

A bit of thinking about data ethics

K Arnold, based on DSBox and others

Logistics

- Project rubric, details, and logistics posted!
- *Unofficial* milestone: make some plot using your data before Thanksgiving

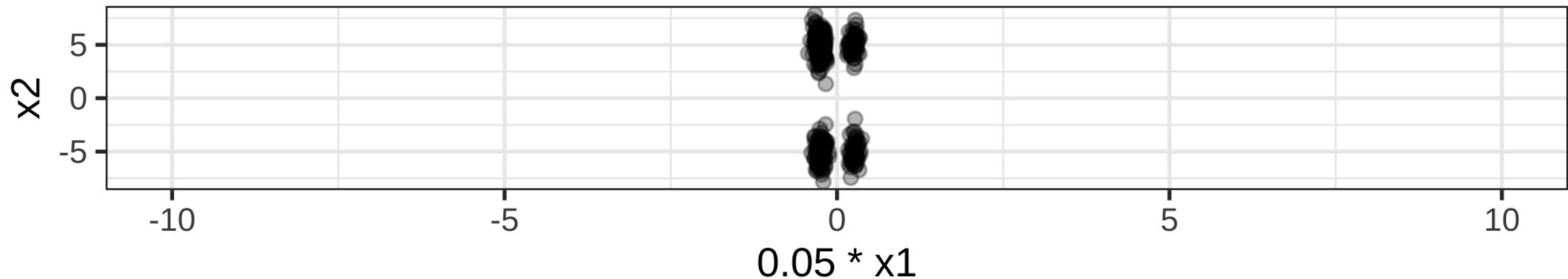
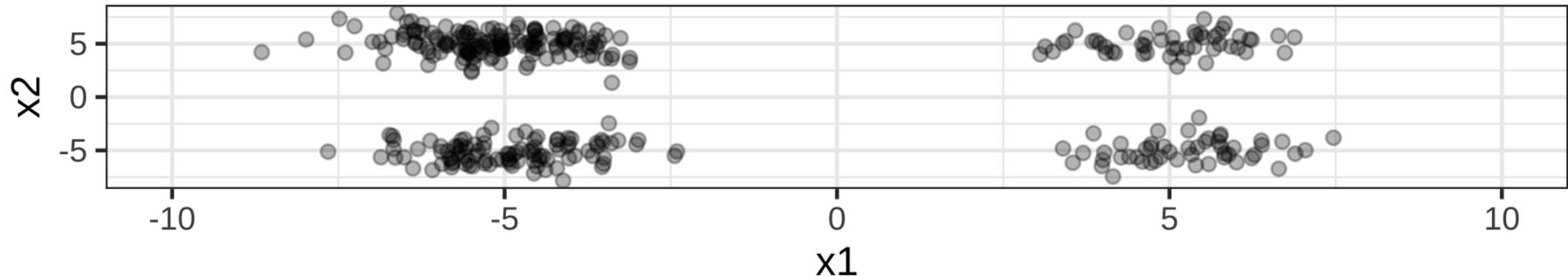
Q&A / Review

| Why did we use such a huge `set.seed` last time?

Today's date!

Q&A / Review

| Why did a bigger `step_range` make the feature more prominent?



Rest of the semester

- Advanced models (random forests, neural nets, ...)
- SQL
- Text
- Techniques for Communication

Today: A Series of Mini Discussions

Preview of questions:

- What characterizes an ethical data scientist or analyst?
 - What are some ways that we have practiced *humility* in this course?
 - What are some ways that we have practiced *integrity*?
 - What are some ways that we have practiced *compassion and justice*?
-
- What are some things we could do that would *compromise* our integrity?
 - What are some ways that data-driven systems can *help* or *hinder* truth?
 - What are some ways that data systems can help *perpetuate*, or *fight*, injustice?

What characterizes an ethical data scientist or analyst?

Post your thoughts in your Cohort channel. I will pick a random cohort.

02 : 00

Recall from Day 1...

Our Goals

- *Skill*: how to do these things
- *Knowledge*: understanding the underlying concepts
- *Character*: wisdom in practicing these skills

Character??

DS is a lens. How can we see rightly through it? Some areas:

- Humility
- Integrity / Honesty
- Hospitality
- Compassion and justice

Humility

Challenge: data feels powerful, people listen to what you use it to say.

What are some ways that we have practiced humility?

02 : 00

Finally, brothers and sisters, whatever is true,
whatever is noble, whatever is right, whatever is
pure, whatever is lovely, whatever is admirable—if
anything is excellent or praiseworthy—think about
such things.

Philippians 4:8

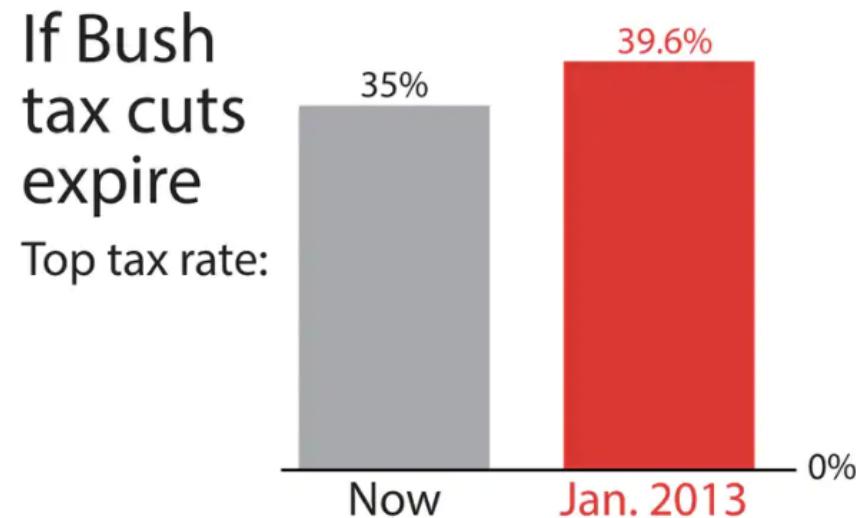
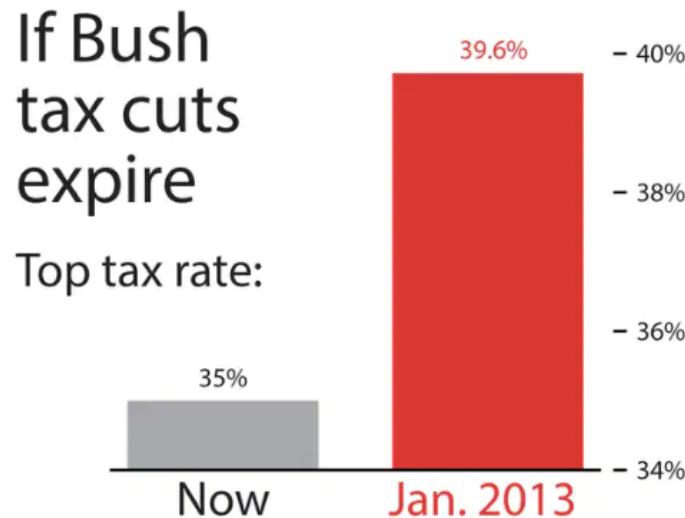
Integrity

It's tempting to say something that isn't entirely true, or to manipulate the collection/analysis/reporting process to yield the answer you want

What are some ways that we have practiced integrity?

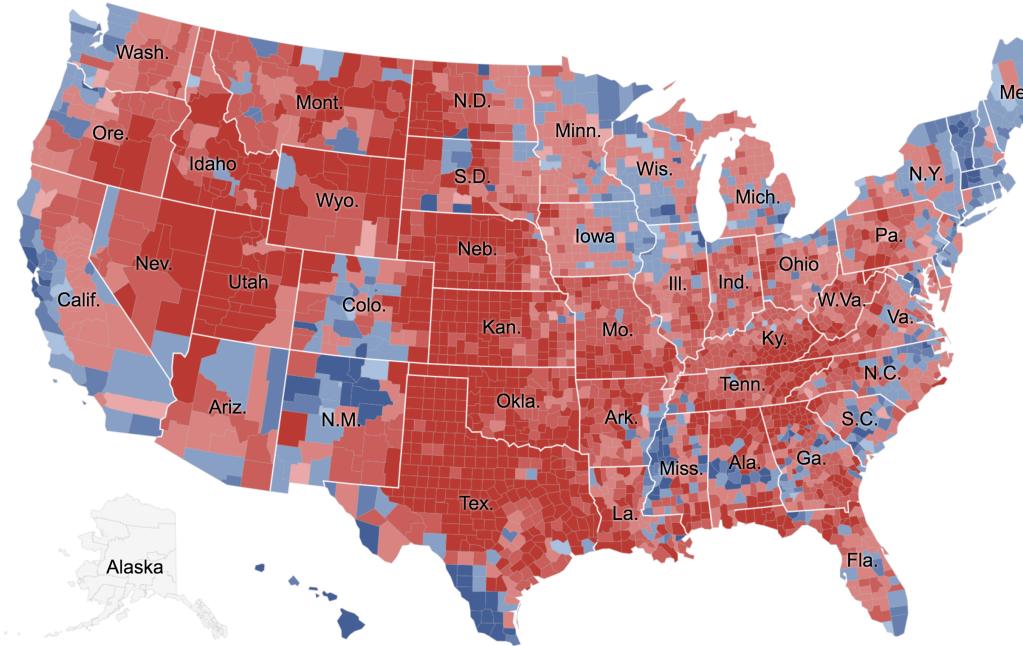
What are some things we could do that would *compromise* our integrity?

03 : 00

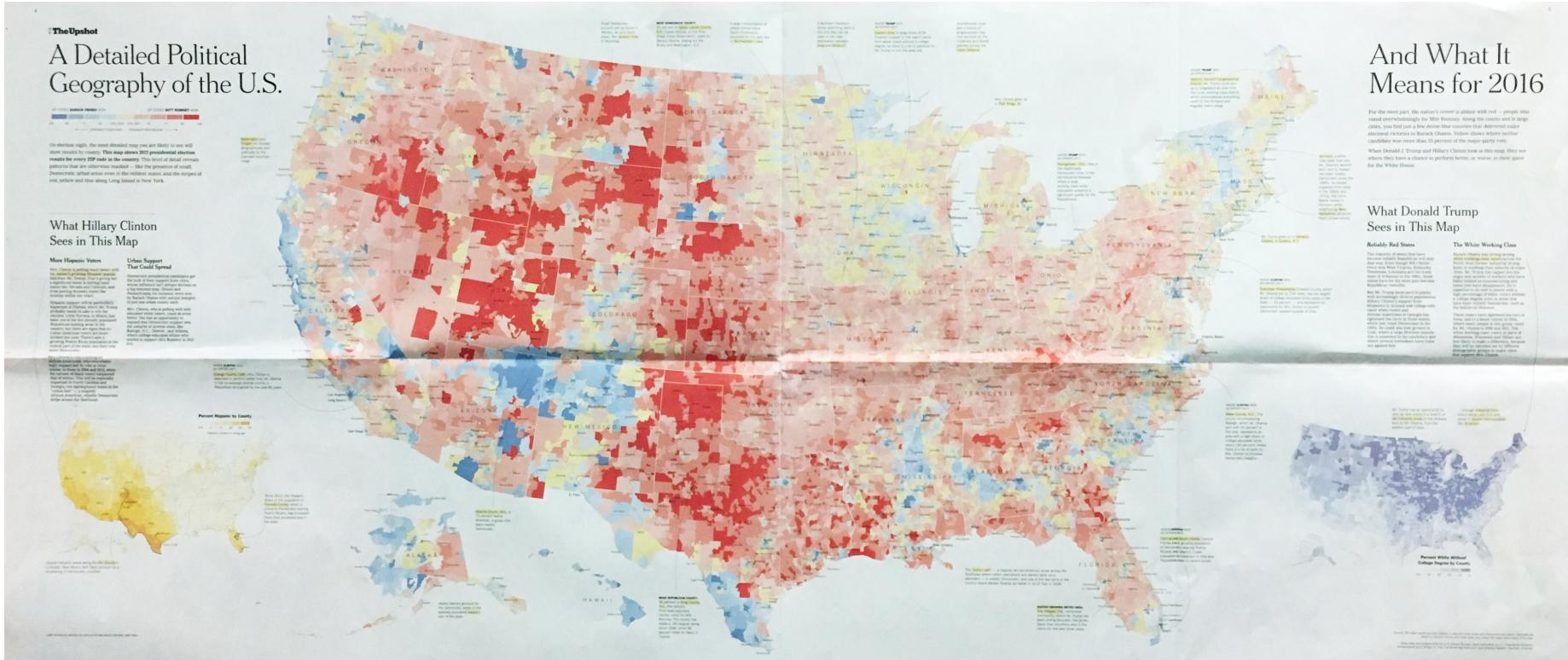


Ingraham, C. (2019) "You've been reading charts wrong. Here's how a pro does it.", The Washington Post, 14 Oct.

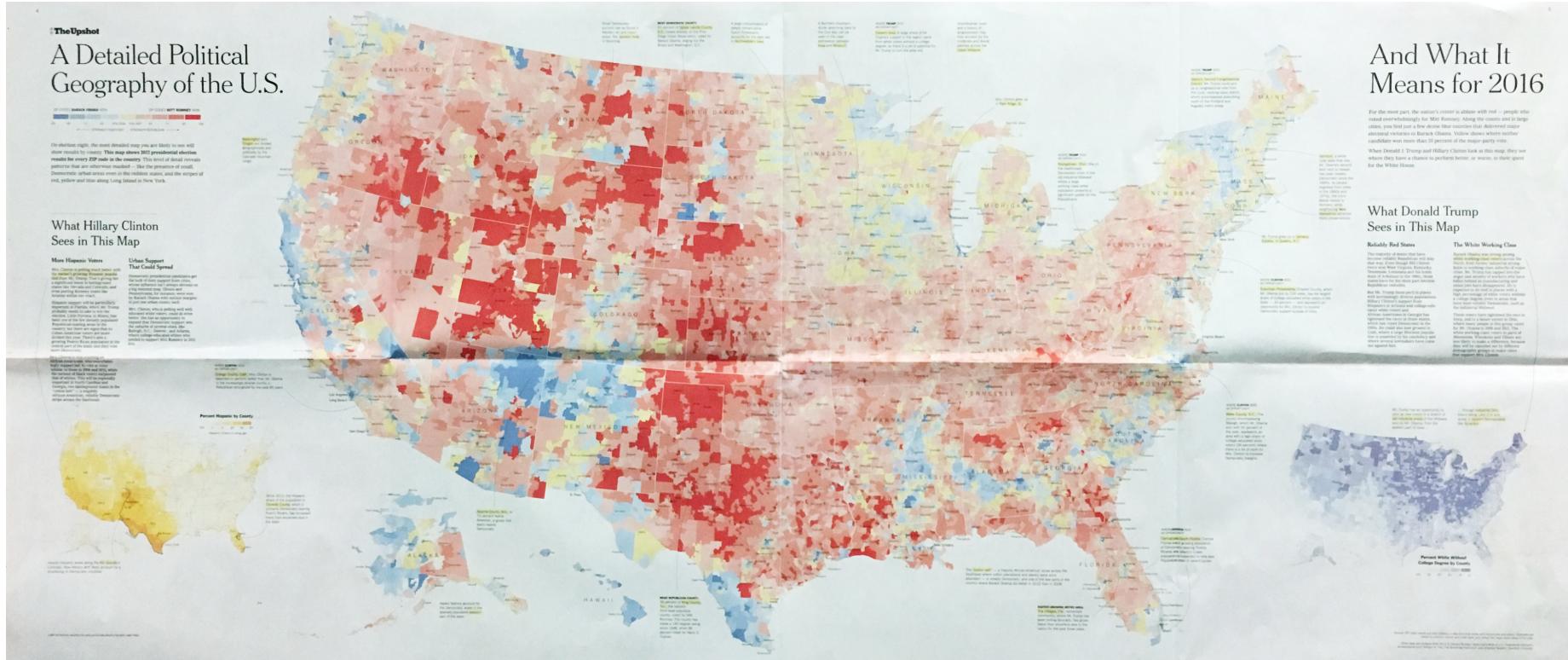
Coloring Maps



Coloring Maps



Coloring Maps



See also: Gamio, L. (2016) "Election maps are telling you big lies about small things", The Washington Post, 1 Nov.

Truth and Falsehood

What are some ways that data-driven systems can *help* or *hinder* truth?

02 : 00

Recommender Systems and Information

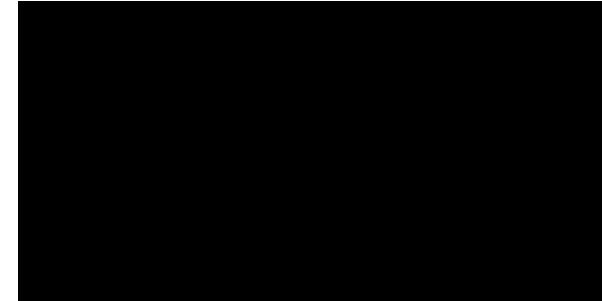
- Why does YouTube recommend some videos rather than others?
- What is YouTube optimizing for? (what brings them money?)

Recommender Systems and Information

- Why does YouTube recommend some videos rather than others?
- What is YouTube optimizing for? (what brings them money?)
- **Time on YouTube**
 - | Any smart AI that optimizes engagement with itself will have a tendency to discourage engagement to other channels.
 - ... which *may* mean that recommendations promote videos that *amplify resentment against other media*.

Guillaume Chaslot, How Algorithms Can Learn to Discredit the Media

Deepfakes



Hospitality

We can choose to use our tools to elucidate and clarify, rather than obscure.
So we will practice:

- Clear visual communication
- Clarity of code and process
- Writing explanations that are accessible and appropriate to audience.

Learn to do right; seek justice.
Defend the oppressed.
Take up the cause of the fatherless;
plead the case of the widow.

Isaiah 1:17 (NIV)

Compassion and Justice

Data Science can both cause harm and reveal it.

So we will:

- Study examples of how data might cause harm
- Study examples of how harm might be mitigated or revealed

What are some ways that data systems can help *perpetuate*, or *fight*, injustice?

02 : 00

Compassion and Justice

Data Science can both cause harm and reveal it.

So we will:

- Study examples of how data might cause harm
- Study examples of how harm might be mitigated or revealed

What are some ways that data systems can help *perpetuate*, or *fight*, injustice?

Examples

- Fairness (vs algorithmic bias)
- Transparency and Accountability
- Data Stewardship
- Environmental impact

02 : 00

Algorithmic bias and gender

Google Translate

The screenshot shows the Google Translate interface with the source language set to Turkish and the target language set to English. A dropdown menu is open on the left, listing various occupation terms in Turkish, each followed by its English translation. The interface includes a toolbar at the top with language selection buttons and a 'Translate' button.

Turkish	English
o bir asker	he is a soldier
o bir öğretmen	She's a teacher
O bir doktor	He is a doctor
o bir hemşire	she is a nurse
o bir yazar	he is a writer
o bir köpek	he is a dog
o bir dadi	she is a nanny
o bir kedi	it is a cat
o bir rektör	he is a rector
o bir başkanı	he is a president
o bir girişimci	he is an entrepreneur
o bir Şarkıcı	she is a singer
o bir Öğrenci	he is a student
o bir Tercüman	he is a translator
o çalışkan	he is hard working
o tembel	she is lazy
o bir ressam	he is a painter
o bir kuaför	he is a hairdresser
o bir garson	he is a waiter
O bir mühendis	He is an engineer
o bir mimar	he is an architect
o bir sanatçı	he is an artist

Amazon's experimental hiring algorithm

- Used AI to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said
- Company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way
- Amazon's system taught itself that male candidates were preferable

Gender bias was not the only issue. Problems with the data that underpinned the models' judgments meant that unqualified candidates were often recommended for all manner of jobs, the people said.

Dastin, J. (2018) [Amazon scraps secret AI recruiting tool that showed bias against women](#), Reuters, 10 Oct.

Algorithmic bias and race



Interview

'A white mask worked better': why algorithms are not colour blind

Ian Tucker

When Joy Buolamwini found that a robot recognised her face better when she wore a white mask, she knew a problem needed fixing

Sun 28 May 2017 13.27 BST

Joy Buolamwini is a graduate researcher at the MIT Media Lab and founder of the Algorithmic Justice League – an organisation that aims to challenge the biases in decision-making software. She grew up in Mississippi, gained a Rhodes scholarship, and she is also a Fulbright fellow, an Astronaut scholar and a Google Anita Borg scholar. Earlier this year she won a \$50,000 scholarship funded by the makers of the film *Hidden Figures* for her work fighting coded discrimination.

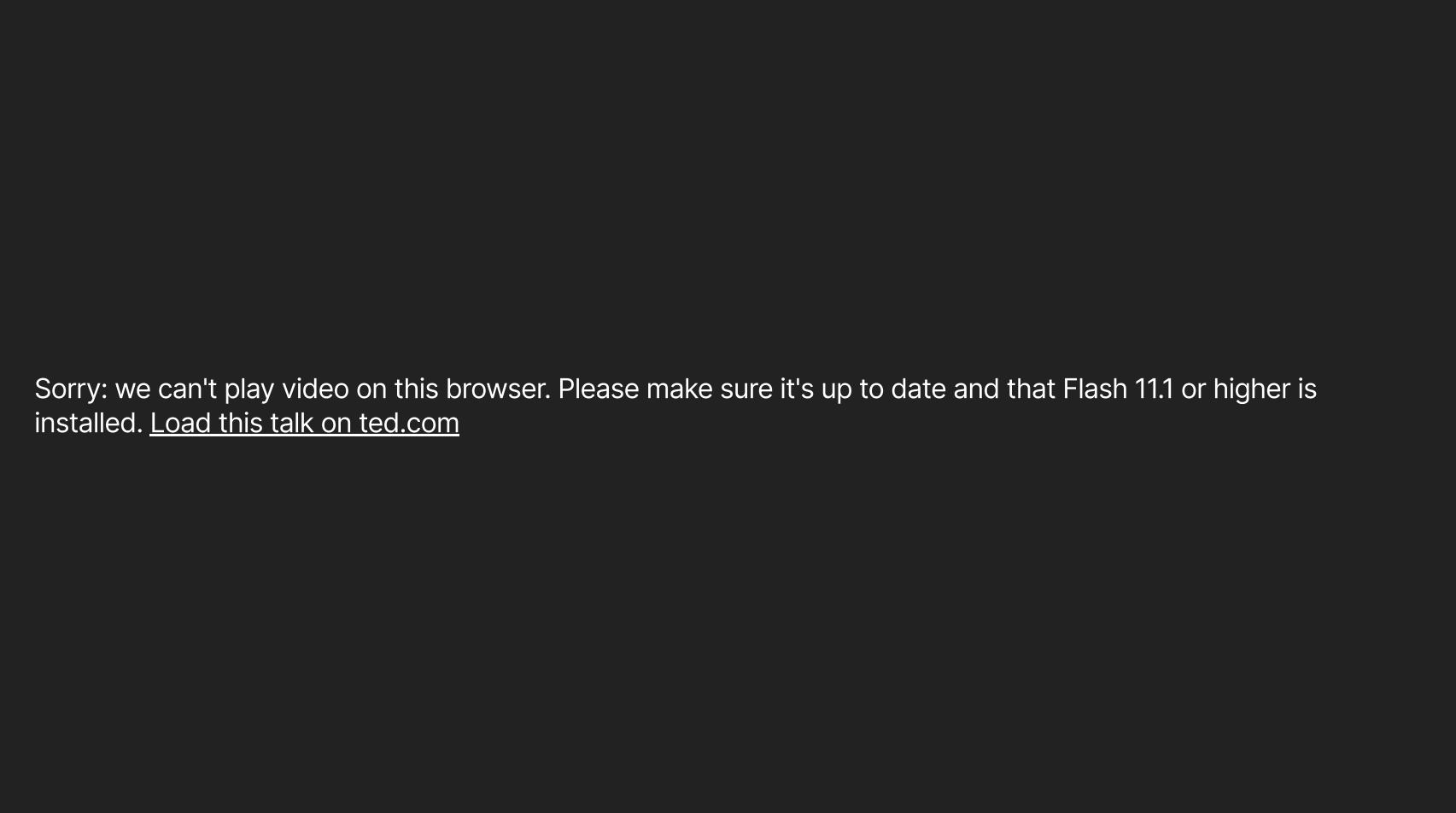
A lot of your work concerns facial recognition technology. How did you become interested in that area?

When I was a computer science undergraduate I was working on social robotics – the robots use computer vision to detect the humans they socialise with. I discovered I had a hard time being detected by the robot compared to lighter-skinned people. At the time I thought this was a one-off thing and that people would fix this.

'A white mask worked better': why algorithms are not colour blind

by Ian Tucker

Further watching



Sorry: we can't play video on this browser. Please make sure it's up to date and that Flash 11.1 or higher is installed. [Load this talk on ted.com](#)

U.S. NEWS

California may replace cash bail with algorithms – but some worry that will be less fair

A fight over replacing bail with "risk assessment tools" has split reform advocates. Some fear the change will worsen anti-Black discrimination.

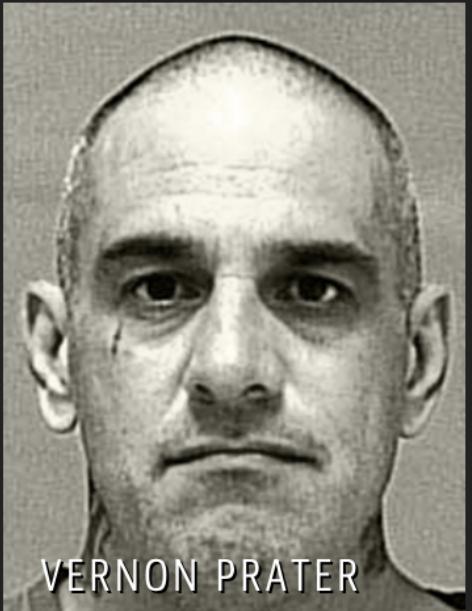


There's software used across the country to predict future criminals. And it's biased against blacks.

propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, May 23, 2016

A tale of two convicts

Two Petty Theft Arrests



VERNON PRATER

LOW RISK

3



BRISHA BORDEN

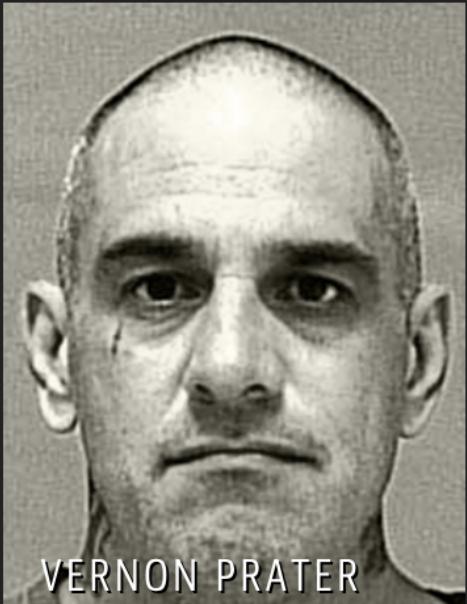
HIGH RISK

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.

A tale of two convicts

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.

Two Petty Theft Arrests

VERNON PRATER

Prior Offenses

2 armed robberies, 1 attempted armed robbery

Subsequent Offenses

1 grand theft

LOW RISK

3

BRISHA BORDEN

Prior Offenses

4 juvenile misdemeanors

Subsequent Offenses

None

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.

“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.”
Then U.S. Attorney General Eric Holder (2014)

ProPublica analysis

Data:

Risk scores assigned to more than 7,000 people arrested in Broward County, Florida, in 2013 and 2014 + whether they were charged with new crimes over the next two years

ProPublica analysis

Results:

- 20% of those predicted to commit violent crimes actually did
- Algorithm had higher accuracy (61%) when full range of crimes taken into account (e.g. misdemeanors)

	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%

- Algorithm was more likely to falsely flag black defendants as future criminals, at almost twice the rate as white defendants
- White defendants were mislabeled as low risk more often than black defendants

Further reading

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

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, "That's my kid's stuff." Borden and her friend immediately dropped the bike and scooter and walked away.

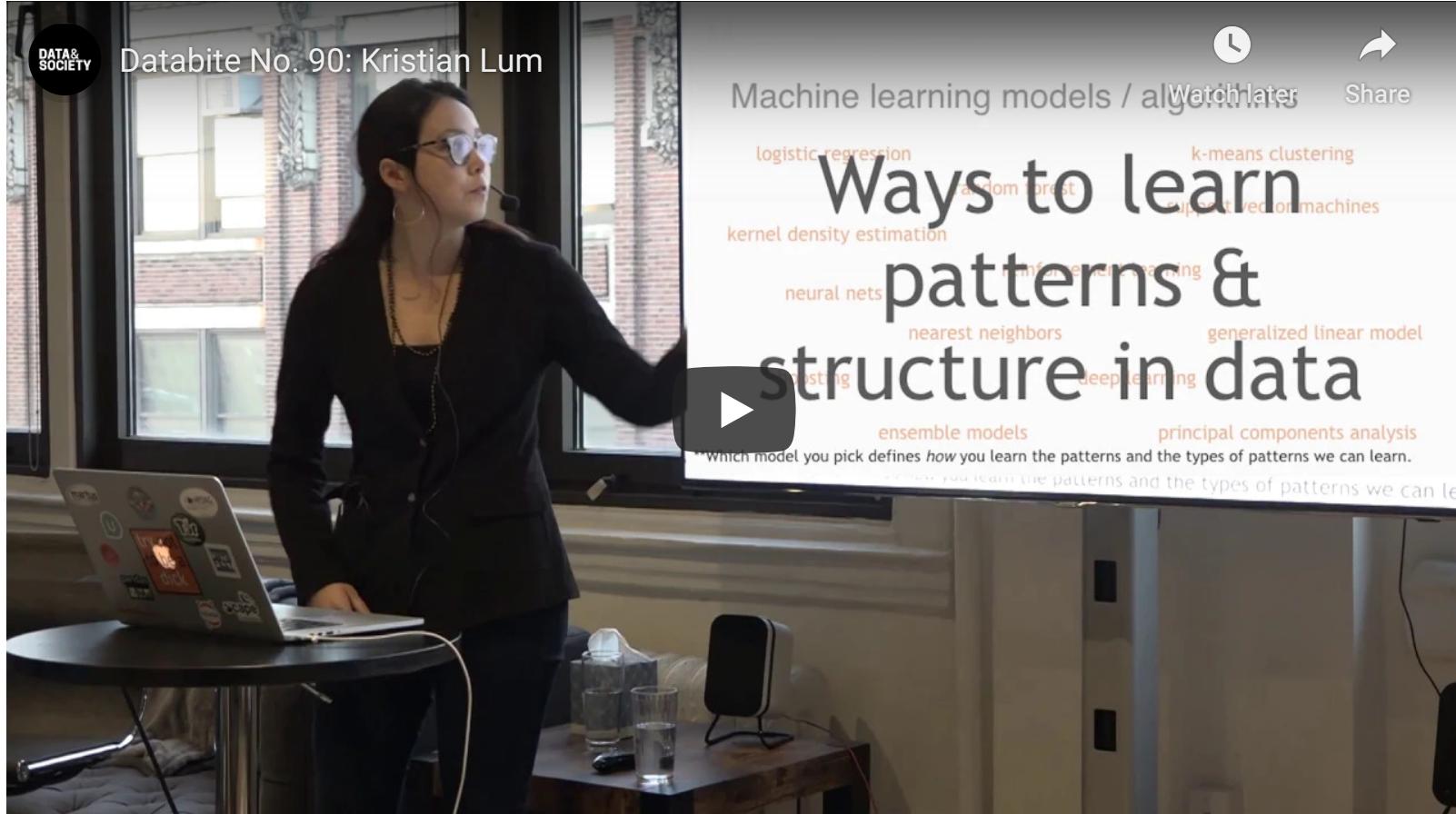
But it was too late — a neighbor who witnessed the heist had already called the police. Borden and her friend were arrested and charged with burglary and petty theft for the items, which were valued at a total of \$80.

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

Further watching



Predictive Policing: Bias In, Bias Out by Kristian Lum

How to make a racist AI without trying

How to make a racist AI without really trying

Robyn Speer — July 13, 2017



A cautionary tutorial.

Let's make a sentiment classifier!

Sentiment analysis is a very frequently-implemented task in NLP, and it's no surprise. Recognizing whether people are expressing positive or negative opinions about things has obvious business applications. It's used in social media monitoring, customer feedback, and even automatic stock trading (leading to bots that buy Berkshire Hathaway when Anne Hathaway gets a good movie review).

It's simplistic, sometimes too simplistic, but it's one of the easiest ways to get measurable results from NLP. In a few steps, you can put text in one end and get positive and negative scores out the

[Link to post](#)

How to write a racist AI in R without really trying

2018/09/27



Last year, Rob Speer wrote a really great post [How to make a racist AI without really trying](#). Go read it.

The idea is to do sentiment analysis with obvious, off-the-shelf tools. As the post says

So that's what we're going to do here, following the path of least resistance at every step, obtaining a classifier that should look very familiar to anyone involved in current NLP.

The original post used Python and I'm teaching an undergraduate data science course using R at the moment, so I wanted an R version. There were two issues in converting the code: my laptop doesn't really have enough memory for the data, and the model was fitted using scikitlearn.

The first problem can be fixed with `dbplyr`.

[Link to post](#)

Review

A company uses a machine learning algorithm to determine which job advertisement to display for users searching for technology jobs. Based on past results, the algorithm tends to display lower paying jobs for women than for men (after controlling for other characteristics than gender).

What ethical considerations might be considered when reviewing this algorithm?

Source: Modern Data Science with R, by Baumer, Kaplan, and Horton

Some Further Reading / Watching

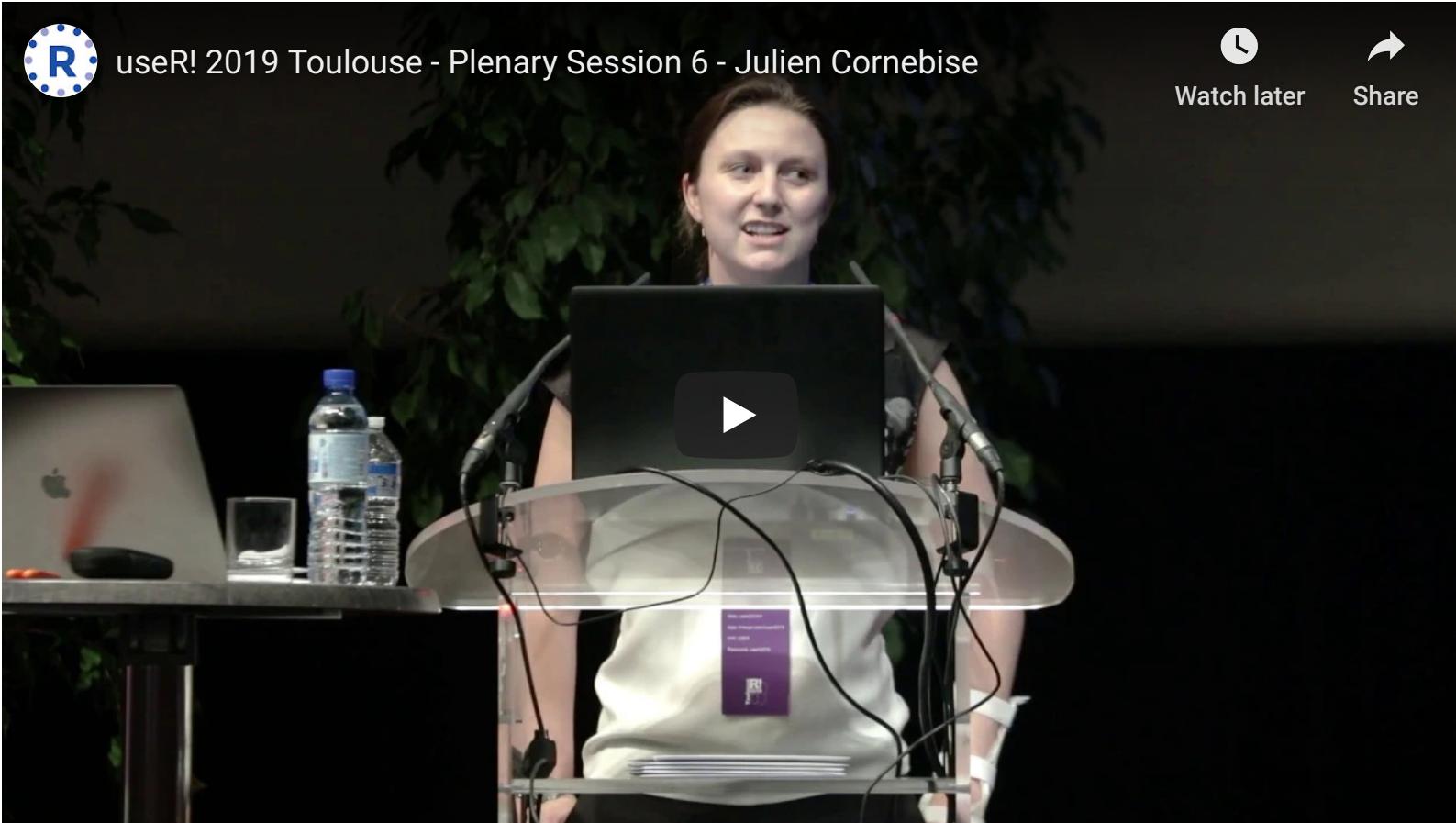
Overall

- Fast.AI Data Ethics course
- Ethics and Data Science
by Mike Loukides, Hilary Mason, DJ Patil
(Free Kindle download)
- Weapons of Math Destruction
How Big Data Increases Inequality and Threatens Democracy
by Cathy O'Neil

Integrity

- How Charts Lie
Getting Smarter about Visual Information
by Alberto Cairo
- How Deceptive are Deceptive Visualizations?
Pandey et al., CHI 2015

Further watching



AI for Good in the R and Python ecosystems
by Julien Cornebise

Parting thoughts

- At some point during your data science learning journey you will learn tools that can be used unethically
- You might also be tempted to use your knowledge in a way that is ethically questionable either because of business goals or for the pursuit of further knowledge (or because your boss told you to do so)

How do you train yourself to make the right decisions (or reduce the likelihood of accidentally making the wrong decisions) at those points?