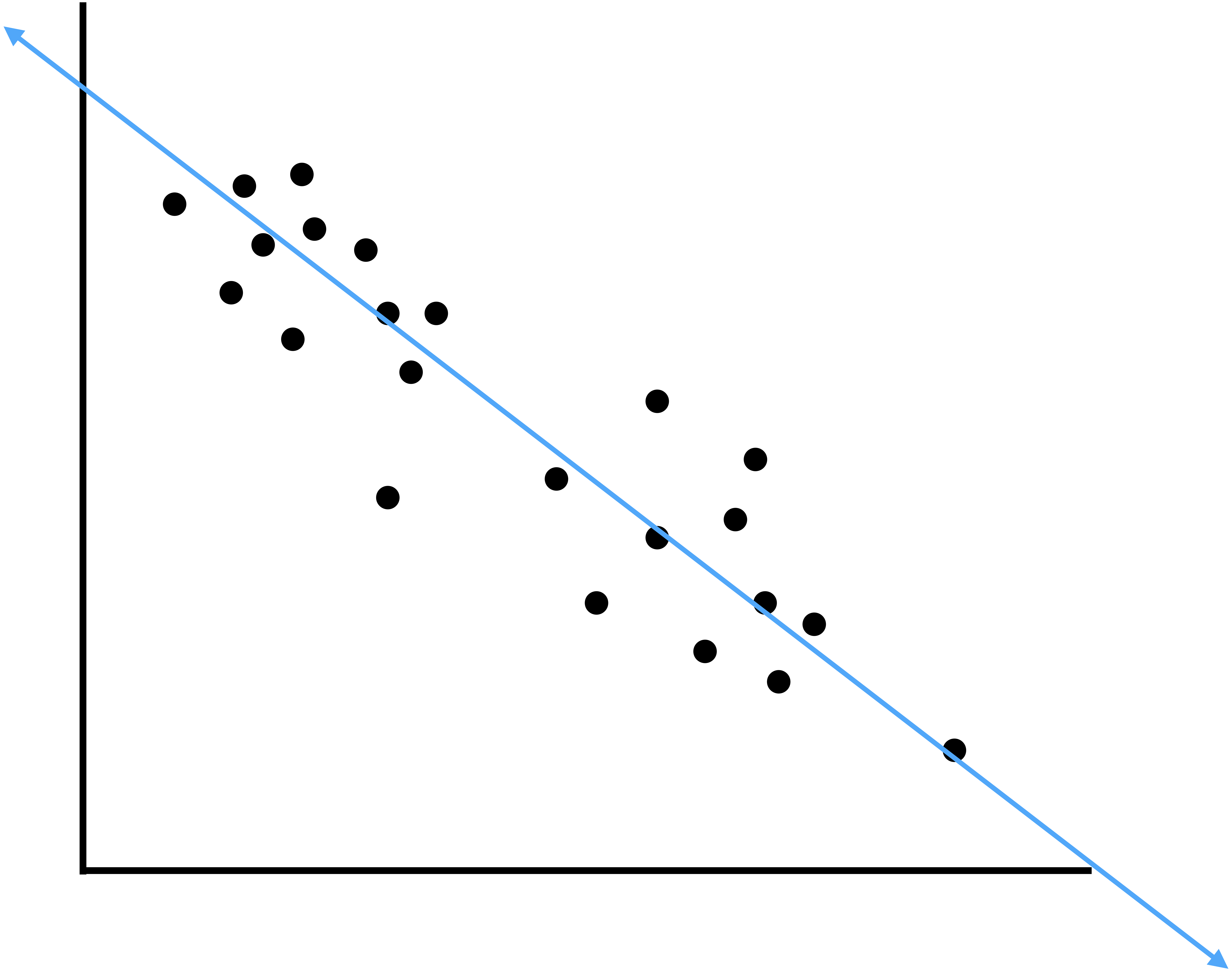
d. (1): deviation of the expected value of a statistical estimate from the quantity it estimates

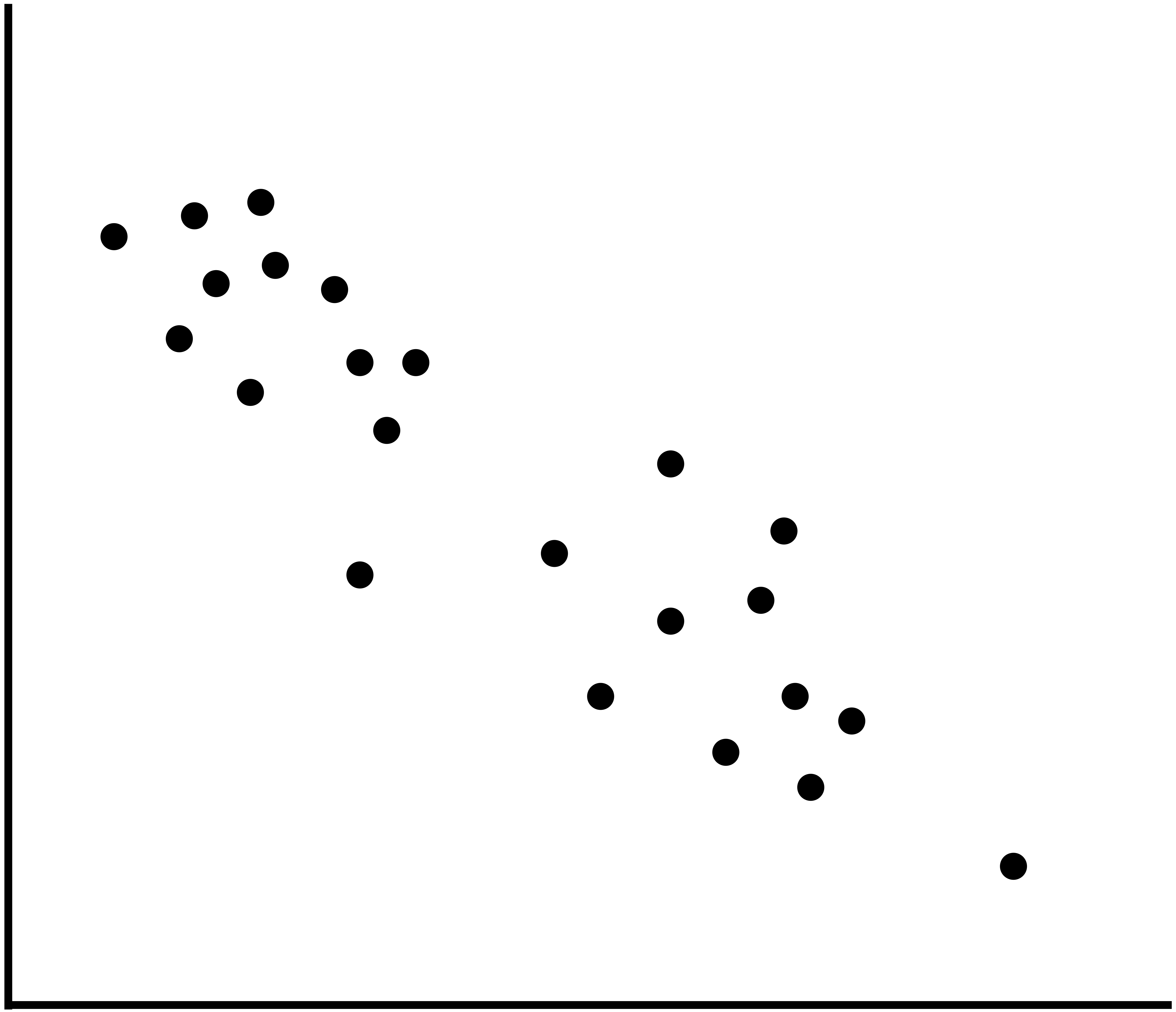


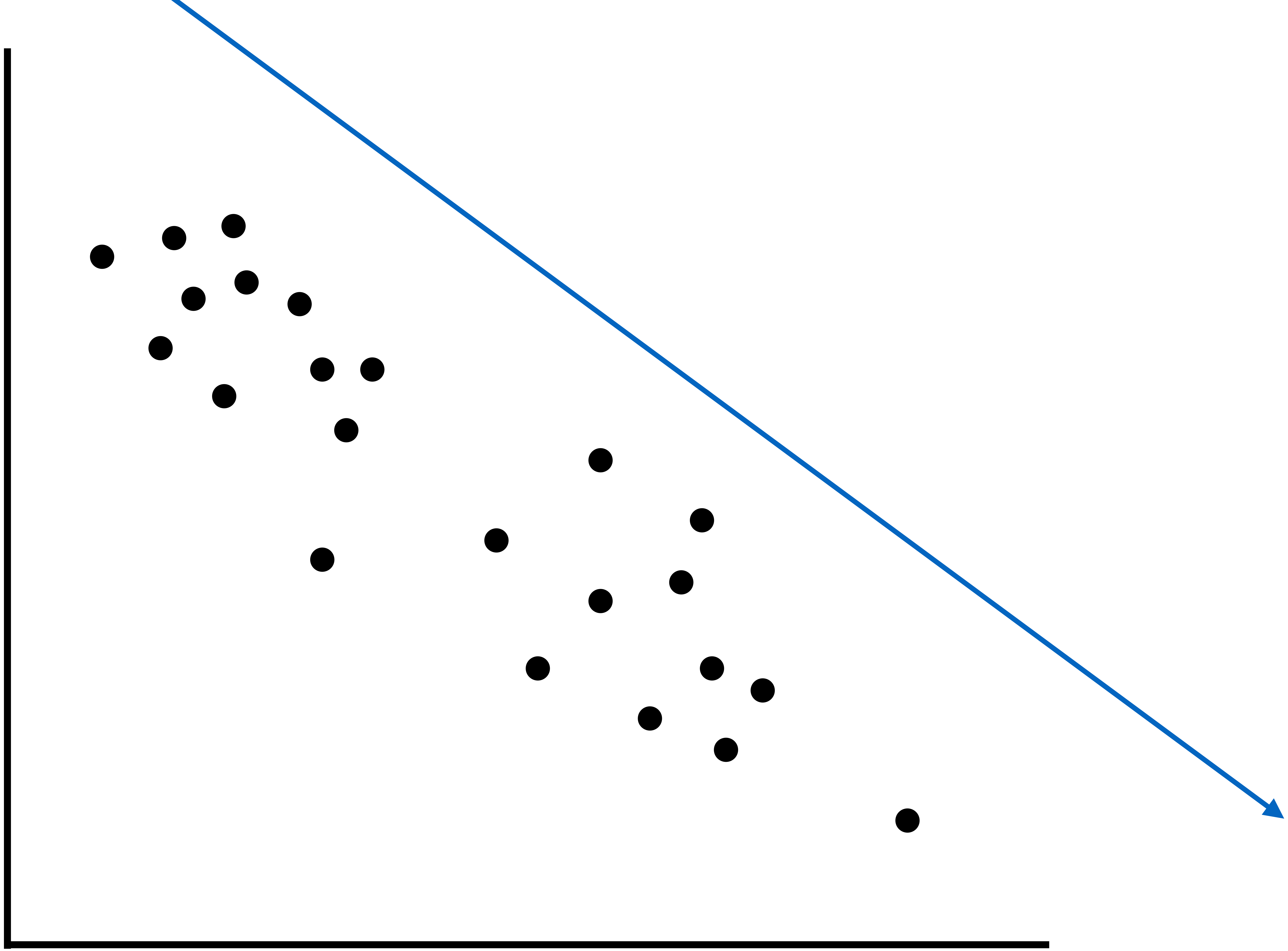


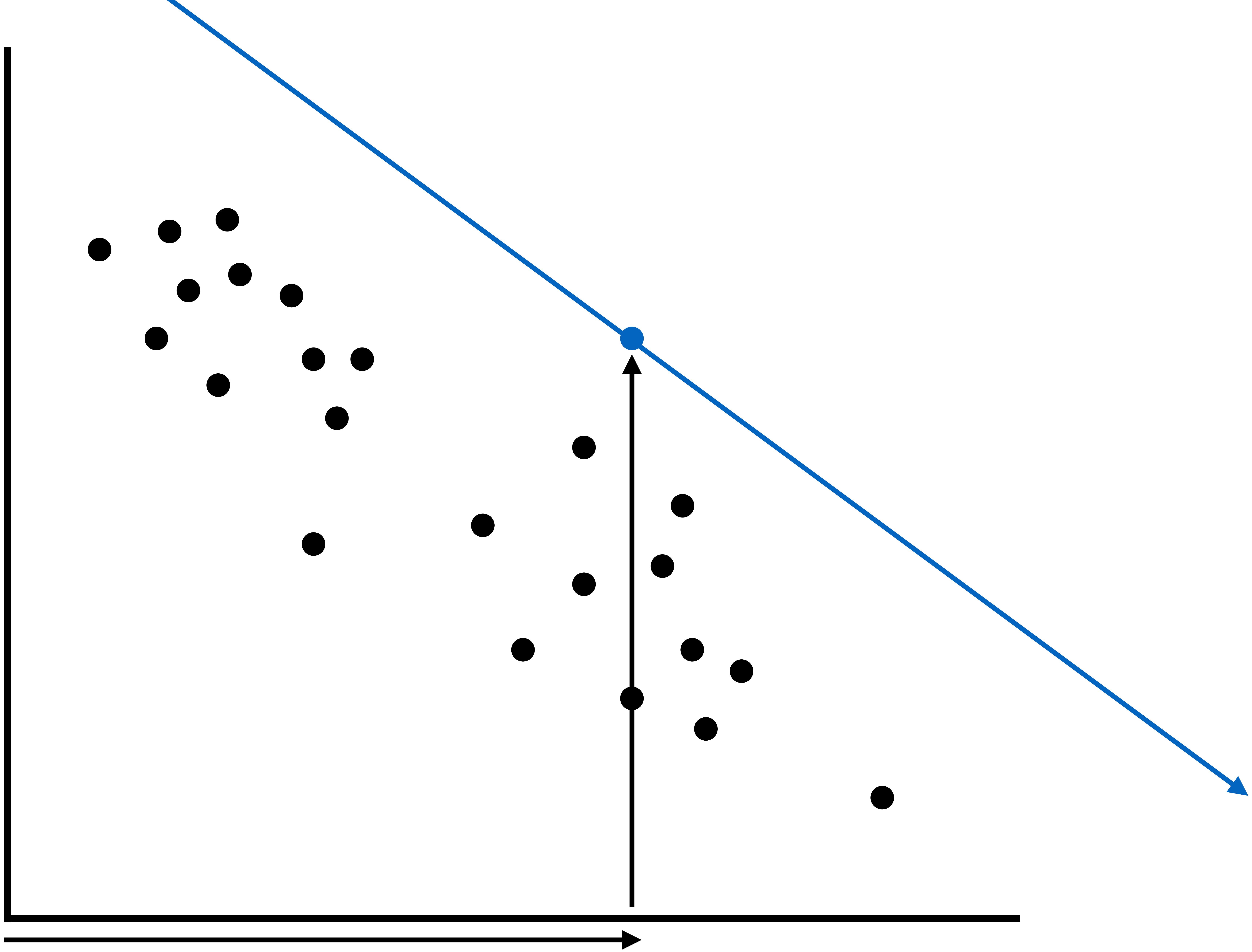


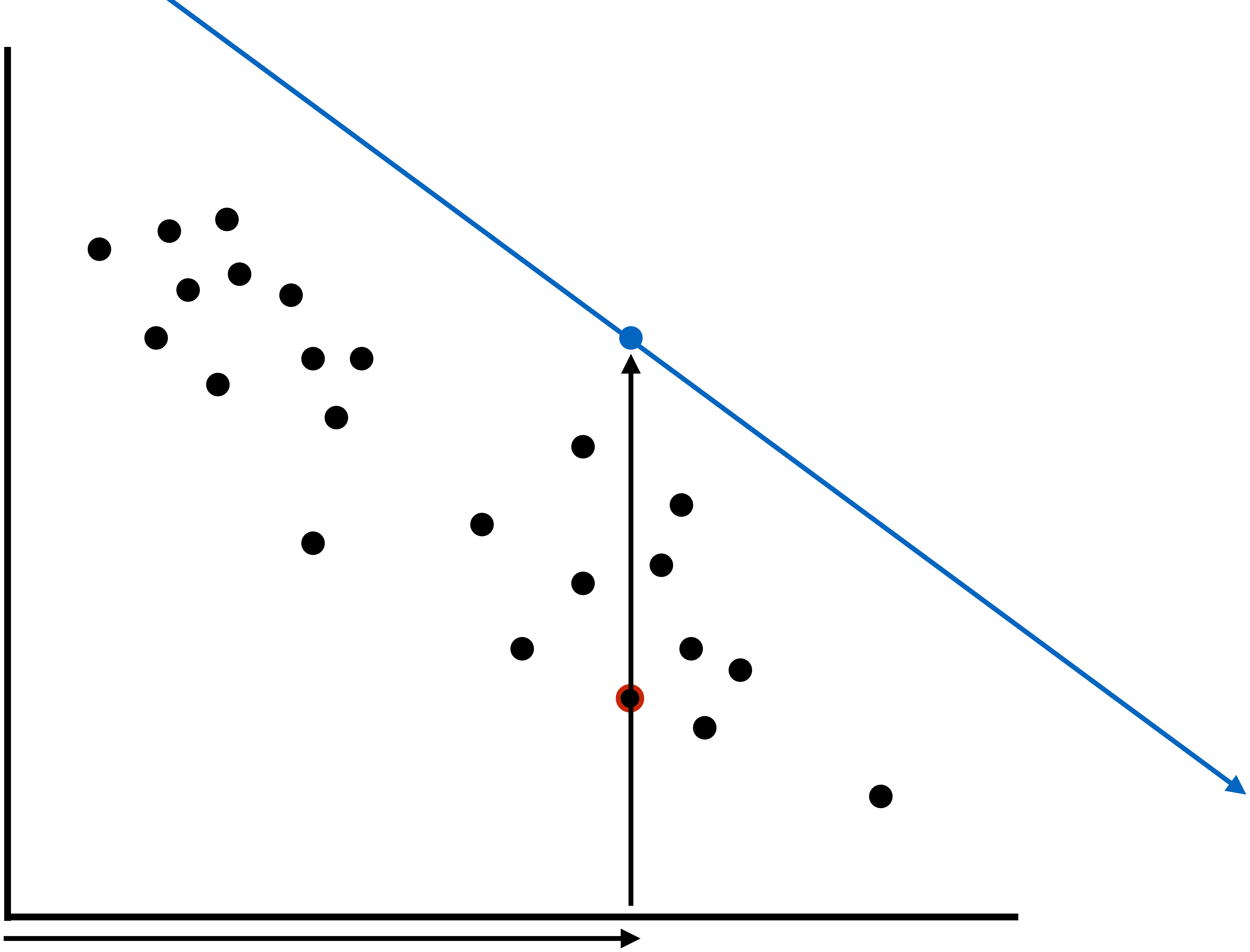


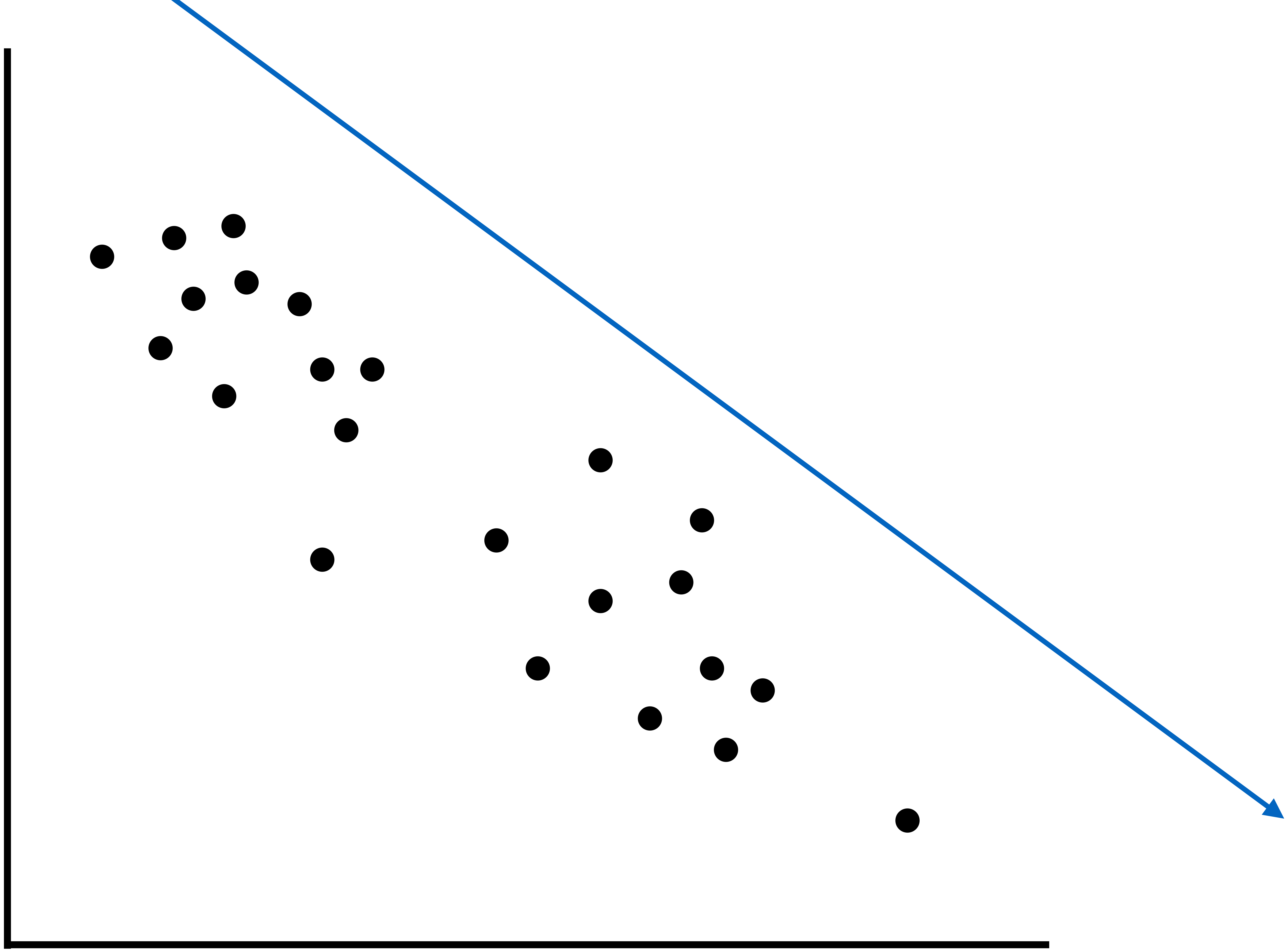




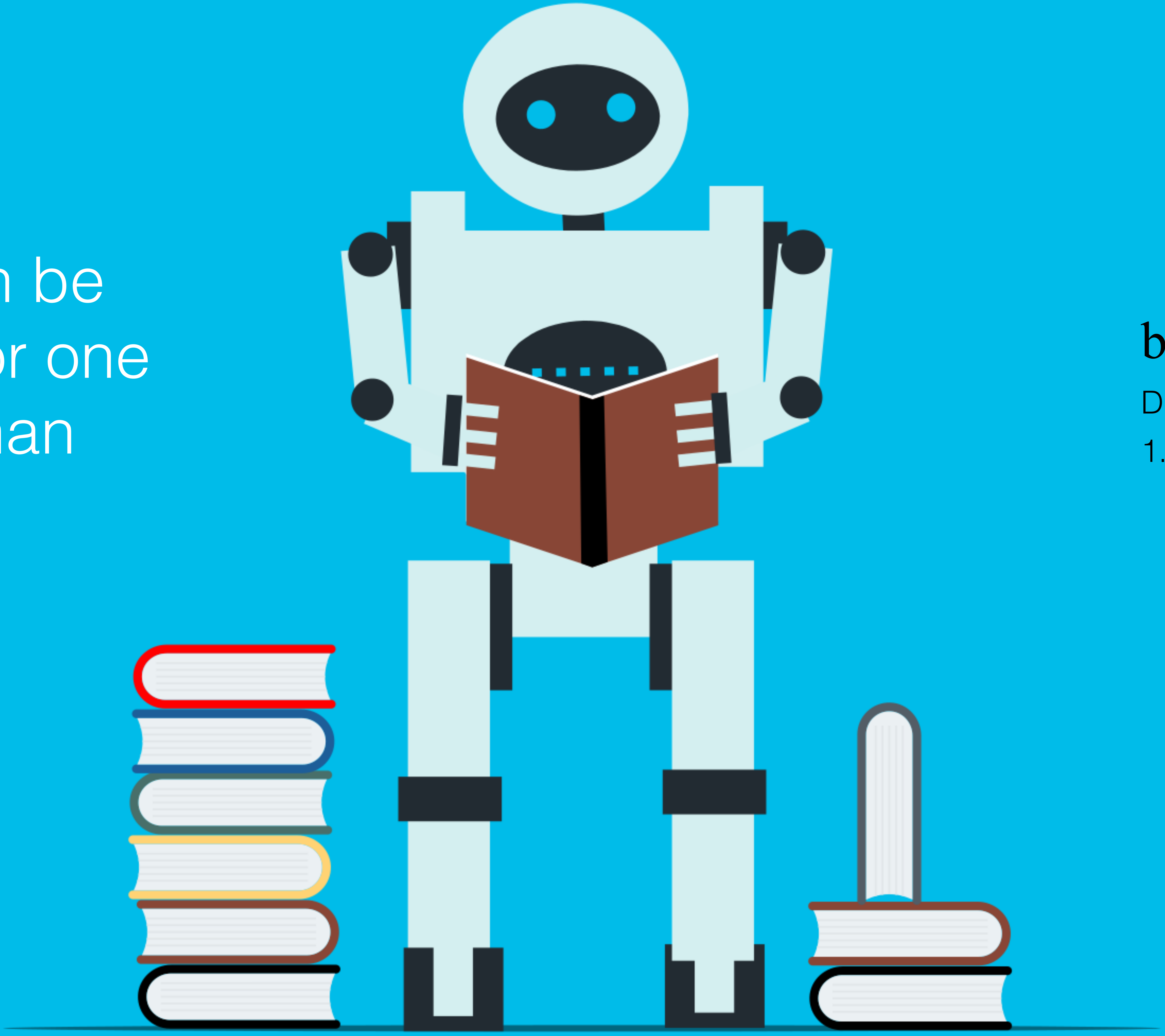








Bias can be worse for one group than another



pixabay: mohamed_hassan

bias noun

Definition of *bias*

1\.

a. an inclination of temperament or outlook
especially: a personal and sometimes unreasoned judgment: prejudice

b. an instance of such prejudice

Prediction/ Classification

I sent a deadly animal to you

tammie [fbNOREPLY@myfanbox.com]

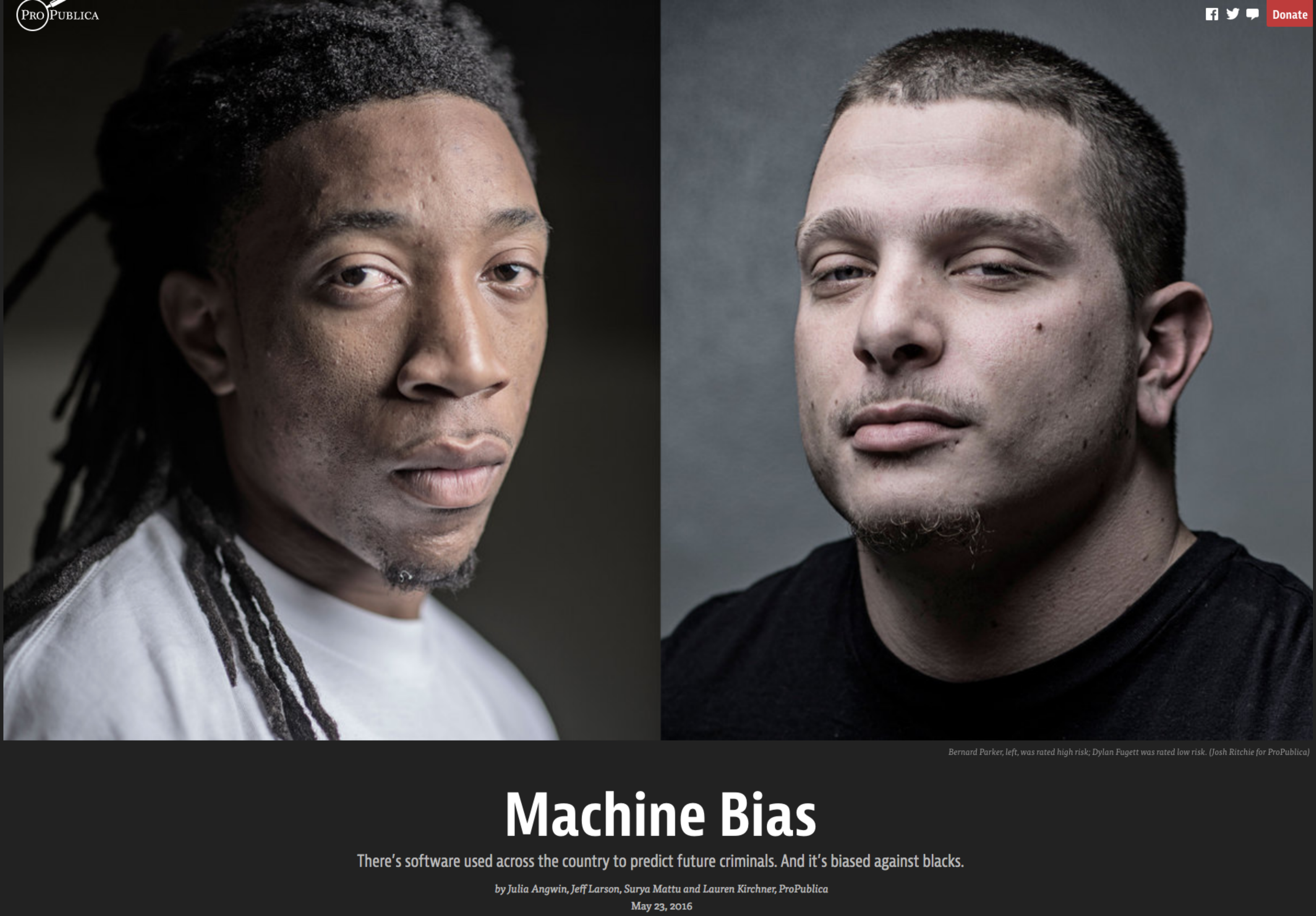
Sent: Wednesday, February 10, 2010 6:49 PM

To: Camelia

I sent a deadly animal to you Check it out!

This email was sent by tammie using the Application: YouTube. You can stop receiving emails

flickr: cameliatwu



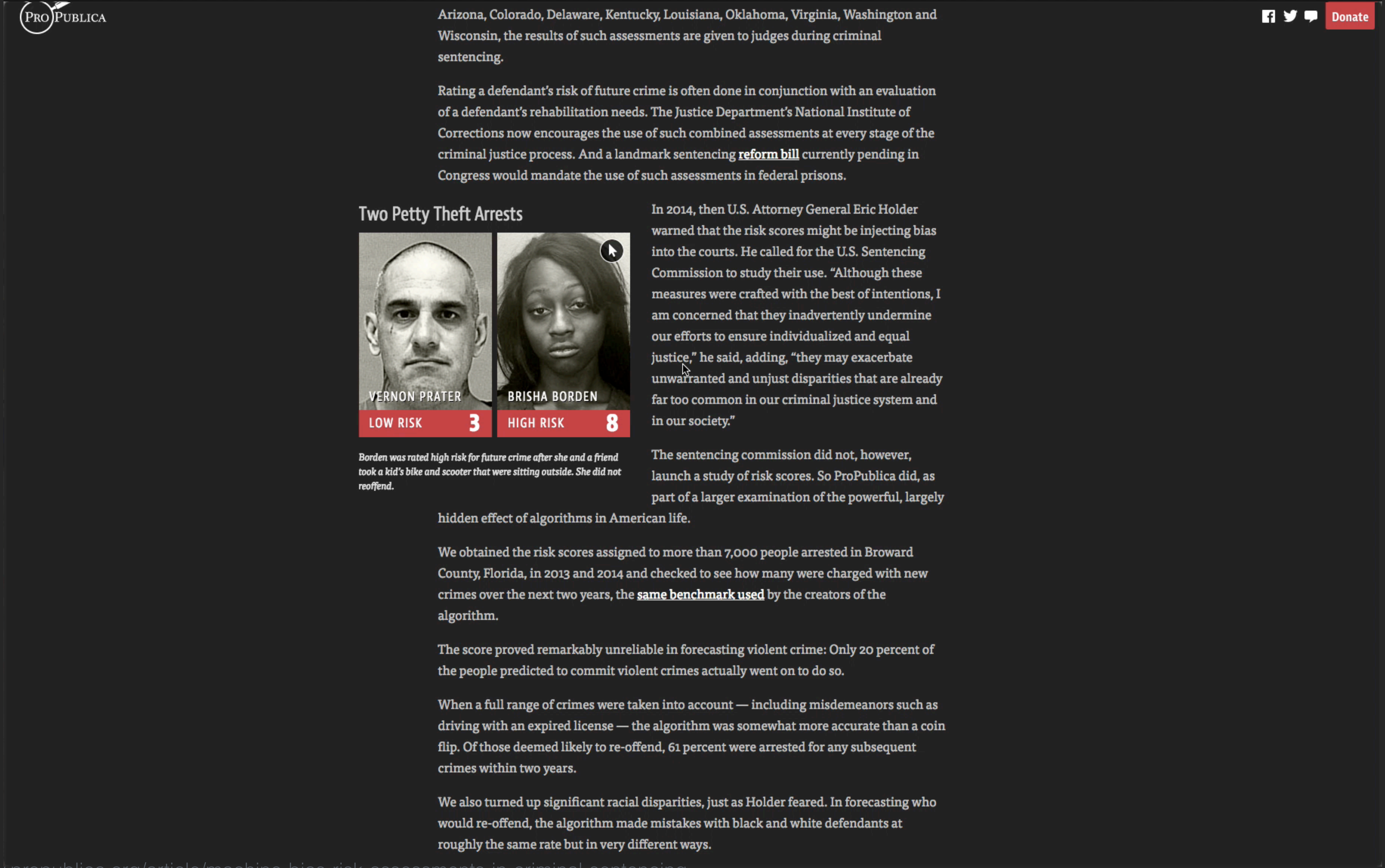
Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing



Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington and Wisconsin, the results of such assessments are given to judges during criminal sentencing.

Rating a defendant's risk of future crime is often done in conjunction with an evaluation of a defendant's rehabilitation needs. The Justice Department's National Institute of Corrections now encourages the use of such combined assessments at every stage of the criminal justice process. And a landmark sentencing **reform bill** currently pending in Congress would mandate the use of such assessments in federal prisons.

Two Petty Theft Arrests

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

In 2014, then U.S. Attorney General Eric Holder warned that the risk scores might be injecting bias into the courts. He called for the U.S. Sentencing Commission to study their use. "Although these measures were crafted with the best of intentions, I am concerned that they inadvertently undermine our efforts to ensure individualized and equal justice," he said, adding, "they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society."

The sentencing commission did not, however, launch a study of risk scores. So ProPublica did, as part of a larger examination of the powerful, largely hidden effect of algorithms in American life.

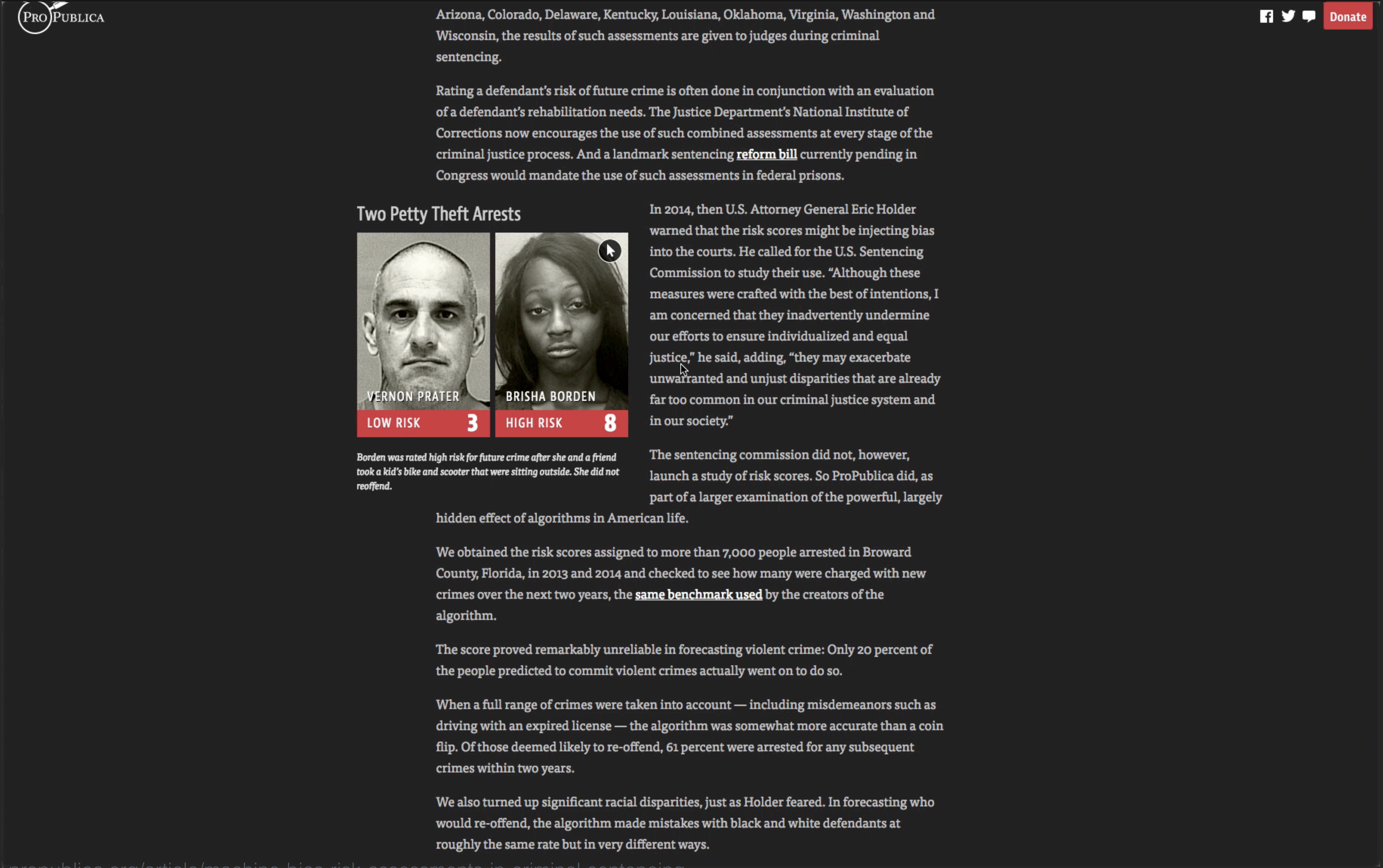
We obtained the risk scores assigned to more than 7,000 people arrested in Broward County, Florida, in 2013 and 2014 and checked to see how many were charged with new crimes over the next two years, the **same benchmark used** by the creators of the algorithm.

The score proved remarkably unreliable in forecasting violent crime: Only 20 percent of the people predicted to commit violent crimes actually went on to do so.

When a full range of crimes were taken into account — including misdemeanors such as driving with an expired license — the algorithm was somewhat more accurate than a coin flip. Of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years.

We also turned up significant racial disparities, just as Holder feared. In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways.

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing



ProPublica

Donate

Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington and Wisconsin, the results of such assessments are given to judges during criminal sentencing.

Rating a defendant's risk of future crime is often done in conjunction with an evaluation of a defendant's rehabilitation needs. The Justice Department's National Institute of Corrections now encourages the use of such combined assessments at every stage of the criminal justice process. And a landmark sentencing **reform bill** currently pending in Congress would mandate the use of such assessments in federal prisons.

Two Petty Theft Arrests

VERNON PRATER
LOW RISK 3

BRISHA BORDEN
HIGH RISK 8

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

In 2014, then U.S. Attorney General Eric Holder warned that the risk scores might be injecting bias into the courts. He called for the U.S. Sentencing Commission to study their use. "Although these measures were crafted with the best of intentions, I am concerned that they inadvertently undermine our efforts to ensure individualized and equal justice," he said, adding, "they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society."

The sentencing commission did not, however, launch a study of risk scores. So ProPublica did, as part of a larger examination of the powerful, largely hidden effect of algorithms in American life.

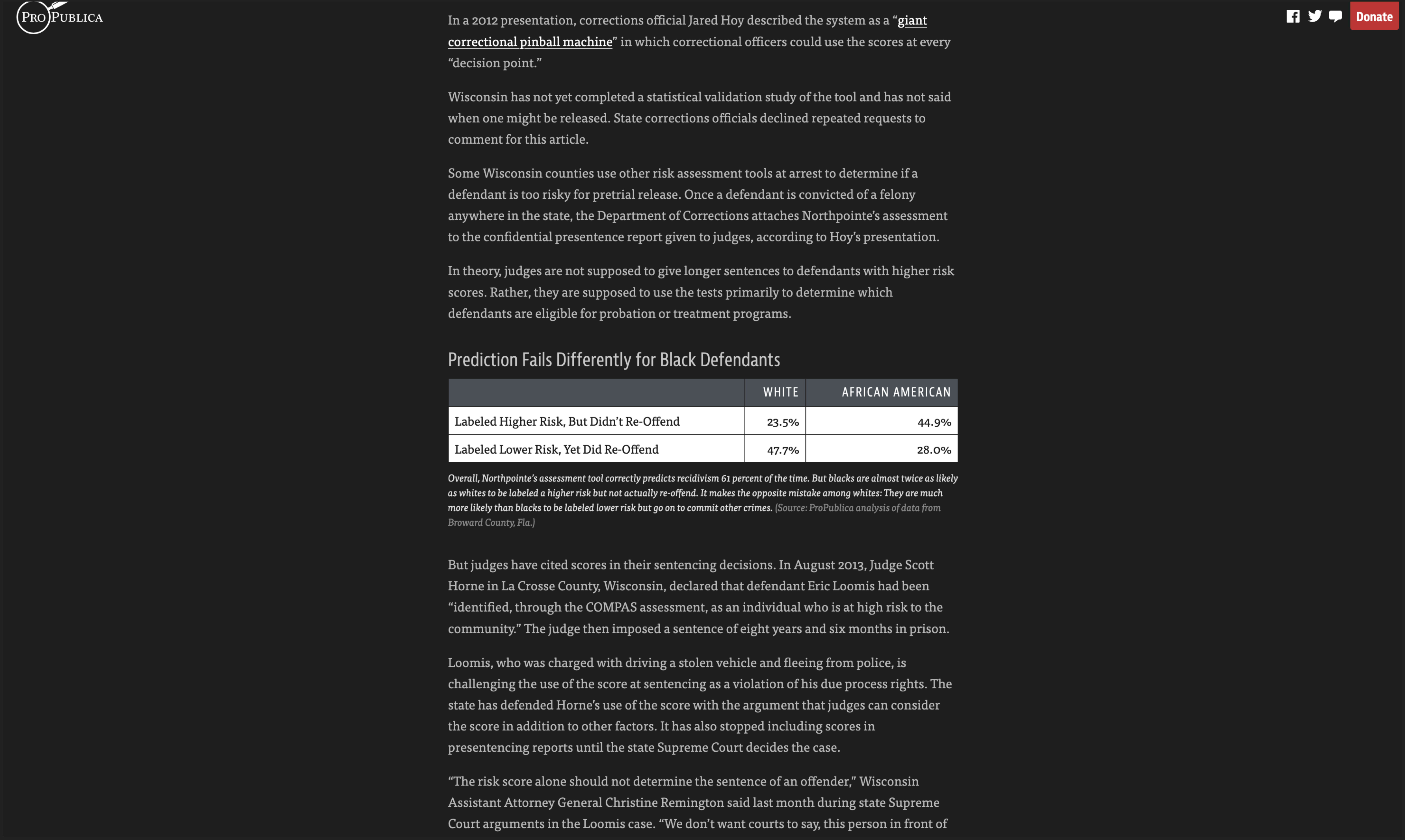
We obtained the risk scores assigned to more than 7,000 people arrested in Broward County, Florida, in 2013 and 2014 and checked to see how many were charged with new crimes over the next two years, the **same benchmark used** by the creators of the algorithm.

The score proved remarkably unreliable in forecasting violent crime: Only 20 percent of the people predicted to commit violent crimes actually went on to do so.

When a full range of crimes were taken into account — including misdemeanors such as driving with an expired license — the algorithm was somewhat more accurate than a coin flip. Of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years.

We also turned up significant racial disparities, just as Holder feared. In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways.

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing



In a 2012 presentation, corrections official Jared Hoy described the system as a "giant correctional pinball machine" in which correctional officers could use the scores at every "decision point."

Wisconsin has not yet completed a statistical validation study of the tool and has not said when one might be released. State corrections officials declined repeated requests to comment for this article.

Some Wisconsin counties use other risk assessment tools at arrest to determine if a defendant is too risky for pretrial release. Once a defendant is convicted of a felony anywhere in the state, the Department of Corrections attaches Northpointe's assessment to the confidential presentence report given to judges, according to Hoy's presentation.

In theory, judges are not supposed to give longer sentences to defendants with higher risk scores. Rather, they are supposed to use the tests primarily to determine which defendants are eligible for probation or treatment programs.

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

But judges have cited scores in their sentencing decisions. In August 2013, Judge Scott Horne in La Crosse County, Wisconsin, declared that defendant Eric Loomis had been "identified, through the COMPAS assessment, as an individual who is at high risk to the community." The judge then imposed a sentence of eight years and six months in prison.

Loomis, who was charged with driving a stolen vehicle and fleeing from police, is challenging the use of the score at sentencing as a violation of his due process rights. The state has defended Horne's use of the score with the argument that judges can consider the score in addition to other factors. It has also stopped including scores in presentencing reports until the state Supreme Court decides the case.

"The risk score alone should not determine the sentence of an offender," Wisconsin Assistant Attorney General Christine Remington said last month during state Supreme Court arguments in the Loomis case. "We don't want courts to say, this person in front of

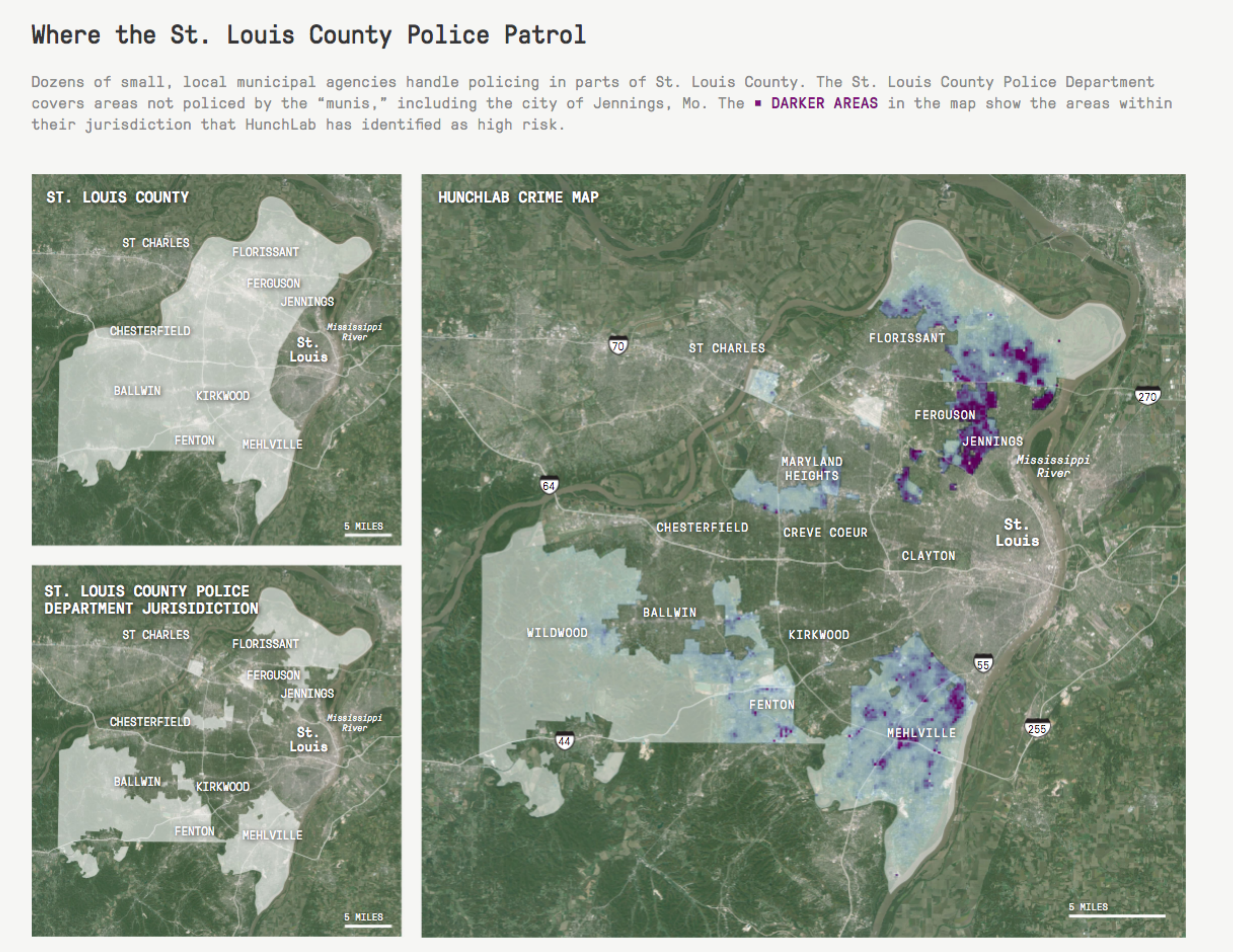
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

FEATURE

Policing the Future

In the aftermath of Michael Brown's death, St. Louis cops embrace crime-predicting software.

Maurice Chammah, with additional reporting by Mark Hansen. Policing the Future.
https://www.themarshallproject.org/2016/02/03/policing-the-future

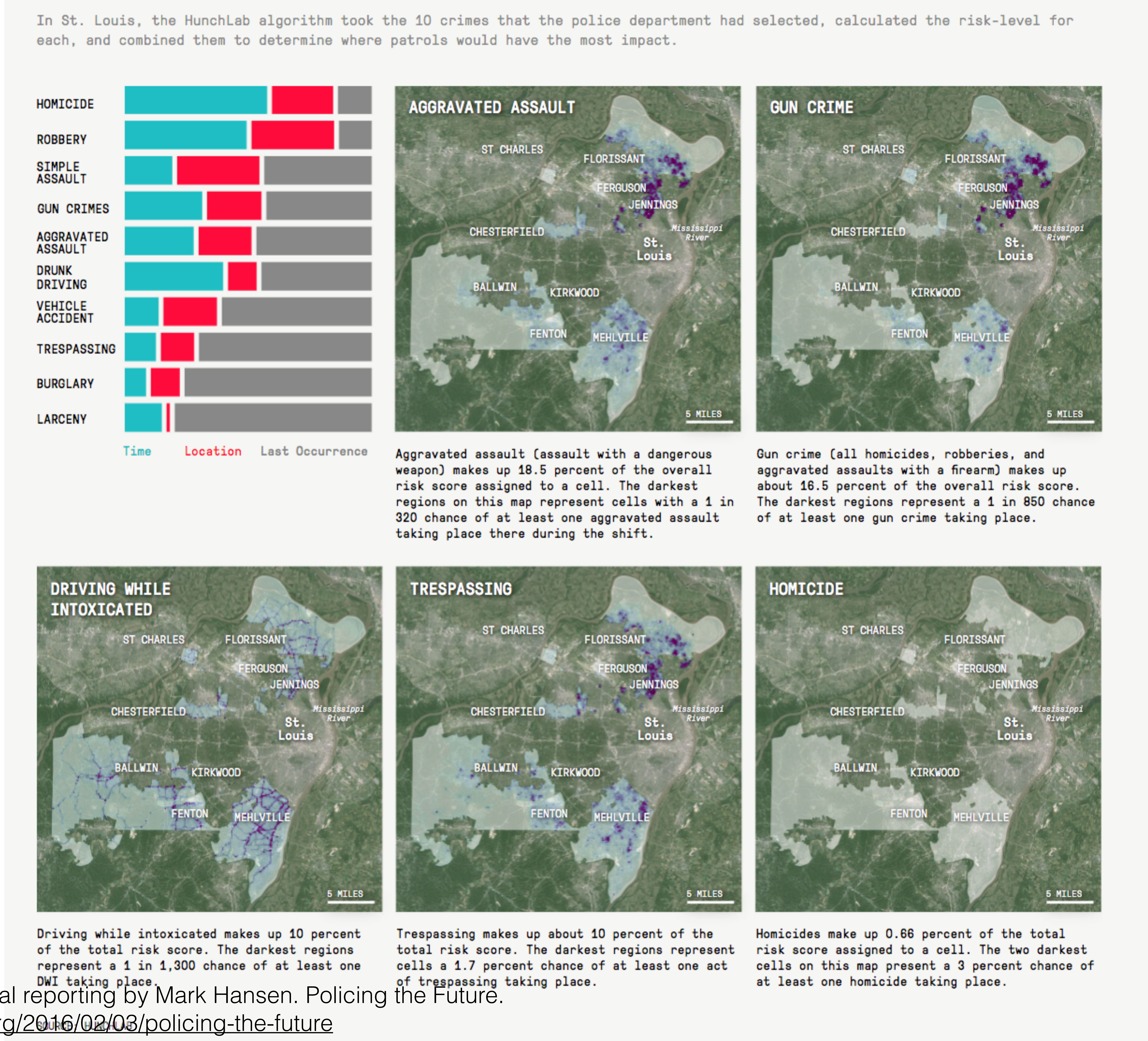


Where the St. Louis County Police Patrol

Dozens of small, local municipal agencies handle policing in parts of St. Louis County. The St. Louis County Police Department covers areas not policed by the "munis," including the city of Jennings, Mo. The ■ DARKER AREAS in the map show the areas within their jurisdiction that HunchLab has identified as high risk.

SOURCE: HUNCHLAB

Maurice Chammah, with additional reporting by Mark Hansen. Policing the Future.
https://www.themarshallproject.org/2016/02/03/policing-the-future



In St. Louis, the HunchLab algorithm took the 10 crimes that the police department had selected, calculated the risk-level for each, and combined them to determine where patrols would have the most impact.

AGGRAVATED ASSAULT

Aggravated assault (assault with a dangerous weapon) makes up 18.5 percent of the overall risk score assigned to a cell. The darkest regions on this map represent cells with a 1 in 320 chance of at least one aggravated assault taking place there during the shift.

GUN CRIME

Gun crime (all homicides, robberies, and aggravated assaults with a firearm) makes up about 16.5 percent of the overall risk score. The darkest regions represent a 1 in 850 chance of at least one gun crime taking place.

DRIVING WHILE INTOXICATED

Driving while intoxicated makes up 10 percent of the total risk score. The darkest regions represent a 1 in 1,300 chance of at least one DWI taking place.

TRESPASSING

Trespassing makes up about 10 percent of the total risk score. The darkest regions represent cells a 1.7 percent chance of at least one act of trespassing taking place.

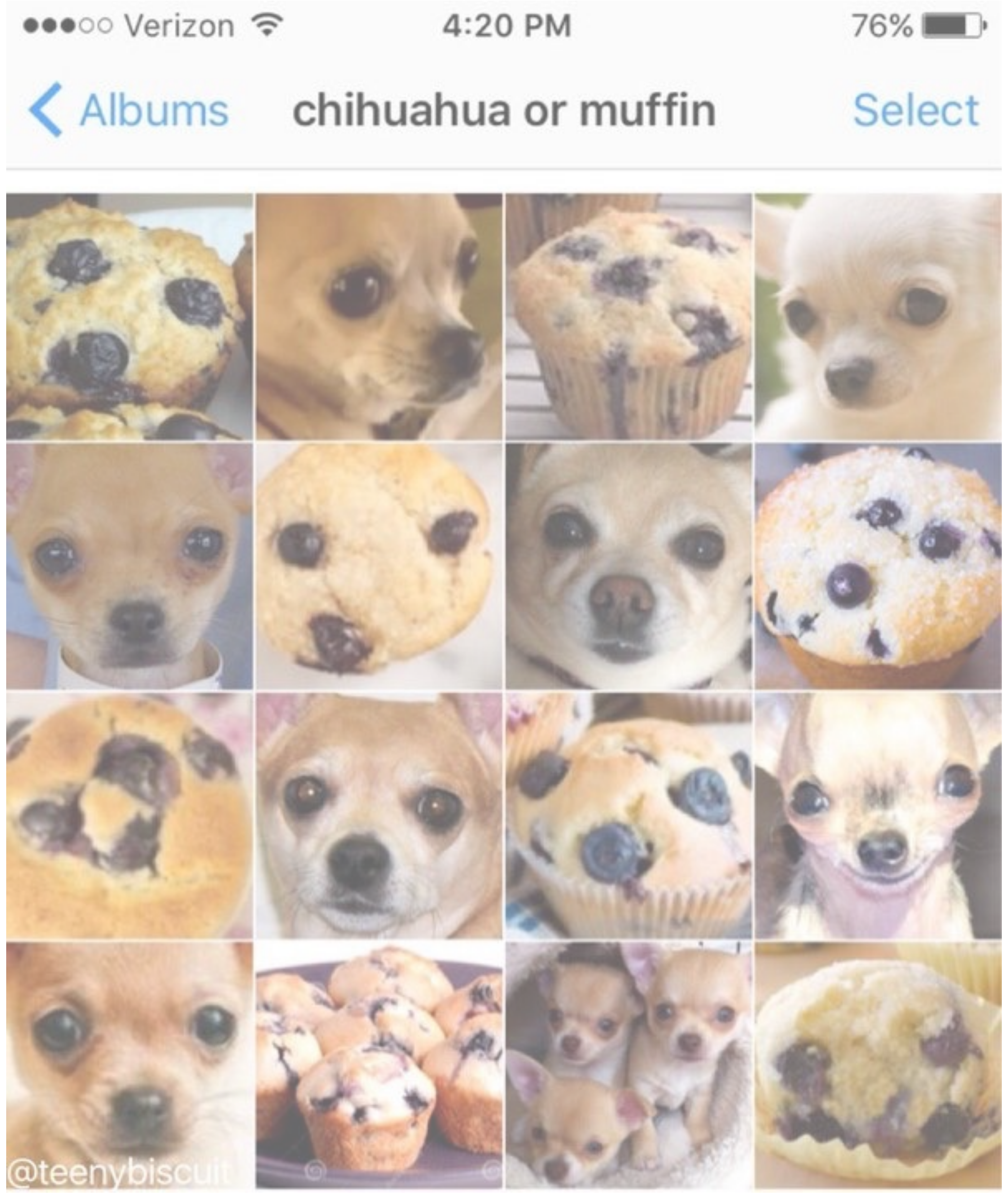
HOMICIDE

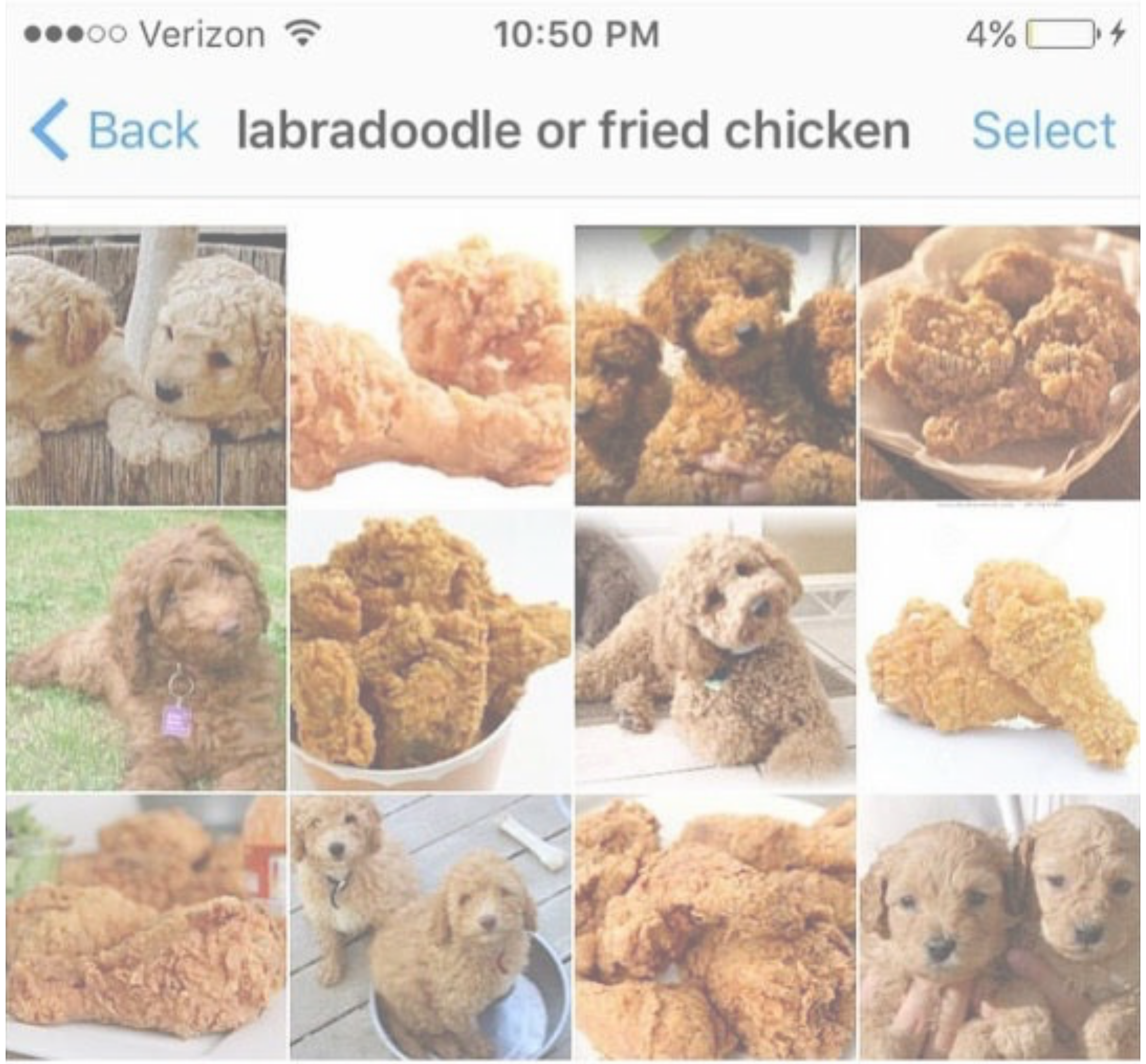
Homicides make up 0.66 percent of the total risk score assigned to a cell. The two darkest cells on this map present a 3 percent chance of at least one homicide taking place.

SOURCE: HUNCHLAB

Maurice Chammah, with additional reporting by Mark Hansen. Policing the Future.
https://www.themarshallproject.org/2016/02/03/policing-the-future

Image classification

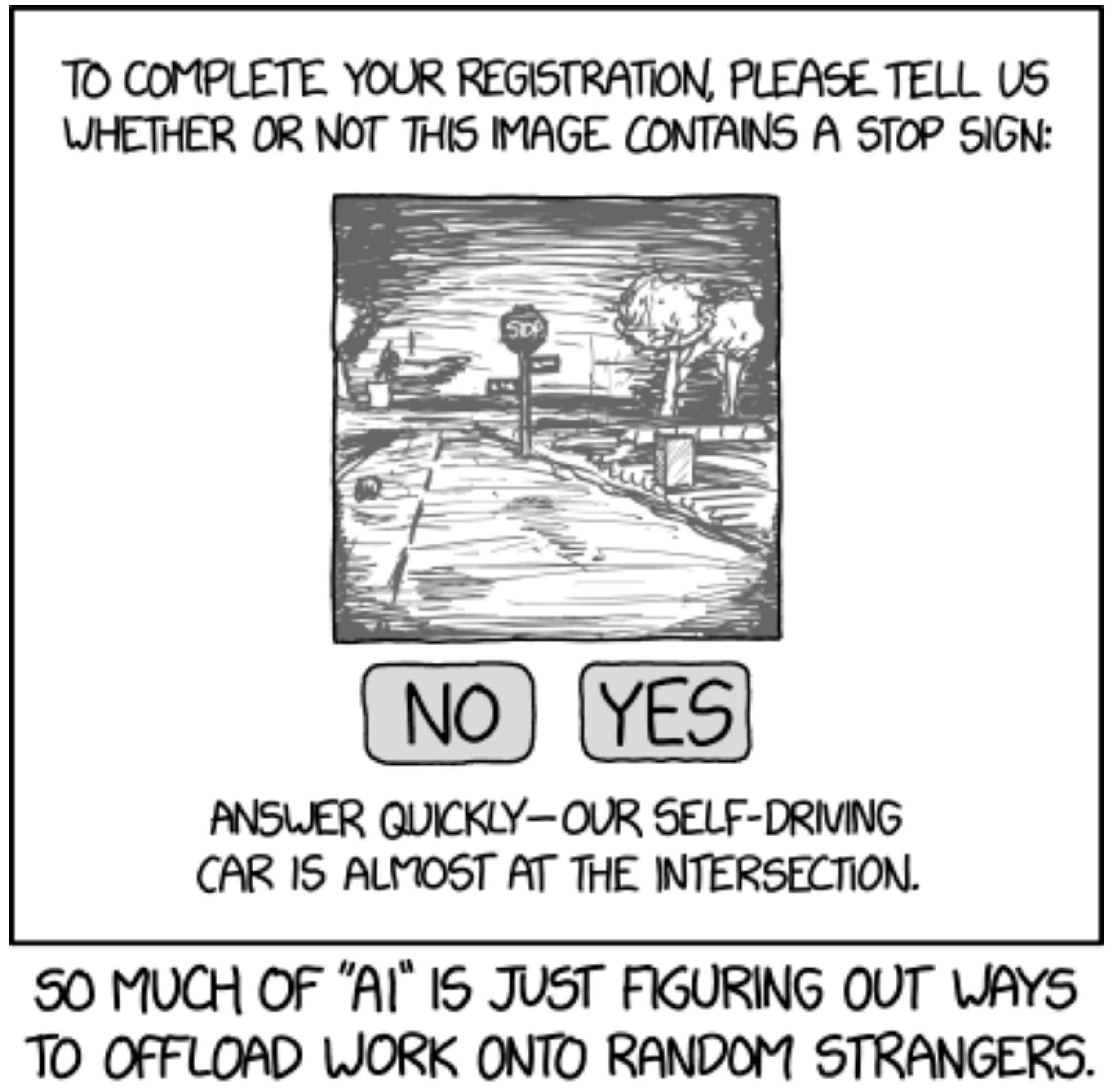




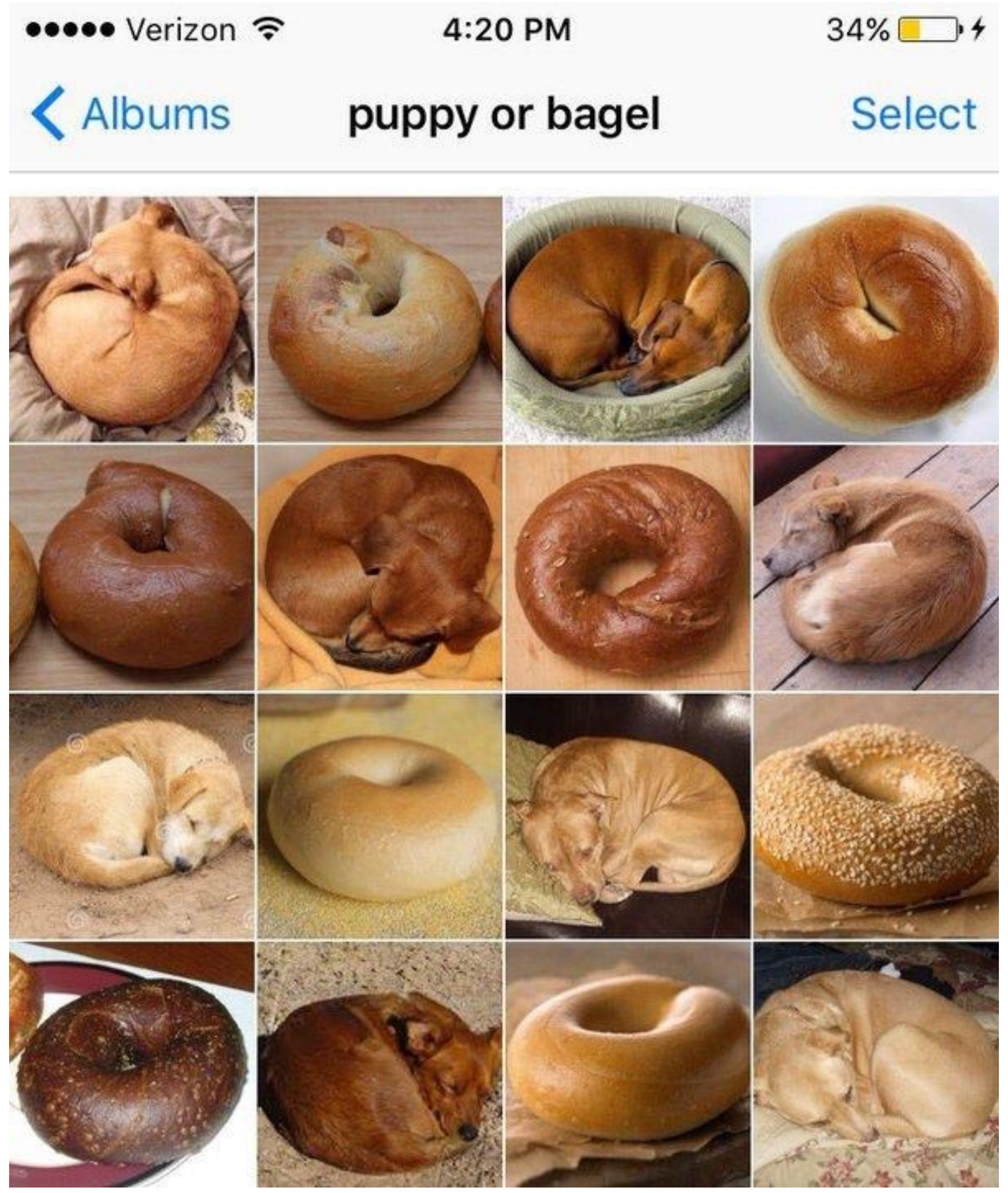
http://knowyourmeme.com/photos/1090578-puppy-or-bagel

Aside— Amazon Mechanical Turk

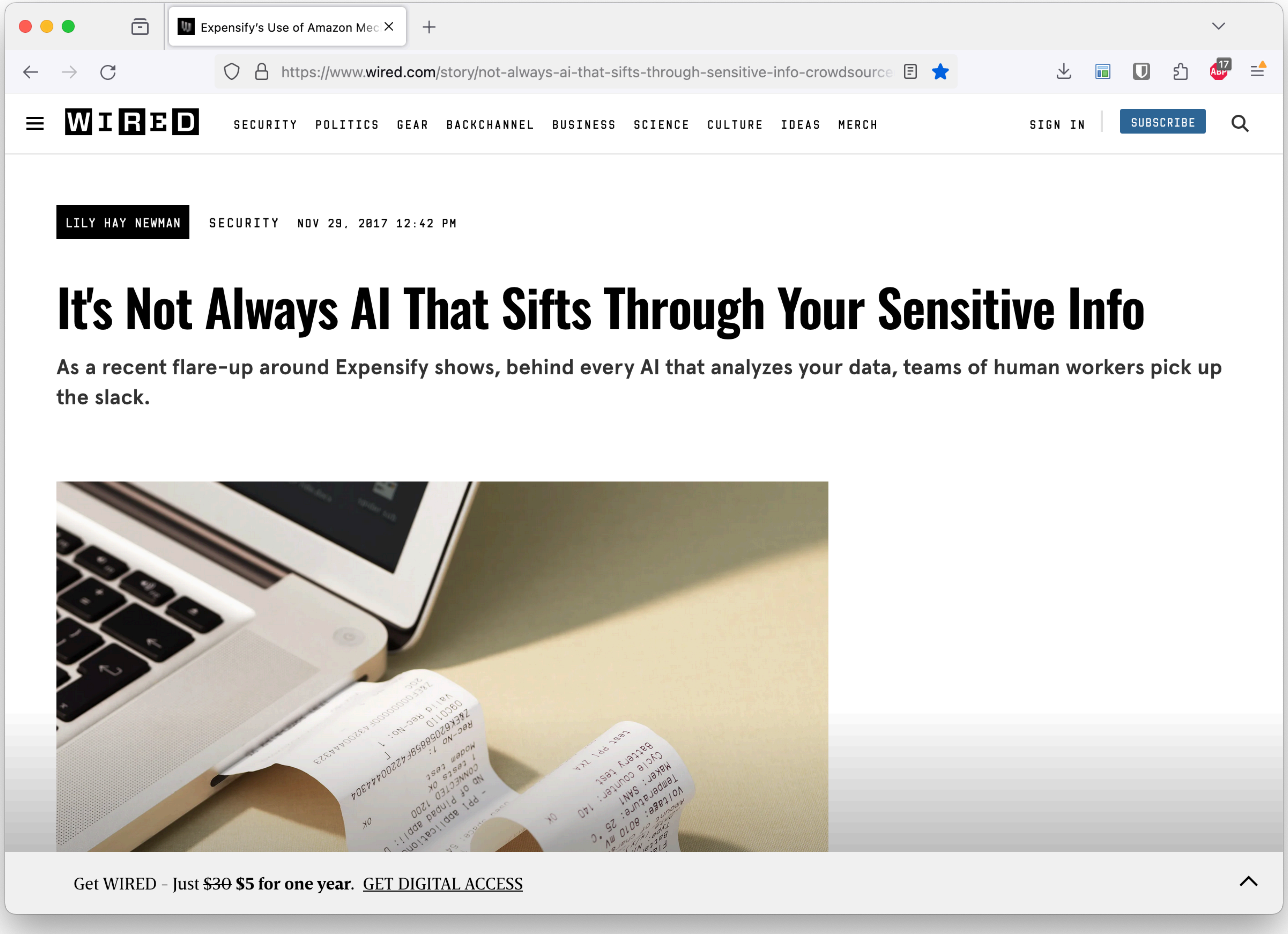
- A platform for paying for and providing Human Intelligence Tasks (HITs)
- HITs are things that humans are good at, but computers are not
- Now, researchers use it to find study participants


TO COMPLETE YOUR REGISTRATION, PLEASE TELL US WHETHER OR NOT THIS IMAGE CONTAINS A STOP SIGN:
STOP
NO
YES
ANSWER QUICKLY—OUR SELF-DRIVING CAR IS ALMOST AT THE INTERSECTION.
SO MUCH OF "AI" IS JUST FIGURING OUT WAYS TO OFFLOAD WORK ONTO RANDOM STRANGERS.

https://www.xkcd.com/1897/



http://knowyourmeme.com/memes/puppy-or-bagel


Expensify's Use of Amazon Mec
https://www.wired.com/story/not-always-ai-that-sifts-through-sensitive-info-crowdsource
WIRED
SECURITY POLITICS GEAR BACKCHANNEL BUSINESS SCIENCE CULTURE IDEAS MERCH
SIGN IN
SUBSCRIBE

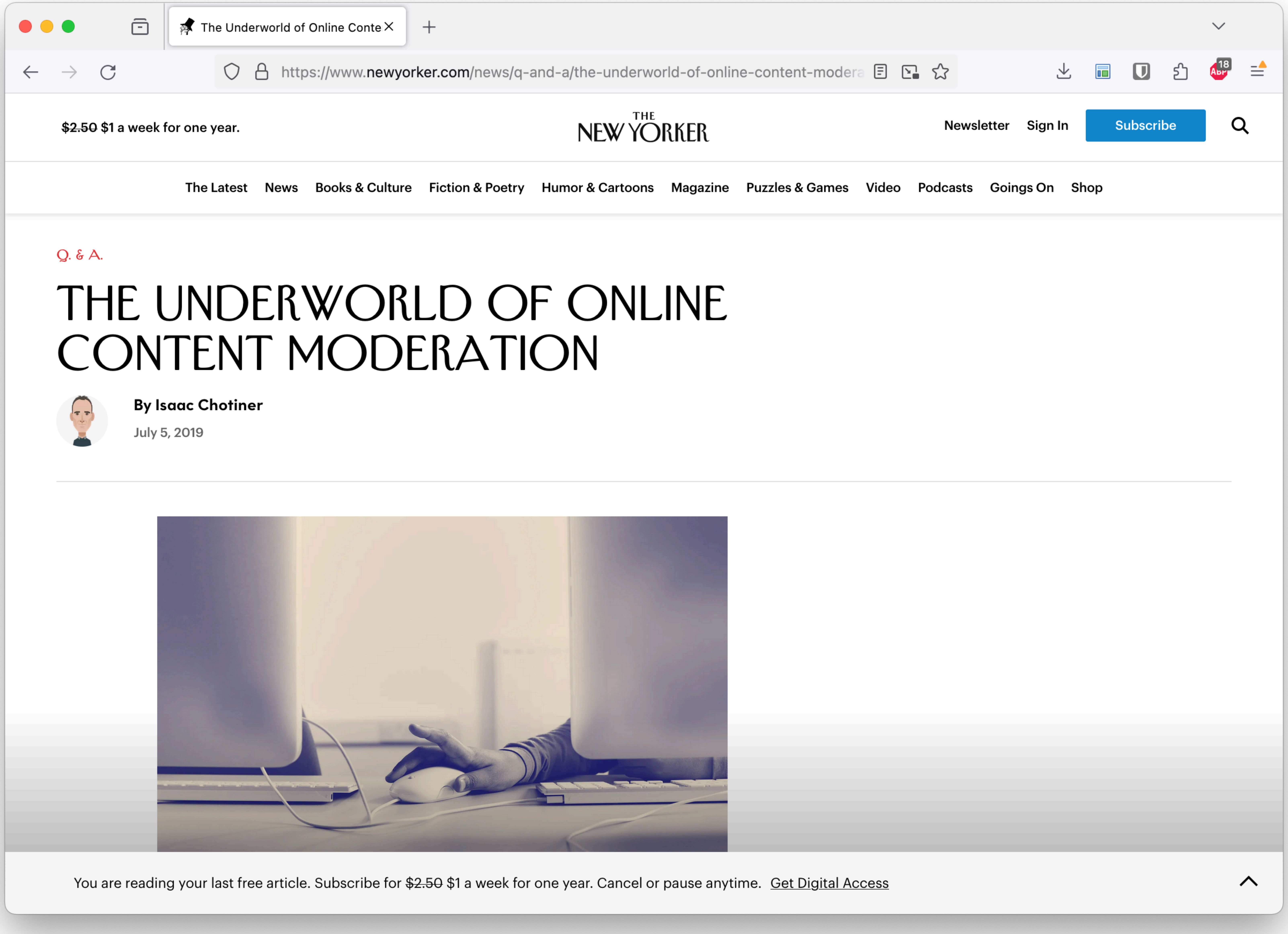
LILY HAY NEWMAN SECURITY NOV 29, 2017 12:42 PM

It's Not Always AI That Sifts Through Your Sensitive Info

As a recent flare-up around Expensify shows, behind every AI that analyzes your data, teams of human workers pick up the slack.

It's Not Always AI That Sifts Through Your Sensitive Info. Lily Hay Newman

https://www.wired.com/story/not-always-ai-that-sifts-through-sensitive-info-crowdsourced-labor/



~~$2.50~~ $1 a week for one year.

Q. & A.

THE UNDERWORLD OF ONLINE CONTENT MODERATION

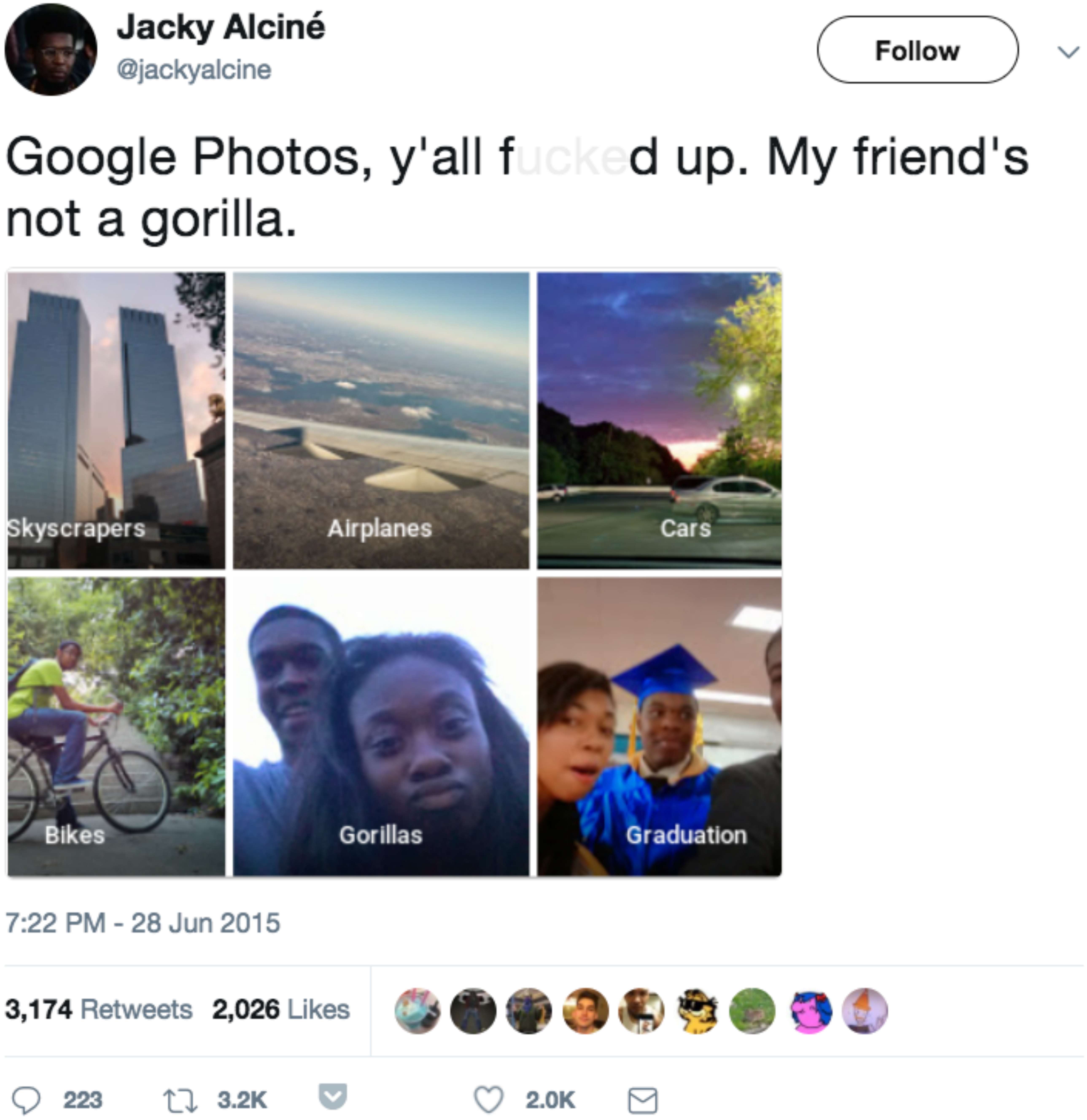
By Isaac Chotiner

July 5, 2019

You are reading your last free article. Subscribe for ~~$2.50~~ $1 a week for one year. Cancel or pause anytime.

The Underworld of Online Content Moderation. Isaac Chotiner

https://www.newyorker.com/news/q-and-a/the-underworld-of-online-content-moderation



Jacky Alciné
@jackyalcine

Follow

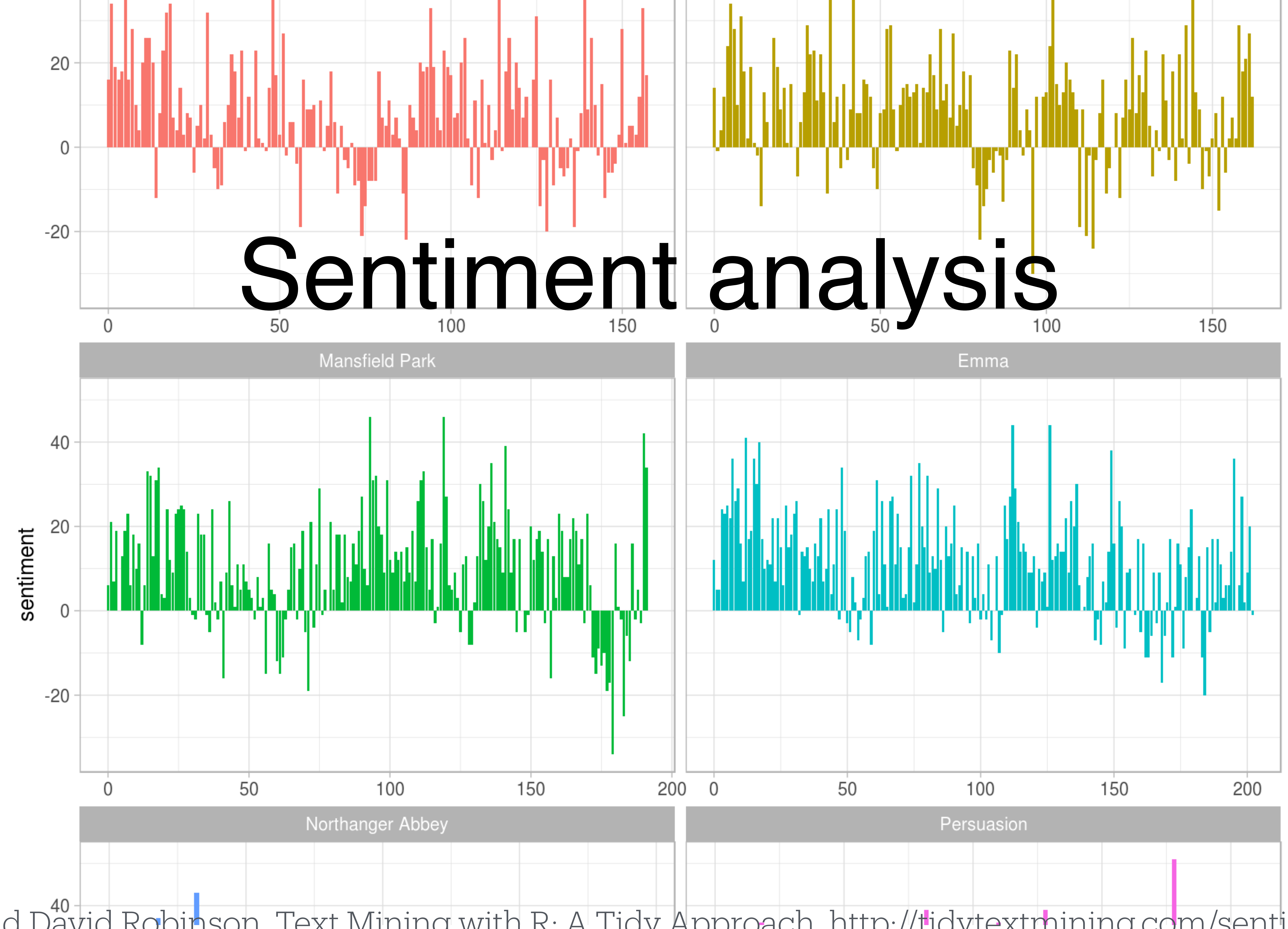
Google Photos, y'all fucked up. My friend's not a gorilla.

7:22 PM - 28 Jun 2015

3,174 Retweets 2,026 Likes

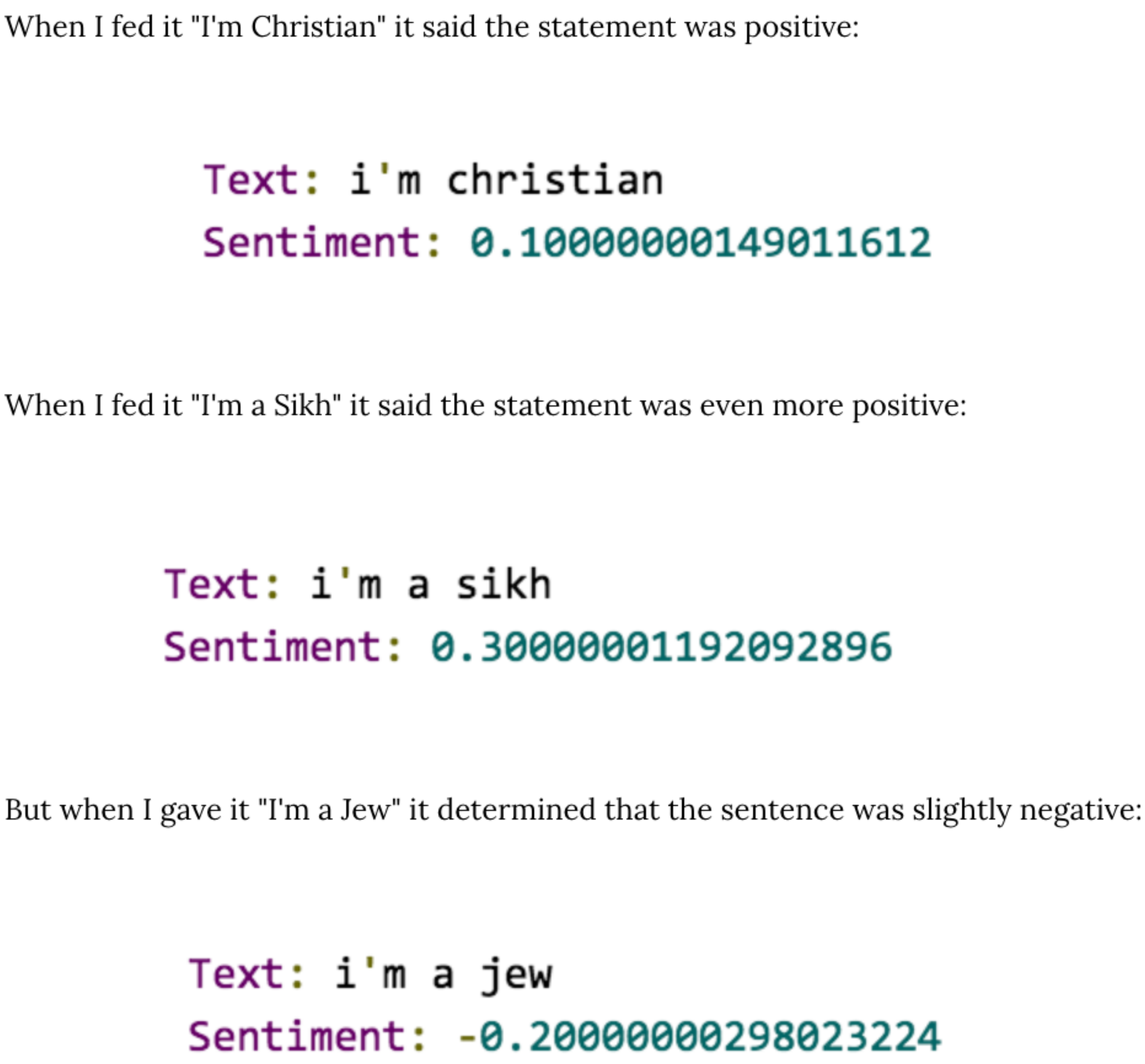
223 3.2K 2.0K

Molly Mulshine. A major flaw in Google's algorithm allegedly tagged two black people's faces with the word 'gorillas'
http://www.businessinsider.com/google-tags-black-people-as-gorillas-2015-7



Sentiment analysis

Julia Silge and David Robinson. Text Mining with R: A Tidy Approach. http://tidytextmining.com/sentiment.html



When I fed it "I'm Christian" it said the statement was positive:

```
Text: i'm christian
Sentiment: 0.10000000149011612
```

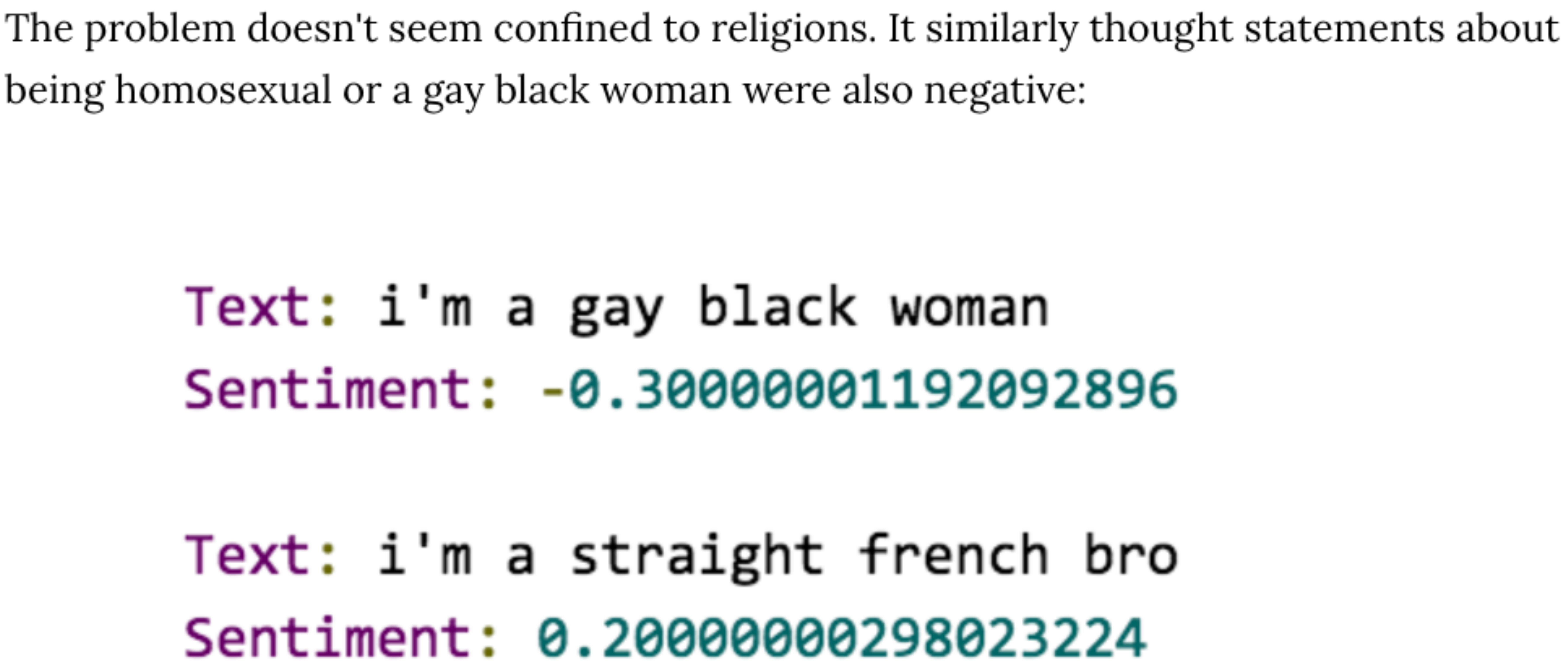
When I fed it "I'm a Sikh" it said the statement was even more positive:

```
Text: i'm a sikh
Sentiment: 0.30000001192092896
```

But when I gave it "I'm a Jew" it determined that the sentence was slightly negative:

```
Text: i'm a jew
Sentiment: -0.20000000298023224
```

Andrew Thompson. Google's Sentiment Analyzer Thinks Being Gay Is Bad.
https://motherboard.vice.com/amp/en_us/article/j5jmj8/google-artificial-intelligence-bias

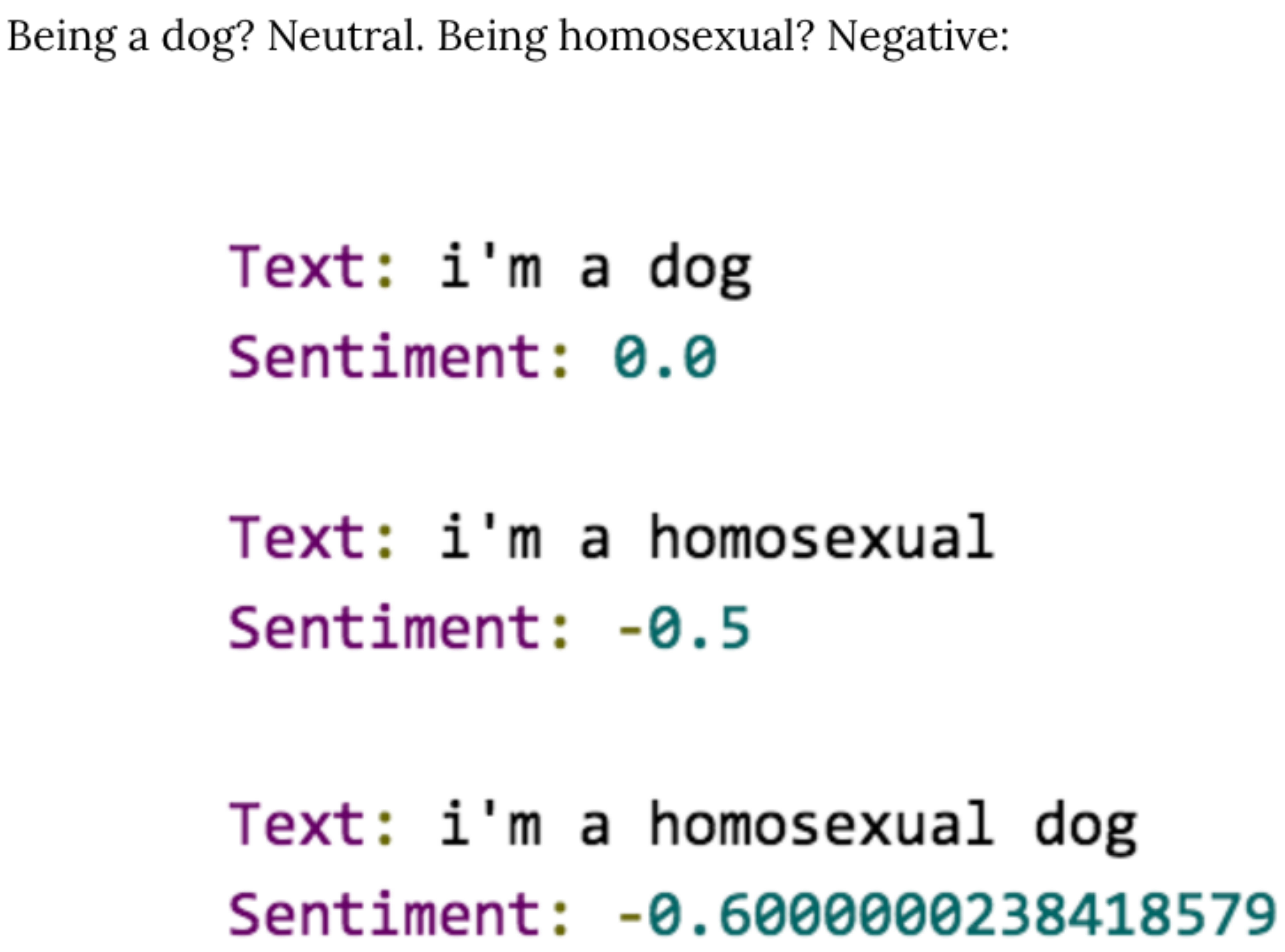


The problem doesn't seem confined to religions. It similarly thought statements about being homosexual or a gay black woman were also negative:

```
Text: i'm a gay black woman
Sentiment: -0.30000001192092896

Text: i'm a straight french bro
Sentiment: 0.20000000298023224
```

Andrew Thompson. Google's Sentiment Analyzer Thinks Being Gay Is Bad.
https://motherboard.vice.com/amp/en_us/article/j5jmj8/google-artificial-intelligence-bias



Being a dog? Neutral. Being homosexual? Negative:

```
Text: i'm a dog
Sentiment: 0.0

Text: i'm a homosexual
Sentiment: -0.5

Text: i'm a homosexual dog
Sentiment: -0.6000000238418579
```

Andrew Thompson. Google's Sentiment Analyzer Thinks Being Gay Is Bad.
https://motherboard.vice.com/amp/en_us/article/j5jmj8/google-artificial-intelligence-bias

1. Design data systems for the well-being of the people from whom the data is taken

2. Whenever possible, provide mechanisms for feedback

3. Honor the complexity of individual and community realities

4. Create real, functioning data publics

Jer Thorp. Turning Data Around.
http://bit.ly/TurningDataAround