TransFAT

translating fairness, accountability, and transparency into data science practice

Prof. Julia Stoyanovich

Computer Science and Engineering &
Center for Data Science
New York University

@stoyanoj

NYU

The power of data science

Power

unprecedented data collection capabilities

enormous computational power

ubiquity and broad acceptance

Opportunity

improve people's lives, e.g., recommendation

accelerate scientific discovery, e.g., medicine

boost innovation, e.g., autonomous cars

transform society, e.g., open government

optimize business, e.g., advertisement targeting

goal - progress

and now some bad news

Online price discrimination

THE WALL STREET JOURNAL.

WHAT THEY KNOW

Websites Vary Prices, Deals Based on Users' Information

By JENNIFER VALENTINO-DEVRIES, JEREMY SINGER-VINE and ASHKAN SOLTANI
December 24, 2012

It was the same Swingline stapler, on the same Staples.com website. But for Kim Wamble, the price was $15.79, while the price on Trude Frizzell's screen, just a few miles away, was $14.29.

A key difference: where Staples seemed to think they were located.

WHAT PRICE WOULD YOU SEE?

lower prices offered to buyers who live in more affluent neighborhoods

https://www.wsj.com/articles/SB10001424127887323777204578189391813881534

Amazon same-day delivery

Bloomberg

Amazon Doesn't Consider the Race of Its Customers. Should It?

"… In six major same-day delivery cities, however, **the service area excludes predominantly black ZIP codes** to varying degrees, according to a Bloomberg analysis that compared Amazon same-day delivery areas with U.S. Census Bureau data."

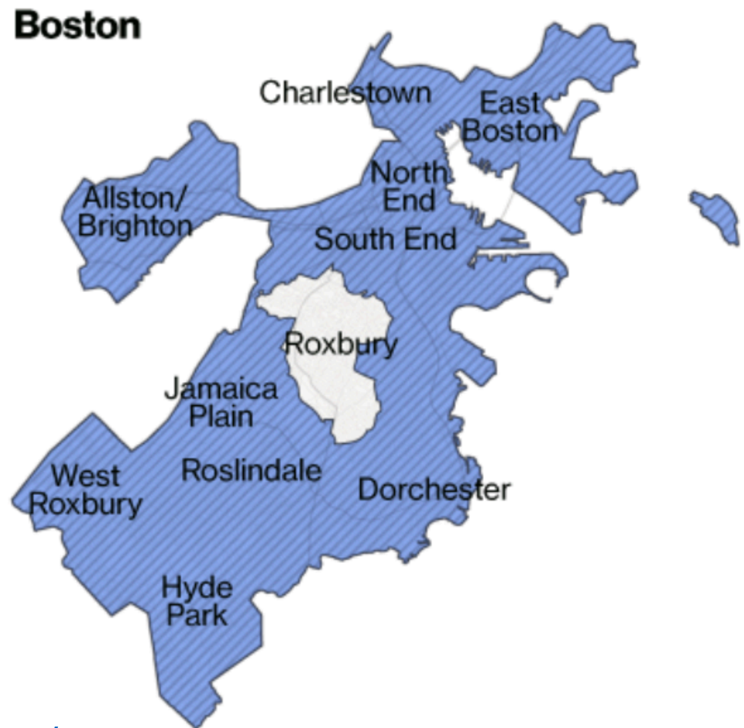
New York City

https://www.bloomberg.com/graphics/2016-amazon-same-day/

Amazon same-day delivery

Bloomberg

Amazon Doesn't Consider the Race of Its Customers. Should It?

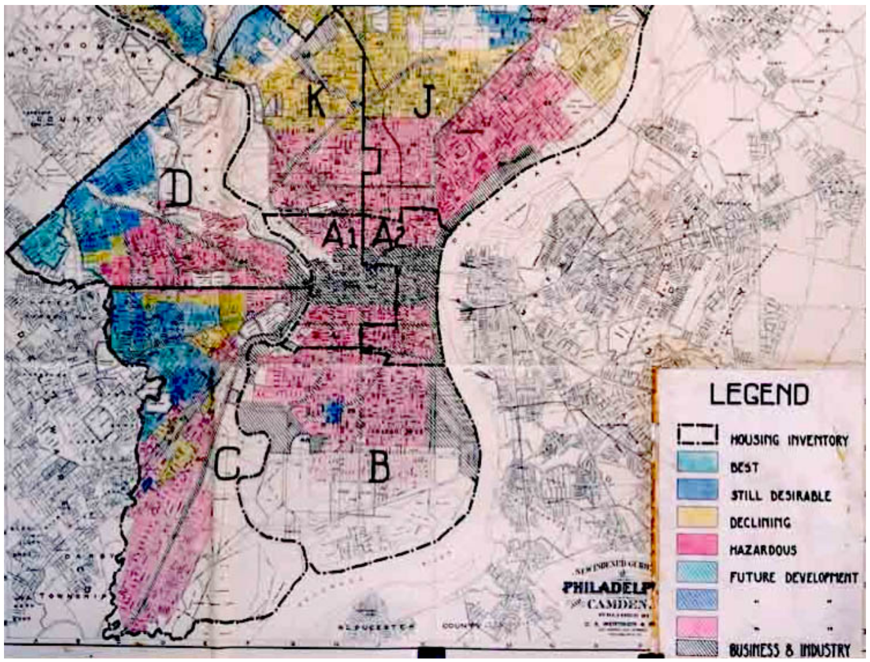


"The most striking gap in Amazon's same-day service is in Boston, where **three ZIP codes encompassing the primarily black neighborhood of Roxbury are excluded** from same-day service, while the neighborhoods that surround it on all sides are eligible."

https://www.bloomberg.com/graphics/2016-amazon-same-day/

Redlining

Redlining is the practice of arbitrarily denying or limiting **financial services** to specific neighborhoods, generally because its residents are people of color or are poor.



A HOLC 1936 security map of Philadelphia showing redlining of lower income neighborhoods

https://en.wikipedia.org/wiki/Redlining

Households and businesses in the **red zones** could not get mortgages or business loans.

Online job ads

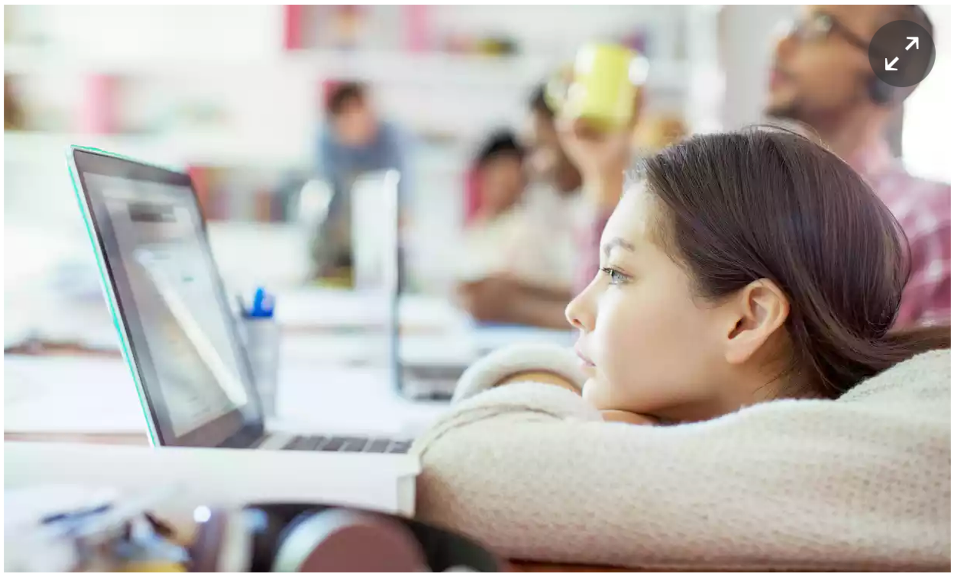
theguardian

Samuel Gibbs

Wednesday 8 July 2015 11.29 BST

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,852 times to the male group and only 318 times to the female group. Photograph: Alamy

The AdFisher tool simulated job seekers that did not differ in browsing behavior, preferences or demographic characteristics, except in gender.

One experiment showed that Google displayed ads for a career coaching service for "$200k+" executive jobs **1,852 times to the male group and only 318 times to the female group**. Another experiment, in July 2014, showed a similar trend but was not statistically significant.

https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study

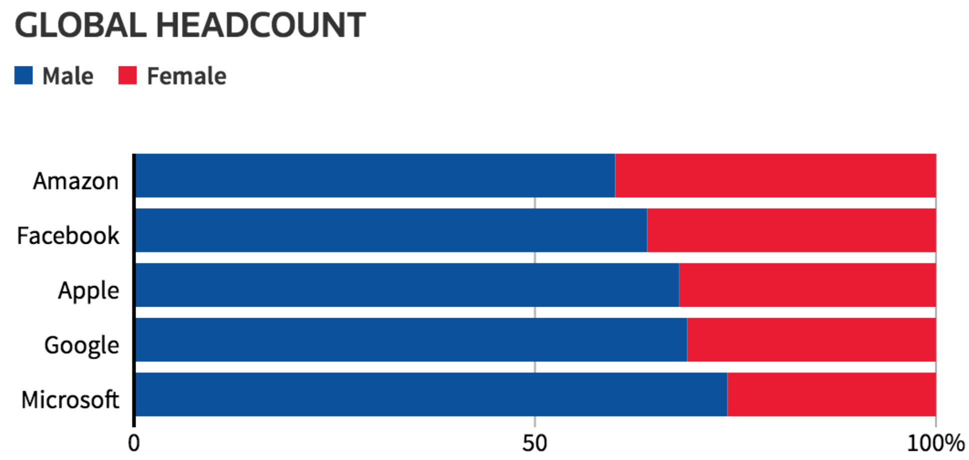
Gender bias in recruiting

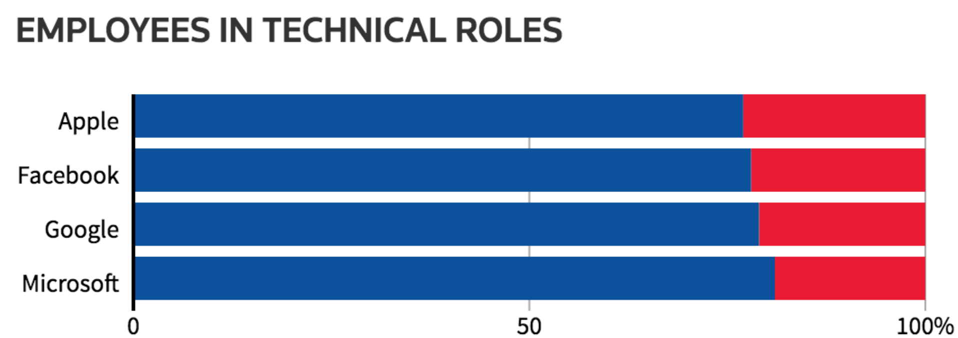
REUTERS Jeffrey Dastin

BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 6 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

"In effect, **Amazon's system taught itself that male candidates were preferable**. It penalized resumes that included the word "women's," as in "women's chess club captain." And it **downgraded graduates of two all-women's colleges**, according to people familiar with the matter. They did not specify the names of the schools."





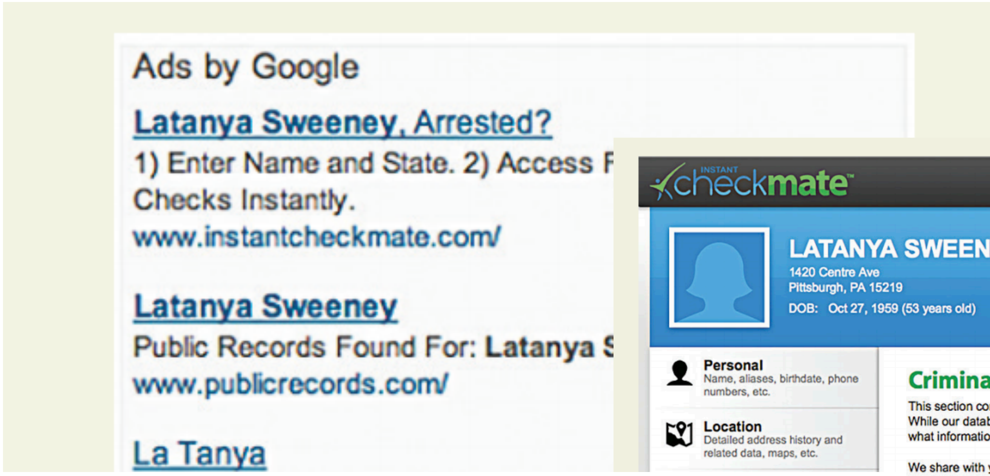
"Note: Amazon does not disclose the gender breakdown of its technical workforce."

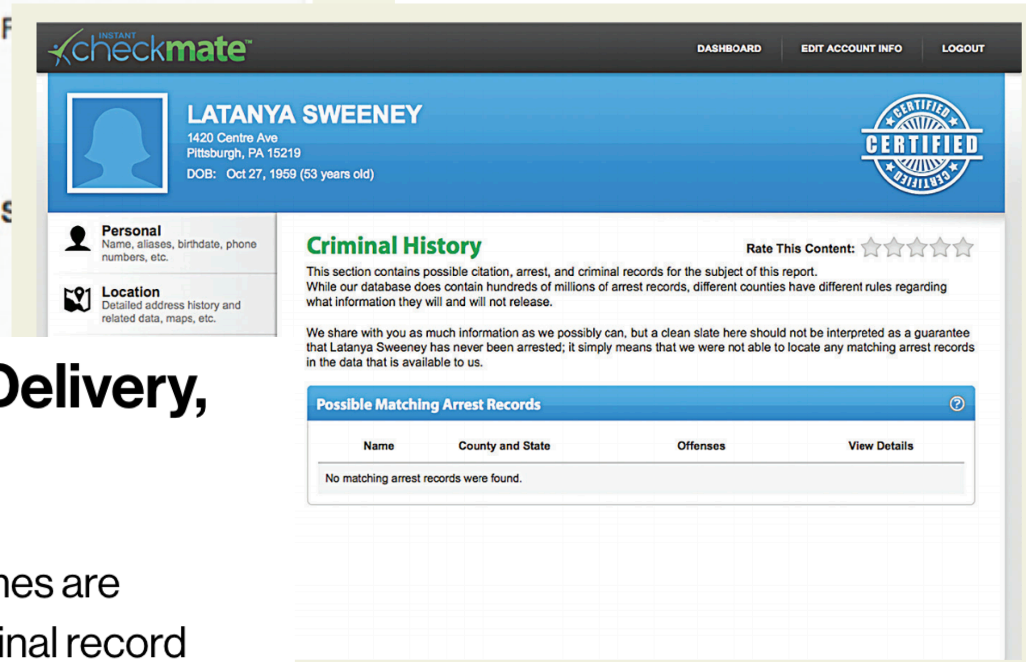
https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Racially identifying names

[Latanya Sweeney; *CACM 2013*]

Google AdSense





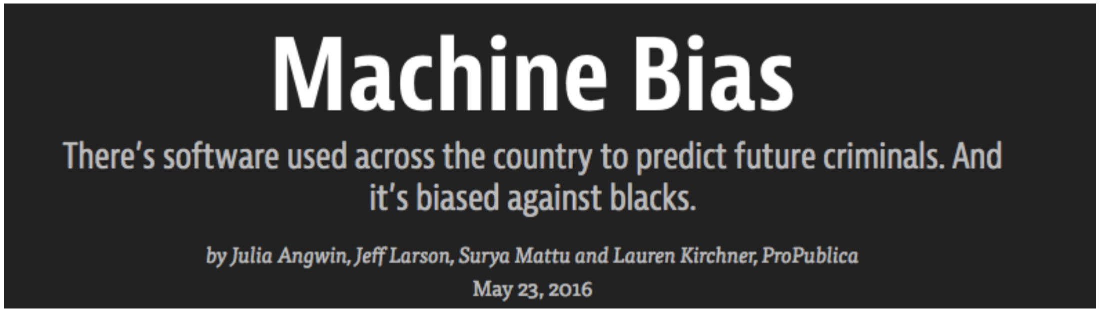
Racism is Poisoning Online Ad Delivery, Says Harvard Professor

Google searches involving black-sounding names are more likely to serve up ads suggestive of a criminal record than white-sounding names, says computer scientist

racially identifying names trigger ads suggestive of a criminal record

https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/

Racial bias in criminal sentencing



A commercial tool **COMPAS** automatically predicts some categories of future crime to assist in bail and sentencing decisions. It is used in courts in the US.

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

The tool correctly predicts recidivism **61% of the time.**

Blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend.

The tool 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.

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

Racial bias in criminal sentencing

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

A commercial tool **COMPAS** automatically predicts some categories of future crime to assist in bail and sentencing decisions. It is used in courts in the US.

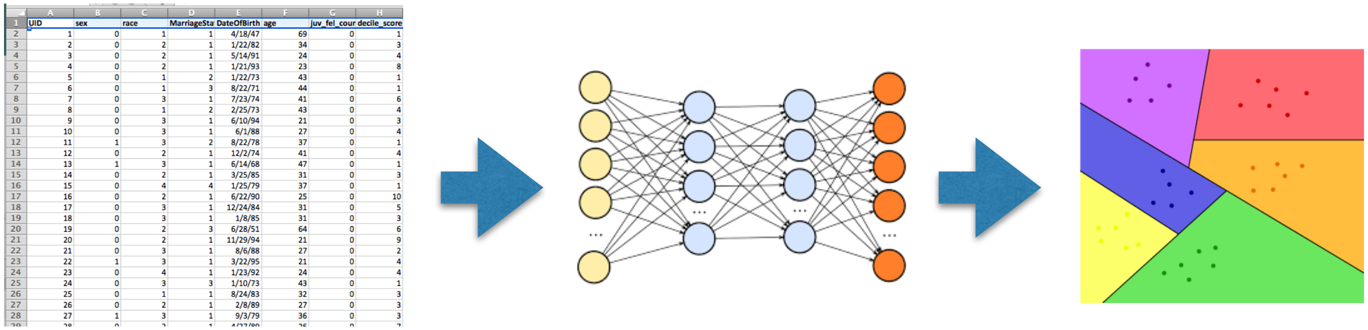
Prediction Fails Differently for Black Defendants

	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.)*

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

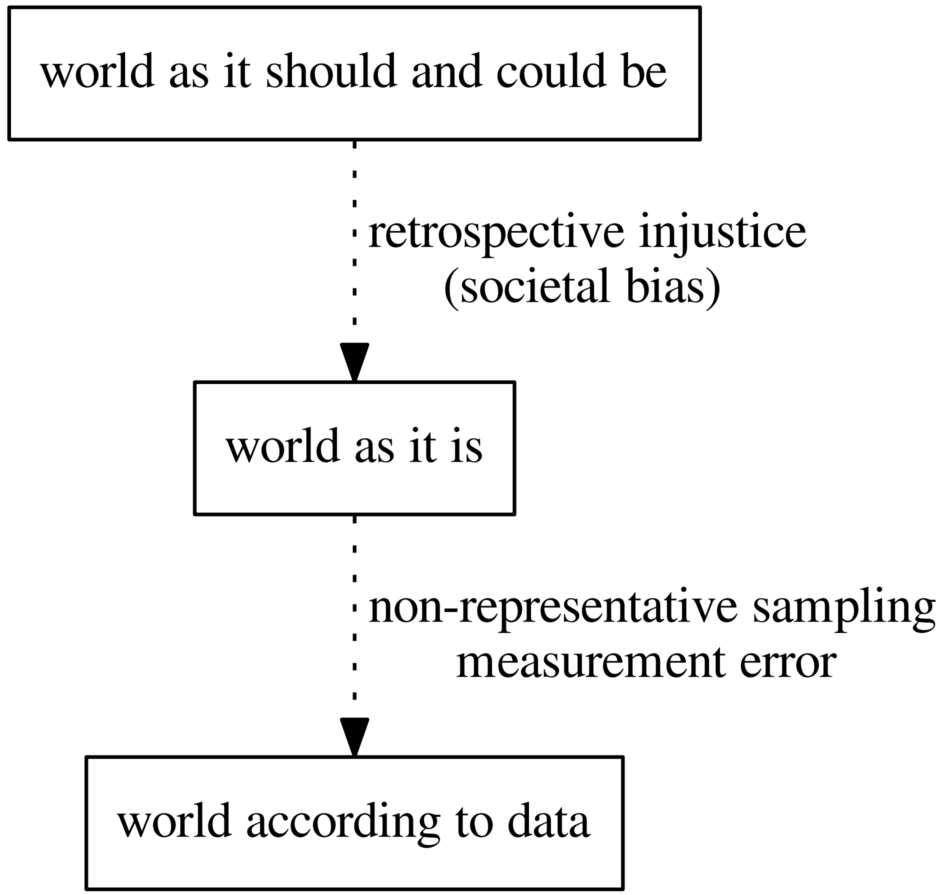
“Bias” in predictive analytics



- **Statistical bias in the model**: a model is biased if it doesn’t summarize the data correctly
- **Societal bias in the data**: a dataset is biased if it does not represent the world “correctly”, e.g., data is not representative, there is measurement error,
- …. or the **world is “incorrect”?**

the world as it is or as it should be?

Biased data reflects the world



from “Prediction-Based Decisions and Fairness” by Mitchell, Potash and Barocas, 2018

when data is about people, bias can lead to discrimination

The evils of discrimination

Disparate treatment is the illegal practice of treating an entity, such as a creditor or employee, differently based on a **protected characteristic** such as race, gender, age, religion, sexual orientation, or national origin.

Disparate impact is the result of systematic disparate treatment, where disproportionate **adverse impact** is observed on members of a **protected class**.

http://www.allenovery.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx

Regulated domains

Credit - Equal Credit Opportunity Act

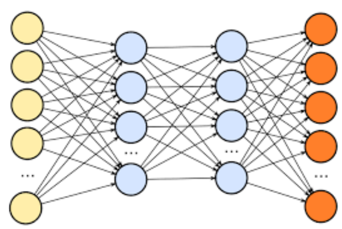
Education - Civil Rights Act of 1964

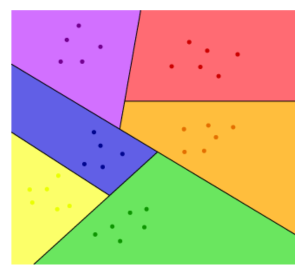
Employment - Civil Rights Act of 1964

Housing - Fair Housing Act

http://www.allenovery.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx

	A	B	C	D	E	F	G	H
1	UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_cour	decile_score
2	1	0	1	1	4/18/47	69	0	1
3	2	0	2	1	1/22/82	34	0	3
4	3	0	2	1	5/14/91	24	0	4
5	4	0	2	1	1/21/93	23	0	8
6	5	0	1	2	1/22/73	43	0	1
7	6	0	1	3	8/22/71	44	0	1
8	7	0	3	1	7/23/74	41	0	6
9	8	0	1	2	2/25/73	43	0	4
10	9	0	3	1	6/10/94	21	0	3
11	10	0	3	1	6/1/88	27	0	4
12	11	1	3	2	8/22/78	37	0	1
13	12	0	2	1	12/2/74	41	0	4
14	13	1	3	1	6/14/68	47	0	1
15	14	0	2	1	3/25/85	31	0	3
16	15	0	4	4	1/25/79	37	0	1
17	16	0	2	1	6/22/90	25	0	10
18	17	0	3	1	12/24/84	31	0	5
19	18	0	3	1	1/8/85	31	0	3
20	19	0	2	3	6/28/51	64	0	6
21	20	0	2	1	11/29/94	21	0	9
22	21	0	3	1	8/6/88	27	0	2
23	22	1	3	1	3/22/95	21	0	4
24	23	0	4	1	1/23/92	24	0	4
25	24	0	3	3	1/10/73	43	0	1
26	25	0	1	1	8/24/83	32	0	3
27	26	0	2	1	2/8/89	27	0	3
28	27	1	3	1	9/3/79	36	0	3





regulation does not immediately apply here!

Regulated domains

Credit - Equal Credit Opportunity Act

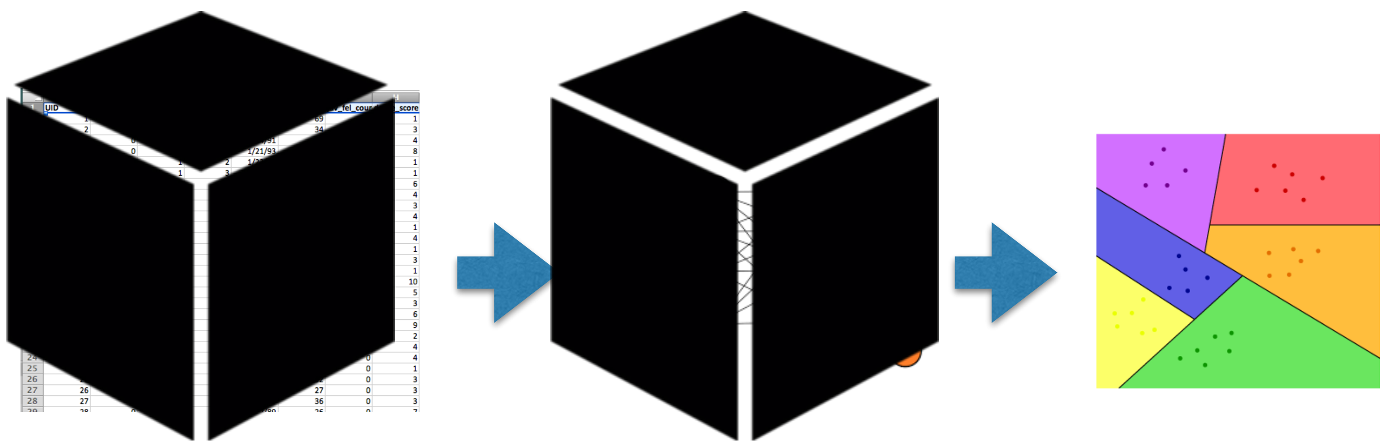
Education - Civil Rights Act of 1964

Employment - Civil Rights Act of 1964

Housing - Fair Housing Act

http://www.allenovery.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx

opacity makes detecting problems difficult



regulation does not immediately apply here!

The punchline

Data science is algorithmic, therefore it cannot be biased! And yet…

- All traditional evils of **discrimination**, and many new ones, exhibit themselves in the data science ecosystem
- **Transparency** helps prevent discrimination, enable public debate, establish **trust**
- Technology alone won't do: also need **regulation** and **civic engagement**

http://www.allenovery.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx

responsible data science is our new frontier!

an (ongoing) attempt at regulation

NYC ADS transparency law

1/11/2018

Local Law 49 of 2018 in relation to automated decision systems used by agencies

THE NEW YORK CITY COUNCIL
Corey Johnson, Speaker

Sign In

LEGISLATIVE RESEARCH CENTER

Council Home | Legislation | Calendar | City Council | Committees

RSS | Alerts

Details | Reports

File #:	Int 1696-2017 Version: A	Name:	Automated decision systems used by agencies.
Type:	Introduction	Status:	Enacted
		Committee:	Committee on Technology
On agenda:	8/24/2017		
Enactment date:	1/11/2018	Law number:	2018/049
Title:	A Local Law in relation to automated decision systems used by agencies		
Sponsors:	James Vacca, Helen K. Rosenthal, Corey D. Johnson, Rafael Salamanca, Jr., Vincent J. Gentile, Robert E. Cornegy, Jr., Jumaane D. Williams, Ben Kallos, Carlos Menchaca		
Council Member Sponsors:	9		
Summary:	This bill would require the creation of a task force that provides recommendations on how information on agency automated decision systems may be shared with the public and how agencies may address instances where people are harmed by agency automated decision systems.		
Indexes:	Oversight		
Attachments:	1. Summary of Int. No. 1696-A, 2. Summary of Int. No. 1696, 3. Int. No. 1696, 4. August 24, 2017 - Stated Meeting Agenda with Links to Files, 5. Committee Report 10/16/17, 6. Hearing Testimony 10/16/17, 7. Hearing Transcript 10/16/17, 8. Proposed Int. No. 1696-A - 12/12/17, 9. Committee Report 12/7/17, 10. Hearing Transcript 12/7/17, 11. December 11, 2017 - Stated Meeting Agenda with Links to Files, 12. Hearing Transcript - Stated Meeting 12-11-17, 13. Int. No. 1696-A (FINAL), 14. Fiscal Impact Statement, 15. Legislative Documents - Letter to the Mayor, 16. Local Law 49, 17. Minutes of the Stated Meeting - December 11, 2017		

The original draft

8/16/2017

Int. No. 1696

By Council Member Vacca

A Local Law to amend the administrative code of the city of New York, in relation to automated processing of data for the purposes of targeting services, penalties, or policing to persons

Be it enacted by the Council as follows:

Section 1. Section 23-502 of the administrative code of the city of New York is amended to add a new subdivision g to read as follows:

g. Each agency that uses, for the purposes of targeting services to persons, imposing penalties upon persons or policing, an algorithm or any other method of automated processing system of data shall:

1. Publish on such agency's website, the source code of such system; and

2. Permit a user to (i) submit data into such system for self-testing and (ii) receive the results of having such data processed by such system.

§ 2. This local law takes effect 120 days after it becomes law.

MAJ
LS# 10948
8/16/17 2:13 PM

this is **NOT** what was adopted

Let’s speak up!

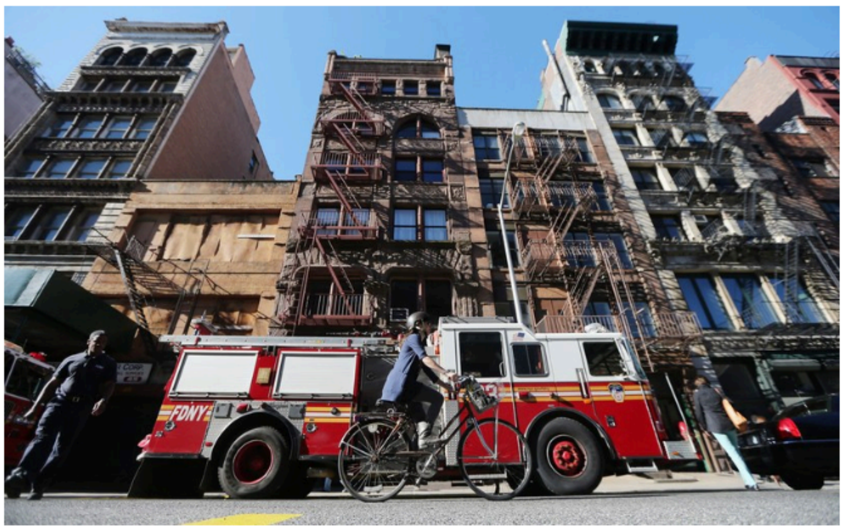
10/16/2017

THE NEW YORKER

By Julia Powles December 20, 2017

ELEMENTS

NEW YORK CITY’S BOLD, FLAWED ATTEMPT TO MAKE ALGORITHMS ACCOUNTABLE



Automated systems guide the allocation of everything from firehouses to food stamps. So why don’t we know more about them?

Photograph by Mario Tama / Getty

Summary of Local Law 49

1/11/2018

Form an automated decision systems (**ADS**) task force that surveys current use of algorithms and data in City agencies and develops procedures for:

- requesting and receiving an **explanation** of an algorithmic decision affecting an individual (3(b))
- interrogating ADS for **bias and discrimination** against members of legally-protected groups (3(c) and 3(d))
- allowing the **public** to **assess** how ADS function and are used (3(e)), and archiving ADS together with the data they use (3(f))

The ADS Task Force

5/16/2018

Visit **alpha.nyc.gov** to help us test out new ideas for NYC's website.

The Official Website of the City of New York

简体中文 Translate ▼ Text Size

NYC Resources | NYC311 | Office of the Mayor | Events | Connect | Jobs | Search

Mayor | First Lady | News | Officials

SHARE

Email

Print

Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City

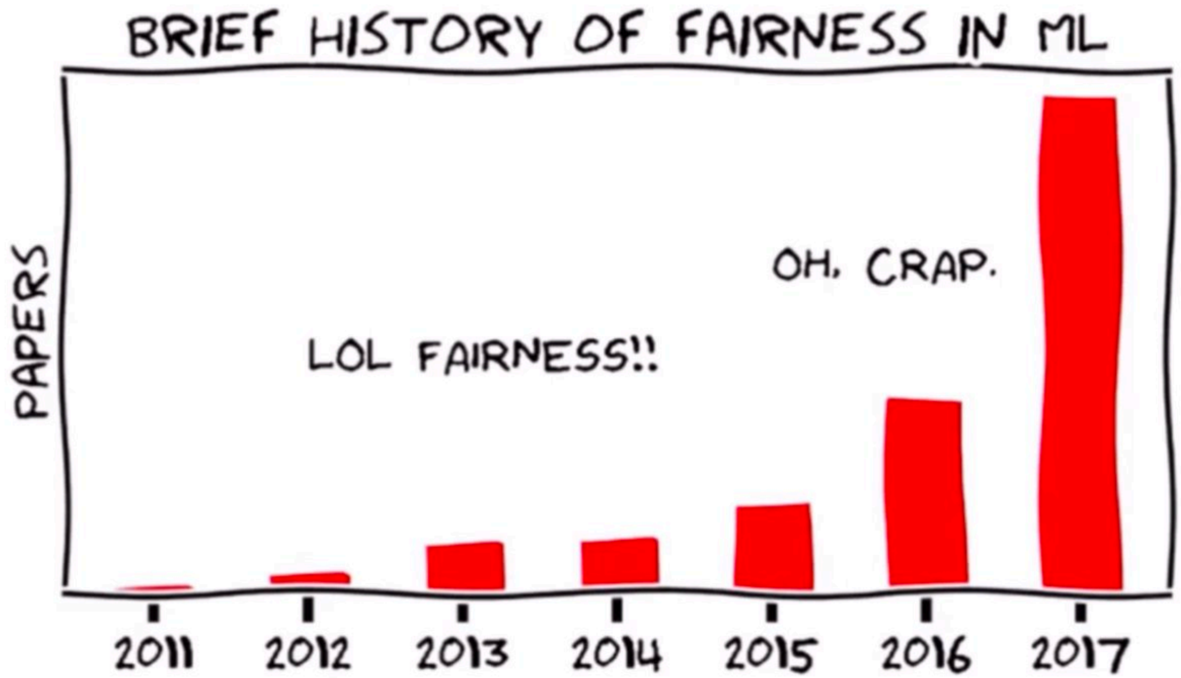
May 16, 2018

NEW YORK— Today, Mayor de Blasio announced the creation of the Automated Decision Systems Task Force which will explore how New York City uses algorithms. The task force, the first of its kind in the U.S., will work to develop a process for reviewing "automated decision systems," commonly known as algorithms, through the lens of equity, fairness and accountability.

"As data and technology become more central to the work of city government, the algorithms we use to aid decision making must be aligned with our goals and values," said **Mayor de Blasio**. "The establishment of the Automated Decision Systems Task Force is an important first step towards greater transparency and equity in our use of technology."

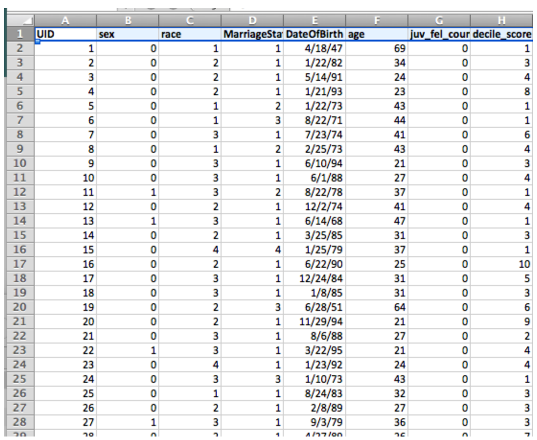
mitigating bias and discrimination

Fairness in machine learning

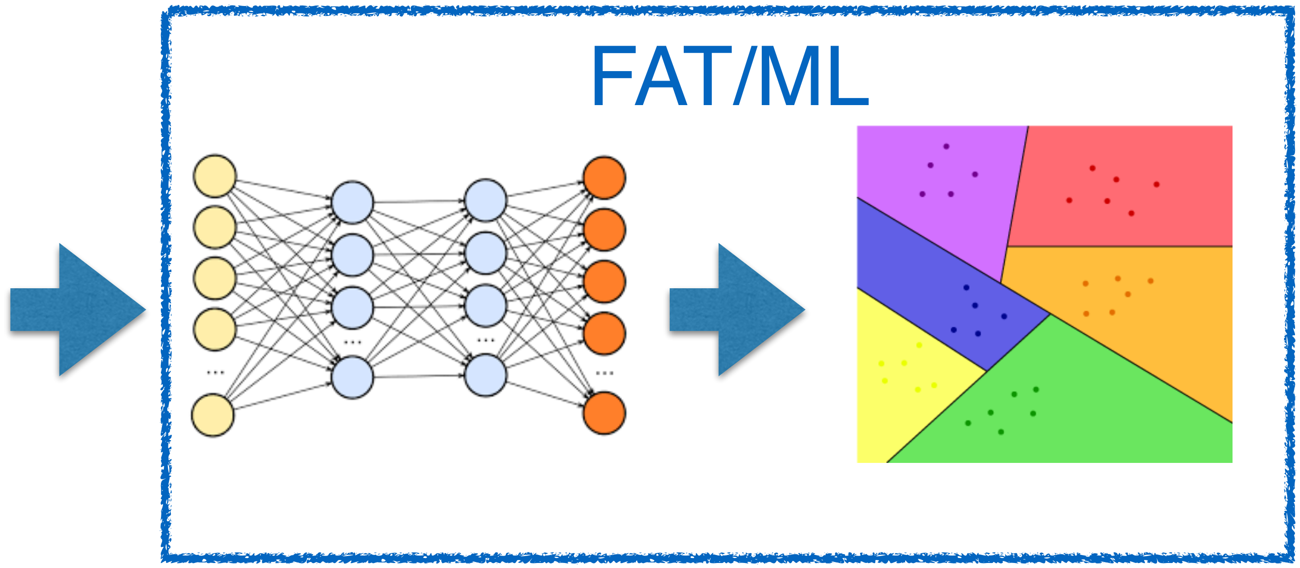


by Moritz Hardt

Fairness in machine learning



	A	B	C	D	E	F	G	H
1	UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_cour	decile_score
2	1	0	1	1	4/18/47	69	0	1
3	2	0	2	1	1/22/82	34	0	3
4	3	0	2	1	5/14/91	24	0	4
5	4	0	2	1	1/21/93	23	0	8
6	5	0	1	2	1/22/73	43	0	1
7	6	0	1	3	8/22/71	44	0	1
8	7	0	3	1	7/23/74	41	0	6
9	8	0	1	2	2/25/73	43	0	4
10	9	0	3	1	6/10/94	21	0	3
11	10	0	3	1	6/1/88	27	0	4
12	11	1	3	2	8/22/78	37	0	1
13	12	0	2	1	12/2/74	41	0	4
14	13	1	3	1	6/14/68	47	0	1
15	14	0	2	1	3/25/85	31	0	3
16	15	0	4	4	1/25/79	37	0	1
17	16	0	2	1	6/22/90	25	0	10
18	17	0	3	1	12/24/84	31	0	5
19	18	0	3	1	1/8/85	31	0	3
20	19	0	2	3	6/28/51	64	0	6
21	20	0	2	1	11/29/94	21	0	9
22	21	0	3	1	8/6/88	27	0	2
23	22	1	3	1	3/22/95	21	0	4
24	23	0	4	1	1/23/92	24	0	4
25	24	0	3	3	1/10/73	43	0	1
26	25	0	1	1	8/24/83	32	0	3
27	26	0	2	1	2/8/89	27	0	3
28	27	1	3	1	9/3/79	36	0	3



Fairness in ranking

Idea: Rankings are relative, fairness measures should be rank-aware

rank	gender	rank	gender	rank	gender
1	M	1	M	1	M
2	M	2	M	2	F
3	M	3	F	3	M
4	M	4	M	4	F
5	M	5	M	5	M
6	F	6	F	6	F
7	F	7	M	7	M
8	F	8	F	8	F
9	F	9	F	9	M
10	F	10	F	10	F
$f = 0$		$f = 0.3$		$f = 0.5$	

[Yang and Stoyanovich, **FATML 2016**]

More fairness in ranking

Designing Fair Ranking Schemes

Abolfazl Asudeh†, H. V. Jagadish†, Julia Stoyanovich‡, Gautam Das††
†University of Michigan, ‡Drexel University, ††University of Texas at Arlington
†{asudeh, jag}@umich.edu, ‡stoyanovich@drexel.edu, ††gdas@uta.edu

ACM SIGMOD 2019

ABSTRACT

Items from a database are often ranked based on a combination of multiple criteria. A user may have the flexibility to accept combinations that weigh these criteria differently, within limits. On the other hand, this choice of weights can greatly affect the fairness of the produced ranking. In this paper, we develop a system that helps users choose criterion weights that lead to greater fairness.

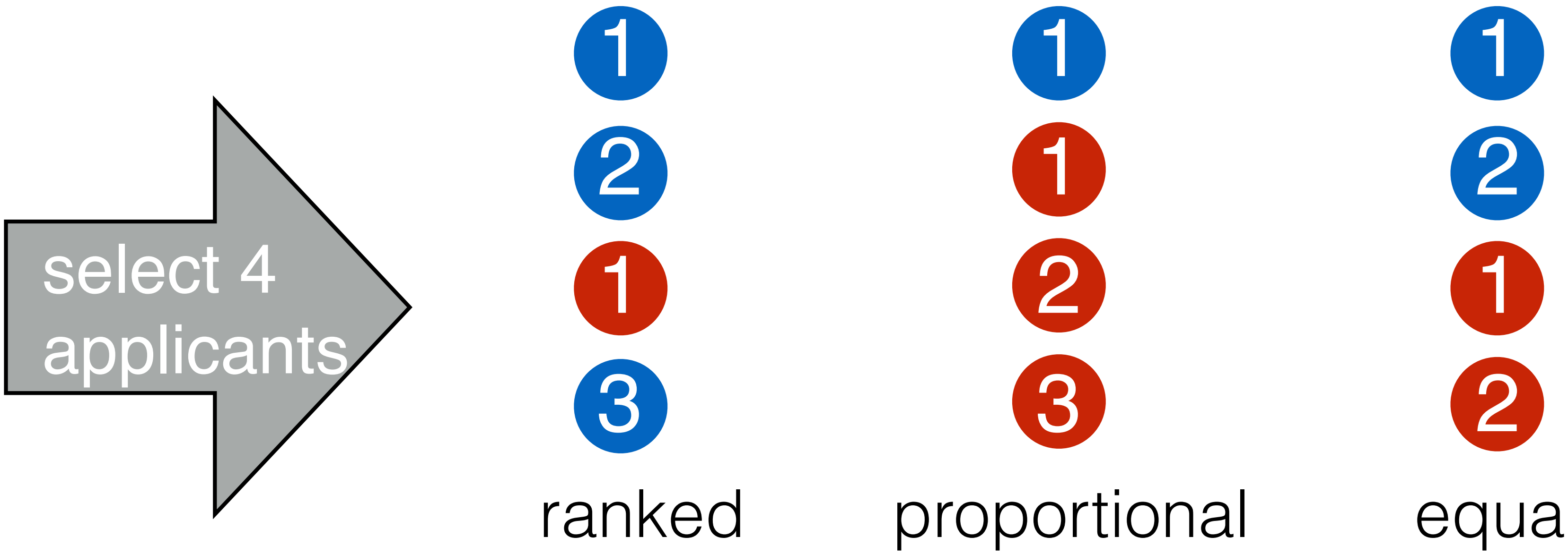
We consider ranking functions that compute the score of each item as a weighted sum of (numeric) attribute values, and then sort items on their score. Each ranking function can be expressed as a vector of weights, or as a point in a multi-dimensional space. For a broad range of fairness criteria, we show how to efficiently identify regions in this space that satisfy these criteria. Using this identification method, our system is able to tell users whether their proposed ranking function satisfies the desired fairness criteria and, if it does not, to suggest the smallest modification that does. We develop user-controllable approximation that and indexing techniques that are applied during preprocessing, and support sub-second response times during the online phase. Our extensive experiments on real datasets demonstrate that our methods are able to find solutions that

impact processes that are directly designed and validated by humans. Perhaps the most immediate example of such a process is a score-based ranker. In this paper we consider the task of *designing a fair score-based ranking scheme*.

Ranking of individuals is ubiquitous, and is used, for example, to establish credit worthiness, desirability for college admissions and employment, and attractiveness as dating partners. A prominent family of ranking schemes are score-based rankers, which compute the score of each individual from some database $\mathcal{D}$, sort the individuals in decreasing order of score, and finally return either the full ranked list, or its highest-scoring sub-set, the top-k. Many score-based rankers compute the score of an individual as a linear combination of attribute values, with non-negative weights. Designing a ranking scheme amounts to selecting a set of weights, one for each feature, and validating the outcome on the database $\mathcal{D}$.

Our goal is to assist the user in designing a ranking scheme that both reflects a user's a priori notion of quality and is fair, in the sense that it mitigates *preexisting bias with respect to a protected feature* that is embodied in the data. In line with prior work [17, 27, 31–33], a protected feature denotes membership of an individual

Fairness and diversity in set selection



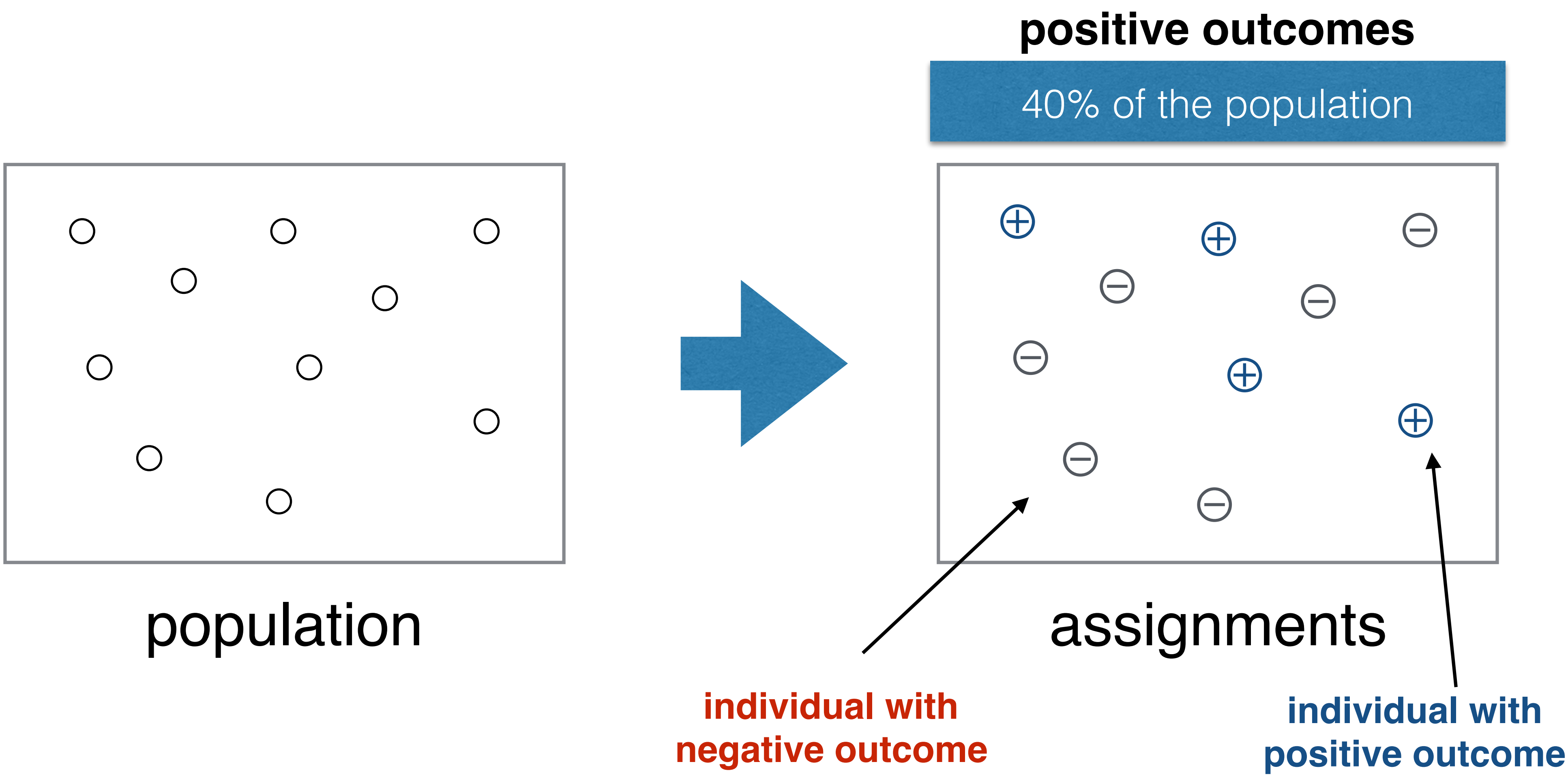
Can state all these as constraints:

for each category ***i***, pick K_i elements, with $floor_i \leq K_i \leq ceil_i$

[Stoyanovich, Yang, Jagadish **EDBT 2018**] and now some bad news

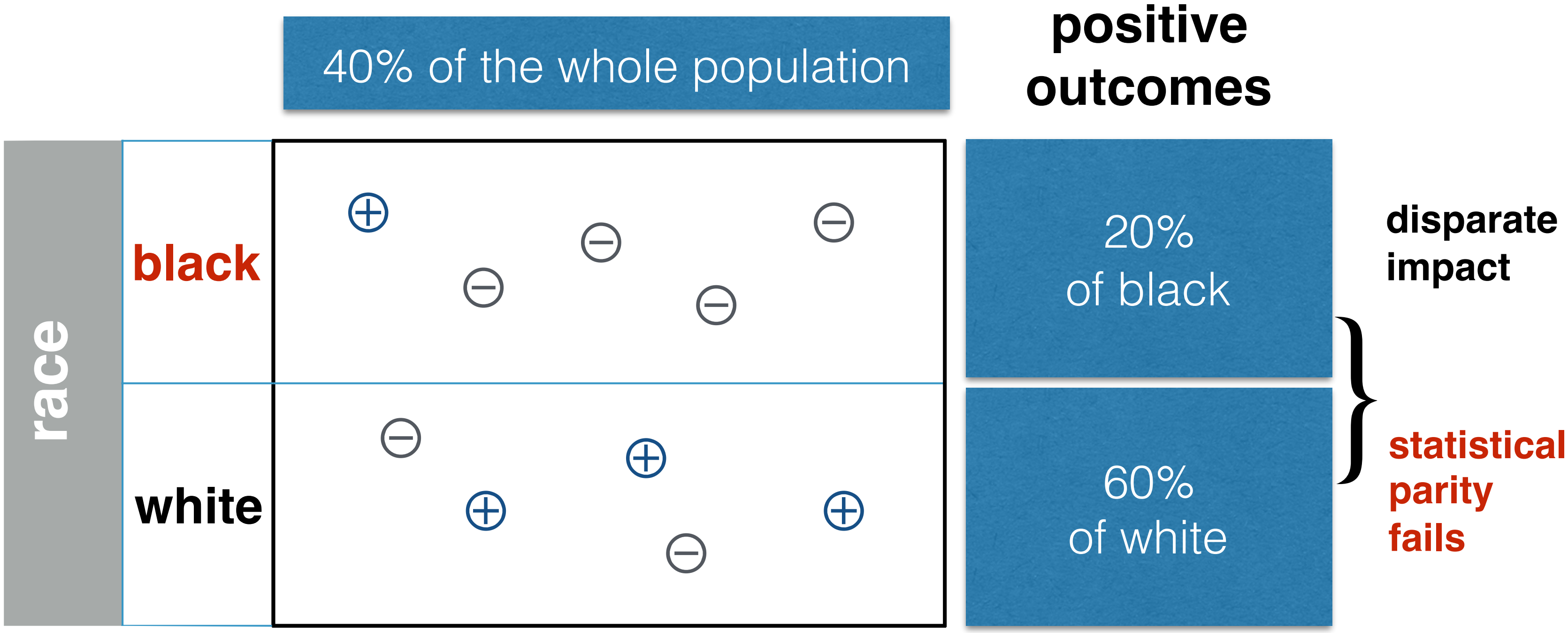
Fairness in classification

Fairness in classification is concerned with how outcomes are assigned to a population



Fairness in classification

Sub-populations are treated differently



Statistical parity

Statistical parity (a popular **group fairness** measure)
demographics of the individuals receiving any outcome are the same as demographics of the underlying population

		SAT score	
		high	**low**
race	**black**	⊕	⊖ ⊖ ⊖ ⊖
	white	⊕ ⊕ ⊕	⊖ ⊖

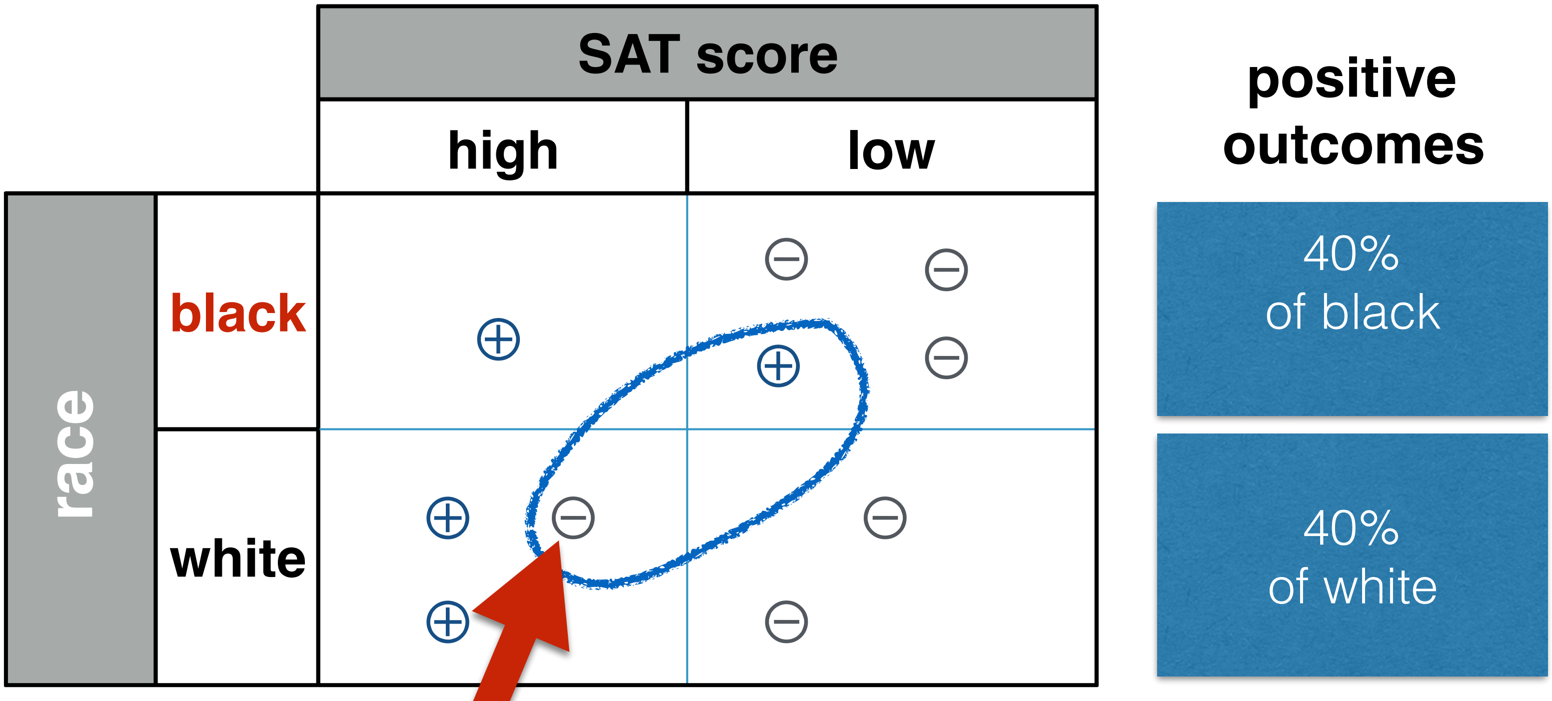
positive outcomes

20%
of black

60%
of white

Is statistical parity sufficient?

Statistical parity (a popular **group fairness** measure)
demographics of the individuals receiving any outcome are the same as demographics of the underlying population



Individual fairness
any two individuals who are similar w.r.t. a particular task should receive similar outcomes

Ricci v. DeStefano (2009)

Supreme Court Finds Bias Against White Firefighters

By ADAM LIPTAK JUNE 29, 2009

The New York Times

Karen Lee Torre, left, a lawyer who represented the New Haven firefighters in their lawsuit, with her clients Monday at the federal courthouse in New Haven. Christopher Capozziello for The New York Times

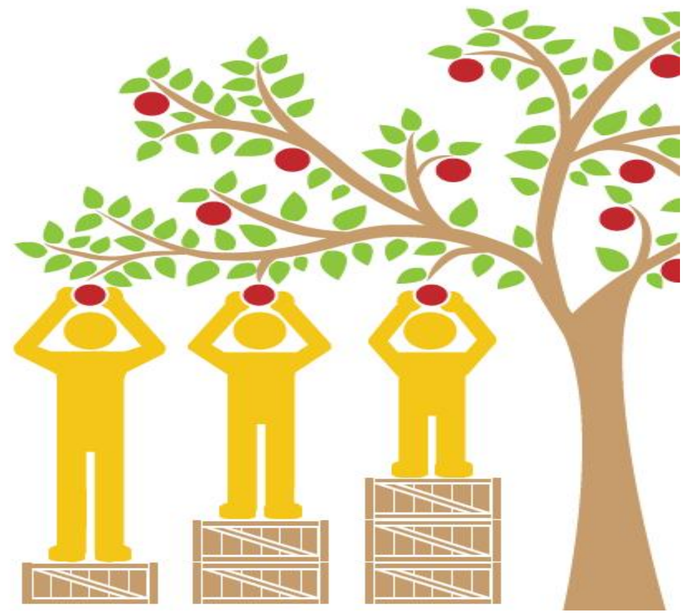
Case opinions	
Majority	Kennedy, joined by Roberts, Scalia, Thomas, Alito
Concurrence	Scalia
Concurrence	Alito, joined by Scalia, Thomas
Dissent	Ginsburg, joined by Stevens, Souter, Breyer
Laws applied	
Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e et seq.	

Two notions of fairness

individual fairness

equality

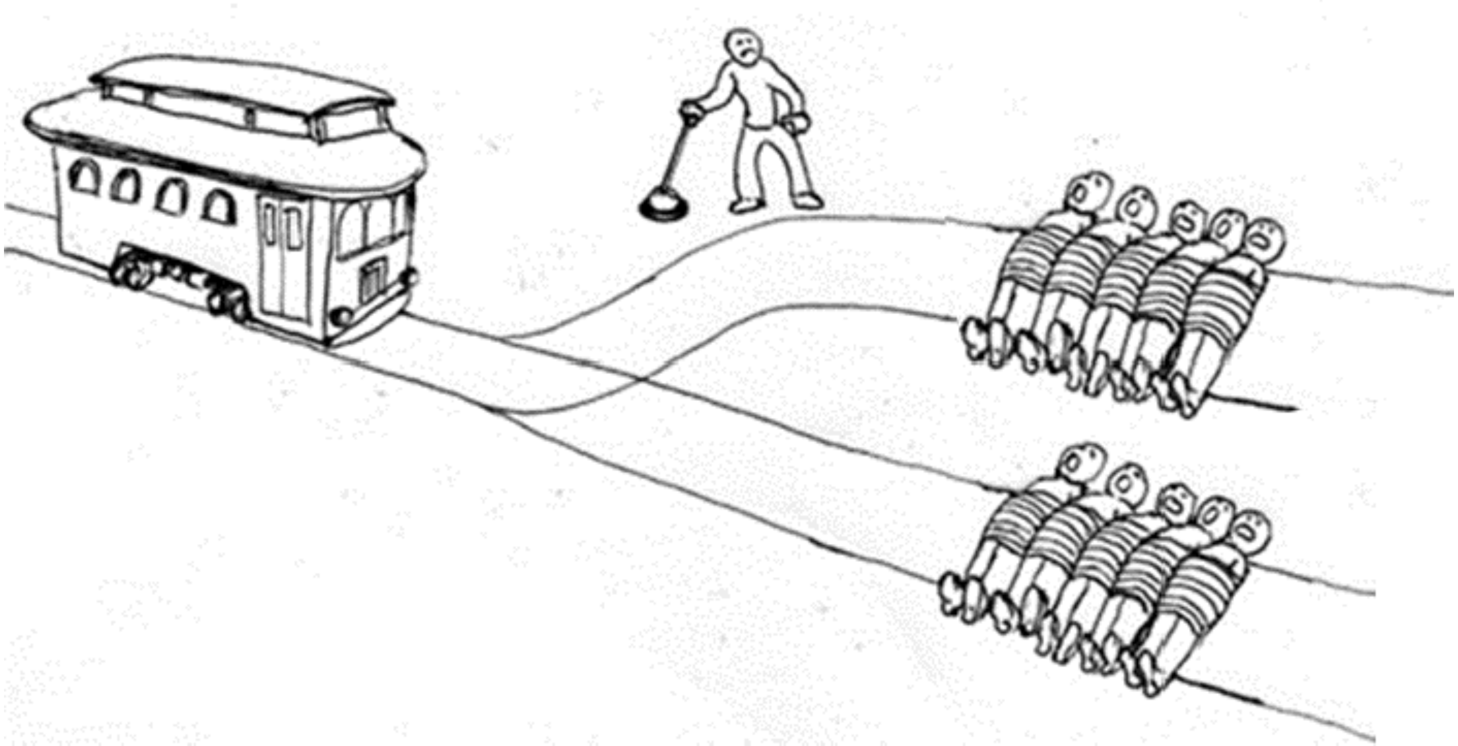
group fairness



equity

two intrinsically different world views

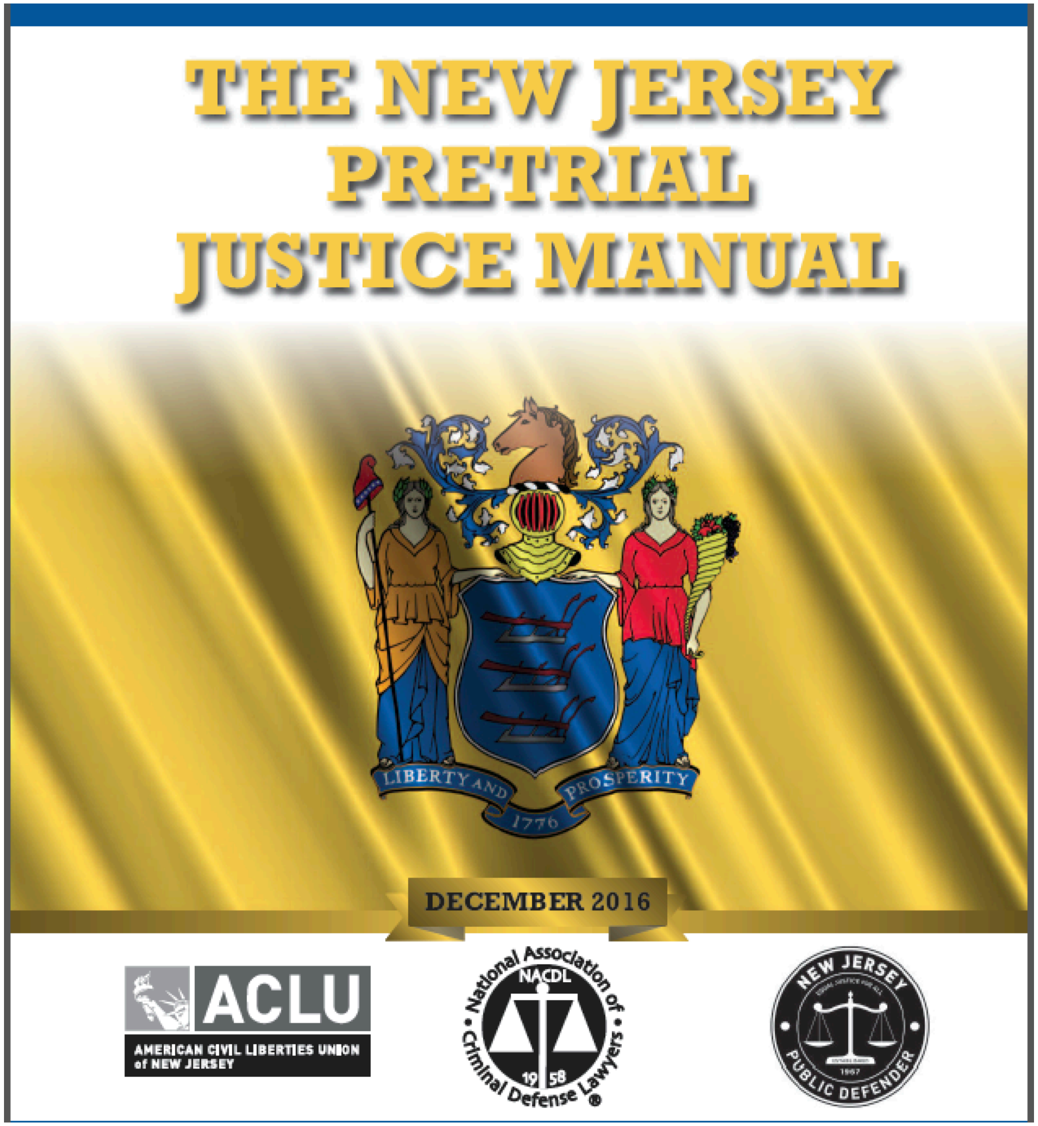
Fairness definitions as “trolley problems”



https://www.helpage.org/silo/images/blogs/16_1391611056.gif

and more bad news

New Jersey bail reform



New Jersey bail reform

THE NEW JERSEY PRETRIAL JUSTICE MANUAL

6 or subjected to onerous conditions of release.

Switching from a system based solely on instinct and experience (often referred to as "gut instinct") to one in which judges have access to scientific, objective risk assessment tools could further the criminal justice system's central goals of increasing public safety, reducing crime, and making the most effective, fair, and efficient use of public resources.

Risk Assessment and Release/Detention Decision Making in New Jersey

DECEMBER 2016

ACLU

AMERICAN CIVIL LIBERTIES UNION of NEW JERSEY

National Association of Criminal Defense Lawyers NACDL 1958

NEW JERSEY PUBLIC DEFENDER

Racial bias in criminal sentencing

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

A commercial tool **COMPAS** automatically predicts some categories of future crime to assist in bail and sentencing decisions. It is used in courts in the US.

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.)*

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

Fairness in risk assessment

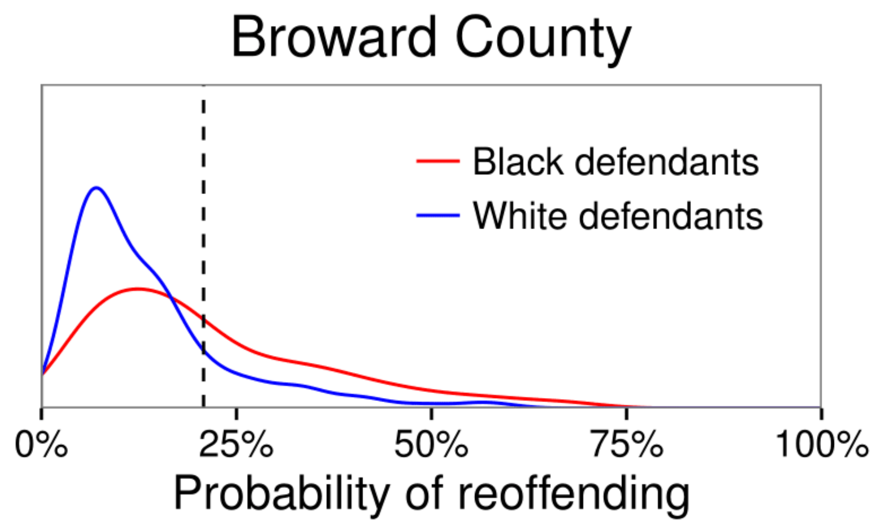
- A risk assessment tool **gives a probability estimate of a future outcome**
- Used in many domains:
  - insurance, criminal sentencing, medical testing, hiring, banking
  - also in less-obvious set-ups, like online advertising
- **Fairness** in risk assessment is concerned with **how different kinds of errors are distributed among sub-populations**

COMPAS as a predictive instrument

[J. Kleinberg, S. Mullainathan, M. Raghavan; *ITCS 2017*]

Predictive parity (also called **calibration**)
an instrument identifies a set of instances as having probability *x* of constituting positive instances, then approximately an *x* fraction of this set are indeed positive instances, over-all and in sub-populations

COMPAS is well-calibrated: in the window around 40%, the fraction of defendants who were re-arrested is ~40%, both over-all and per group.



[plot from Corbett-Davies et al.; *KDD 2017*]

Group fairness impossibility result

[A. Chouldechova; arXiv:1610.07524v1 *(2017)*]

If a predictive instrument **satisfies predictive parity**, but the **prevalence** of the phenomenon **differs between groups**, then the instrument **cannot achieve** equal false positive rates and equal false negative rates across these groups

Recidivism rates in the ProPublica dataset are higher for the black group than for the white group

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.)*

Fairness for whom?

based on a slide by Arvind Narayanan

Decision-maker: of those I've labeled high-risk, how many will recidivate?

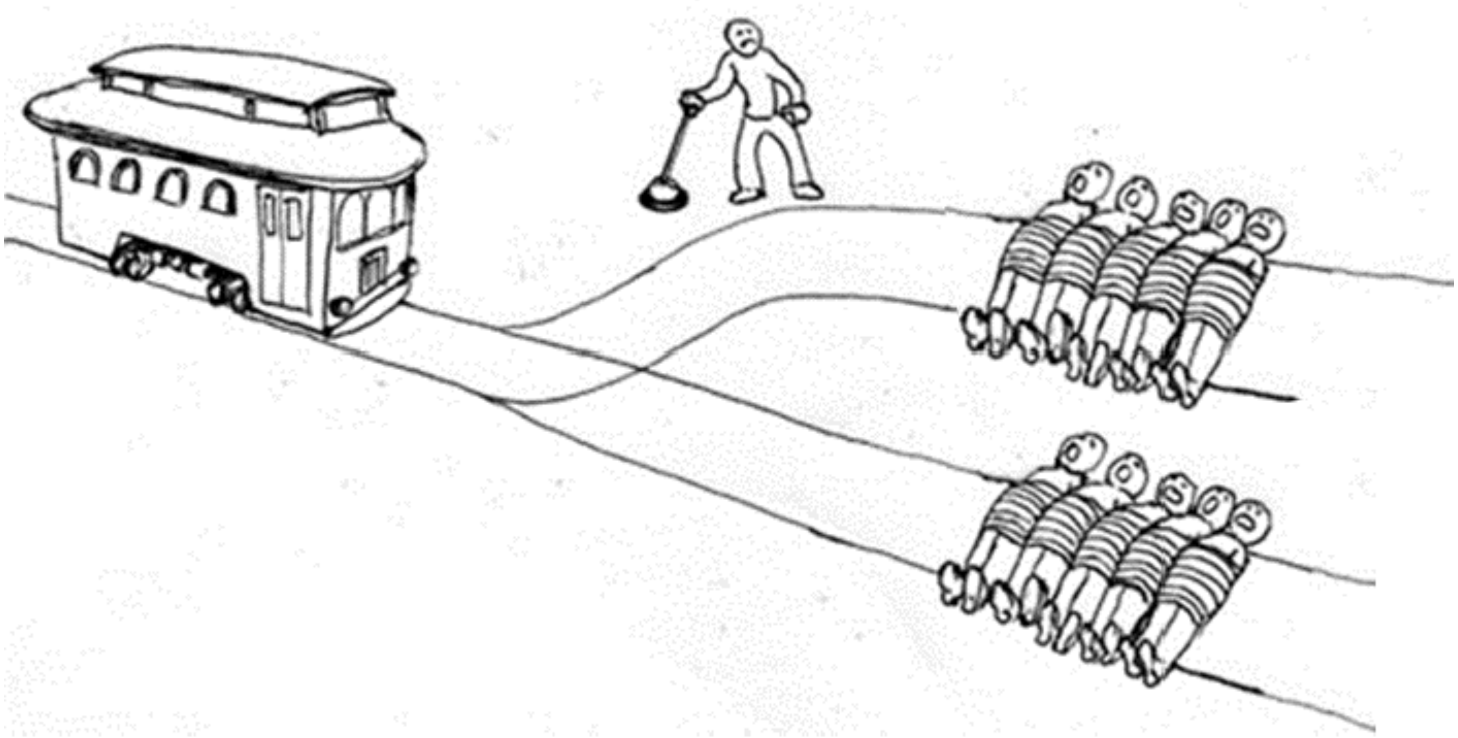
Defendant: how likely am I to be incorrectly classified high-risk?

	labeled low-risk	labeled high-risk
did not recidivate	TN	**FP**
recidivated	**FN**	TP

different metrics matter to different stakeholders

https://www.propublica.org/article/propublica-responds-to-companys-critique-of-machine-bias-story

Fairness definitions as “trolley problems”



https://www.helpage.org/silo/images/blogs/16_1391611056.gif

towards algorithmic transparency

Point 1

algorithmic transparency is not synonymous with releasing the source code

publishing source code helps, but it is sometimes unnecessary and often insufficient

Point 2

algorithmic transparency requires data transparency

data is used in training, validation, deployment

validity, accuracy, applicability can only be understood in the data context

data transparency is necessary for all ADS, not only for ML-based systems

Point 3

data transparency is not synonymous with making all data public

release data whenever possible;

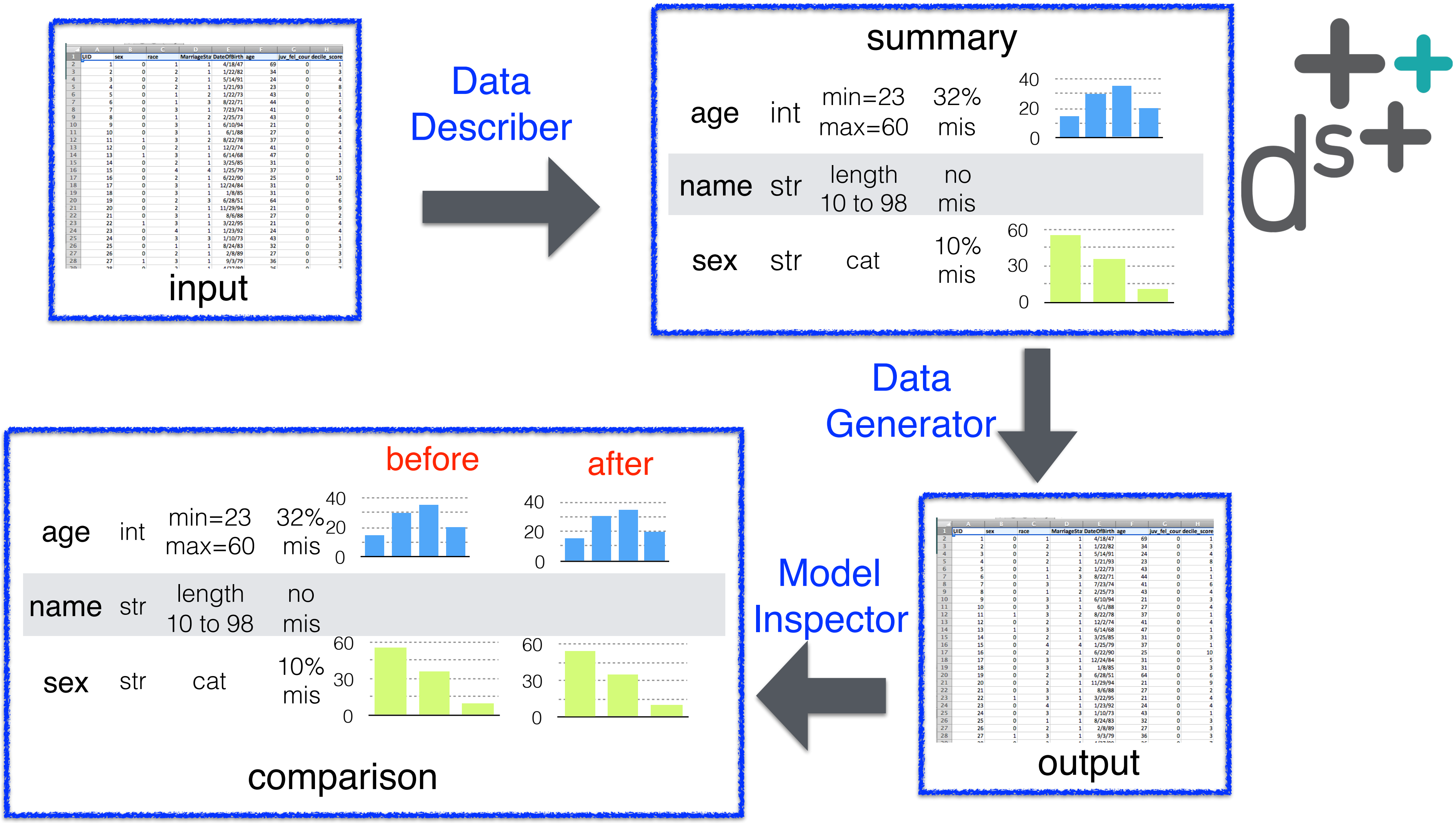
also release:

data selection, collection and pre-processing methodologies; data provenance and quality information; known sources of bias; privacy-preserving statistical summaries of the data

Data Synthesizer

[Ping, Stoyanovich, Howe **SSDBM 2017**]

http://demo.dataresponsibly.com/synthesizer/



Data Synthesizer

MAGAZINE EVENTS PAPERS TOPICS GOVTECH BIZ NAVIGATOR

gt government technology

MetroLab "Innovation of the Month"

SECURITY

University Researchers Use 'Fake' Data for Social Good

BY BEN LEVINE / NOVEMBER 7, 2017

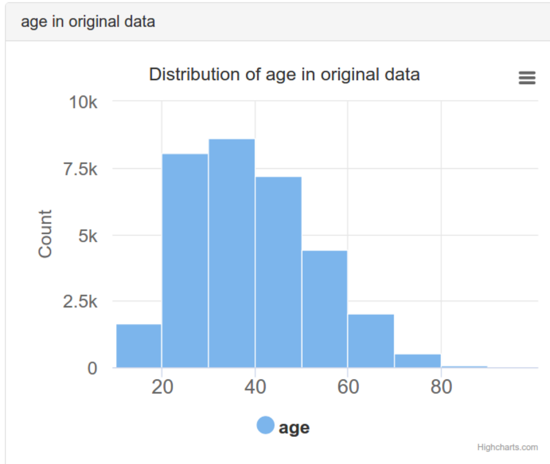
Virtually every interaction we have with a public agency creates a data point. Amass enough data points and they can tell a story. However, factors like privacy, data storage and usability present challenges for local governments and researchers interested in helping improve services. In this installment of MetroLab's Innovation of the Month series, we highlight how researchers at Data Responsibly are addressing those challenges by creating synthetic data sets for social good.

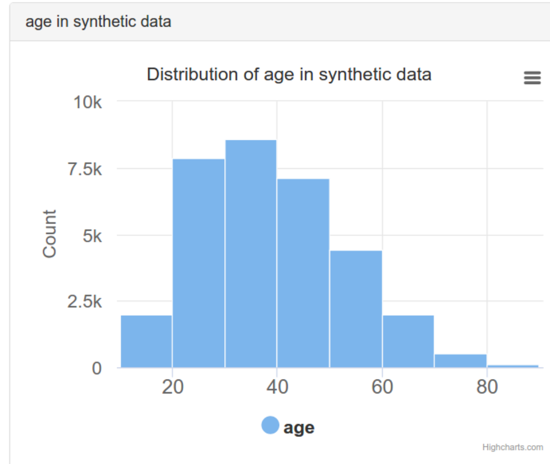
Since its development, the tool has been receiving a lot of attention. For example: T-Mobile is interested in generating synthetic data to better engage with researchers and improve transparency for customers, the Colorado Department of Education has asked relevant agencies to use the tool to experiment with sharing sensitive data, and Elsevier is interested in using the tool to generate synthetic citation networks for research.

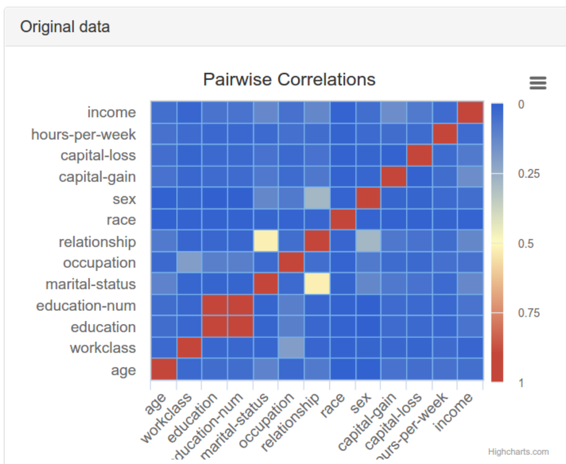
http://www.govtech.com/security/University-Researchers-Use-Fake-Data-for-Social-Good.html

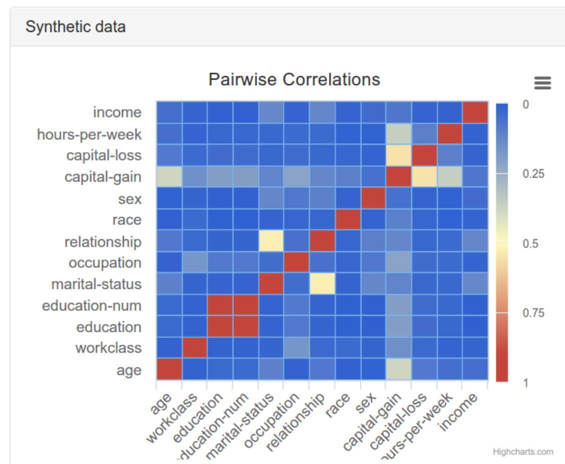
Data Synthesizer

[Ping, Stoyanovich, Howe **SSDBM 2017**]









http://demo.dataresponsibly.com/synthesizer/

Point 4

actionable transparency requires interpretability

explain assumptions and effects, not details of operation

engage the public - technical and non-technical

Transparency with “nutritional labels”

[Yang, Stoyanovich et al. ACM **SIGMOD 2018**]

Recipe

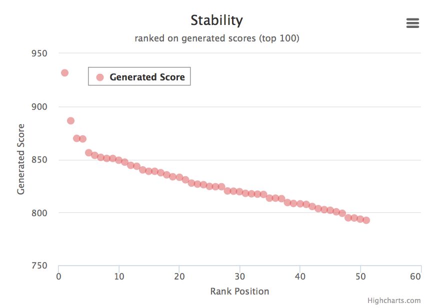
Top 10:

Attribute	Maximum	Median	Minimum
PubCount	18.3	9.6	6.2
Faculty	122	52.5	45
GRE	800.0	796.3	771.9

Overall:

Attribute	Maximum	Median	Minimum
PubCount	18.3	2.9	1.4
Faculty	122	32.0	14
GRE	800.0	790.0	757.8

Stability



Slope at top-10: -6.91. Slope overall: -1.61.

Unstable when absolute value of slope of fit line in scatter plot <= 0.25 (slope threshold). Otherwise it is stable.

Ranking Facts

Recipe

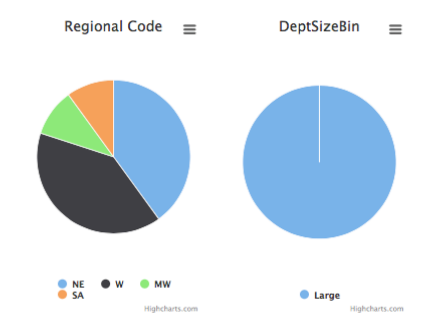
Attribute	Weight
PubCount	1.0
Faculty	1.0
GRE	1.0

Ingredients

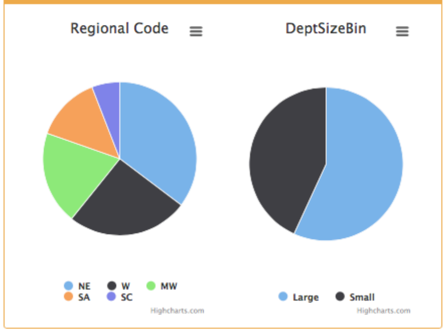
Attribute	Correlation
PubCount	1.0
CSRankingAllArea	0.24
Faculty	0.12

Correlation strength is based on its absolute value. Correlation over 0.75 is high, between 0.25 and 0.75 is medium, under 0.25 is low.

Diversity at top-10



Diversity overall



Stability

Top-K	Stability
Top-10	Stable
Overall	Stable

Fairness

DeptSizeBin	FA*IR	Pairwise	Proportion
Large	Fair	Fair	Fair
Small	Unfair	Unfair	Unfair

Unfair when p-value of corresponding statistical test <= 0.05.

Ingredients

Top 10:

Attribute	Maximum	Median	Minimum
PubCount	18.3	9.6	6.2
CSRankingAllArea	13	6.5	1
Faculty	122	52.5	45

Overall:

Attribute	Maximum	Median	Minimum
PubCount	18.3	2.9	1.4
CSRankingAllArea	48	26.0	1
Faculty	122	32.0	14

Fairness

	FA*IR		Pairwise		Proportion	
DeptSizeBin	p-value	adjusted α	p-value	α	p-value	α
Large	1.0	0.87	0.99	0.05	1.0	0.05
Small	0.0	0.71	0.0	0.05	0.0	0.05

Top K = 26 in FA*IR and Proportion oracles. Setting of top K: In FA*IR and Proportion oracle, if N > 200, set top K =100. Otherwise set top K = 50%N. Pairwise oracle takes whole ranking as input. FA*IR is computed as using code in FA*IR codes. Proportion is implemented as statistical test 4.1.3 in Proportion paper.

http://demo.dataresponsibly.com/rankingfacts/nutrition_facts/

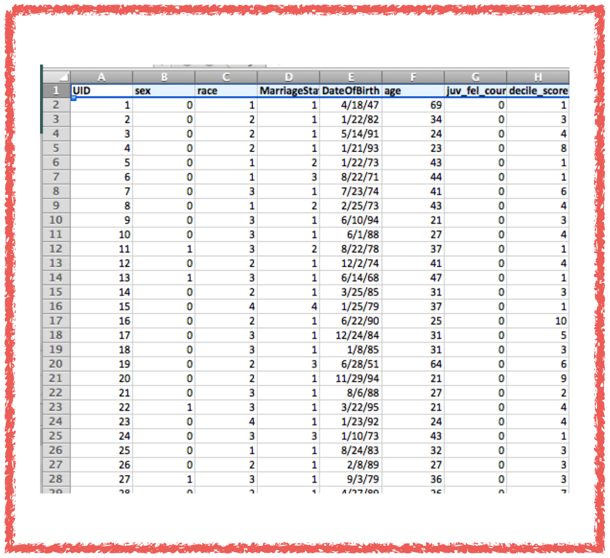
Point 5

transparency by design, not as an afterthought

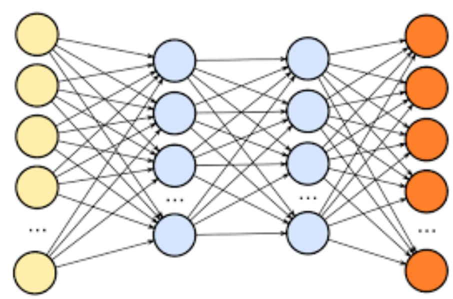
provision for transparency and interpretability at every stage of the data lifecycle

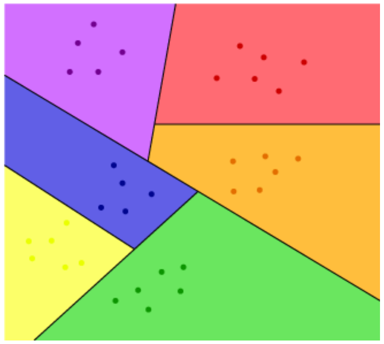
useful internally during development, for communication and coordination between agencies, and for accountability to the public

Frog’s eye view



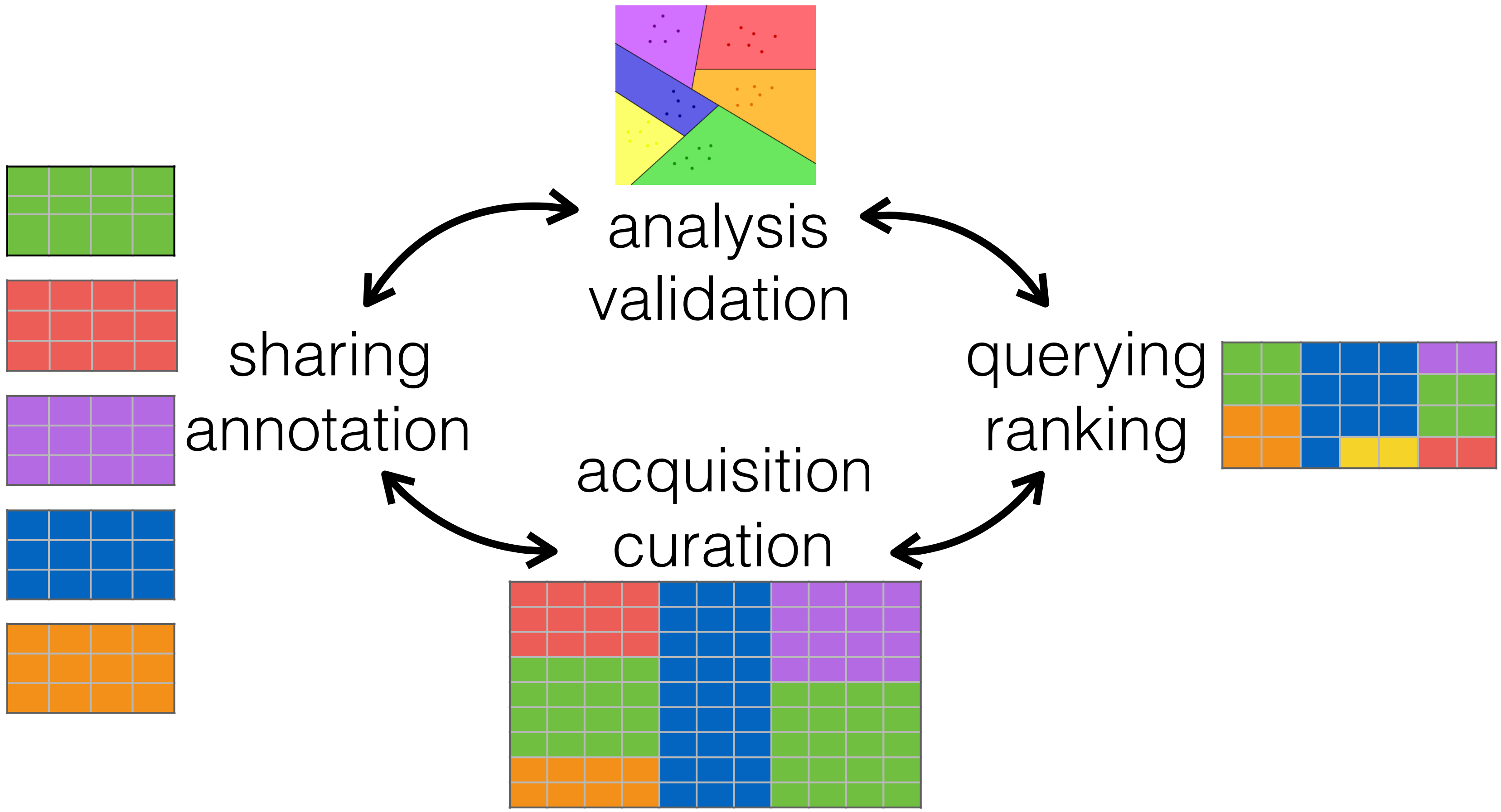
	A	B	C	D	E	F	G	H
1	UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_cour	decile_score
2	1	0	1	1	4/18/47	69	0	1
3	2	0	2	1	1/22/82	34	0	3
4	3	0	2	1	5/14/91	24	0	4
5	4	0	2	1	1/21/93	23	0	8
6	5	0	1	2	1/22/73	43	0	1
7	6	0	1	3	8/22/71	44	0	1
8	7	0	3	1	7/23/74	41	0	6
9	8	0	1	2	2/25/73	43	0	4
10	9	0	3	1	6/10/94	21	0	3
11	10	0	3	1	6/1/88	27	0	4
12	11	1	3	2	8/22/78	37	0	1
13	12	0	2	1	12/2/74	41	0	4
14	13	1	3	1	6/14/68	47	0	1
15	14	0	2	1	3/25/85	31	0	3
16	15	0	4	4	1/25/79	37	0	1
17	16	0	2	1	6/22/90	25	0	10
18	17	0	3	1	12/24/84	31	0	5
19	18	0	3	1	1/8/85	31	0	3
20	19	0	2	3	6/28/51	64	0	6
21	20	0	2	1	11/29/94	21	0	9
22	21	0	3	1	8/6/88	27	0	2
23	22	1	3	1	3/22/95	21	0	4
24	23	0	4	1	1/23/92	24	0	4
25	24	0	3	3	1/10/73	43	0	1
26	25	0	1	1	8/24/83	32	0	3
27	26	0	2	1	2/8/89	27	0	3
28	27	1	3	1	9/3/79	36	0	3





but where does the data come from?

The data science lifecycle



responsible data science requires a holistic view of the data lifecycle

Responsibility by design

Fides

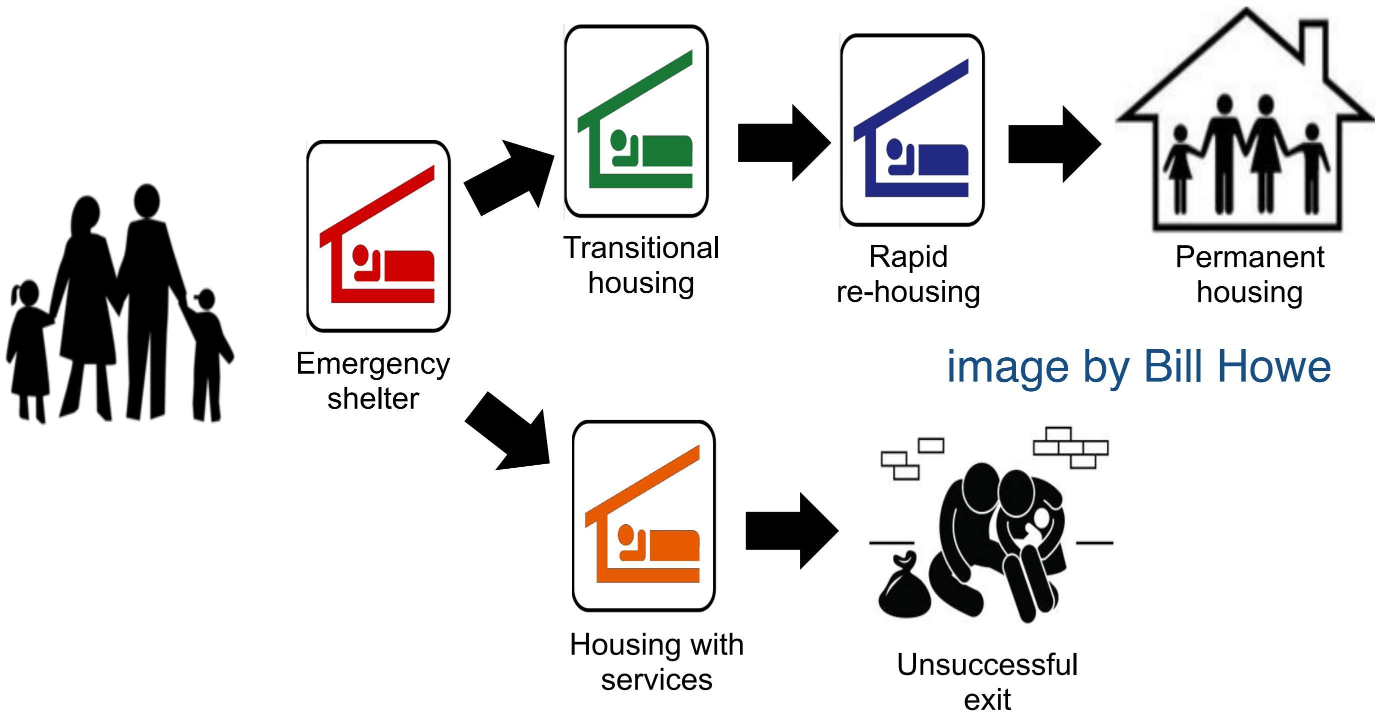
Sharing and Curation	Annotation Anonymization
Integration	Triage Alignment Transformation
Processing	Querying Ranking Analytics
Verification and compliance	Provenance Explanations

Systems support for responsible data science

Responsibility by design, managed at all stages of the lifecycle of data-intensive applications

[BIGDATA] Foundations of responsible data management 09/2017-

ADS example



- **Allocate** interventions: services and support mechanisms
- **Recommend** pathways through the system
- **Evaluate** effectiveness of interventions, pathways, over-all system

How do we get the data?

- A multitude of datasets gathered from local communities, data is **weakly structured**: inconsistencies, missing values, hidden and apparent bias
- Some data was **anonymized**, other data was **not shared** in fear of violating regulations or the trust of participants
- Shared data was **triaged**, **aligned**, **integrated** (ETL + SQL)
- Integrated data was then **filtered** (SQL) and **prioritized** (sorted/ranked), and only then passed as input to the learning module

Mayor de Blasio Scrambles to Curb Homelessness After Years of Not Keeping Pace

By J. DAVID GOODMAN and NIKITA STEWART JAN. 13, 2017

Volunteers during the homeless census in February 2015. In a decision made by Mayor Bill de Blasio, New York City stopped opening shelters for much of that year. Stephanie Keith for The New York Times

The New York Times

https://www.nytimes.com/2017/01/13/nyregion/mayor-de-blasio-scrambles-to-curb-homelessness-after-years-of-not-keeping-pace.html

Ms. Glen emphasized that the construction of new housing takes several years, a long-term solution whose effect on homelessness could not yet be evaluated.

Homeless Young People of New York, Overlooked and Underserved

By NIKITA STEWART FEB. 5, 2016

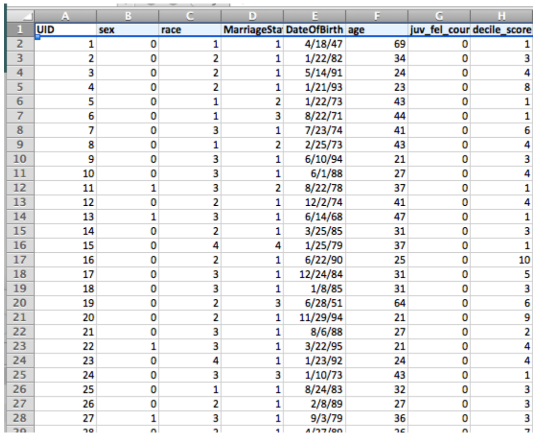
Abdul, 23, at Safe Horizon in Harlem, has been homeless since 2010. Jake Naughto

The New York Times

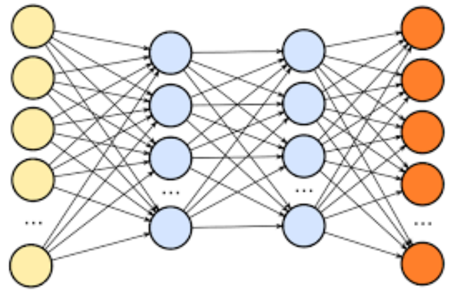
https://www.nytimes.com/2016/02/06/nyregion/young-and-homeless-in-new-york-overlooked-and-underserved.html

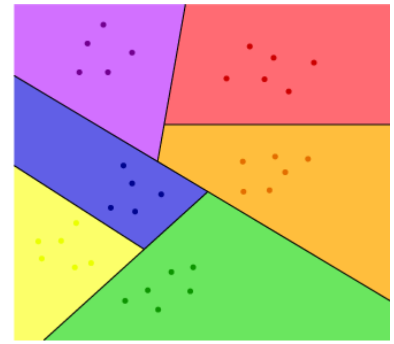
Last year, the total number of sheltered and unsheltered homeless people in the city was 75,323, which included 1,706 people between ages 18 and 24. The actual number of young people is significantly higher, according to the service providers, who said the census mostly captured young people who received social services. The census takers were not allowed to enter private businesses, including many of the late-night spots where young people often create an ad hoc shelter by pretending to be customers.

Mitigating urban homelessness



	A	B	C	D	E	F	G	H
1	UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_cour	decile_score
2	1	0	1	1	4/18/47	69	0	1
3	2	0	2	1	1/22/82	34	0	3
4	3	0	2	1	5/14/91	24	0	4
5	4	0	2	1	1/21/93	23	0	8
6	5	0	1	2	1/22/73	43	0	1
7	6	0	1	3	8/22/71	44	0	1
8	7	0	3	1	7/23/74	41	0	6
9	8	0	1	2	2/25/73	43	0	4
10	9	0	3	1	6/10/94	21	0	3
11	10	0	3	1	6/1/88	27	0	4
12	11	1	3	2	8/22/78	37	0	1
13	12	0	2	1	12/2/74	41	0	4
14	13	1	3	1	6/14/68	47	0	1
15	14	0	2	1	3/25/85	31	0	3
16	15	0	4	4	1/25/79	37	0	1
17	16	0	2	1	6/22/90	25	0	10
18	17	0	3	1	12/24/84	31	0	5
19	18	0	3	1	1/8/85	31	0	3
20	19	0	2	3	6/28/51	64	0	6
21	20	0	2	1	11/29/94	21	0	9
22	21	0	3	1	8/6/88	27	0	2
23	22	1	3	1	3/22/95	21	0	4
24	23	0	4	1	1/23/92	24	0	4
25	24	0	3	3	1/10/73	43	0	1
26	25	0	1	1	8/24/83	32	0	3
27	26	0	2	1	2/8/89	27	0	3
28	27	1	3	1	9/3/79	36	0	3





finding: women are underrepresented in the favorable outcome groups (group fairness)

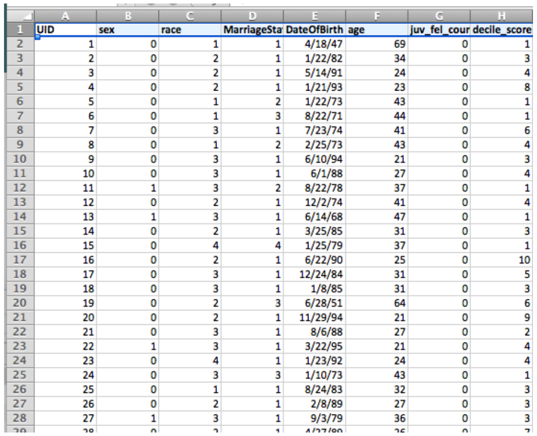
fix the model!

of course, but maybe… the input was generated with:

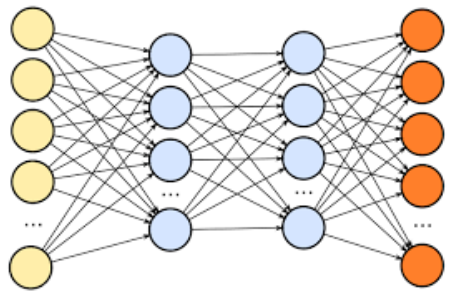
select * from R
where status = 'unsheltered'
and length > 2 month

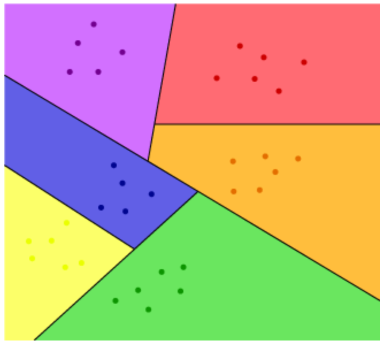
10% female

Mitigating urban homelessness



UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_cour	decile_score
1	0	1	1	4/18/47	69	0	1
2	0	2	1	1/22/82	34	0	3
3	0	2	1	5/14/91	24	0	4
4	0	2	1	1/21/93	23	0	8
5	0	1	2	1/22/73	43	0	1
6	0	1	3	8/22/71	44	0	1
7	0	3	1	7/23/74	41	0	6
8	0	1	2	2/25/73	43	0	4
9	0	3	1	6/10/94	21	0	3
10	0	3	1	6/1/88	27	0	4
11	1	3	2	8/22/78	37	0	1
12	0	2	1	12/2/74	41	0	4
13	1	3	1	6/14/68	47	0	1
14	0	2	1	3/25/85	31	0	3
15	0	4	4	1/25/79	37	0	1
16	0	2	1	6/22/90	25	0	10
17	0	3	1	12/24/84	31	0	5
18	0	3	1	1/8/85	31	0	3
19	0	2	3	6/28/51	64	0	6
20	0	2	1	11/29/94	21	0	9
21	0	3	1	8/6/88	27	0	2
22	1	3	1	3/22/95	21	0	4
23	0	4	1	1/23/92	24	0	4
24	0	3	3	1/10/73	43	0	1
25	0	1	1	8/24/83	32	0	3
26	0	2	1	2/8/89	27	0	3
27	1	3	1	9/3/79	36	0	3





finding: women are underrepresented in the favorable outcome groups (group fairness)

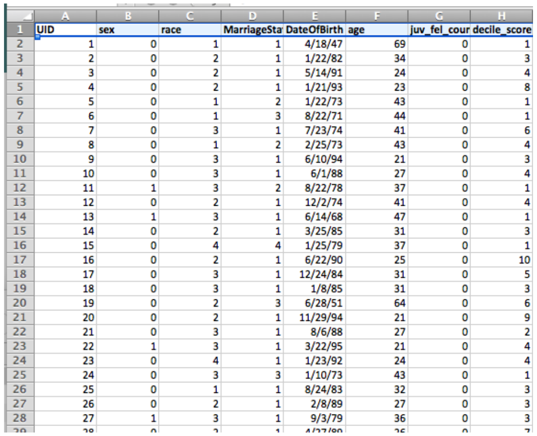
fix the model!

of course, but maybe… the input was generated with:

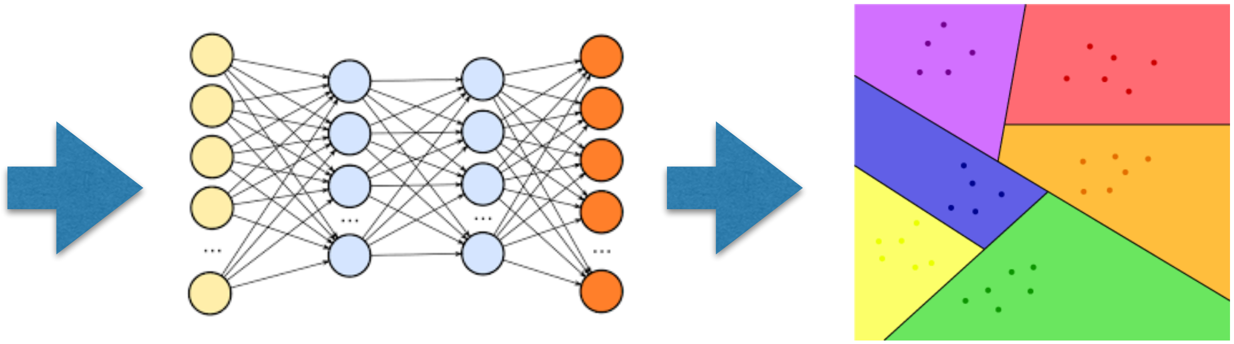
select * from R
where status = 'unsheltered'
and length > **1 month**

40% female

Mitigating urban homelessness



	A	B	C	D	E	F	G	H
1	UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_cour	decile_score
2	1	0	1	1	4/18/47	69	0	1
3	2	0	2	1	1/22/82	34	0	3
4	3	0	2	1	5/14/91	24	0	4
5	4	0	2	1	1/21/93	23	0	8
6	5	0	1	2	1/22/73	43	0	1
7	6	0	1	3	8/22/71	44	0	1
8	7	0	3	1	7/23/74	41	0	6
9	8	0	1	2	2/25/73	43	0	4
10	9	0	3	1	6/10/94	21	0	3
11	10	0	3	1	6/1/88	27	0	4
12	11	1	3	2	8/22/78	37	0	1
13	12	0	2	1	12/2/74	41	0	4
14	13	1	3	1	6/14/68	47	0	1
15	14	0	2	1	3/25/85	31	0	3
16	15	0	4	4	1/25/79	37	0	1
17	16	0	2	1	6/22/90	25	0	10
18	17	0	3	1	12/24/84	31	0	5
19	18	0	3	1	1/8/85	31	0	3
20	19	0	2	3	6/28/51	64	0	6
21	20	0	2	1	11/29/94	21	0	9
22	21	0	3	1	8/6/88	27	0	2
23	22	1	3	1	3/22/95	21	0	4
24	23	0	4	1	1/23/92	24	0	4
25	24	0	3	3	1/10/73	43	0	1
26	25	0	1	1	8/24/83	32	0	3
27	26	0	2	1	2/8/89	27	0	3
28	27	1	3	1	9/3/79	36	0	3



finding: young people are recommended pathways of lower effectiveness (high error rate)

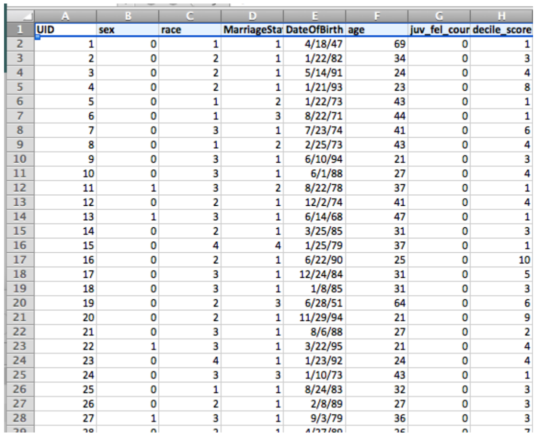
fix the model!

of course, but maybe…

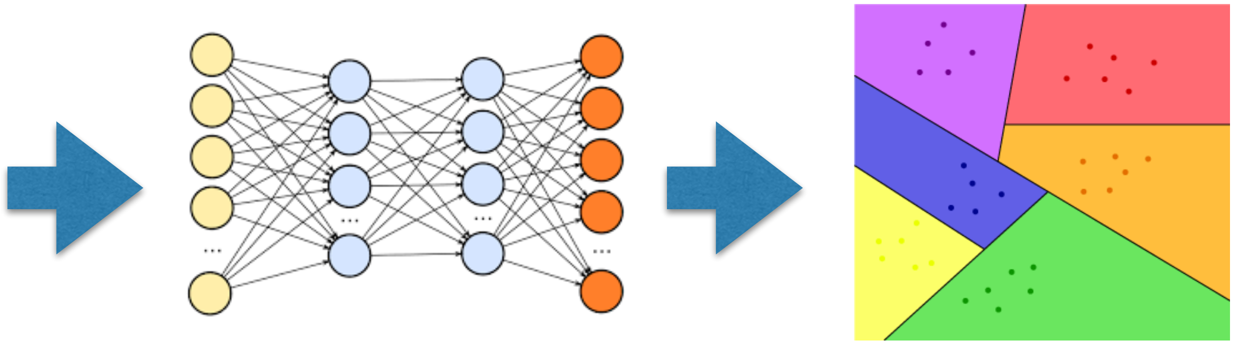
mental health info was missing for this population

go back to the data acquisition step, look for additional datasets

Mitigating urban homelessness



	A	B	C	D	E	F	G	H
1	UID	sex	race	MarriageSta	DateOfBirth	age	juv_fel_cour	decile_score
2	1	0	1	1	4/18/47	69	0	1
3	2	0	2	1	1/22/82	34	0	3
4	3	0	2	1	5/14/91	24	0	4
5	4	0	2	1	1/21/93	23	0	8
6	5	0	1	2	1/22/73	43	0	1
7	6	0	1	3	8/22/71	44	0	1
8	7	0	3	1	7/23/74	41	0	6
9	8	0	1	2	2/25/73	43	0	4
10	9	0	3	1	6/10/94	21	0	3
11	10	0	3	1	6/1/88	27	0	4
12	11	1	3	2	8/22/78	37	0	1
13	12	0	2	1	12/2/74	41	0	4
14	13	1	3	1	6/14/68	47	0	1
15	14	0	2	1	3/25/85	31	0	3
16	15	0	4	4	1/25/79	37	0	1
17	16	0	2	1	6/22/90	25	0	10
18	17	0	3	1	12/24/84	31	0	5
19	18	0	3	1	1/8/85	31	0	3
20	19	0	2	3	6/28/51	64	0	6
21	20	0	2	1	11/29/94	21	0	9
22	21	0	3	1	8/6/88	27	0	2
23	22	1	3	1	3/22/95	21	0	4
24	23	0	4	1	1/23/92	24	0	4
25	24	0	3	3	1/10/73	43	0	1
26	25	0	1	1	8/24/83	32	0	3
27	26	0	2	1	2/8/89	27	0	3
28	27	1	3	1	9/3/79	36	0	3



finding: minors are underrepresented in the input, compared to their actual proportion in the population (insufficient data)

unlikely to help!

fix the model??

minors data was not shared

go back to the data sharing step, help data providers share their data while adhering to laws and upholding the trust of the participants

wrapping up

The punchline

Data science is algorithmic, therefore it cannot be biased! And yet…

- All traditional evils of **discrimination**, and many new ones, exhibit themselves in the data science ecosystem
- **Transparency** helps prevent discrimination, enable public debate, establish **trust**
- Technology alone won't do: also need **regulation** and **civic engagement**

http://www.allenovery.com/publications/en-gb/Pages/Protected-characteristics-and-the-perception-reality-gap.aspx

responsible data science is our new frontier!

Codes of ethics

Digital Library CACM Queue TechNews Learning Center Career Center

Association for Computing Machinery

Advancing Computing as a Science & Profession

Join Volunteer myACM Search

ABOUT ACM MEMBERSHIP PUBLICATIONS SPECIAL INTEREST GROUPS CONFERENCES CHAPTERS AWARDS EDUCATION PUBLIC POLICY GOVERNANCE

Home > Code Of Ethics

ACM Code of Ethics and Professional Conduct

ACM Code of Ethics and Professional Conduct

Preamble

Computing professionals' actions change the world. To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good. The ACM Code of Ethics and Professional Conduct ("the Code") expresses the conscience of the profession.

The Code is designed to inspire and guide the ethical conduct of all computing professionals, including current and aspiring practitioners, instructors, students, influencers, and anyone who uses computing technology in an impactful way. Additionally, the Code serves as a basis for remediation when violations occur. The Code includes principles formulated as statements of responsibility, based on the understanding that the public good is always the primary consideration. Each principle is supplemented by guidelines, which provide explanations to assist computing professionals in understanding and applying the principle.

Section 1 outlines fundamental ethical principles that form the basis for the remainder of the Code. Section 2 addresses additional, more specific considerations of professional responsibility. Section 3 guides individuals who have a leadership role, whether in the workplace or in a volunteer professional capacity. Commitment to ethical conduct is required of every ACM member, and principles involving compliance with the Code are given in Section 4.

The Code as a whole is concerned with how fundamental ethical principles apply to a computing professional's conduct. The Code is not an algorithm for solving ethical problems; rather it serves as a basis for ethical decision-making. When thinking through a particular issue, a computing professional may find that multiple principles should be taken into account, and that different principles will have different relevance to the issue. Questions related to these kinds of issues can best be answered by thoughtful consideration of the fundamental ethical principles, understanding that the public good is the paramount consideration. The entire computing profession benefits when the ethical decision-making process is accountable to and transparent to all stakeholders. Open discussions about ethical issues promote this accountability and transparency.

PDF of the ACM Code of Ethics

On This Page

Codes of ethics

BACK

SUBSCRIBE

Community Principles on Ethical Data Practices

OVERVIEW

BACKGROUND

VALUES

PRINCIPLES

AUTHORS

SIGNATORIES

SIGN

JOIN

This code of ethics for data sharing is created and proposed for adoption by the data science community to reflect the behaviors and principles for the responsible and ethical use and sharing of data by data scientists.

As a community-driven crowdsourced effort, you can join the the discussion and contribute to the next version of the Community Principles on Ethical Data Sharing.

OVERVIEW

The Community Principles on Ethical Data Practices are being developed by people from the data science community in conjunction with data science organizations. These principles focus on defining ethical and responsible behaviors for sourcing, sharing and implementing data in a manner that will cause no harm and maximize positive impact. The goal of this initiative is to develop a community-driven code of ethics for data collection, sharing and utilization that provides people in the data science community a standard set of easily digestible, recognizable principles for guiding their behaviors.

This code is not intended to be all encompassing. Rather, these principles will provide academia, industry, and individual data scientists a common set of guidelines for driving the development of standards, curriculums, and best practices for the ethical use and sharing of data, ultimately advancing the responsible and ethical use of data as a collective force for good.

Three principles

THE BELMONT REPORT

Office of the Secretary

Ethical Principles and Guidelines for the Protection of Human Subjects of Research

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research

April 18, 1979

Respect for persons

Beneficence

Justice

DataResponsibly - Course | Da

https://dataresponsibly.github.io/courses/spring19/

Apps Dropbox Getting Started 05 NSF contacts - G... research administ... Google Calendar -... Other Bookmarks

DS-GA 3001.009: Special Topics in Data Science: Responsible Data Science

New York University, Center for Data Science, Spring 2019

Lecture: Mondays from 11am-12:40pm; **Lab:** Thursdays from 5:20pm-6:10pm

Location: 60 5th Avenue, Room 110

Instructor: Julia Stoyanovich, Assistant Professor of Data Science, Computer Science and Engineering.
Office hours Mondays 1:30-3pm or by appointment, at 60 5th Avenue, Room 605.

Section Leader: Udita Gupta. Office hours Thursdays 4-5pm at 60 5th Avenue, Room 663.

Syllabus: pdf

Course Description:

The first wave of data science focused on accuracy and efficiency – on what we *can* do with data. The second wave focuses on responsibility – on what we *should* and *shouldn't* do. Irresponsible use of data science can cause harm on an unprecedented scale. Algorithmic changes in search engines can sway elections and incite violence; irreproducible results can influence global economic policy; models based on biased data can legitimize and amplify racist policies in the criminal justice system; algorithmic hiring practices can silently and scalably violate equal opportunity laws, exposing companies to lawsuits and reinforcing the feedback loops that lead to lack of diversity. Therefore, as we develop and deploy data science methods, we are compelled to think about the effects these methods have on individuals, population groups, and on society at large.

DataResponsibly - Press

https://dataresponsibly.github.io/press/

Apps Dropbox Getting Started 05 NSF contacts - G... research administ... Google Calendar -... Other Bookmai

data RESPONSIBLY

Home People Publications Press Talks Tools Course

Press

Follow the Data! Algorithmic Transparency Starts with Data Transparency

The data revolution that is transforming every sector of science and industry has been slow to reach the local and municipal governments and NGOs that deliver vital human services in health, housing, and mobility. Urbanization has made the issue acute in 2016, more than half of North Americans lived in cities with at least 500,000 inhabitants.

Julia Stoyanovich and Bill Howe

The Ethical Machine, November 27, 2018

An Algorithmic Approach to Correct Bias in Urban Transportation Datasets

While a significant amount of attention and research has addressed individual privacy concerns in private companies' datasets, data owners and publishers also want to avoid revealing certain patterns—even in anonymized datasets—that might compromise a competitive advantage or perpetuate discrimination against any group of people. Data published by urban transportation companies is highly valuable for research, policy, and public accountability.

NYU Center for Data Science, October 30, 2018

Thank you!

dataresponsibly.github.io

@stoyanoj

NYU

General Data Protection Regulation
GDPR

Welcome to gdpr-info.eu. Here you can find the official PDF of the Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018 as a neatly arranged website. All Articles of the GDPR are linked with suitable recitals. The European Data Protection Regulation is applicable as of May 25th, 2018 in all member states to harmonize data privacy laws across Europe. If you find the page useful, feel free to support us by sharing the project.

Quick Access