Data science ethics

Prof. Maria Tackett

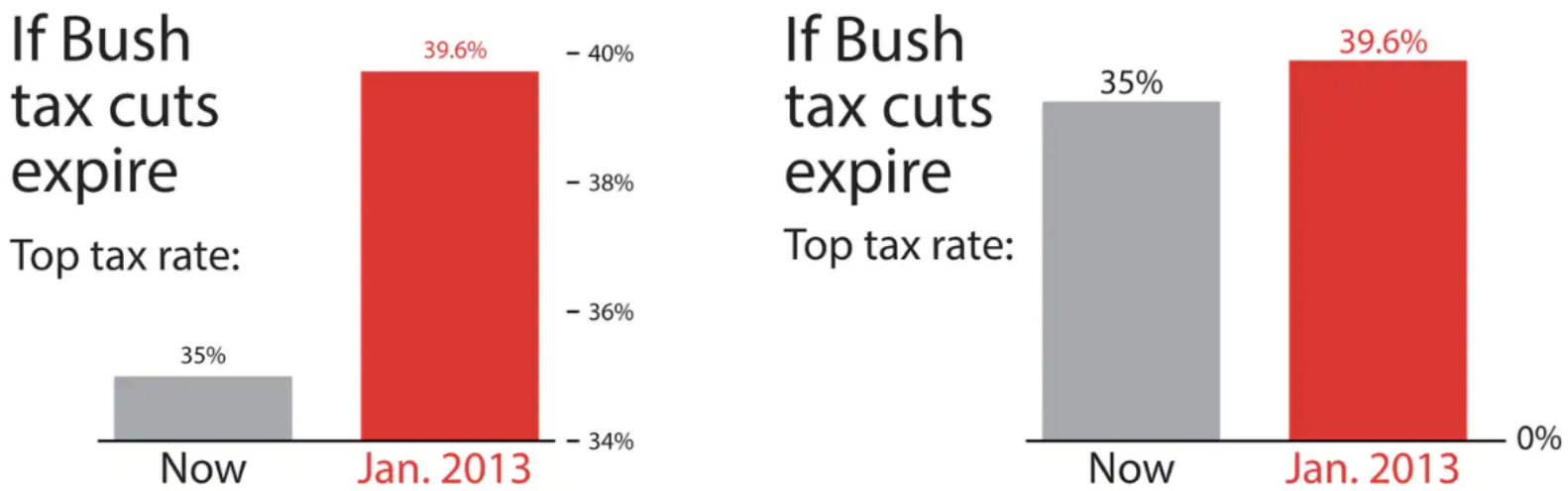
Click for PDF of slides

Topics

- 🚫 Misrepresenting data
- 🚫 Privacy
- 🚫 Algorithmic bias

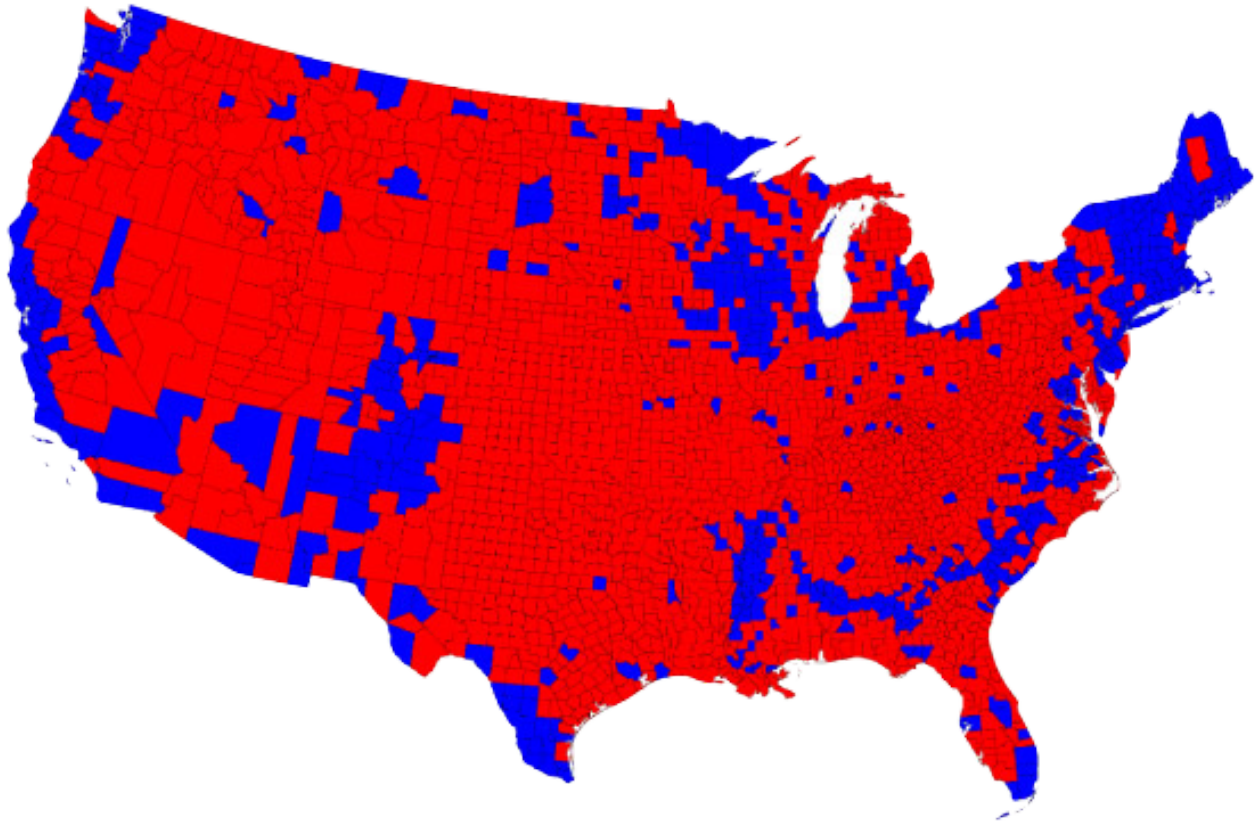
Misrepresenting data

What is the difference between these two pictures? Which presents a better way to represent these data?

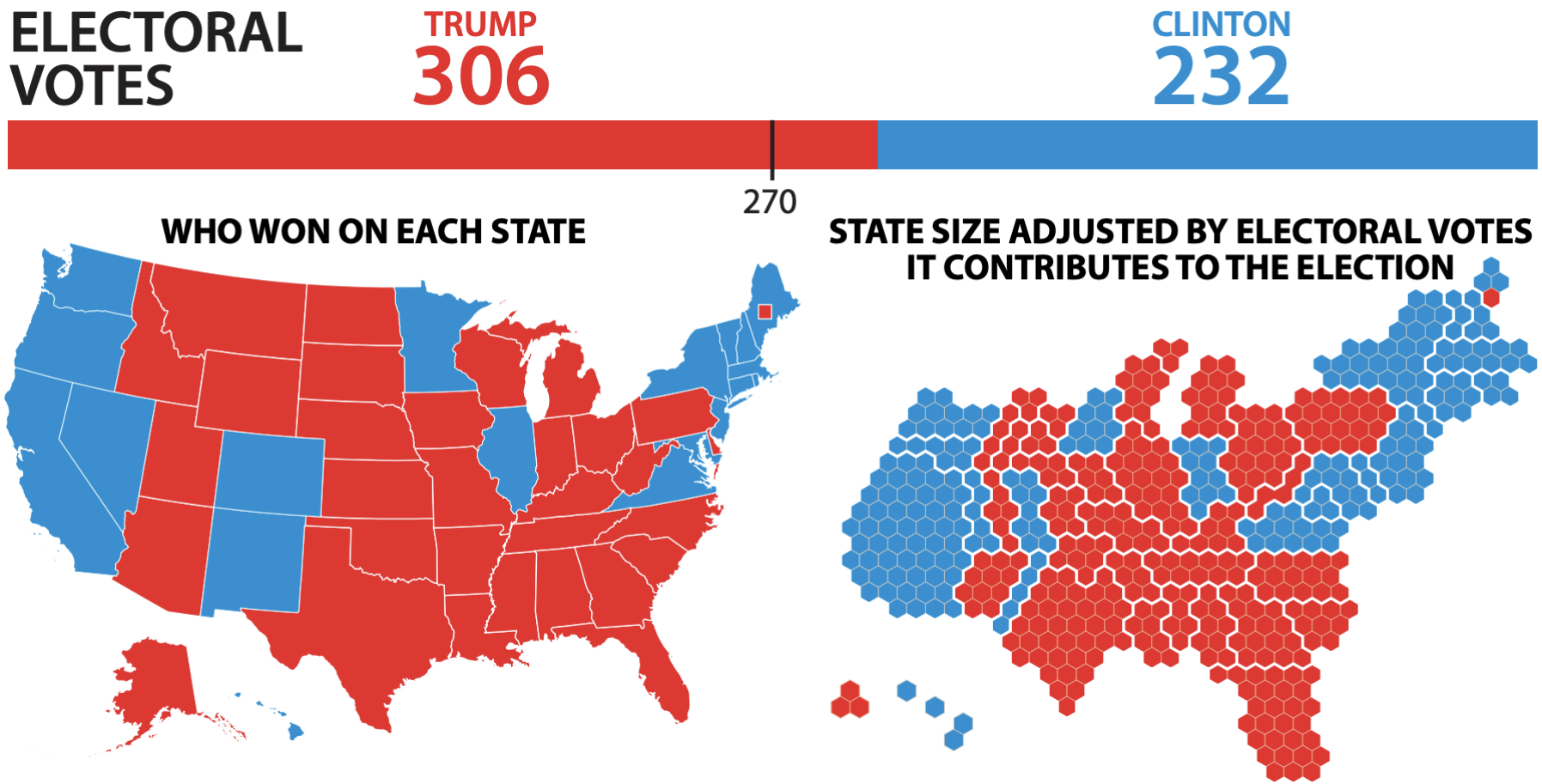


Ingraham, C. (2019) "You've been reading charts wrong. Here's how a pro does it.", The Washington Post, 14 Oct.

Do you recognize this map? What does it show?



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



Credit: Alberto Cairo, Visual Trumpery talk.

Privacy

OK Cupid Data Breach

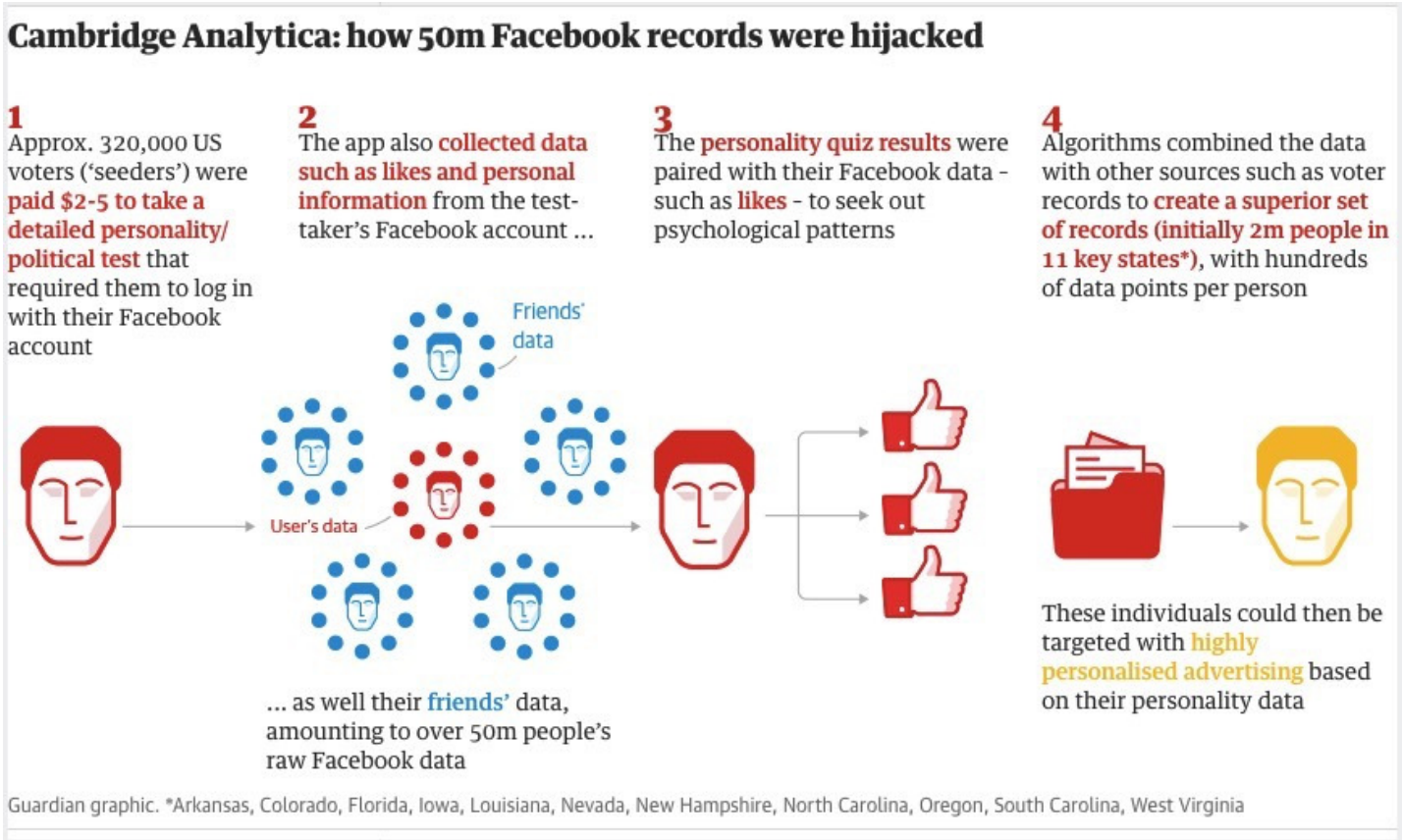
- In 2016, researchers published data of 70,000 OkCupid users—including usernames, political leanings, drug usage, and intimate sexual details.

> "Some may object to the ethics of gathering and releasing this data. However, all the data found in the dataset are or were already publicly available, so releasing this dataset merely presents it in a more useful form.""
>
> Researchers Emil Kirkegaard and Julius Daugbjerg Bjerrekær

- Although the researchers did not release the real names and pictures of the OkCupid users, critics noted that their identities could easily be uncovered from the details provided—such as from the usernames.

In analysis of data individuals willingly shared publicly on a given platform (e.g. social media data), how do you make sure you don't violate reasonable expectations of privacy?

Facebook & Cambridge Analytica



Algorithmic bias

The Hathaway Effect

TECHNOLOGY

Does Anne Hathaway News Drive Berkshire Hathaway's Stock?

ALEXIS C. MADRIGAL MAR 18, 2011

Given the awesome correlating powers of today's stock trading computers, the idea may not be as far-fetched as you think.

"Does Anne Hathaway News Drive Berkshire Hathaway's Stock?"

The Hathaway Effect

- **Oct. 3, 2008** - Rachel Getting Married opens: **BRK.A up .44%**
- **Jan. 5, 2009** - Bride Wars opens: **BRK.A up 2.61%**
- **Feb. 8, 2010** - Valentine's Day opens: **BRK.A up 1.01%**
- **March 5, 2010** - Alice in Wonderland opens: **BRK.A up .74%**
- **Nov. 24, 2010** - Love and Other Drugs opens: **BRK.A up 1.62%**
- **Nov. 29, 2010** - Anne announced as co-host of the Oscars: **BRK.A up .25%**

The Hathaway Effect: How Anne Gives Warren Buffet a Rise

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.

Bias in algorithms used for sentencing

Bias in algorithms used for sentencing

There’s software used across the country to predict future criminal activity. And it's biased...

> “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%

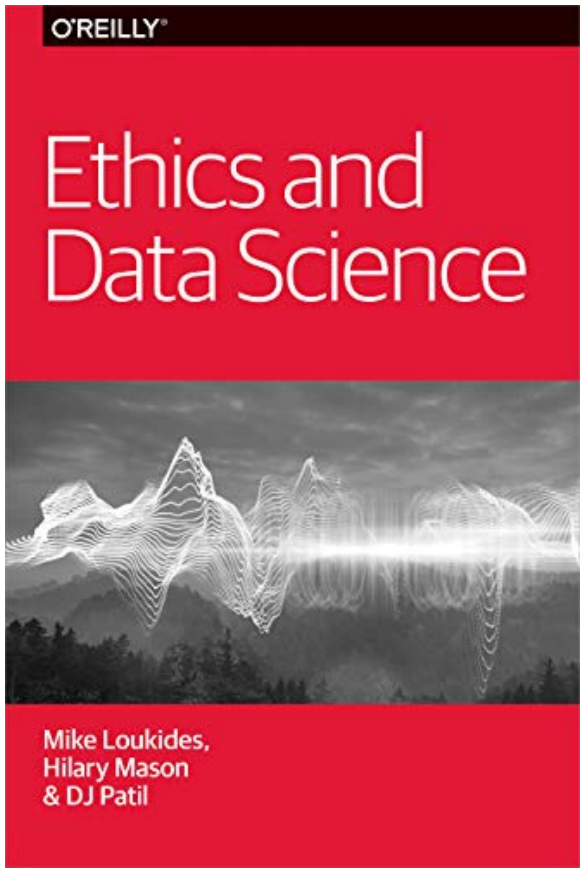
- Algorithm was more likely to falsely flag African American defendants as higher risk, at almost twice the rate as Caucasian defendants

Read more at

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

Further study on data science ethics

Further reading

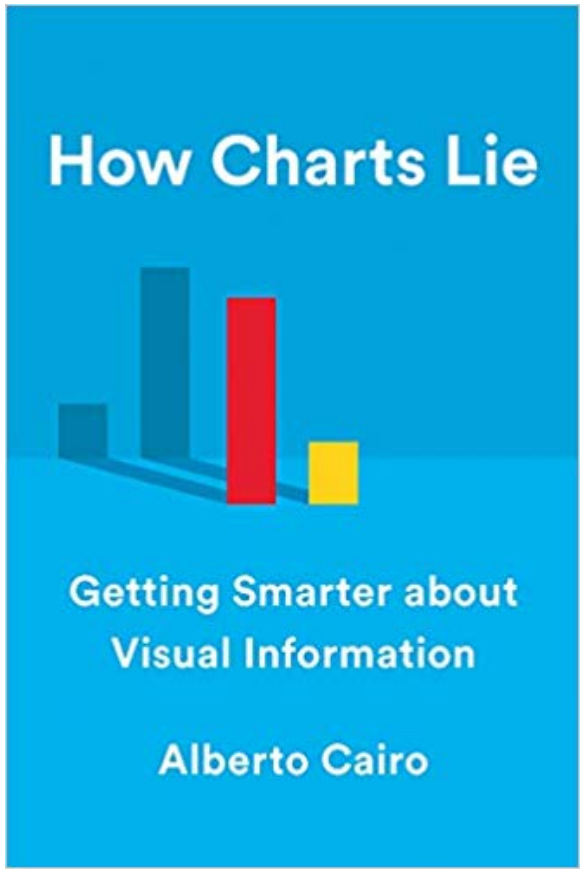


Ethics and Data Science

by Mike Loukides, Hilary Mason, DJ Patil

(free Kindle download)

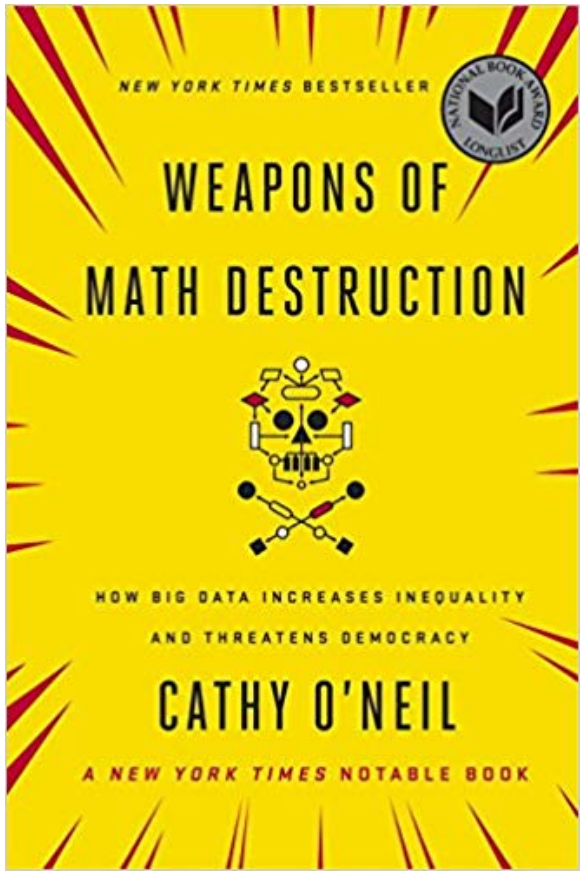
Further reading



How Charts Lie: Getting Smarter About Visual Information

by Alberto Cairo

Further reading



Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy

by Cathy O'Neil

Further watching

Predictive Policing: Bias In, Bias Out
by Kristian Lum

Parting thoughts

- At some point during your data science 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?

Do good with data

- Data Science for Social Good:
  - at the University of Chicago
  - at the Alan Turing Institute
- DataKind: DataKind brings high-impact organizations together with leading data scientists to use data science in the service of humanity
- Pledge to promote data values & practices