# 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

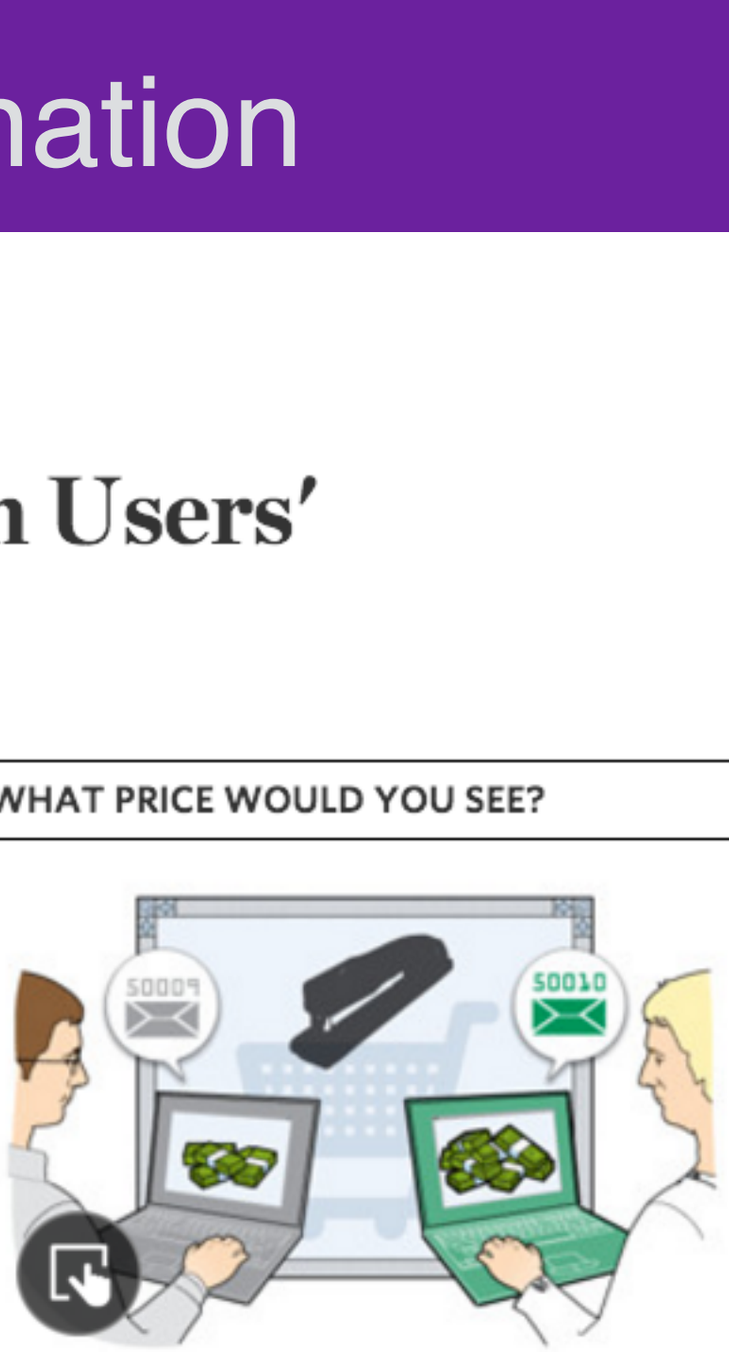
By JENNIFER VALENTINO-DEVRIES, JEREMY SINGER-VINE and ASHKAN SOLTANI

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/SB10001424127887323772045781893918138815 34

Bloomberg

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

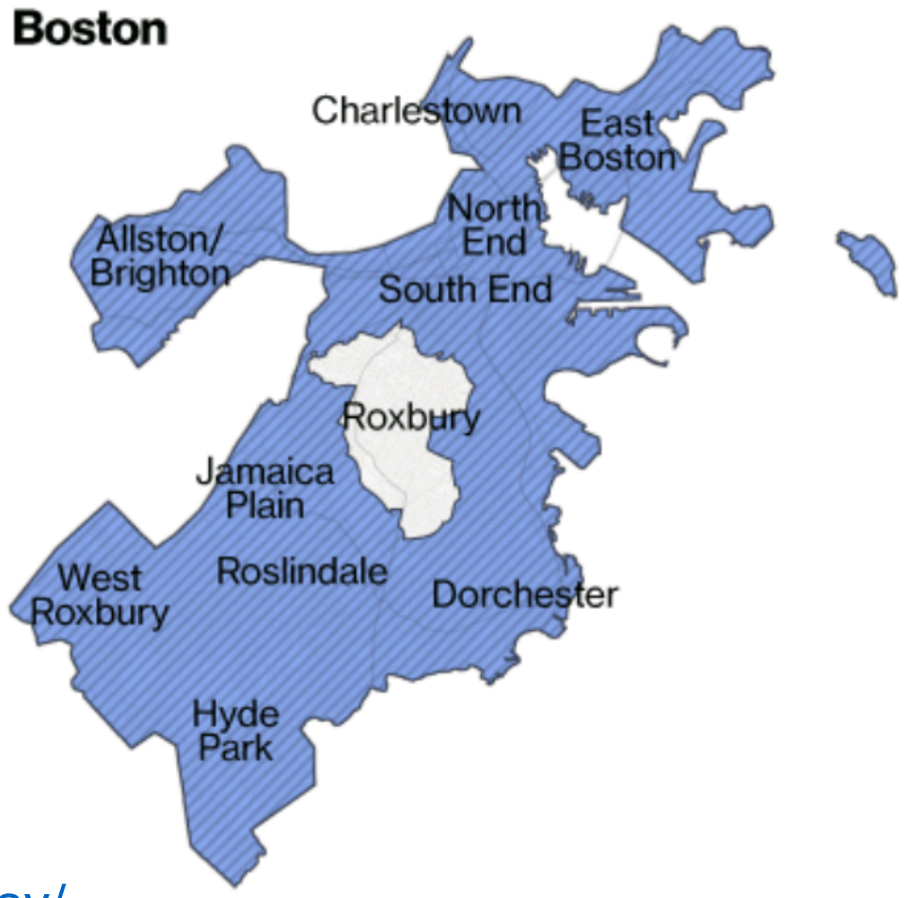
"… In six major same-day delivery cities, however, **the service area excludes predominantly black ZIP codes** to varying degrees, according to a Bloomberg analysis that compared Amazon same-day delivery areas with U.S. Census Bureau data."

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

Bloomberg

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

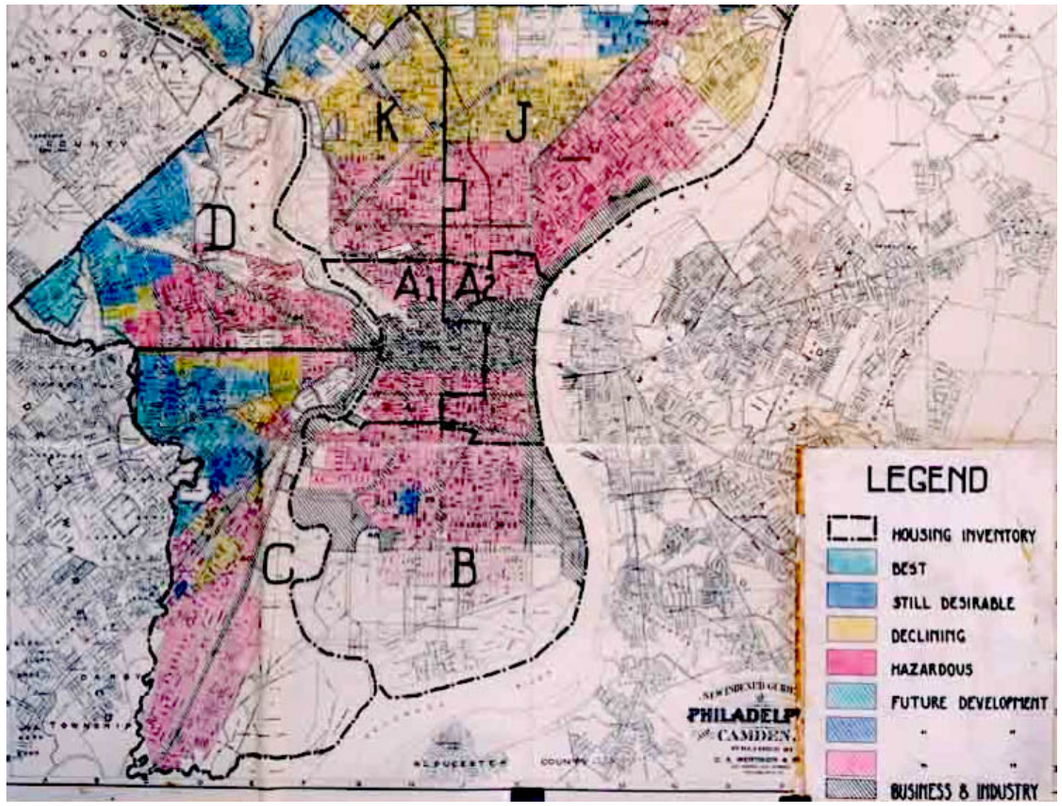
"The most striking gap in Amazon's same-day service is in Boston, where **three ZIP codes encompassing the primarily black neighborhood of Roxbury are excluded** from same-day service, while the neighborhoods that surround it on all sides are eligible."




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.

# Job-screening tests

THE WALL STREET JOURNAL.

## Are Workplace Personality Tests Fair?

Growing Use of Tests Sparks Scrutiny Amid Questions of Effectiveness and Workplace Discrimination

By LAUREN WEBER and ELIZABETH DWOSKIN

Sept. 29, 2014 10:30 p.m. ET

Kyle Behm accused Kroger and six other companies of discrimination against the mentally ill through their use of personality tests. *TROY STAINS FOR THE WALL STREET JOURNAL*

The Equal Employment Opportunity commission is **investigating whether personality tests discriminate against people with disabilities**.

As part of the investigation, officials are trying to determine if the tests **shut out people suffering from mental illnesses** such as depression or bipolar disorder, even if they have the right skills for the job.

http://www.wsj.com/articles/are-workplace-personality-tests-fair-1412044257

# Online job ads

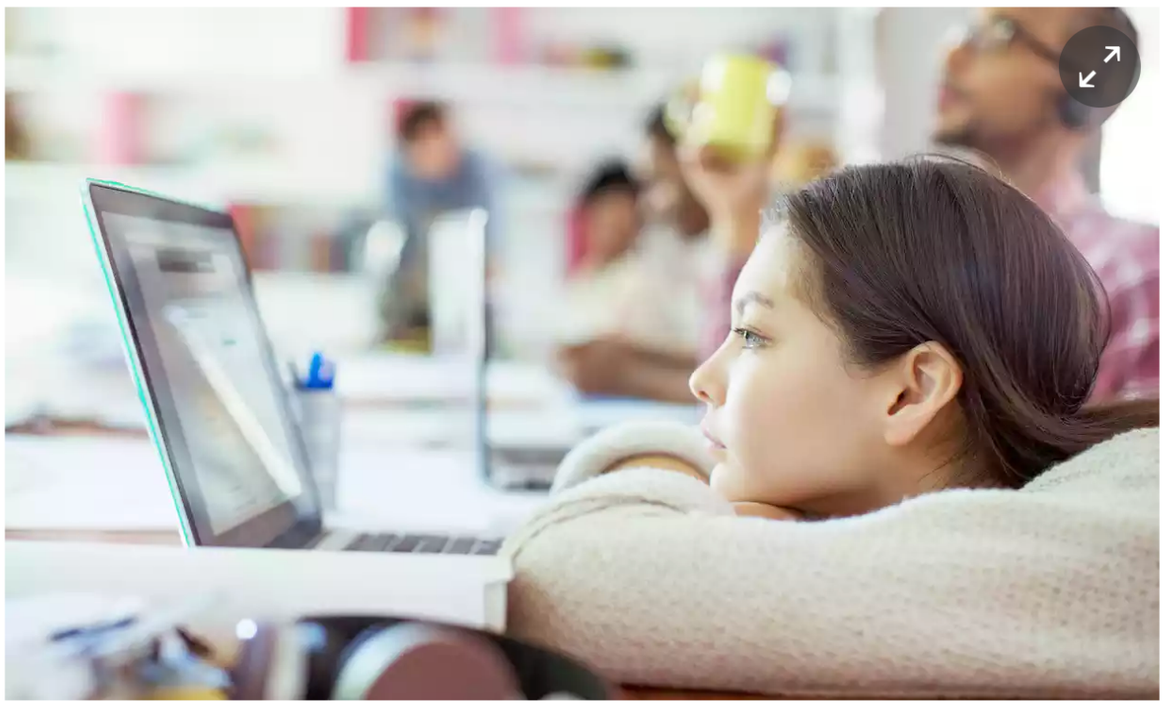
theguardian

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

# 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

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

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.

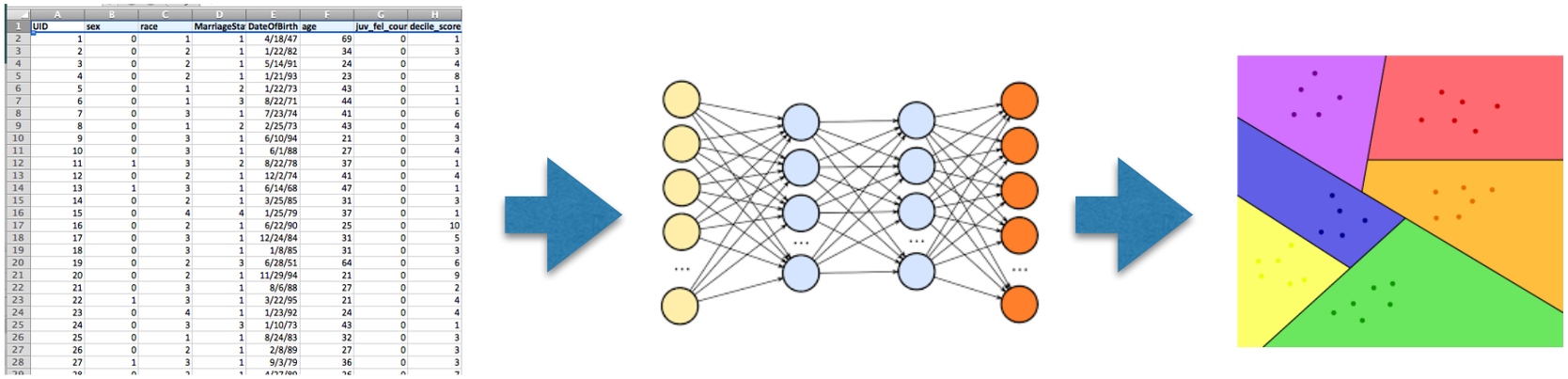
The tool correctly predicts recidivism **61% of the time.**

**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

# “Bias” in predictive analytics



| | 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 |

- **Statistical bias in the model**: a model is biased if it does not 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?**

when data is about people, bias can lead to discrimination

# The evils of discrimination

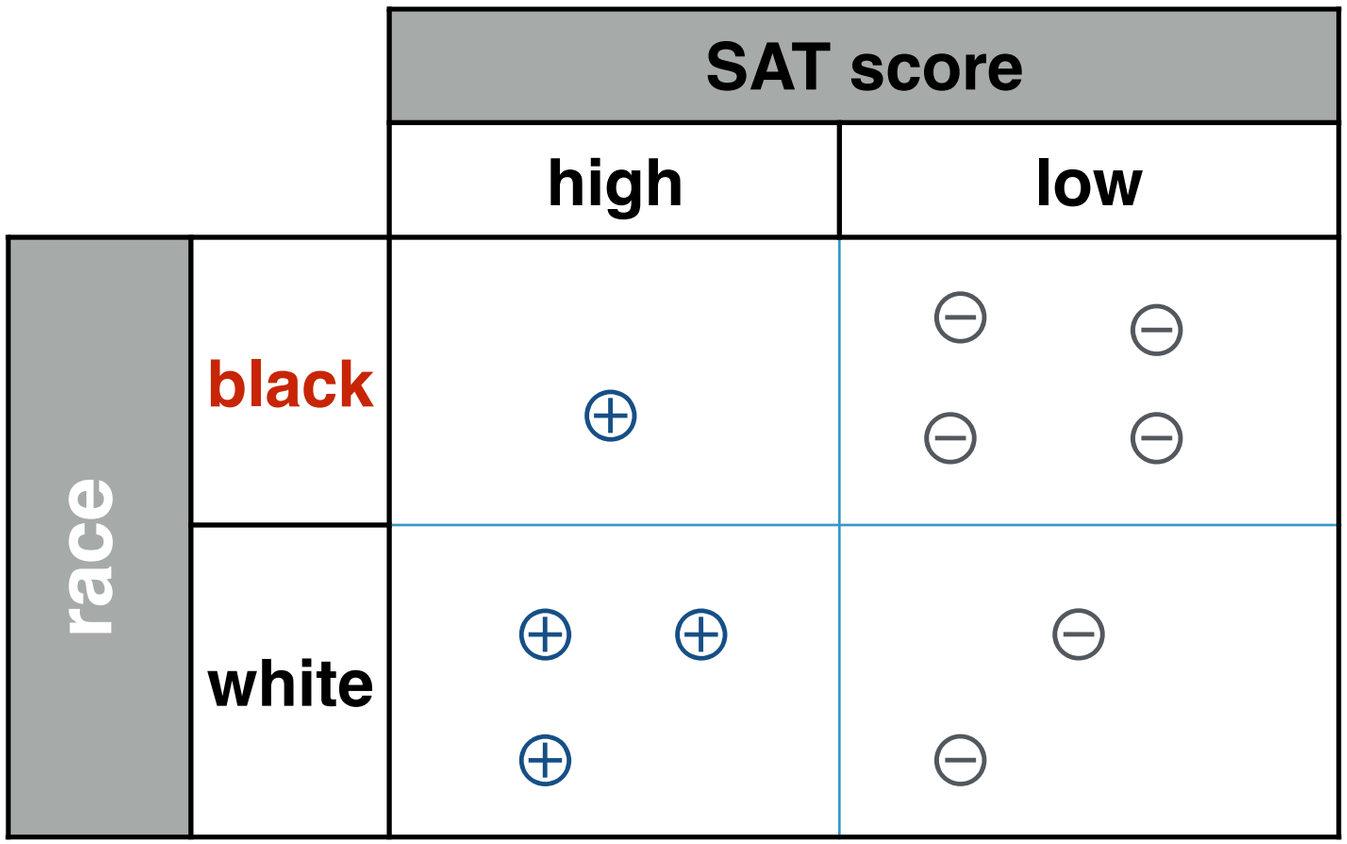
**Disparate treatment** is the illegal practice of treating an entity, such as a creditor or employee, differently based on a **protected characteristic** such as race, gender, age, religion, sexual orientation, or national origin.

**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

# 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



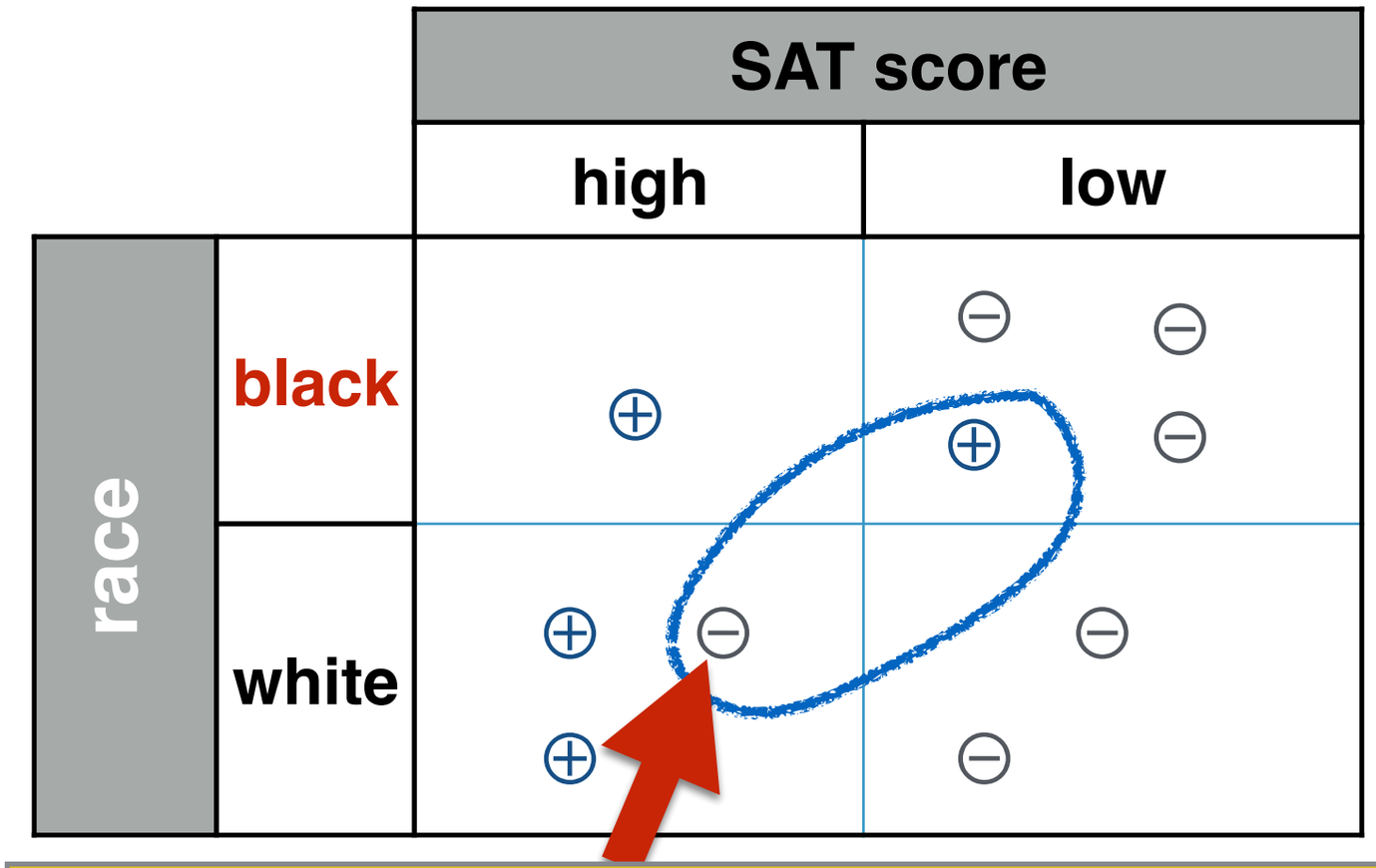

**positive outcomes**

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




**positive outcomes**

40% of black

40% of white

**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

| 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. | |

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

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

# Summary of Local Law 49

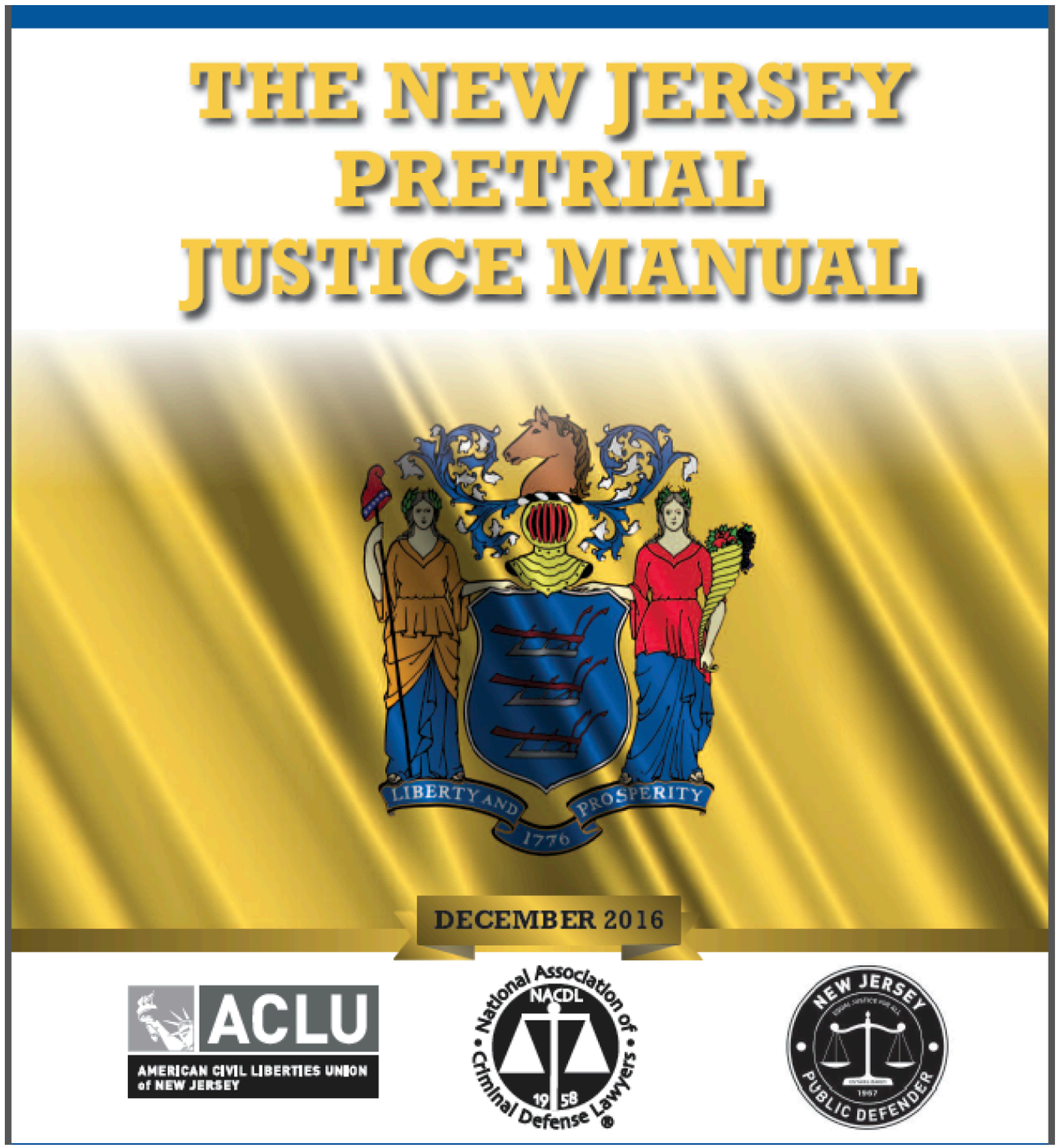
1/11/2018

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

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

# fairness is risk assessment

# 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

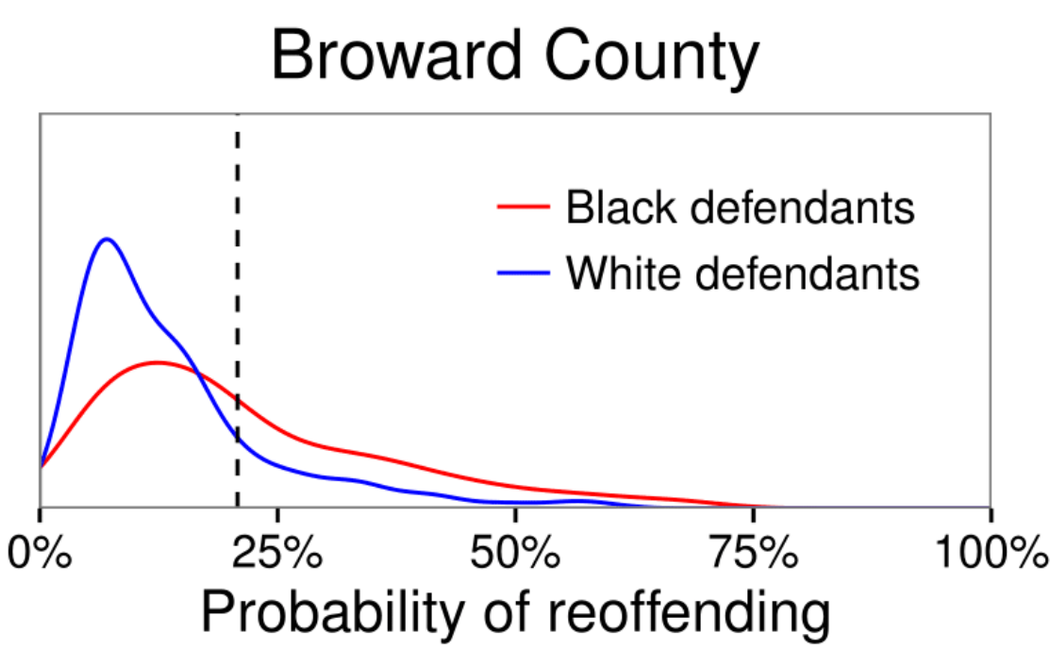
- A risk assessment tool **gives a probability estimate of a future outcome**
- Used in many domains:
  - insurance, criminal sentencing, **medical testing**, hiring, banking
  - also in less-obvious set-ups, like online advertising
- **Fairness** in risk assessment is concerned with **how different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

# An 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?

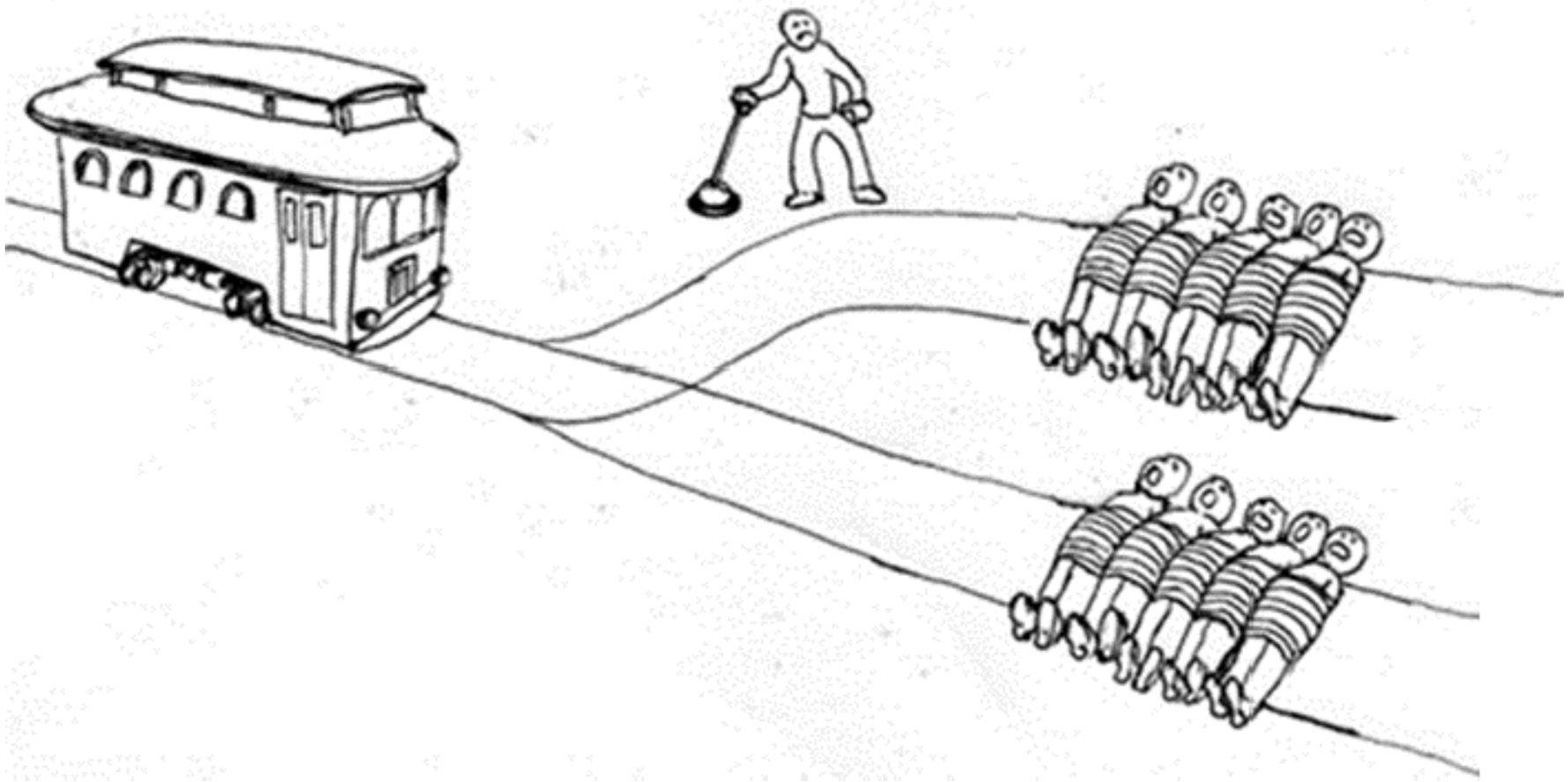
**Defendant**: how likely am I to be incorrectly classified high-risk?

| | 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

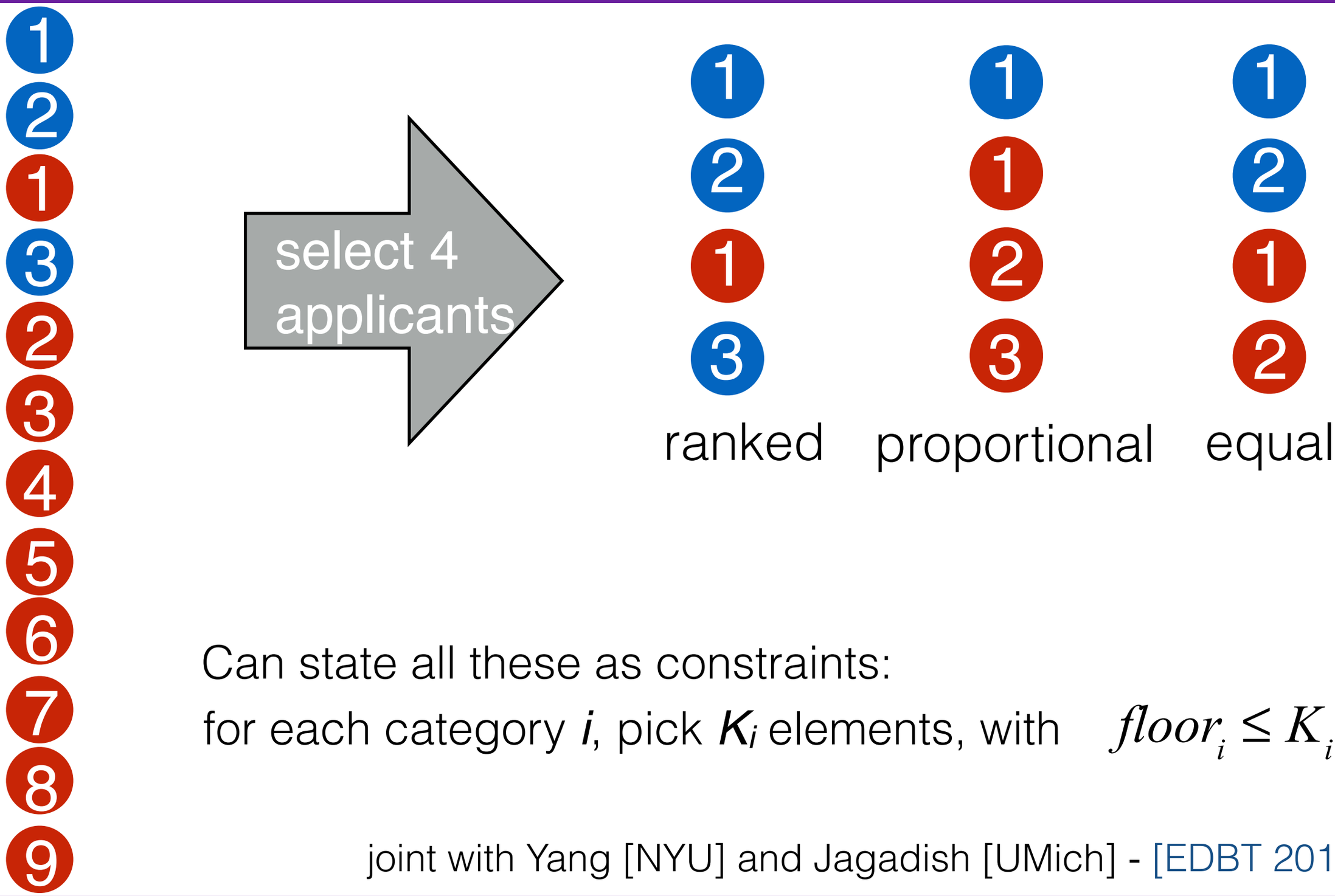
# Fairness & diversity: zooming out

Many reasonable fairness and diversity notions have been considered in philosophy, law, social science, …. , computer science, data science

- Yes, we can state fairness and diversity measures mathematically and operationalize them in code
- **No, there does not exist a single measure to rule them all, so we will always need to consider trade-offs**
- We should be careful to decouple our **beliefs** about what is or is not fair from **mechanisms** consistent with those beliefs

# technical highlights

# Diversity in set selection



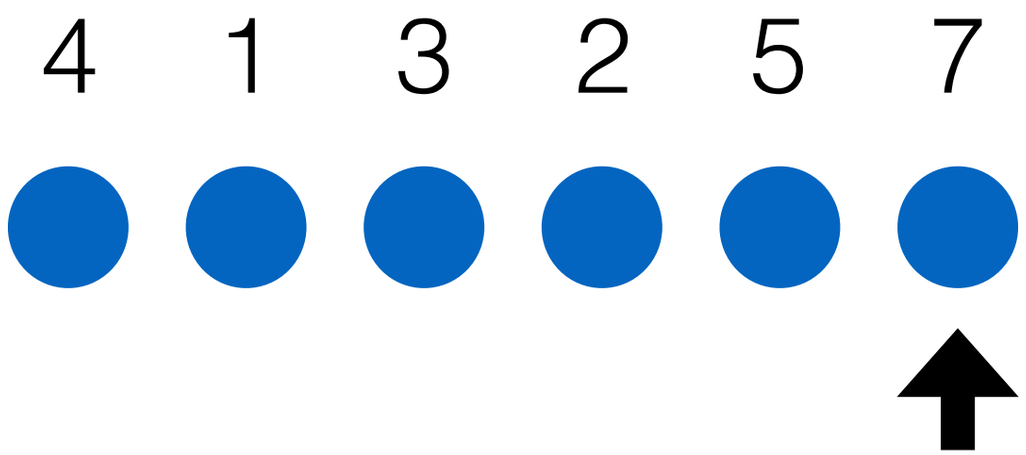

Can state all these as constraints:

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

joint with Yang [NYU] and Jagadish [UMich] - [EDBT 2018]

# Hiring a job candidate

**Goal**: Hire a candidate with a high score

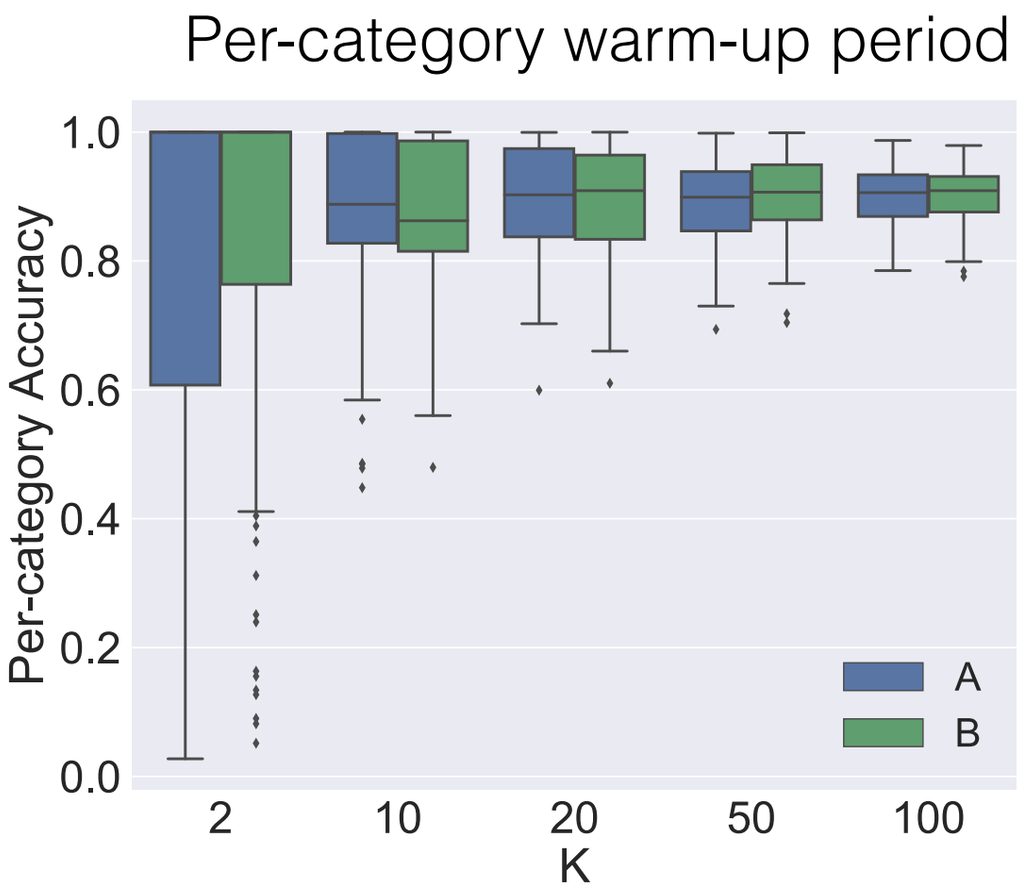


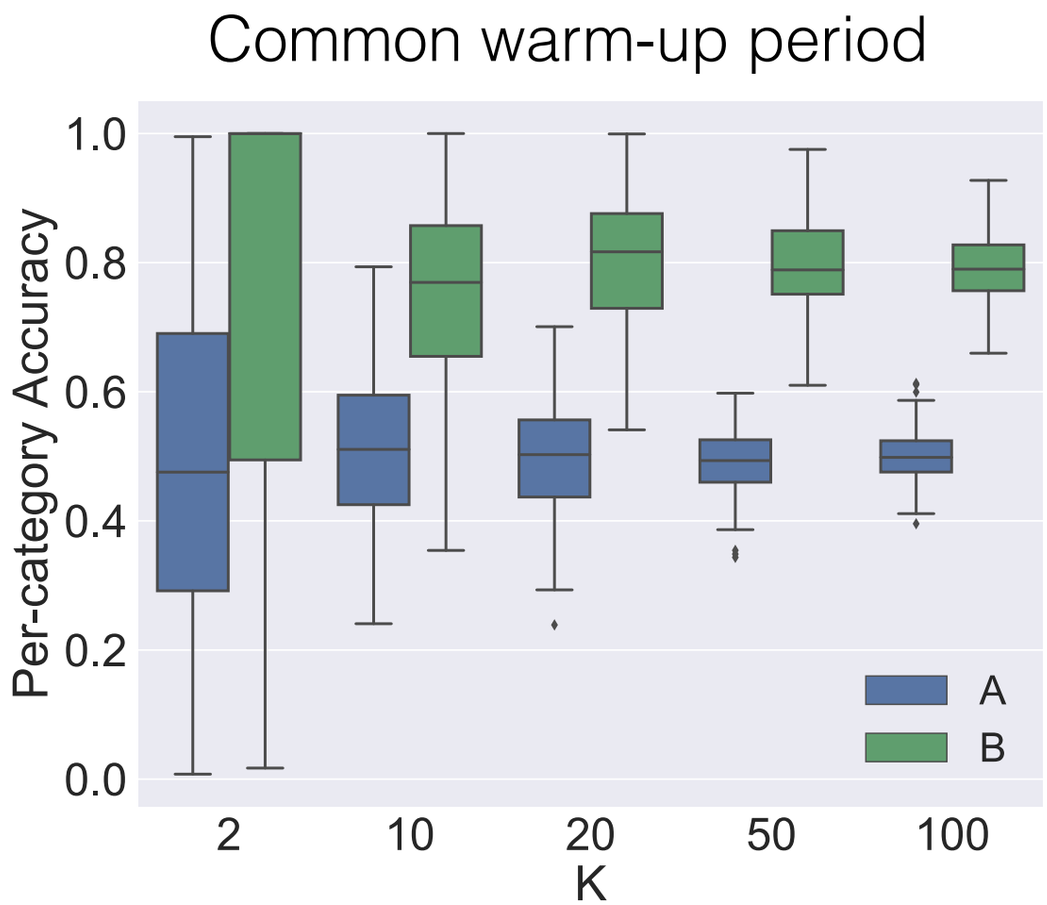

Candidates arrive one-by-one

A candidate's score is revealed when the candidate arrives

Decision to accept or reject a candidate made on the spot

# Per-category warm-up is crucial




synthetic data with categories A and B, score depends on category, lower for A

diversity by design

joint with Yang [NYU] and Jagadish [UMich] - [EDBT 2018]

# Fairness in ranking

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

| rank | gender |
|---|---|
| 1 | M |
| 2 | M |
| 3 | M |
| 4 | M |
| 5 | M |
| 6 | F |
| 7 | F |
| 8 | F |
| 9 | F |
| 10 | F |

$f = 0$

| rank | gender |
|---|---|
| 1 | M |
| 2 | M |
| 3 | F |
| 4 | M |
| 5 | M |
| 6 | F |
| 7 | M |
| 8 | F |
| 9 | F |
| 10 | F |

$f = 0.3$

| rank | gender |
|---|---|
| 1 | M |
| 2 | F |
| 3 | M |
| 4 | F |
| 5 | M |
| 6 | F |
| 7 | M |
| 8 | F |
| 9 | M |
| 10 | F |

$f = 0.5$

**parity in outcomes**

joint with Yang [NYU] - [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

# Stability in ranking

## On Obtaining Stable Rankings*

Abolfazl Asudeh†, H. V. Jagadish†, Gerome Miklau††, Julia Stoyanovich‡
†University of Michigan, ††University of Massachusetts Amherst, ‡New York University
†{asudeh, jag}@umich.edu, ††miklau@cs.umass.edu, ‡stoyanovich@nyu.edu

**VLDB 2019**

**ABSTRACT**

Decision making is challenging when there is more than one criterion to consider. In such cases, it is common to assign a goodness score to each item as a weighted sum of its attribute values and rank them accordingly. Clearly, the ranking obtained depends on the weights used for this summation. Ideally, one would want the ranked order not to change if the weights are changed slightly. We call this property *stability* of the ranking. A consumer of a ranked list may trust the ranking more if it has high stability. A producer of a ranked list prefers to choose weights that result in a stable ranking, both to earn the trust of potential consumers and because a stable ranking is intrinsically likely to be more meaningful.

In this paper, we develop a framework that can be used to assess the stability of a provided ranking and to obtain a stable ranking within an "acceptable" range of weight values (called "the region of interest"). We address the case where the user cares about the rank order of the entire set of items, and also the case where the user cares only about the top-$k$ items. Using a geometric interpretation, we propose algorithms that produce stable rankings. In addition

measure and in their data collection methodology, not of relevance to our paper now. Our concern is that even if these deficiencies were addressed, we are compelled to obtain a single score/rank for a department by combining multiple objective measures, such as publications, citations, funding, and awards. Different ways of combining values for these attributes can lead to very different rankings. There are similar problems when we want to rank/seed sports teams, rank order cars or other products, as Malcolm Gladwell has nicely described [1].

Differences in rank order can have significant consequences. For example, a company may promote high-ranked employees and fire low-ranked employees. In university rankings, it is well-documented that the ranking formula has a significant effect on policies adopted by universities [2, 3]. In other words, it matters how we choose to combine values of multiple attributes into a scoring formula. Even when there is lack of consensus on a specific way to combine attributes, we should make sure that the method we use is robust: it should not be the case that small perturbations, such as small changes in parameter values, can change the rank order.

# Looking at trade-offs

## Balanced Ranking with Diversity Constraints

**Ke Yang**[1]*, **Vasilis Gkatzelis**[2], **Julia Stoyanovich**[1]
[1]New York University, Department of Computer Science and Engineering
[2]Drexel University, Department of Computer Science
ky630@nyu.edu, gkatz@drexel.edu, stoyanovich@nyu.edu

**IJCAI 2019**

### Abstract

Many set selection and ranking algorithms have recently been enhanced with *diversity constraints* that aim to explicitly increase representation of historically disadvantaged populations, or to improve the overall representativeness of the selected set. An unintended consequence of these constraints, however, is reduced *in-group fairness*: the selected candidates from a given group may not be the best ones, and this unfairness may not be well-balanced across groups. In this paper we study this phenomenon using datasets that comprise multiple sensitive attributes. We then introduce additional constraints, aimed at balancing the in-group fairness across groups, and formalize the induced optimization problems as integer linear programs. Using these programs, we conduct an experimental evaluation with real datasets, and quantify the feasible trade-offs between balance and overall performance in the presence of diversity constraints.

are increasingly recognized by sociologists and political scientists [Page, 2008; Surowiecki, 2005]. Last but not least, diversity constraints can be used to ensure dataset representativeness, for example when selecting a group of patients to study the effectiveness of a medical treatment, or to understand the patterns of use of medical services [Cohen *et al.*, 2009], an example we will revisit in this paper.

Our goal in this paper is to evaluate and mitigate an unintended consequence that such diversity constraints may have on the outcomes of set selection and ranking algorithms. Namely, we want to ensure that these algorithms do not systematically select lower-quality items in particular groups. In what follows, we make our set-up more precise.

Given a set of items, each associated with multiple sensitive attribute labels and with a quality score (or utility), a set selection algorithm needs to select $k$ of these items aiming to maximize the overall utility, computed as the sum of *utility scores* of selected items. The score of an item is a single scalar that may be pre-computed and stored as a physical attribute, or it may be computed on the fly. The output of traditional set selection algorithms, however, may lead to

**parity in outcomes**

**loss balance**

# Looking at trade-offs

**Balanced Ranking with Diversity Constraints**

**Ke Y**

[1]New York Un
[2]Dre
ky630@

gineering

du

**IJCAI 2019**

## Abstract

Many set selection and ranking
cently been enhanced with *diver*
aim to explicitly increase repres
cally disadvantaged populations
overall representativeness of th
unintended consequence of thes
ever, is reduced *in-group fairnes*
didates from a given group m
ones, and this unfairness may n
across groups. In this paper
nomenon using datasets that con
sitive attributes. We then introd
straints, aimed at balancing th
across groups, and formalize the
tion problems as integer linear
these programs, we conduct an
uation with real datasets, and
ble trade-offs between balance
mance in the presence of diversi

sociologists and political sci-
cki, 2005]. Last but not least,
ed to ensure dataset represen-
electing a group of patients to
edical treatment, or to under-
edical services [Cohen *et al.*,
isit in this paper.
evaluate and mitigate an unin-
diversity constraints may have
tion and ranking algorithms.
t these algorithms do not sys-
items in particular groups. In
-up more precise.
associated with multiple sen-
a quality score (or utility), a
o select $k$ of these items aim-
tility, computed as the sum of
s. The score of an item is a
omputed and stored as a phys-
nputed on the fly. The output

**parity in outcomes**

**loss balance**

# Ranking with diversity constraints

**Goal**: pick k=4 candidates, including 2 of each gender, and at least one candidate per ethnicity, maximizing the total score of the selected candidates.

| | Male | | Female | |
|---|---|---|---|---|
| White | A (99) | B (98) | C (96) | D (95) |
| Black | E (91) | F (91) | G (90) | H (89) |
| Asian | I (87) | J (87) | K (86) | L (83) |

score=373

Table 1: A set of 12 individuals with sensitive attributes `race` and `gender`. Each cell lists an individual's ID, and score in parentheses.

**Problem**: **In-group fairness fails** for Female (C and D not picked, which G and K are), Black (E and F are not picked, while G is), and Asian (I and J are not picked, while K is). **In-group fairness holds** for White and Male groups though (those with higher scores)!

# Ranking with diversity constraints

**Goal**: pick k=4 candidates, including 2 of each gender, and at least one candidate per ethnicity, maximizing the total score of the selected candidates.

| | Male | | Female | |
|---|---|---|---|---|
| White | A (99) | B (98) | C (96) | D (95) |
| Black | E (91) | F (91) | G (90) | H (89) |
| Asian | I (87) | J (87) | K (86) | L (83) |

score=372

Table 1: A set of 12 individuals with sensitive attributes `race` and `gender`. Each cell lists an individual's ID, and score in parentheses.

**Problem**: **In-group fairness fails** for Female (C and D not picked, which G and K are), Black (E and F are not picked, while G is), and Asian (I and J are not picked, while K is). **In-group fairness holds** for White and Male groups though (those with higher scores)!

**Insight**: while in-group fairness will inevitably fail to some extent because of diversity constraints, this loss should be **balanced** across groups.

# 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

# The Vacca bill

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

# 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;

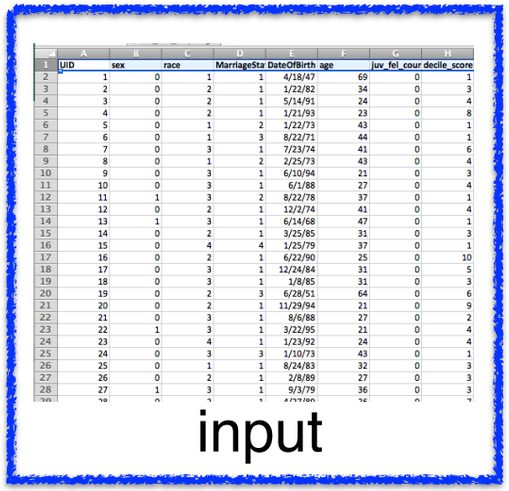
also release:

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

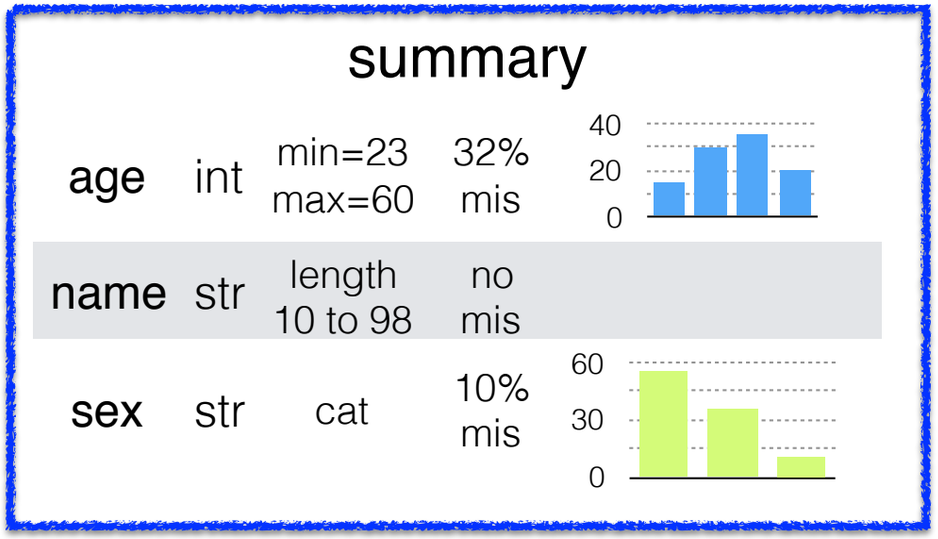
# Data Synthesizer

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

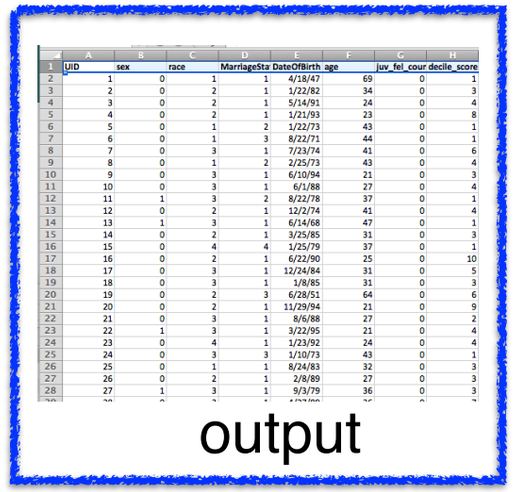
http://demo.dataresponsibly.com/synthesizer/



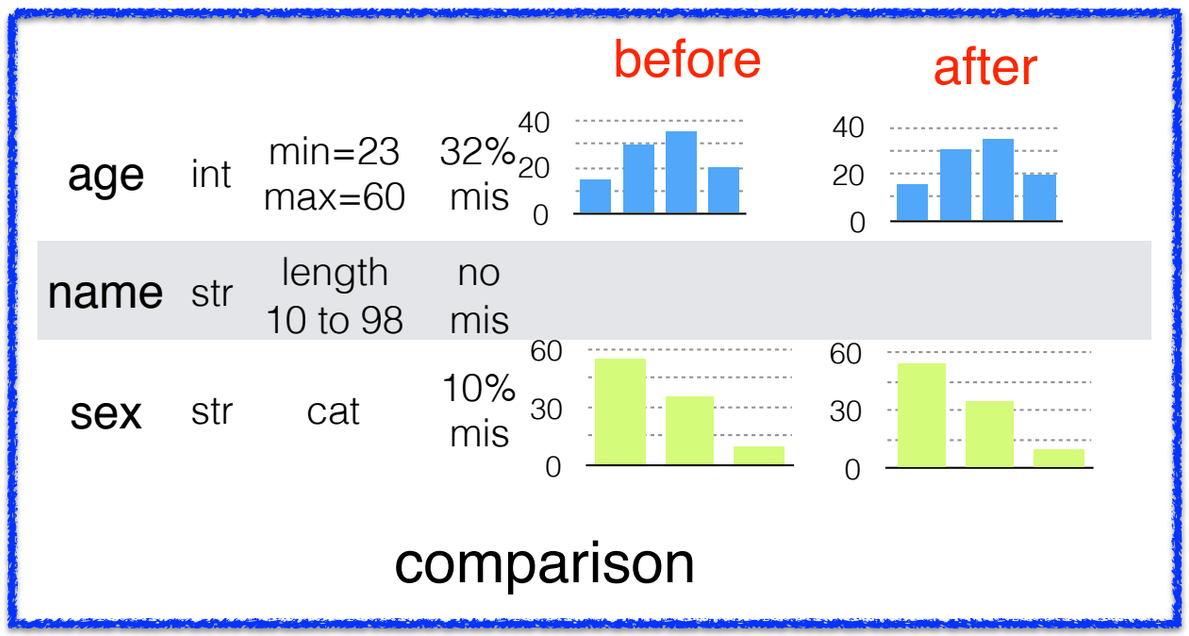

Data Describer




Data Generator




Model Inspector



# Data Synthesizer



government technology

MetroLab "Innovation of the Month"

SECURITY

## University Researchers Use 'Fake' Data for Social Good

BY BEN LEVINE / NOVEMBER 7, 2017

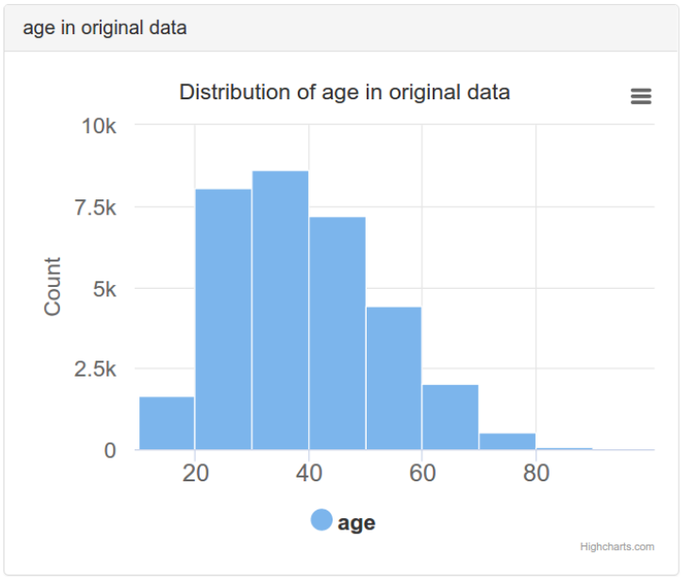
Virtually every interaction we have with a public agency creates a data point. Amass enough data points and they can tell a story. However, factors like privacy, data storage and usability present challenges for local governments and researchers interested in helping improve services. In this installment 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.

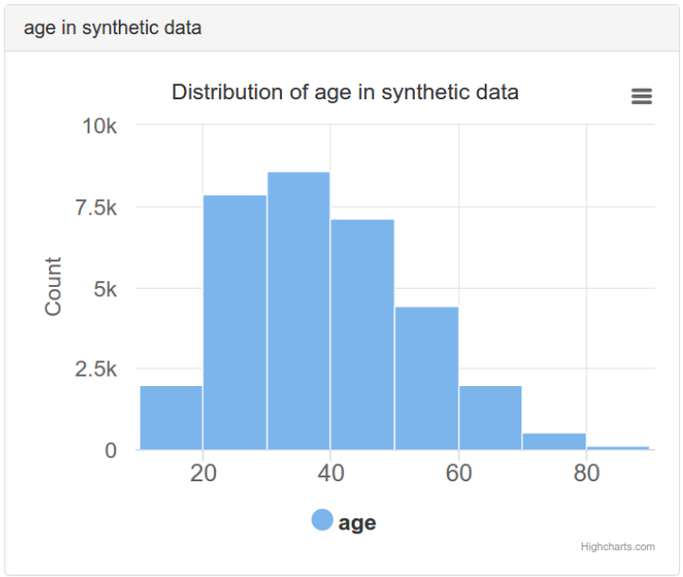
Since its development, the tool has been receiving a lot of attention. For example: T-Mobile is interested in generating synthetic data to better engage with researchers and improve transparency for customers, the Colorado Department of Education has asked relevant agencies to use the tool to experiment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

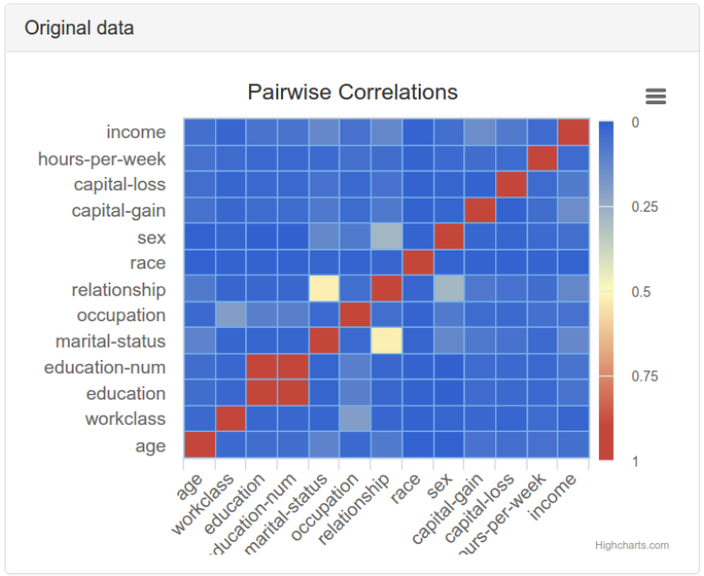
http://www.govtech.com/security/University-Researchers-Use-Fake-Data-for-Social-Good.html

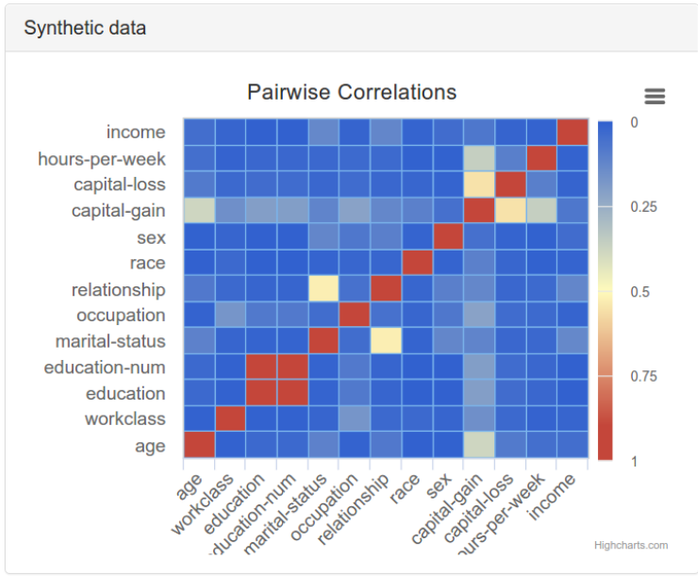
# Data Synthesizer

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




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

# An Algorithmic Approach to Correct Bias in Urban Transportation Datasets

To assuage these concerns for private urban transportation companies, researchers have developed an algorithm that removes selected biases from datasets while retaining the utility of the data. The researchers include **Julia Stoyanovich***, CDS Assistant Professor of Data Science, Computer Science and Engineering, along with **Luke Rodriguez**, **Babak Salimi**, and **Bill Howe** from University of Washington, and **Haoye Ping** from Drexel University.

# 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**]

## Ranking Facts

### 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 |

### 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.

### Ingredients

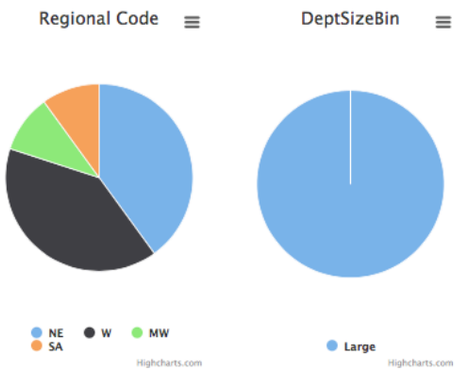
**Top 10:**

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

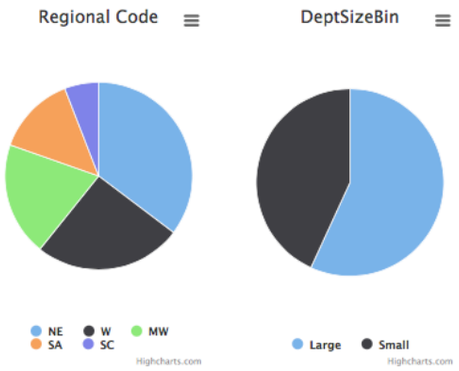
**Overall:**

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

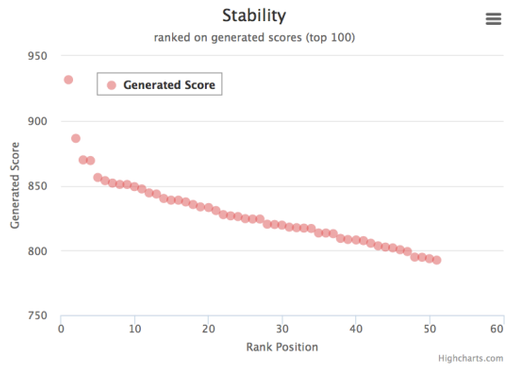
### Diversity at top-10




### Diversity overall




### 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.

### 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.

### 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/

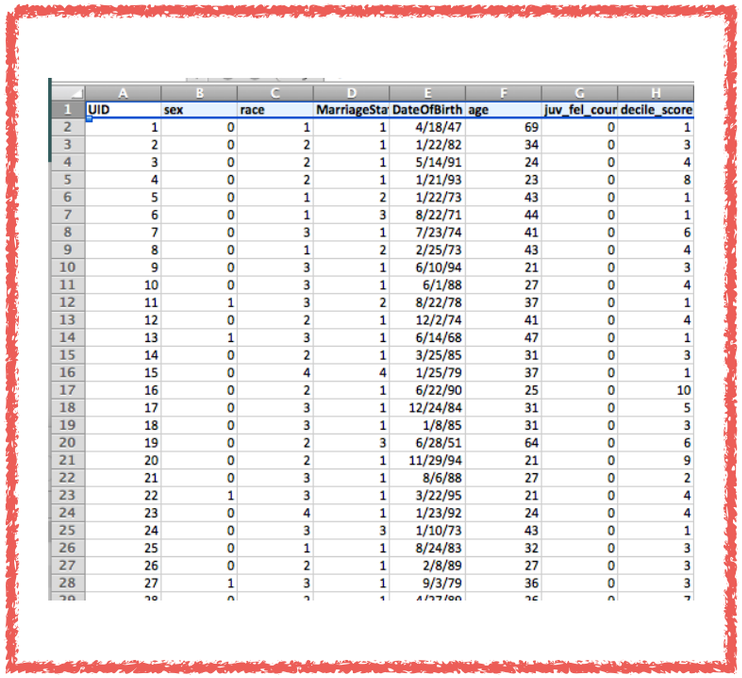
# Point 5

transparency by design, not as an afterthought

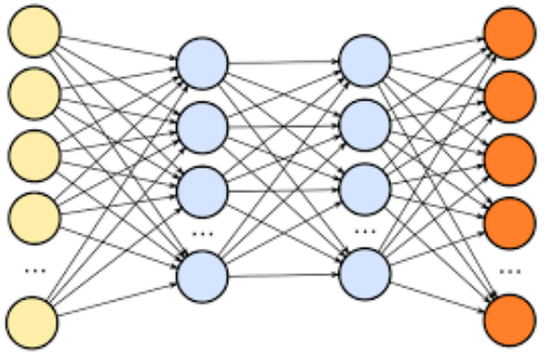
provision for transparency and interpretability at every stage of the data lifecycle

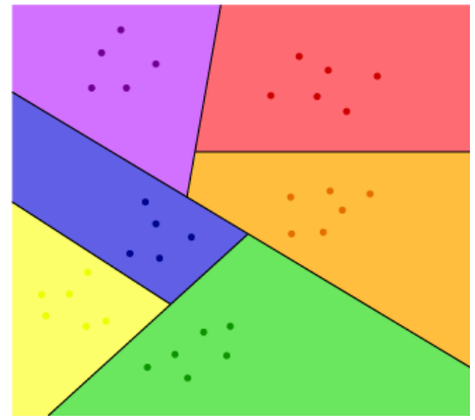
useful internally during development, for communication and coordination between agencies, and for accountability to the public

# Frog's eye view



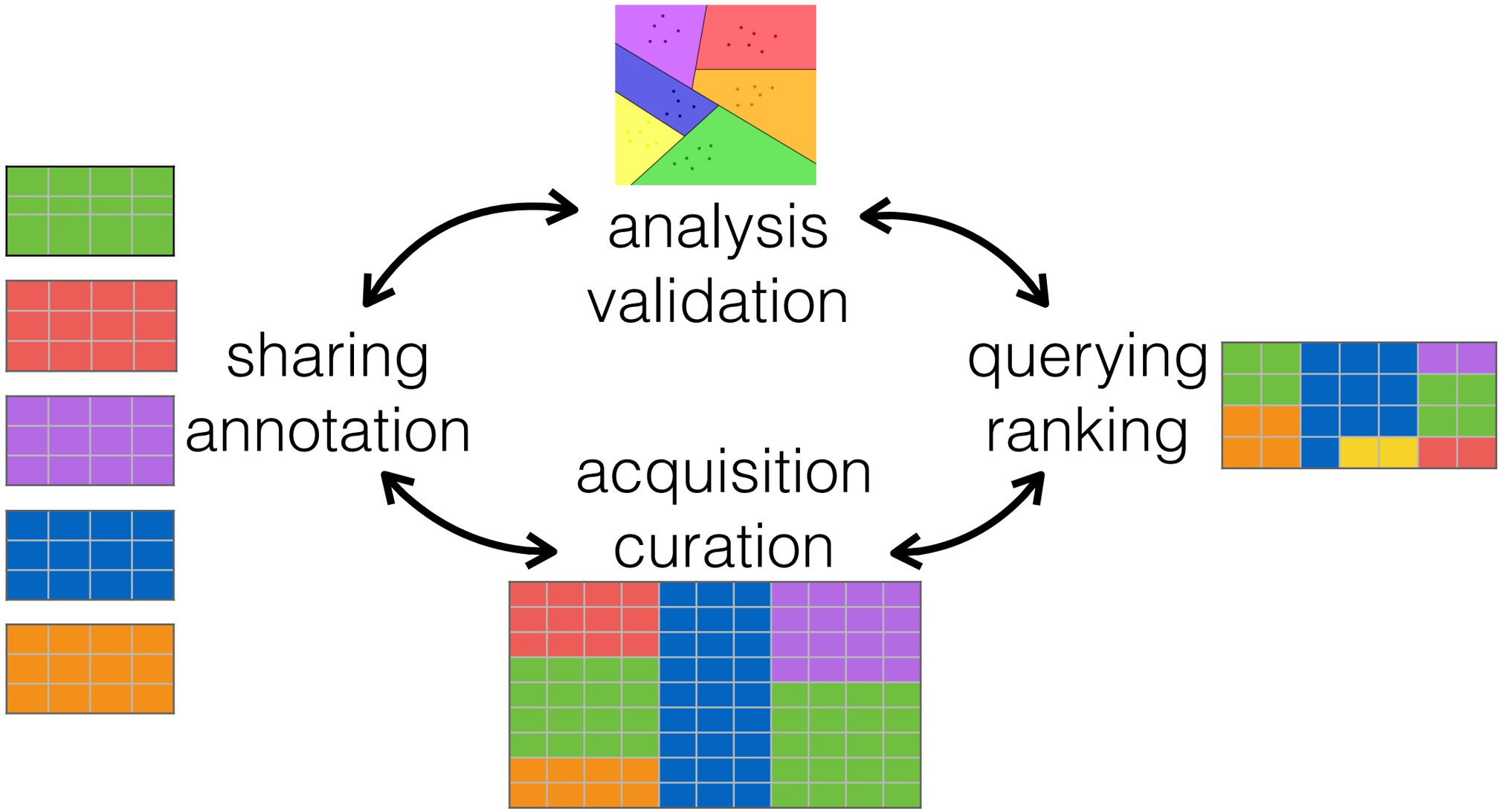
| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | UID | sex | race | MarriageSta | DateOfBirth | age | juv_fel_cour | decile_score |
| 2 | 1 | 0 | 1 | 1 | 4/18/47 | 69 | 0 | 1 |
| 3 | 2 | 0 | 2 | 1 | 1/22/82 | 34 | 0 | 3 |
| 4 | 3 | 0 | 2 | 1 | 5/14/91 | 24 | 0 | 4 |
| 5 | 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 |





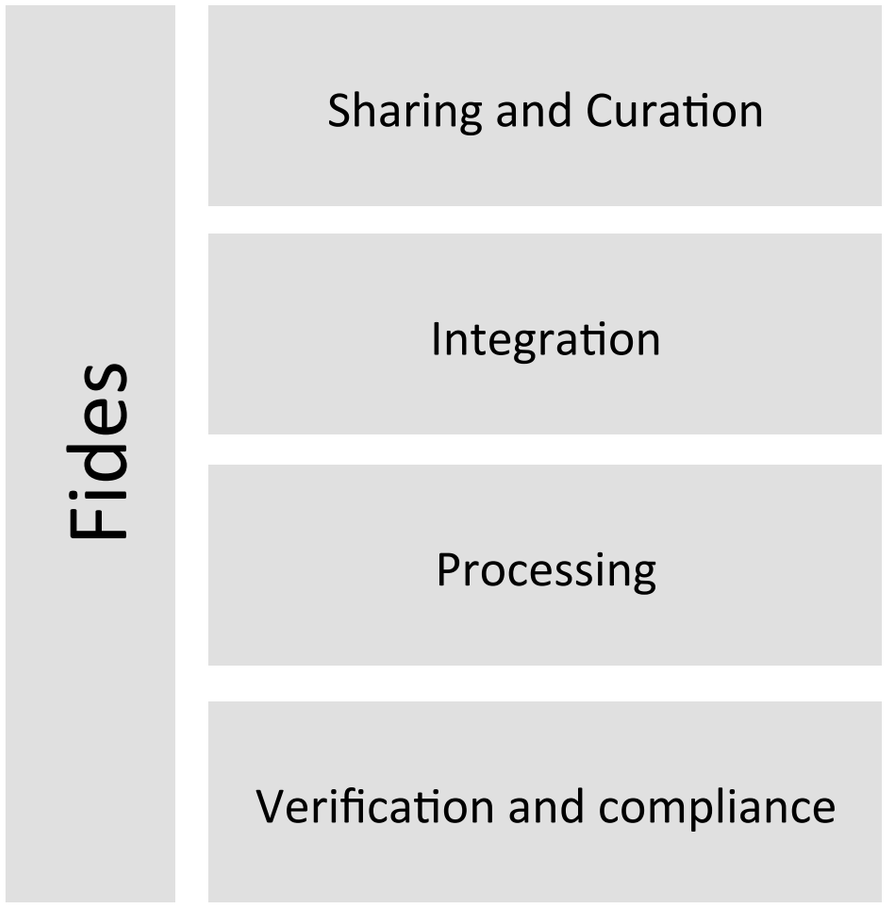
but where does the data come from?

# The data science lifecycle




**responsible data science** requires a holistic view of the data lifecycle

# Fides: responsibility by design




**Systems support** for responsible data science

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

**Applications**: data science for social good

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

[HDR DIRSE-FW] Framework for integrative data equity systems 09/2019-

New!

# teaching responsible data science

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

## New York University, Center for Data Science, Spring 2019

**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.

Responsible Data Science is a technical course that tackles the issues of ethics, legal compliance, data quality, algorithmic fairness and diversity, transparency of data and algorithms, privacy, and data protection. The course is developed and taught by Julia Stoyanovich, Assistant Professor at the Center for Data Science and at the Tandon School of Engineering, and member of the NYC Automated Decision Systems Task Force.

**Prerequisites:** Introduction to Data Science, Introduction to Computer Science, or similar courses.

# Teaching RDS

- **Topic cover the data science lifecycle**, not only the final mile of data analysis
  - fairness, diversity, transparency, interpretability
  - privacy, data protection
  - data profiling, data cleaning, *plan to add data integration*
  - legal frameworks, codes of ethics, professional responsibility

# Teaching RDS

- **Stand-alone technical courses**
  - Designed and delivered a **graduate** course at the Center for Data Science (CDS) at NYU, Spring 2019
  - Designed an **undergraduate** course at NYU CDS, a requirement of the new BS in DS, will be offered in Spring 2020
- **Modules** (readings, slides, assignments) integrated into undergraduate and graduate courses:
  - **(M1)** Intro to RDS, **(M2)** Fairness, **(M3)** Transparency, **(M4)** Data Protection
- All materials available at **dataresponsibly.github.io**

# Data Science offerings at NYU

The Center for Data Science (CDS) @ NYU - an independent Provostial unit, established in 2012

Degrees: MS, PhD (started in 2017), **BS** (started in 2019)

**Gender diversity**: 38-49% female MS classes, 25-50% female PhD classes (representative of the proportion of female applicants)


[NSF NRT] FUTURE: Foundations, Translation, Responsibility for Data Science Impact, 09/2019-

# 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

Association for Computing Machinery

*Advancing Computing as a Science & Profession*

Digital Library CACM Queue TechNews Learning Center Career Center

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

# 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**

Thank you!

**dataresponsibly.github.io**

**@stoyanoj**

NYU

# ADS example

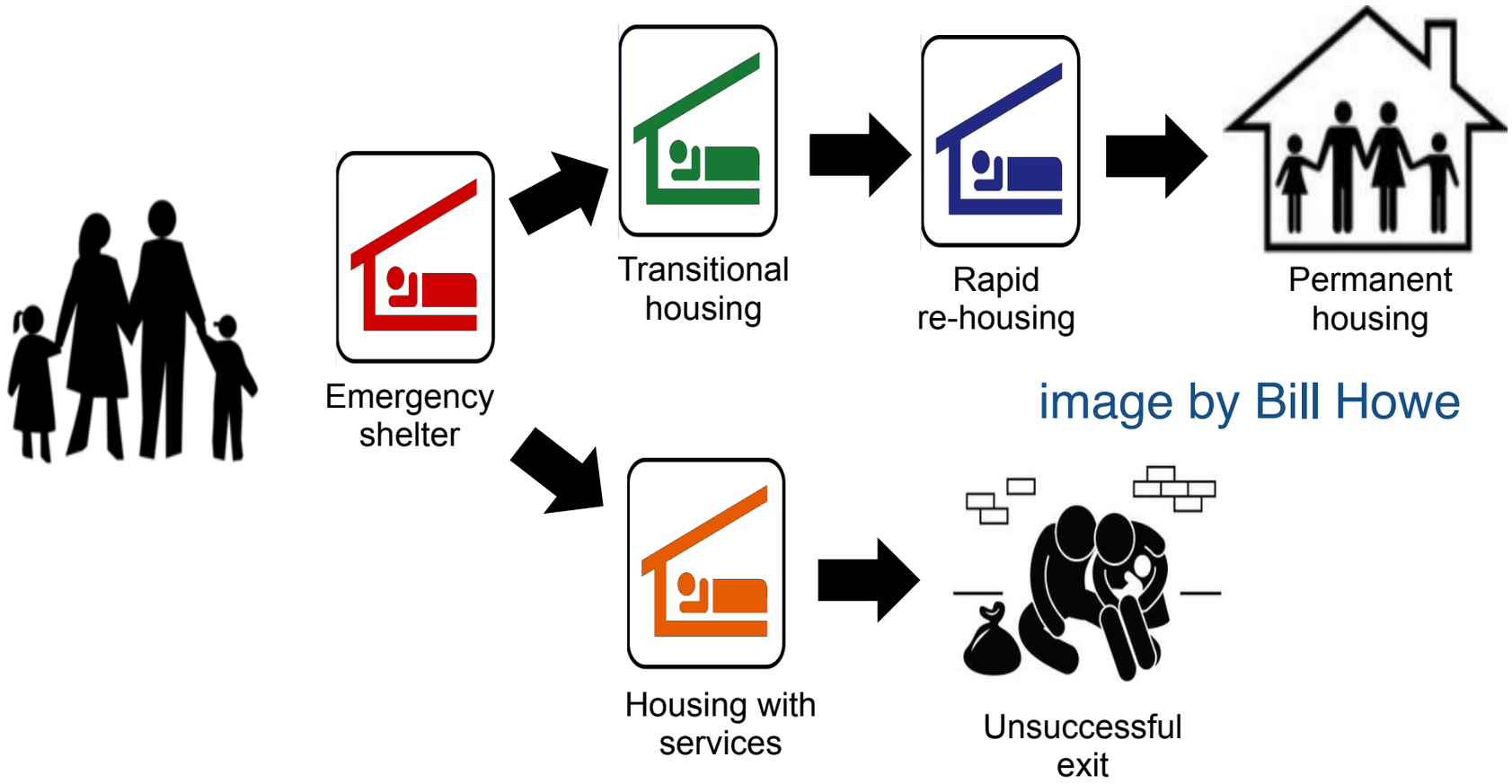



image by Bill Howe

- **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

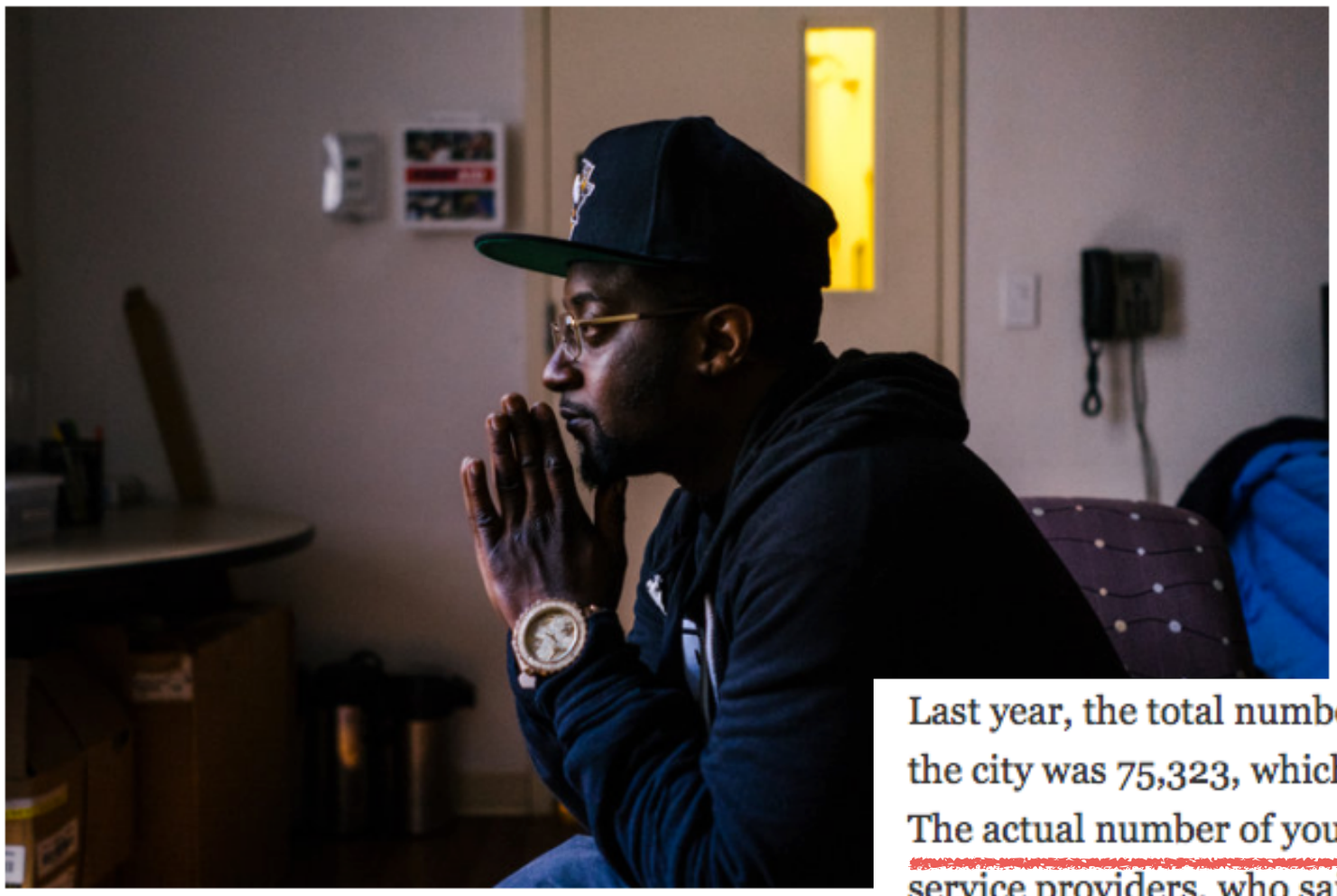
The New York Times

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

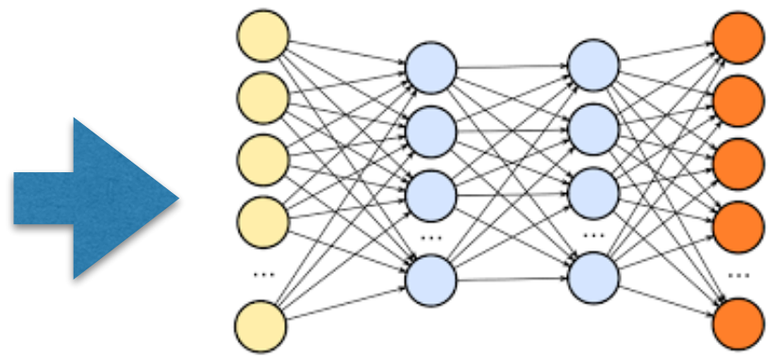
The New York Times

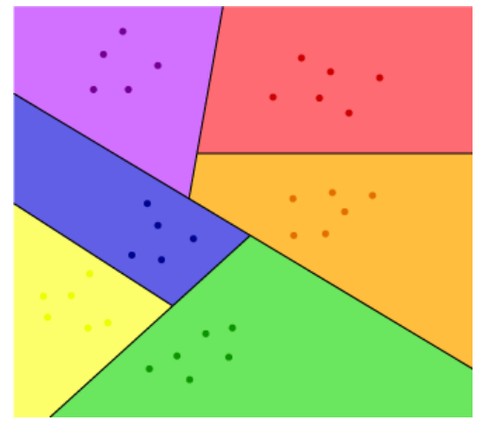
https://www.nytimes.com/2016/02/06/nyregion/young-and-homeless-in-new-york-overlooked-and-underserved.html

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:**

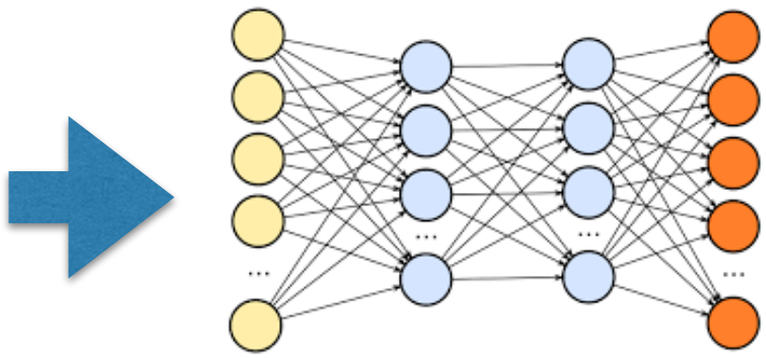
```
select * from R
where status = 'unsheltered'
and length > 2 month
```

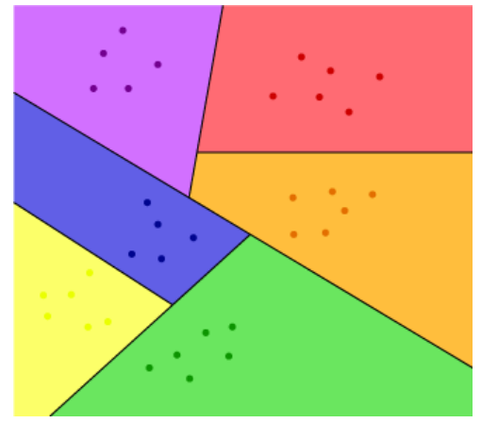
10% female

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!**

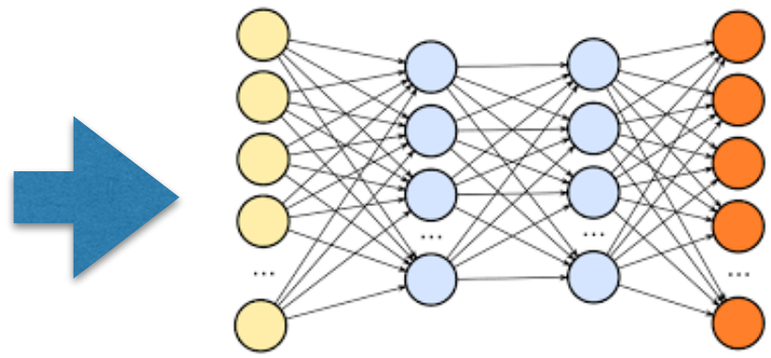
**of course, but maybe…**

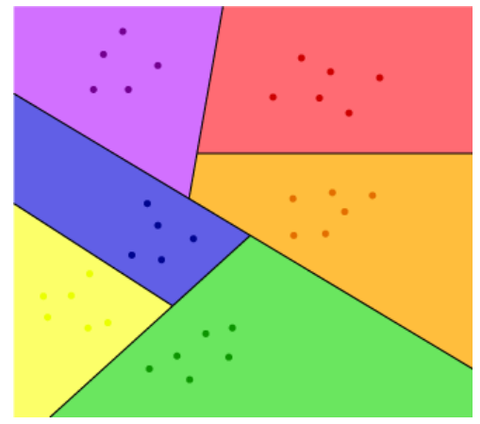
**mental health info was missing for this population**

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