Towards responsible data science

Julia Stoyanovich
Drexel University & Princeton CITP

@stoyanoj

data RESPONSIBLY

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

Online job ads

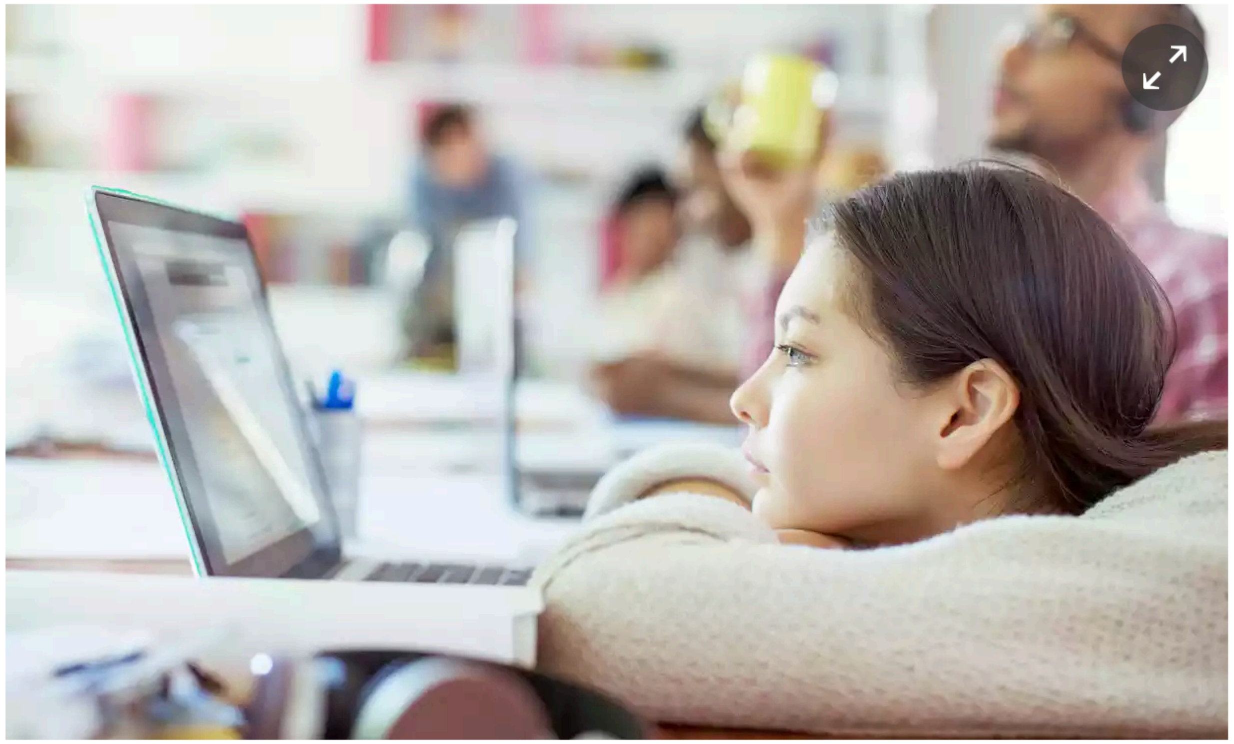
theguardian

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

Job-screening personality tests

THE WALL STREET JOURNAL.

Are Workplace Personality Tests Fair?

By **LAUREN WEBER** and **ELIZABETH DWOSKIN**

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

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

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

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

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

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

Is algorithmic decision-making impartial?

Claim: **Data-driven algorithms cannot be biased!** And yet…

- Algorithms **discriminate** just like humans do, but at a larger scale
- Processes are **opaque**, and defy public scrutiny
- It is our responsibility to understand the issues and offer **technological solutions** that address them
- Technology must be informed by **ethical** and **legal considerations**

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

The evils of discrimination

Disparate treatment is the illegal practice of treating an entity 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

NYC Algorithmic Transparency Law

1/11/2018

Int. No. 1696-A: A Local Law 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		

NYC Algorithmic Transparency Law

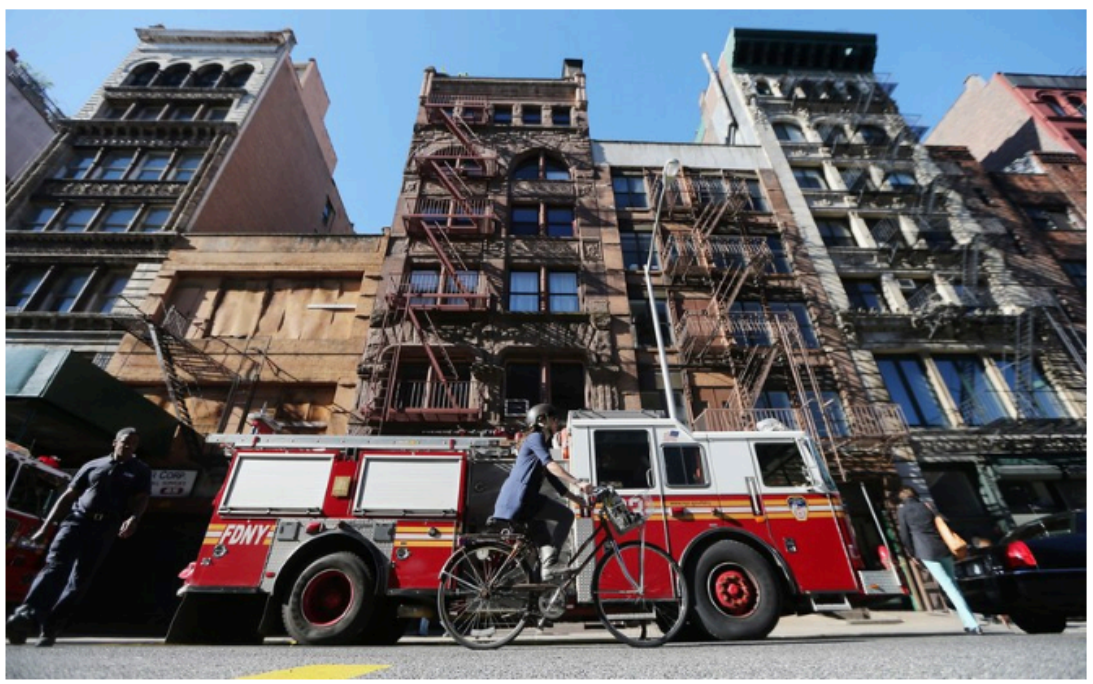
10/16/2017

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

data RESPONSIBLY

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 Int. No. 1696-A

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

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

The Official Website of the City of New York | NYC | 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.”

General Data Protection Regulation
GDPR

Chapter 1 (Art. 1 – 4)
General provisions

Chapter 2 (Art. 5 – 11)
Principles

Chapter 3 (Art. 12 – 23)
Rights of the data subject

Chapter 4 (Art. 24 – 43)
Controller and processor

Chapter 5 (Art. 44 – 50)
Transfers of personal data to third countries or international organisations

Chapter 6 (Art. 51 – 59)
Independent supervisory authorities

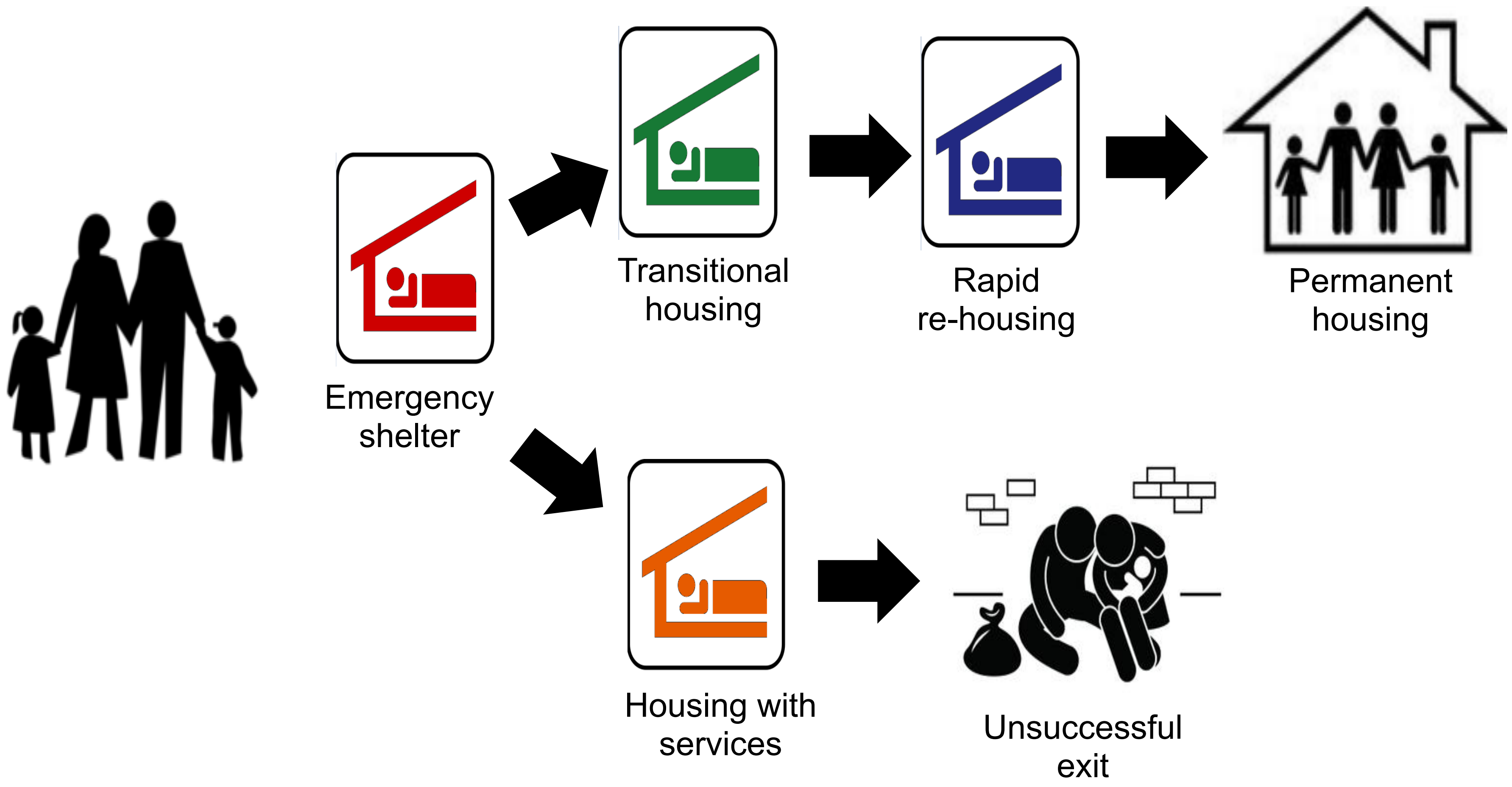
Chapter 7 (Art. 60 – 76)
Cooperation and consistency

Chapter 8 (Art. 77 – 84)
Remedies, liability and penalties

Chapter 9 (Art. 85 – 91)
Provisions relating to specific processing situations

Chapter 10 (Art. 92 – 93)
Delegated acts and implementing acts

ADS example: urban homelessness



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

data RESPONSIBLY

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.

Responsible data science

- Be **transparent** and **accountable**
- Achieve **equitable** resource distribution
- Be cognizant of the **rights** and **preferences** of individuals

fairness

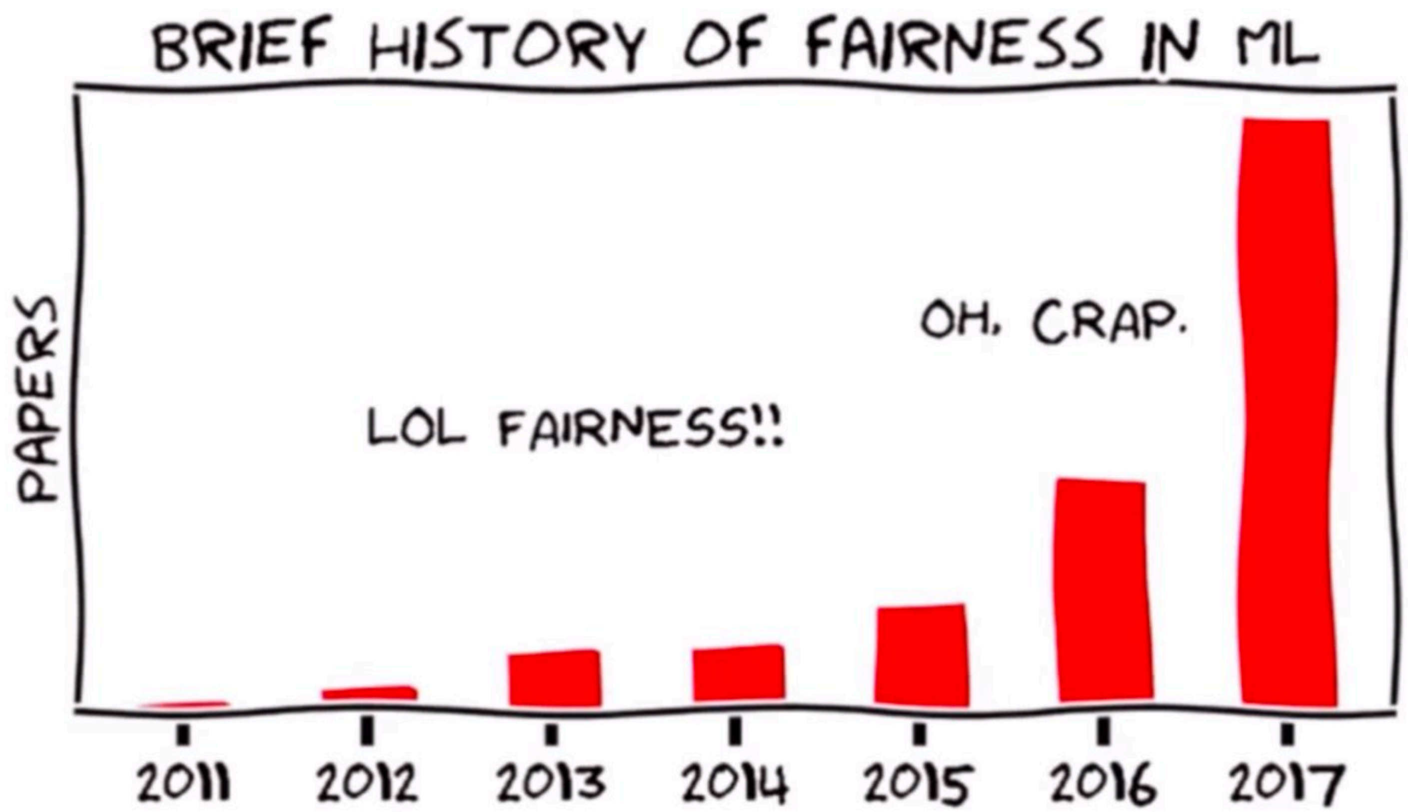
diversity

transparency

data protection

Responsible data science

- Be **transparent** and **accountable**
- Achieve **equitable** resource distribution
- Be cognizant of the **rights** and **preferences** of individuals

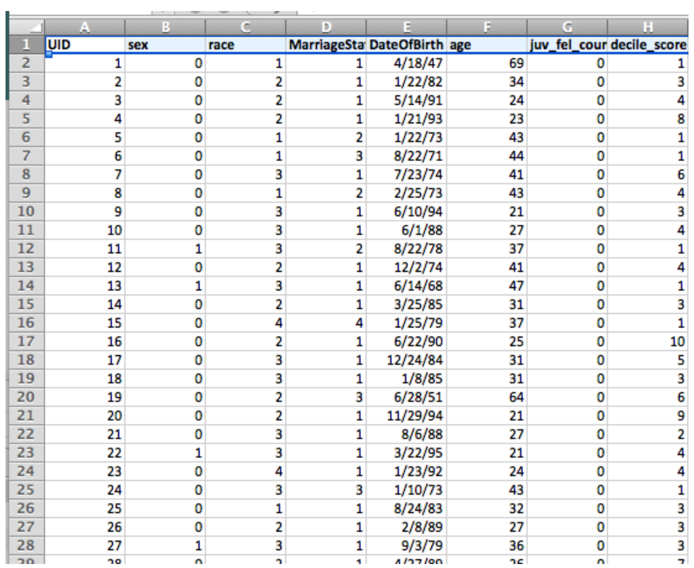


by Moritz Hardt

dataRESPONSIBLY

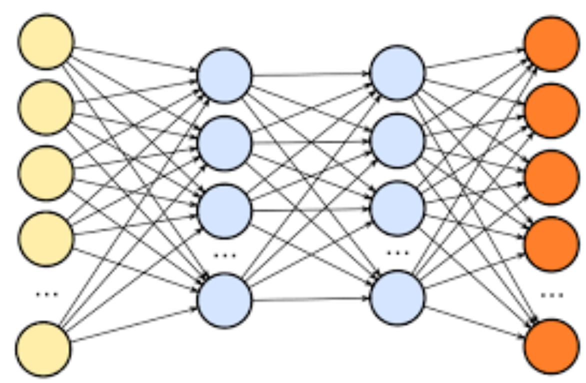
Responsible data science

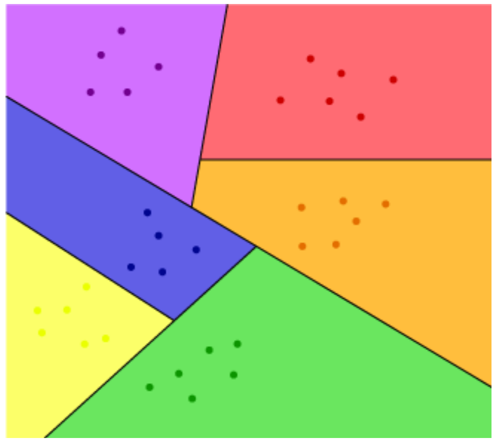
- Be **transparent** and **accountable**
- Achieve **equitable** resource distribution
- Be cognizant of the **rights** and **preferences** of individuals



	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

FAT/ML

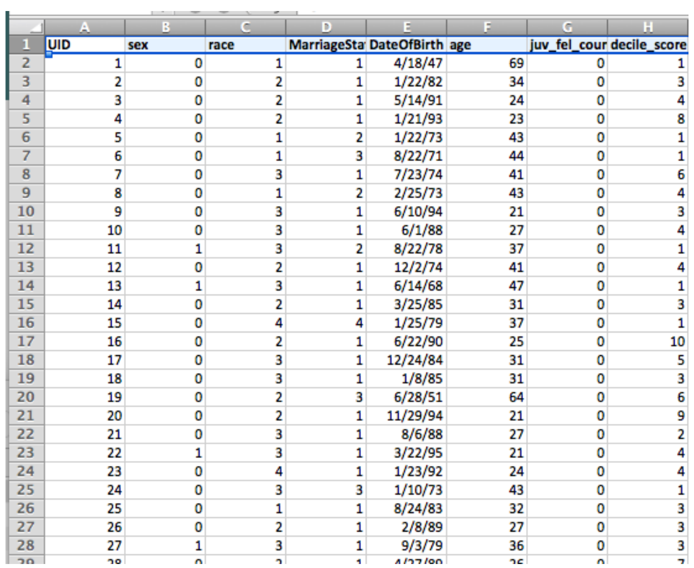




done?

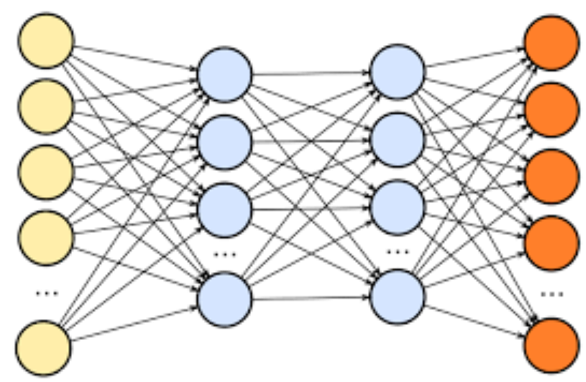
Responsible data science

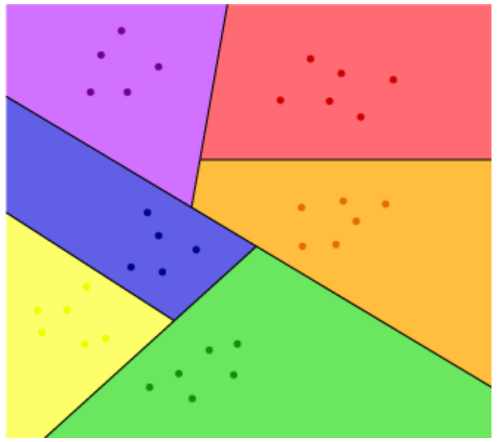
- Be **transparent** and **accountable**
- Achieve **equitable** resource distribution
- Be cognizant of the **rights** and **preferences** of individuals



	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

FAT/ML

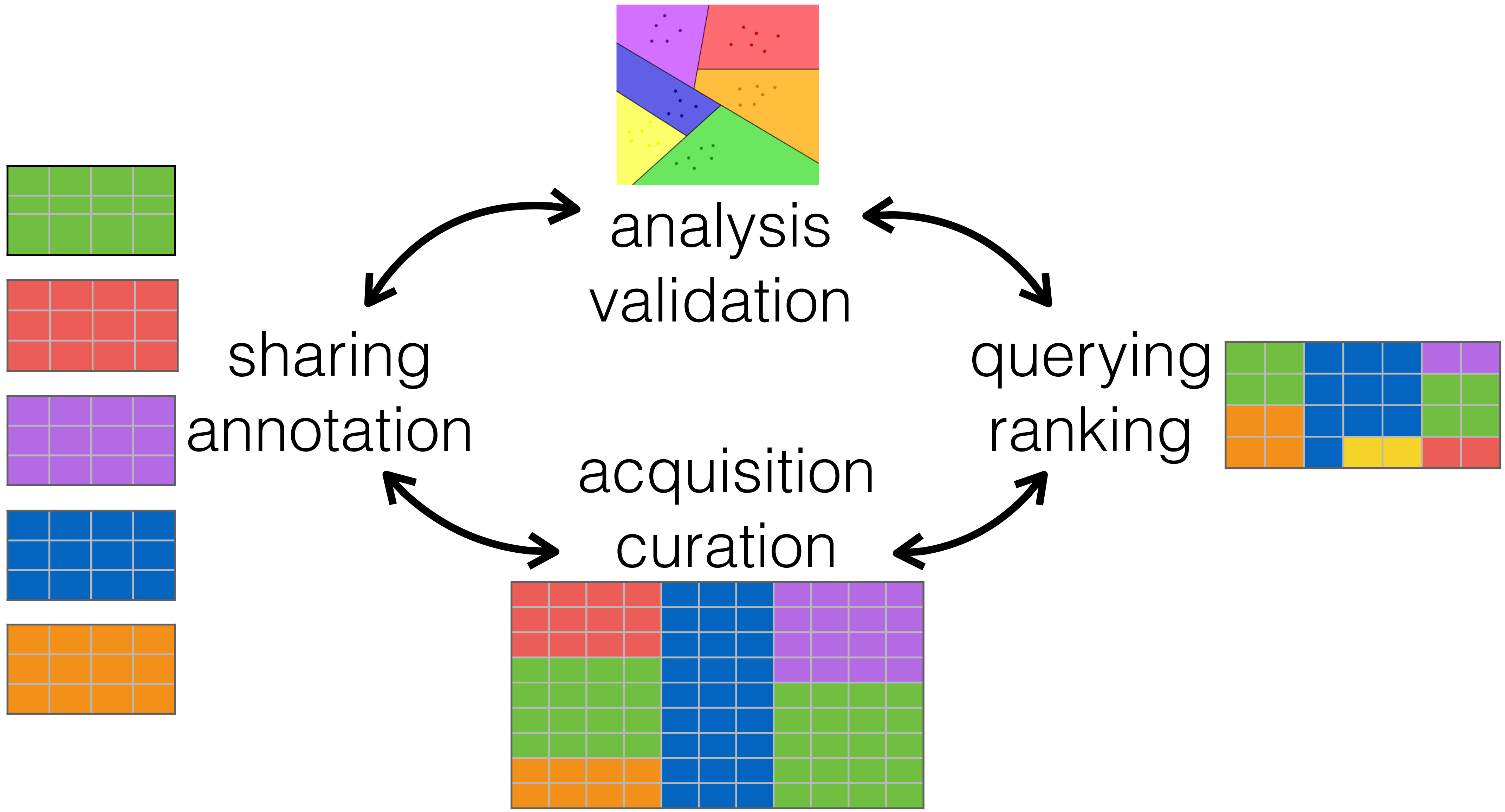




but where does the data come from?

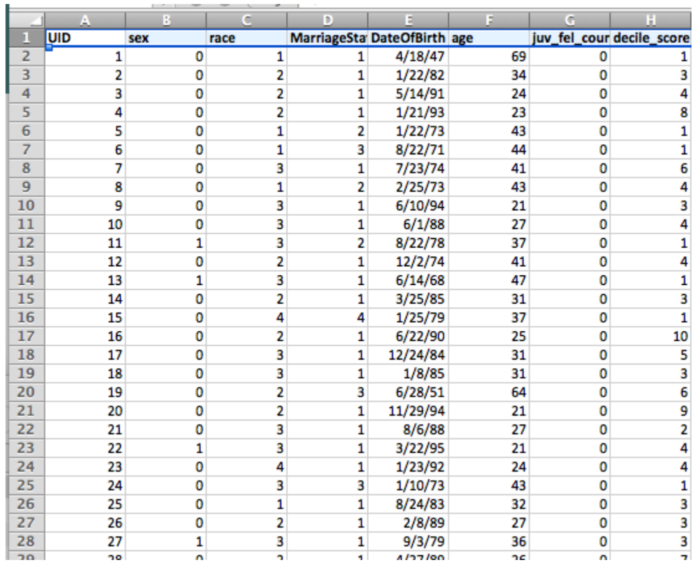
data RESPONSIBLY

The data science lifecycle

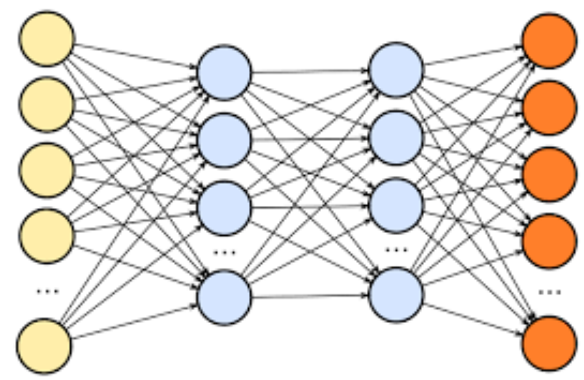


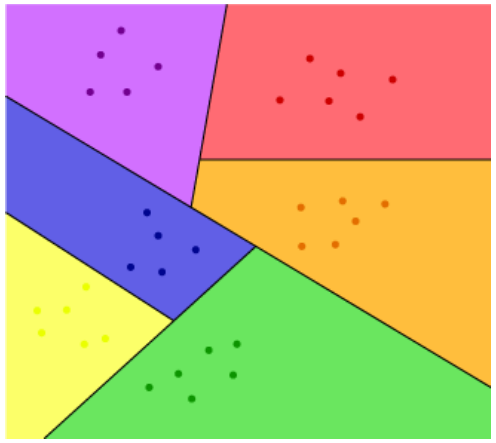
responsible data science requires a holistic view of the data lifecycle

Revisiting the analytics step



	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 some outcome groups (group fairness)

fix the model!

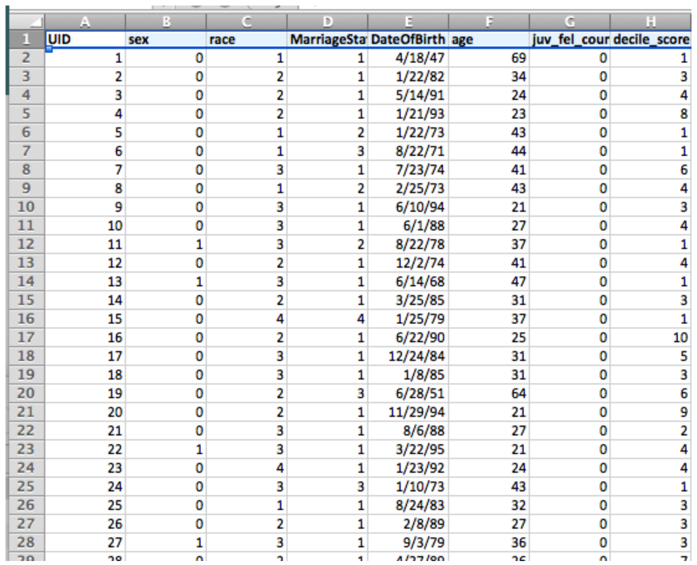
of course, but maybe… the input was generated with:

select * from R
where status = ‘unsheltered’
and length > 2 1 month

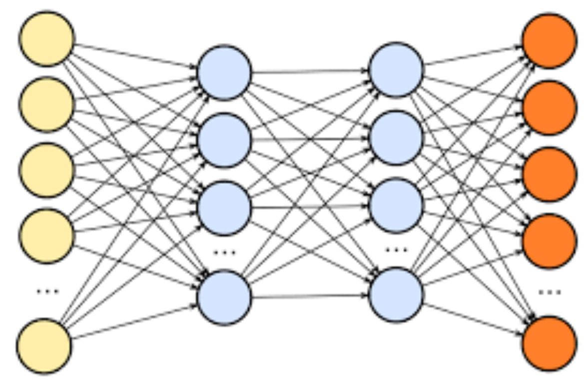
10% female

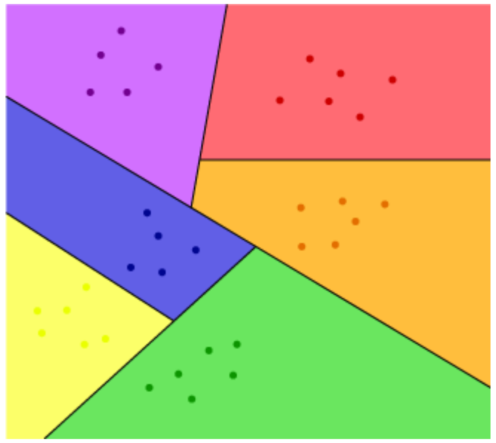
40% female

Revisiting the analytics step



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

fix the model!

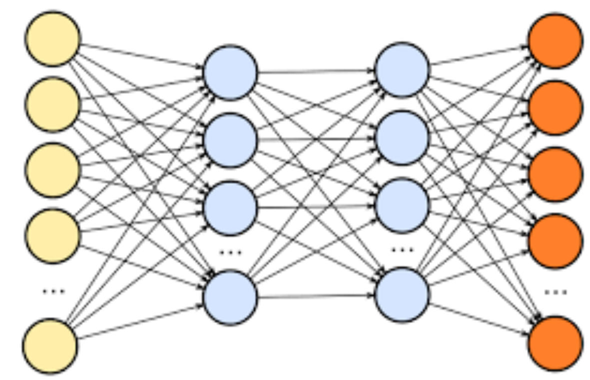
of course, but maybe…

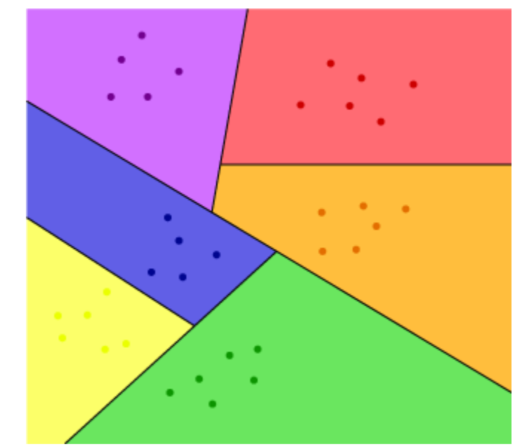
mental health info was missing for this population

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

Revisiting the analytics step

	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

data RESPONSIBLY

Fides: 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

Applications: data science for social good

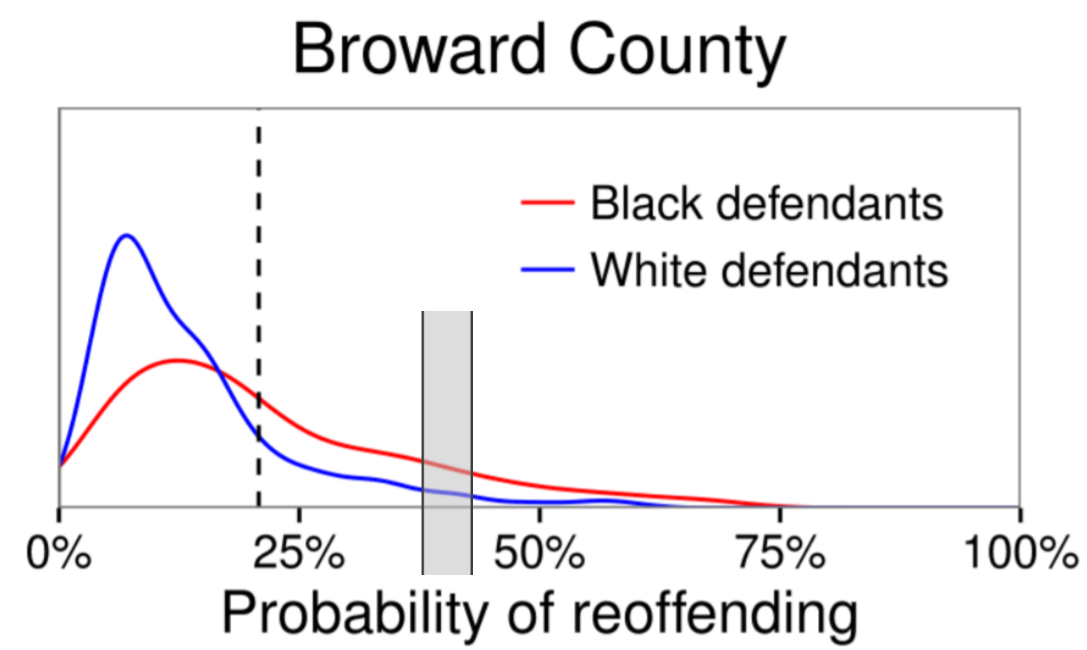
responsible data science requires a holistic view of the data lifecycle

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

What is fairness?

Re-arrest probabilities derived from COMPAS risk scores.

In the window around 40%, about 40% of the defendants are re-arrested

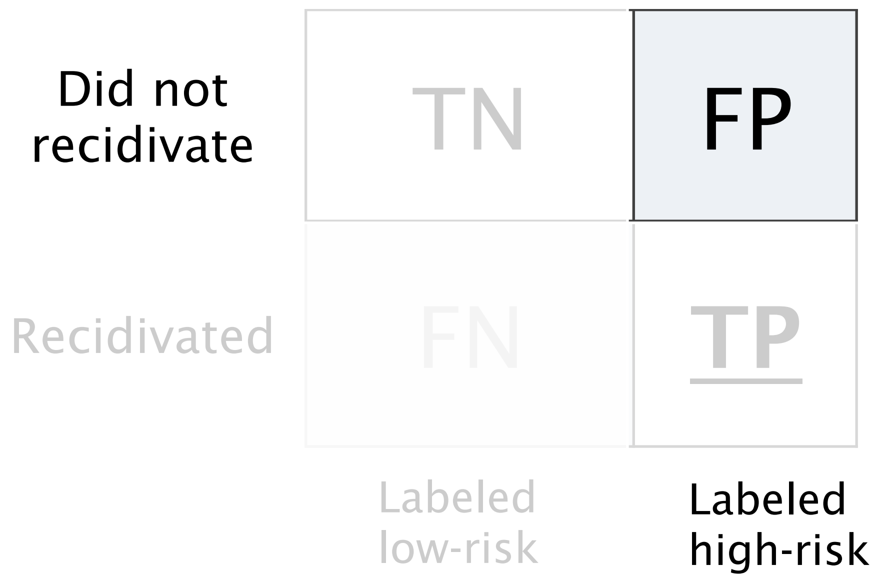


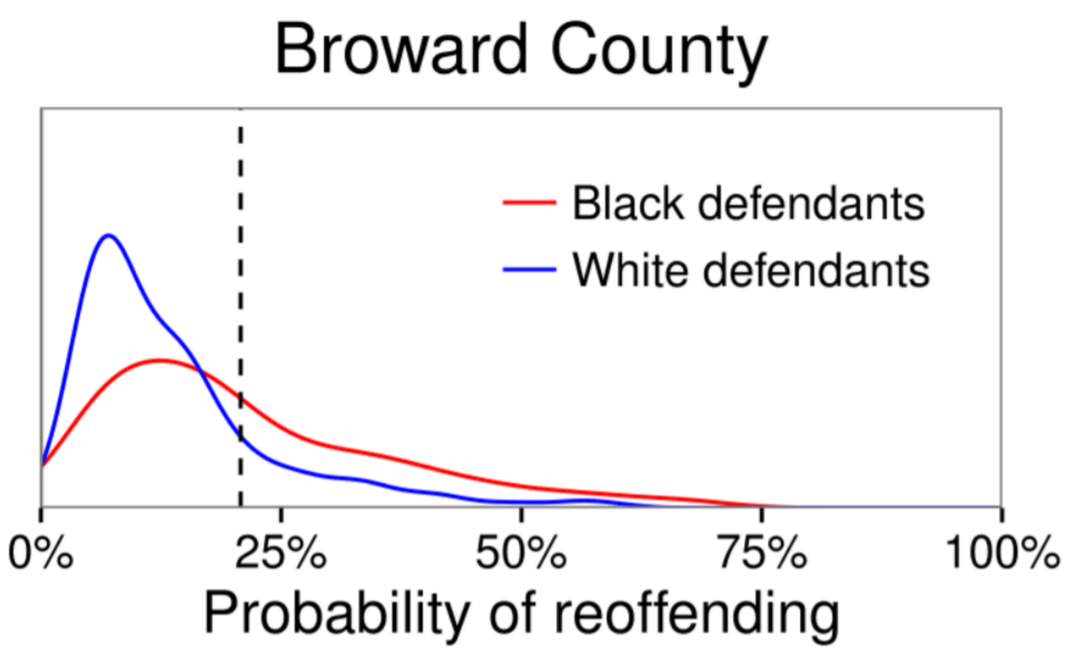
[Fig. from Corbett-Davies et al., "Algorithmic Decision Making and the Cost of Fairness"]

ProPublica: "things look very different when analyzed from the perspective of the defendants, particularly those **wrongly classified as future criminals**."

inspired by Arvind Naranayan's FAT* tutorial

What is fairness?





[Fig. from Corbett-Davies et al., “Algorithmic Decision Making and the Cost of Fairness”]

(**Predictive value**) Decision maker: of those I labeled high-risk, how many will re-offend?

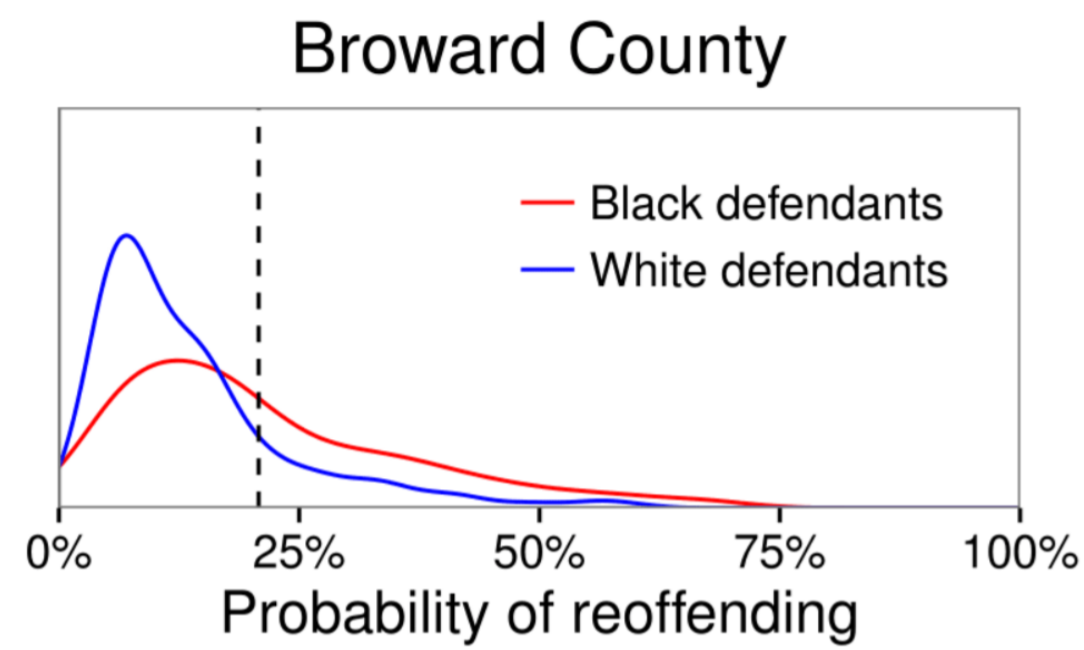
(**False positive rate**) Defendant: how likely am I to be incorrectly classified high-risk?

inspired by Arvind Naranayan’s FAT* tutorial

data RESPONSIBLY

What is fairness?

	Labeled low-risk	Labeled high-risk
Did not recidivate	TN	FP
Recidivated	FN	**TP**



[Fig. from Corbett-Davies et al., “Algorithmic Decision Making and the Cost of Fairness”]

If an instrument satisfies **predictive parity** … but the **prevalence differs between groups**, the instrument **cannot achieve equal false positive rates and equal false negative** rates across those groups.

[Chouldechova. “Fair Prediction with Disparate Impact”, 2018]

inspired by Arvind Naranayan’s FAT* tutorial

data RESPONSIBLY

What is fairness?

	Labeled low-risk	Labeled high-risk
Did not recidivate	TN	FP
Recidivated	FN	**TP**

Metric	Equalized under
Selection probability	Demographic parity
Pos. predictive value	Predictive parity *
Neg. predictive value	
False positive rate	Error rate balance *
False negative rate	Error rate balance *
Accuracy	Accuracy equity

[Chouldechova, 2018]

Different metrics matter to different stakeholders

Different definitions are “right” in different contexts

Often impossible to achieve simultaneously, impose **trade-offs**

inspired by Arvind Naranayan’s FAT* tutorial

Individual vs. group fairness

Group fairness: Do outcomes systematically differ between demographic groups?

Individual fairness: Are similar** individuals being treated similarly?

Another take

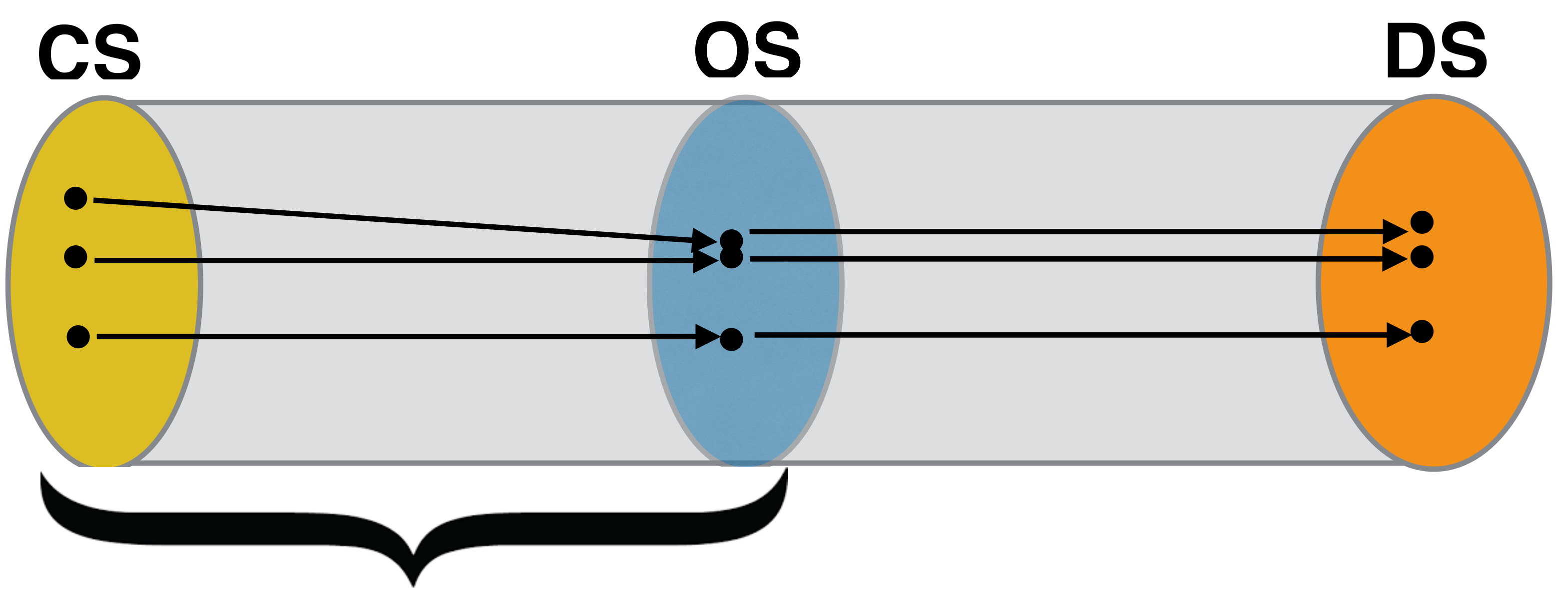
[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]

<table>
<tr><th>Construct Space</th><th>Observed Space</th><th>Decision Space</th></tr>
<tr><td>intelligence</td><td>SAT score</td><td rowspan="2">performance in college</td></tr>
<tr><td>grit</td><td>high-school GPA</td></tr>
<tr><td>propensity to commit crime</td><td>family history</td><td rowspan="2">recidivism</td></tr>
<tr><td>risk-averseness</td><td>age</td></tr>
</table>

fairness through mappings

Individual fairness

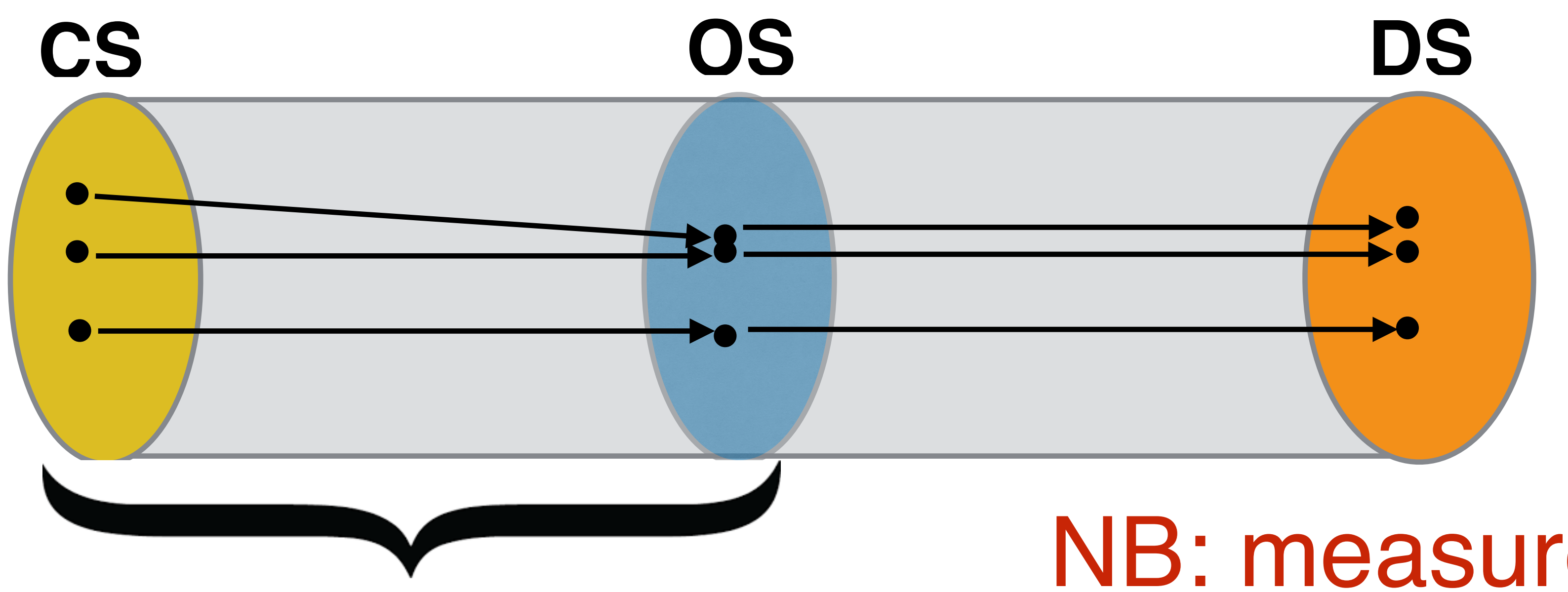
[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]



What you see is what you get (**WYSIWYG**): there exists a mapping from **CS** to **OS** that has low distortion. That is, we might believe that OS faithfully represents CS. **This is the individual fairness world view.**

Group fairness

[S. Friedler, C. Scheidegger and S. Venkatasubramanian, arXiv:1609.07236v1 (2016)]



NB: measurement bias

We are all equal (**WAE**): the mapping from **CS** to **OS** introduces **structural bias** - there is a distortion that aligns with the group structure of CS. **This is the group fairness world view.**

Structural bias examples: SAT verbal questions function differently in the African-American and in the Caucasian subgroups in the US.

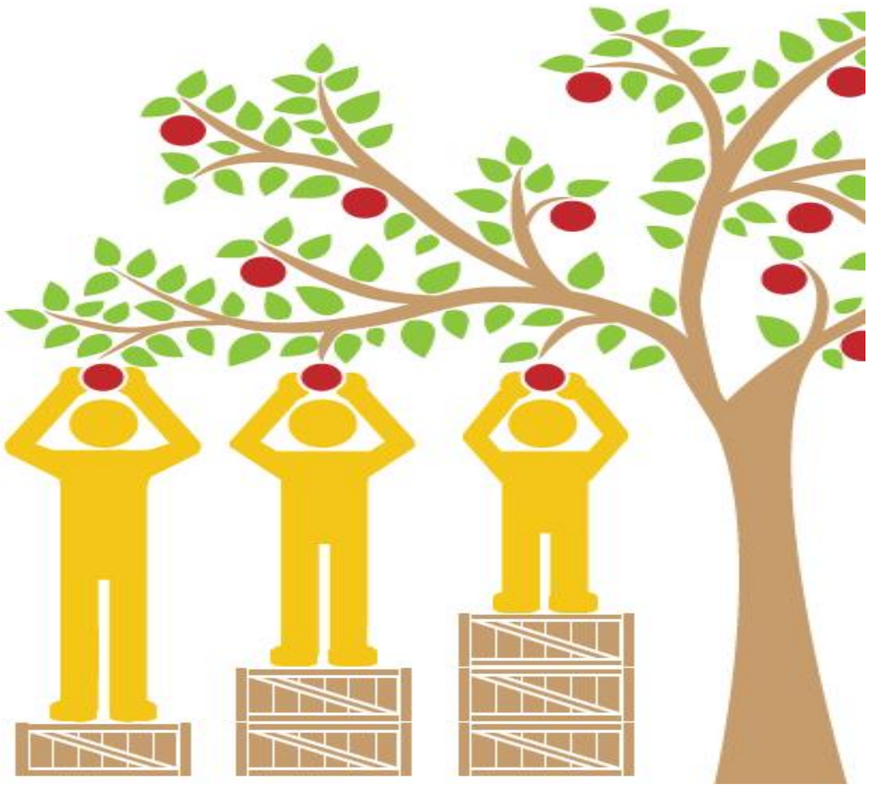
data RESPONSIBLY

Two notions of fairness

individual fairness

group fairness

equality



equity

two intrinsically different world views

Diversity

The New York Times

Artificial Intelligence's White Guy Problem

By KATE CRAWFORD JUNE 25, 2016

Like all technologies before it, artificial intelligence will reflect the values of its creators. So **inclusivity matters** — from who designs it to who sits on the company boards and which ethical perspectives are included.

Otherwise, **we risk constructing machine intelligence that mirrors a narrow and privileged vision of society**, with its old, familiar biases and stereotypes.

REVIEW

Diversity in Big Data: A Review

Marina Drosou,[1] H.V. Jagadish,[2] Evaggelia Pitoura,[1] and Julia Stoyanovich[3,*]

Big Data
Volume 5 Number 2, 2017
© Mary Ann Liebert, Inc.
DOI: 10.1089/big.2016.0054

Abstract

Big data technology offers unprecedented opportunities to society as a whole and also to its individual members. At the same time, this technology poses significant risks to those it overlooks. In this article, we give an overview of recent technical work on diversity, particularly in selection tasks, discuss connections between diversity and fairness, and identify promising directions for future work that will position diversity as an important component of a data-responsible society. We argue that diversity should come to the forefront of our discourse, for reasons that are both ethical—to mitigate the risks of exclusion—and utilitarian, to enable more powerful, accurate, and engaging data analysis and use.

Keywords: data; diversity; empirical studies; models and algorithms; responsibly

+ Fairness in ranked outputs, joint with Yang [Drexel] [FATML 2016] [SSDBM 2017]

data RESPONSIBLY

SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

Report from Dagstuhl Seminar 16291

Data, Responsibly

Edited by

Serge Abiteboul[1], Gerome Miklau[2], Julia Stoyanovich[3], and Gerhard Weikum[4]

1 ENS – Cachan, FR, serge.abiteboul@inria.fr
2 University of Massachusetts – Amherst, US, miklau@cs.umass.edu
3 Drexel University – Philadelphia, US, stoyanovich@drexel.edu
4 MPI für Informatik – Saarbrücken, DE, weikum@mpi-inf.mpg.de

The goals of the seminar were to assess the state of data analysis in terms of fairness, transparency and diversity, identify new research challenges, and derive an agenda for computer science research and education efforts in responsible data analysis and use.

An important goal of the seminar was to **identify opportunities for high-impact contributions to this important emergent area specifically from the data management community**.

http://drops.dagstuhl.de/opus/volltexte/2016/6764/pdf/dagrep_v006_i007_p042_s16291.pdf

Research Directions for Principles of Data Management (Dagstuhl Perspectives Workshop 16151)

Edited by
Serge Abiteboul, Marcelo Arenas, Pablo Barceló, Meghyn Bienvenu, Diego Calvanese, Claire David, Richard Hull, Eyke Hüllermeier, Benny Kimelfeld, Leonid Libkin, Wim Martens, Tova Milo, Filip Murlak, Frank Neven, Magdalena Ortiz, Thomas Schwentick, Julia Stoyanovich, Jianwen Su, Dan Suciu, Victor Vianu, and Ke Yi

1 Introduction

In April 2016, a community of researchers working in the area of Principles of Data Management (PDM) joined in a workshop at the Dagstuhl Castle in Germany. The workshop was organized jointly by the Executive Committee of the ACM Symposium on Principles of Database Systems (PODS) and the Council of the International Conference on Database Theory (ICDT). The mission of this workshop was to identify and explore some of the most important research directions that have high relevance to society and to Computer Science today, and where the PDM community has the potential to make significant contributions. This report describes the family of research directions that the workshop focused on from three perspectives: potential practical relevance, results already obtained, and research questions that appear surmountable in the short and medium term. This report organizes the identified research challenges for PDM around seven core themes, namely *Managing Data at Scale*, *Multi-model Data*, *Uncertain Information*, *Knowledge-enriched Data*, *Data Management and Machine Learning*, *Process and Data*, and *Ethics and Data Management*. Since new challenges in PDM arise all the time, we note that this list of themes is not intended to be exclusive.

Dagstuhl Manifestos 7(1): 1-29 (2018)

INTERNATIONAL | POLITIQUE | SOCIÉTÉ | ÉCO | CULTURE | IDÉES | PLANÈTE | SPORT | SCIENCES

M Sciences

SCIENCES | Vidéos | Archéologie | Affaire de logique | Astronomie | Biologie | Cerveau | Géophysic

ÉDITION ABONNÉS

Plaidoyer pour une analyse « responsable » des données

Face aux risques d'atteinte à la vie privée, les chercheurs en informatique Serge Abiteboul et Julia Stoyanovich plaident pour une collecte et une analyse des données impartiales, transparentes et accessibles à tous.

LE MONDE SCIENCE ET TECHNO | 12.10.2015 à 20h47 • Mis à jour le 19.10.2015 à 16h16

Association For Computing Machinery acm ACM SIGMOD Blog

Serge Abiteboul and Julia Stoyanovich

NOVEMBER 20, 2015

DATA, RESPONSIBLY

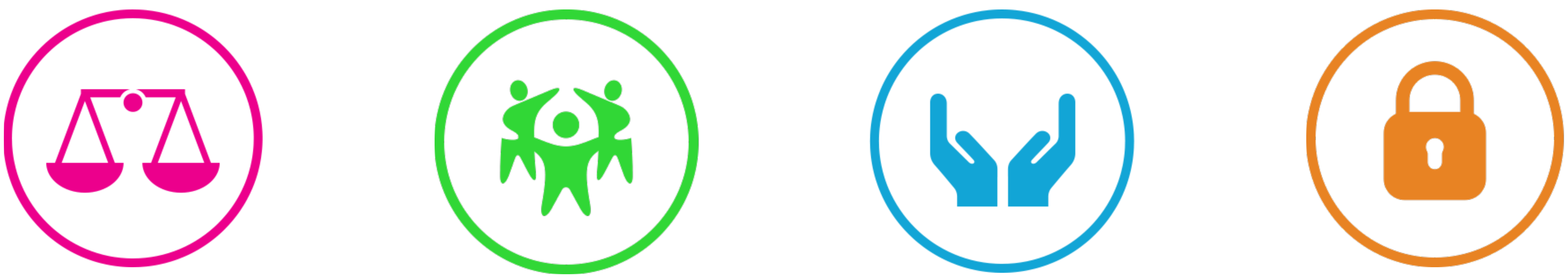
Big Data

(This blog post is an extended version of an October 12, 2015 Le Monde op-ed article (in French))

Our society is increasingly relying on algorithms in all aspects of its operation. We trust algorithms not only *to help carry out routine tasks*, such as accounting and automatic manufacturing, but also *to make decisions on our behalf*. The sorts of decisions with which we now casually entrust algorithms range from unsettling (killer drones), to tedious (automatic trading), or deeply personal (online dating). Computer technology has tremendous power, and with that power comes immense responsibility. Nowhere is the need to control the power and to judiciously use technology more apparent than in massive data analysis, known as big data.

Responsible data science

- Be **transparent** and **accountable**
- Achieve **equitable** resource distribution
- Be cognizant of the **rights** and **preferences** of individuals



fairness diversity transparency data protection

dataresponsibly.com

FAT is the new ACID

Thank you!