UNIVERSITAT DE BARCELONA

MSc in Fundamental Principles of Data Science

Ethical Data Science

Bias and Discrimination

Jordi Vitrià

2020-2021

Definitions

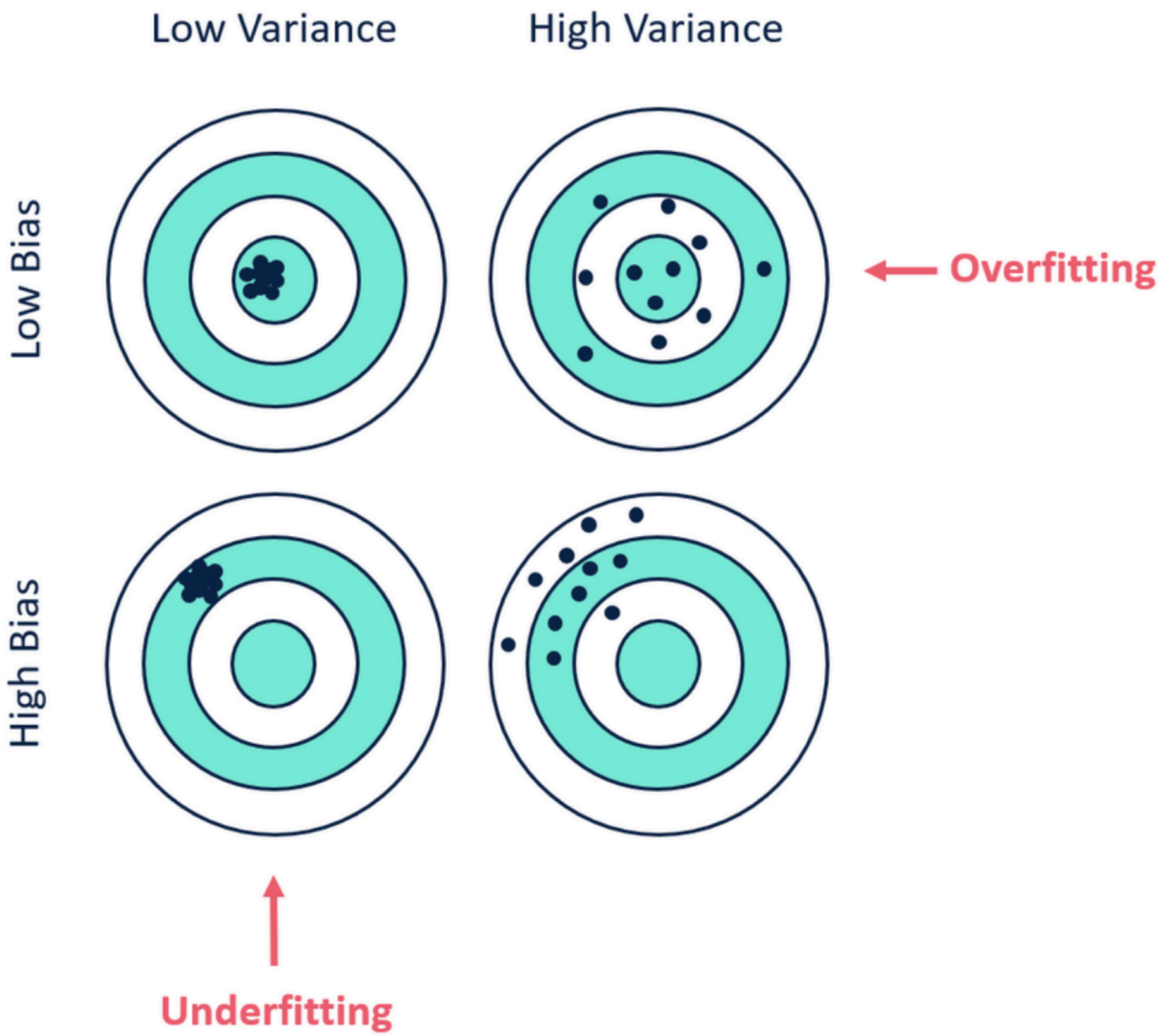
In ML, ideally, one wants to choose a model that both accurately captures the **regularities in its training data**, but also **generalizes well** to unseen data.

Unfortunately, it is typically impossible to do both simultaneously. High-variance learning methods may be able to represent their training set well but are at risk of overfitting to noisy or unrepresentative training data. In contrast, algorithms with **high bias** typically produce simpler models that may fail to capture important regularities (i.e. underfit) in the data.

The **bias error** is an error from erroneous assumptions in the learning algorithm. High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).

Source: Wikipedia

Definitions



https://community.alteryx.com/t5/Data-Science/Bias-Versus-Variance/ba-p/351862

Definitions

But bias is a need to generalize!

The Need for Biases in Learning Generalizations

Tom M. Mitchell

1. Introduction

Learning involves the ability to generalize from past experience in order to deal with new situations that are "related to" this experience. The inductive leaap needed to deal with new situations seems to be possible only under certain biases for choosing one generalization of the situation over another. This paper defines precisely the notion of bias in generalization problems, then shows that biases are necessary for the inductive leap. Classes of justifiable biases are considered, and the relationship between bias and domain-independence is considered.

We restrict the scope of this discussion to the problem of generalizing from training instances, defined as follows:

The Generalization Problem
Given:

1. Language of instances.
2. Language of generalizations.
3. Matching predicate for matching generalizations to instances.
4. Sets of positive and negative training instances.

Determine:

⇒ Generalization(s) consistent with the training instances.

As a concrete example of the above generalization problem, consider the task addressed by Winston's program for learning classes of block structures (Winston 1975). Here, the language of instances is the representation used to describe example block structures. The language of generalizations is the language in which learned concepts (e.g., arch, tower) are described. The matching predicate specifies whether a given generalization applies to a given instance (e.g., whether the inferred description of an arch is satisfied by a specific block structure).

This paper addresses a deep difficulty with the generalization problem as defined above: If consistency with the training instances is taken as the sole determiner of appropriate generalizations, then a program can never make the inductive leap necessary to classify instances beyond those it has observed. Only if the program has other sources of information, or biases for choosing one generalization over the other, can it non-arbitrarily classify instances beyond those in the training set.

Converted to electronic version by: Roby Joehanes, Kansas State University

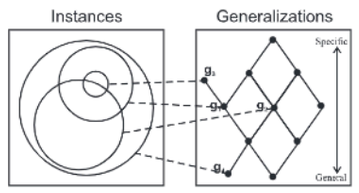


Figure 1: Relationships among Instances and Generalizations (This figure was missing from the original publication and added in 1990.)

In this paper, we use the term *bias* to refer to *any basis for choosing one generalization over another, other than strict consistency with the observed training instances*.

2. What is an UNbiased Generalizer

If generalization is the problem of guessing the class of instances to which the positive training instances belong, then an unbiased generalizer is one that makes no a priori assumptions about which classes of instances are most likely, but bases all its choices on the observed data. Two common sources of bias in existing learning systems are (1) the generalization language is not capable of expressing all possible classes of instances, and (2) the generalization procedure that searches through the space of expressible generalizations is itself biased.

2.1. An Unbiased Generalization Language

In considering bias in the generalization language, it is useful to view each generalization as denoting the set of instances that it matches. In figure 1, for example, g_1 and g_2 are two generalizations expressible in some generalization language, and each matches a different subset of the instances.

Relative to a given langauge of instances, an unbiased generalization language is then one which allows describing every possible subset of these instances. In short, an

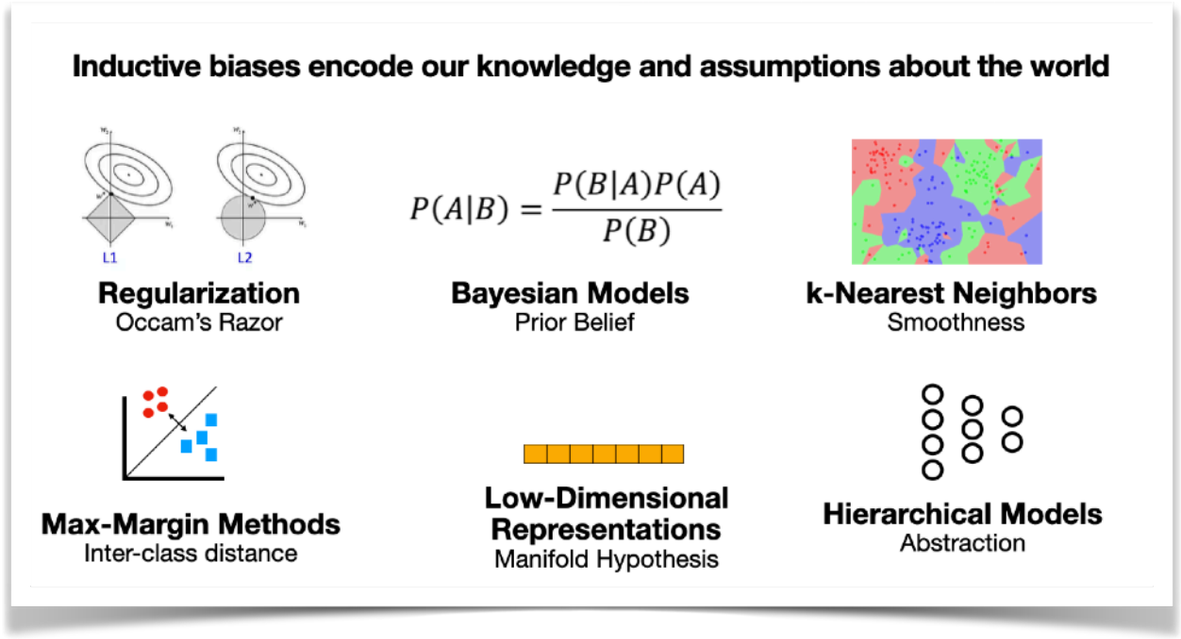
1980:
Bias in ML does help us generalize better and make our model less sensitive to some single data point.

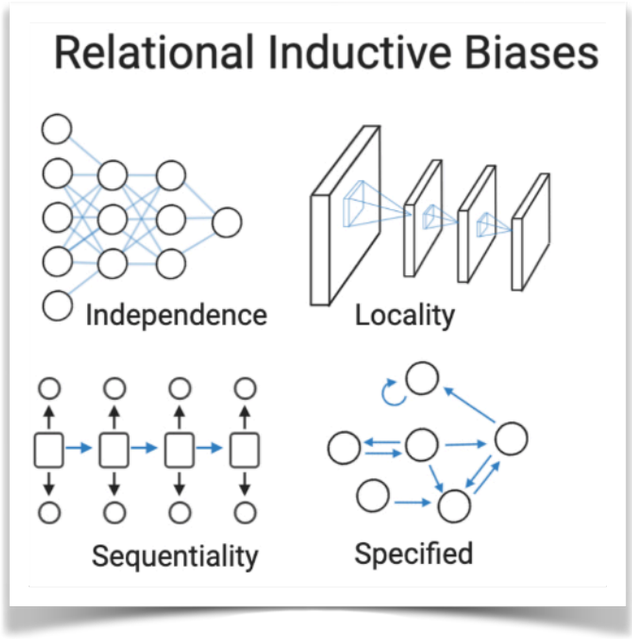
Definitions

> Our goal in building machine learning systems is to create algorithms whose utility extends beyond the dataset in which they are trained. The process of leveraging observations to draw inferences about the unobserved is the principle of induction.

The inductive **bias** (also known as learning bias) of a learning algorithm is the set of assumptions that the learner uses to predict outputs of given inputs that it has not encountered.

Source: Wikipedia





Definitions

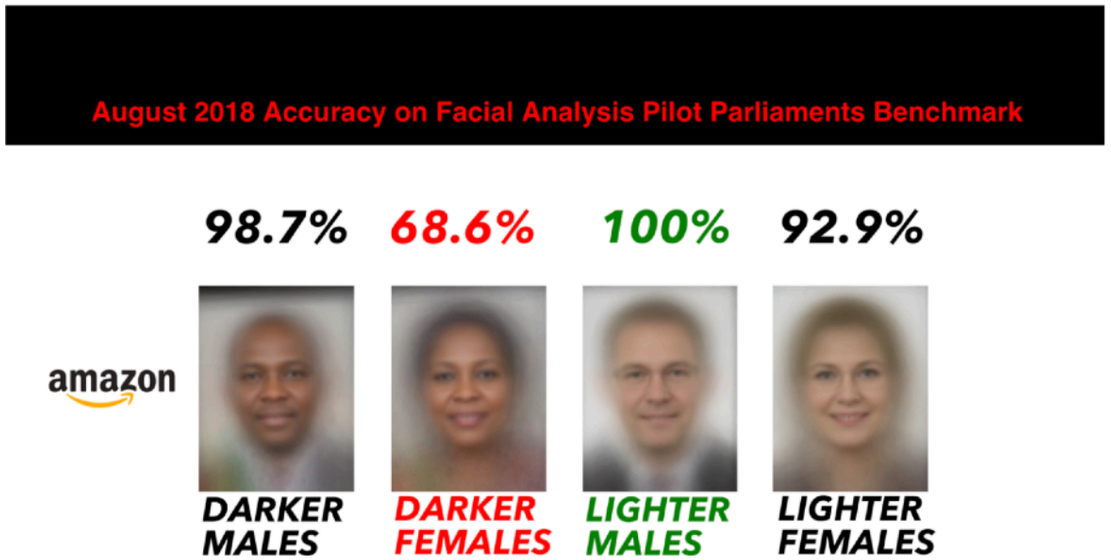
All machine learning systems use patterns in datasets to make predictions, but we have to determine which patterns should qualify as “undesirable biases” (that we shouldn’t use), as opposed to “valid patterns” which we should.

Undesirable bias are related to protected features of the data.

Motivation

Response: Racial and Gender bias in Amazon Rekognition — Commercial AI System for Analyzing Faces.

Joy Buolamwini Jan 25, 2019 · 15 min read



Motivation

NBER | NATIONAL BUREAU of ECONOMIC RESEARCH

‹ Working Papers

Consumer-Lending Discrimination in the FinTech Era

Robert Bartlett, Adair Morse, Richard Stanton & Nancy Wallace

Discrimination: Unjustified basis of differentiation between individuals

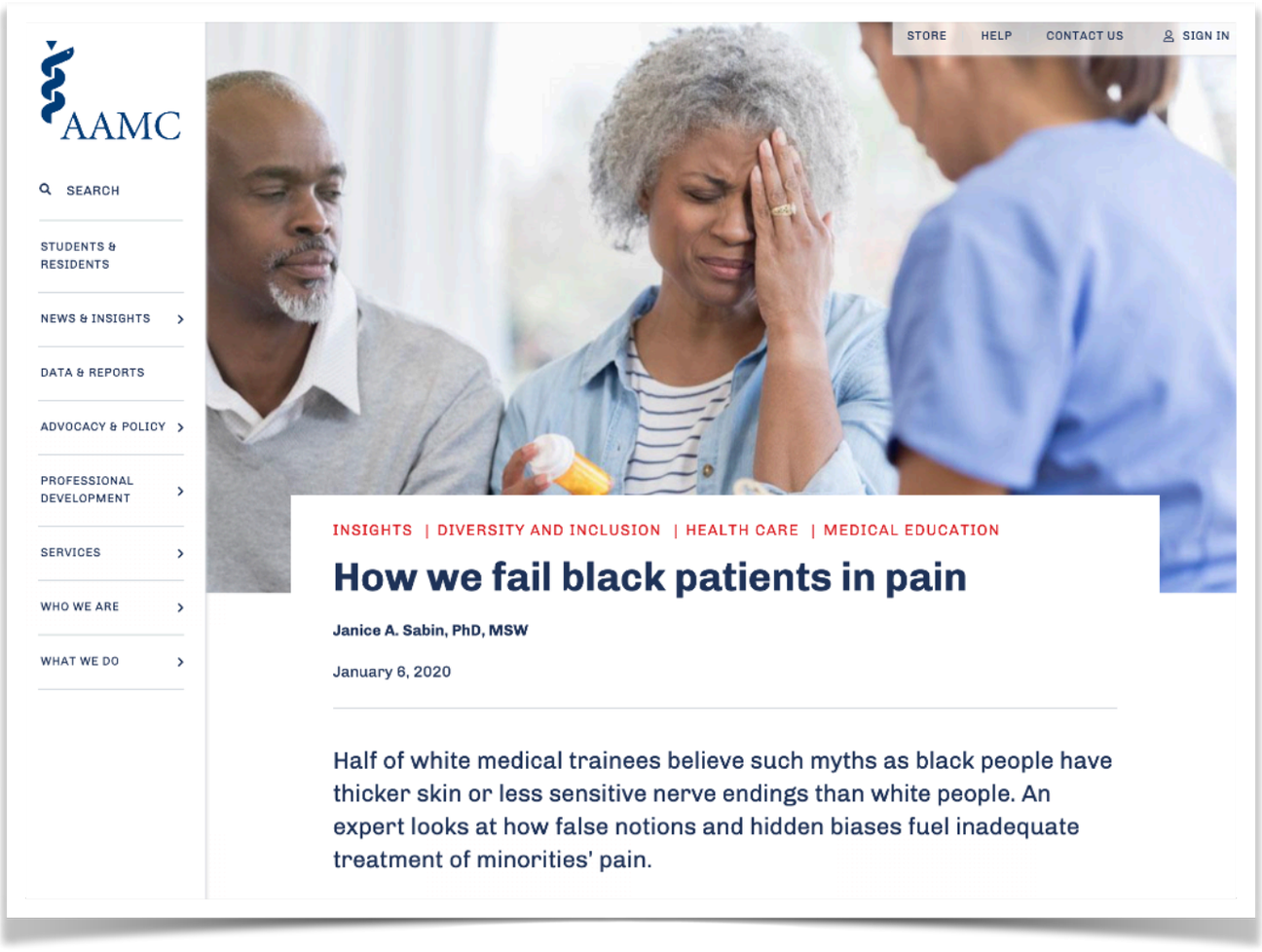
We find that **lenders charge Latin/African-American borrowers 7.9 and 3.6 basis points more for purchase and refinance mortgages respectively**, costing them $765M in aggregate per year in extra interest.

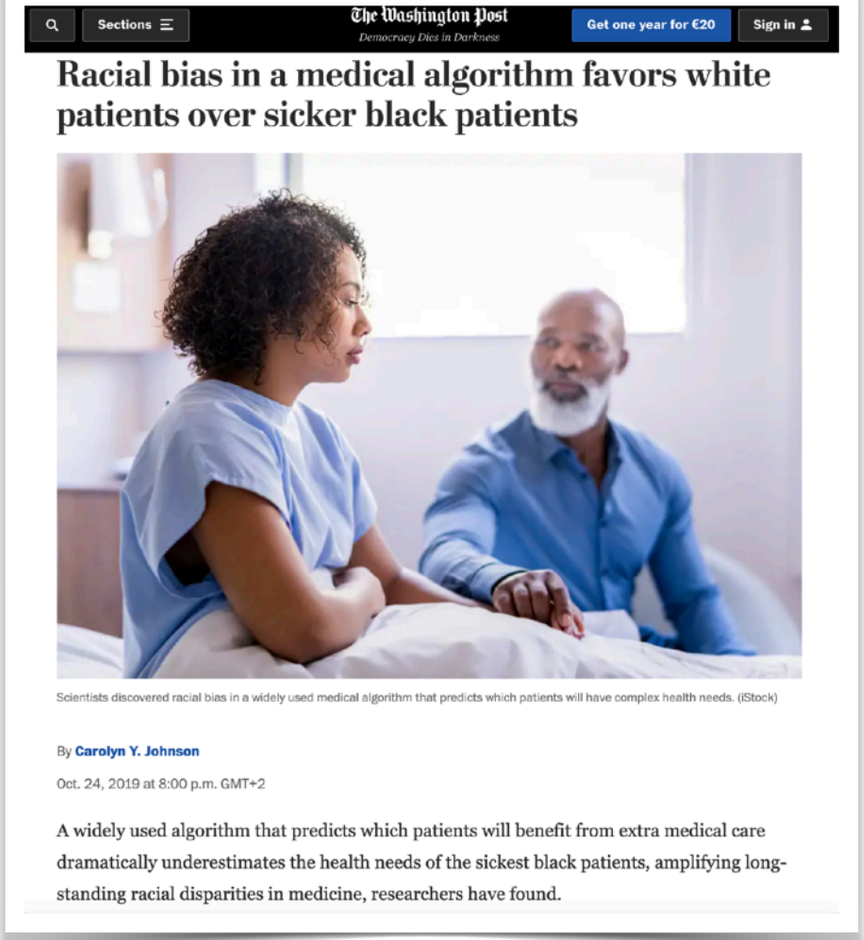
Bad News!

FinTech algorithms also discriminate, but 40% less than face-to-face lenders. The lower levels of price discrimination by algorithms suggests that removing face-to-face interactions can reduce discrimination.

Bias amplification

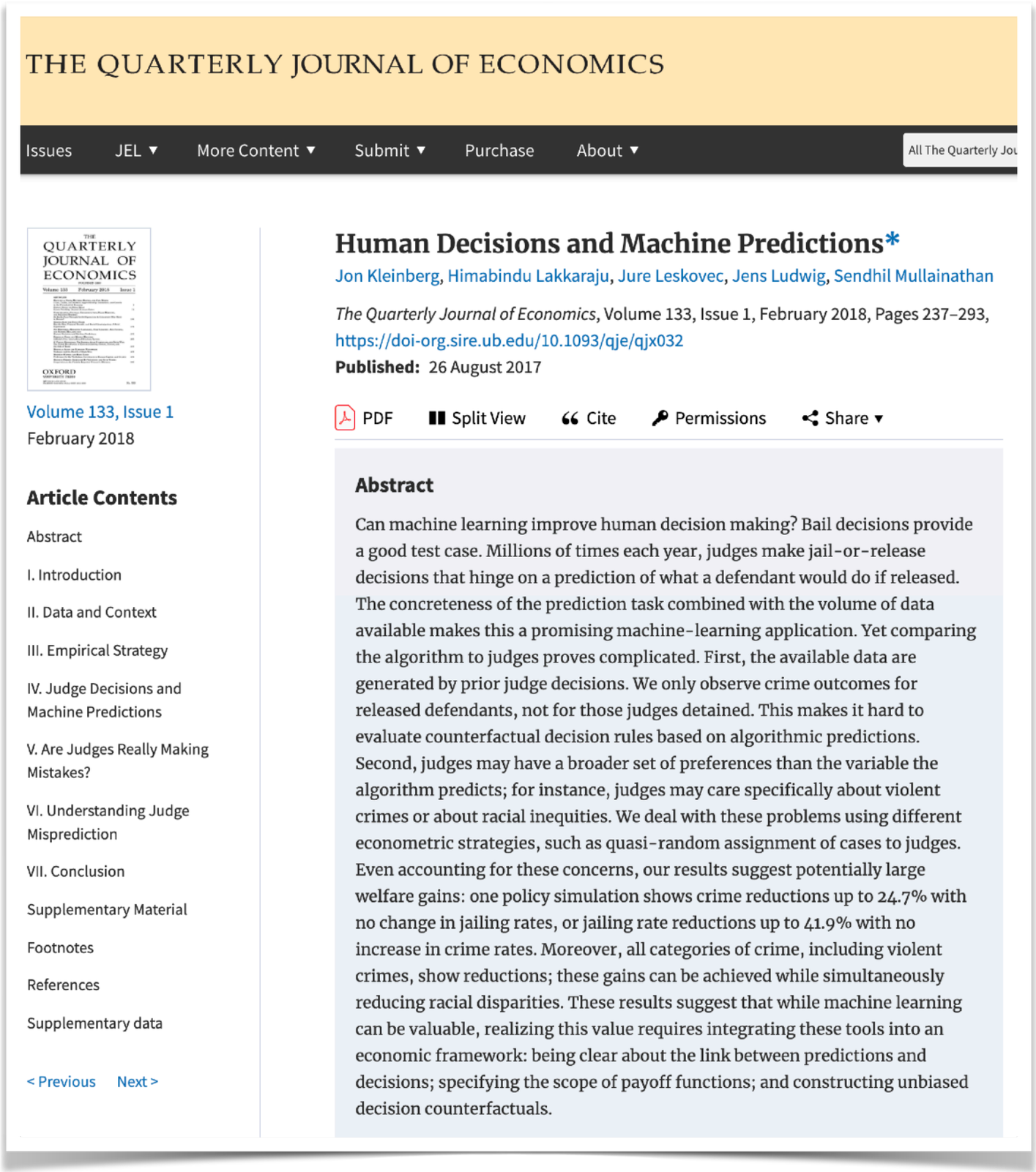
Motivation


AAMC
SEARCH
STUDENTS & RESIDENTS
NEWS & INSIGHTS
DATA & REPORTS
ADVOCACY & POLICY
PROFESSIONAL DEVELOPMENT
SERVICES
WHO WE ARE
WHAT WE DO
STORE
HELP
CONTACT US
SIGN IN
INSIGHTS | DIVERSITY AND INCLUSION | HEALTH CARE | MEDICAL EDUCATION
How we fail black patients in pain
Janice A. Sabin, PhD, MSW
January 6, 2020
Half of white medical trainees believe such myths as black people have thicker skin or less sensitive nerve endings than white people. An expert looks at how false notions and hidden biases fuel inadequate treatment of minorities' pain.


Sections
The Washington Post
Democracy Dies in Darkness
Get one year for €20
Sign in
Racial bias in a medical algorithm favors white patients over sicker black patients
Scientists discovered racial bias in a widely used medical algorithm that predicts which patients will have complex health needs. (iStock)
By Carolyn Y. Johnson
Oct. 24, 2019 at 8:00 p.m. GMT+2
A widely used algorithm that predicts which patients will benefit from extra medical care dramatically underestimates the health needs of the sickest black patients, amplifying long-standing racial disparities in medicine, researchers have found.

Motivation

X-ray image datasets used to diagnose various thoracic diseases


THE QUARTERLY JOURNAL OF ECONOMICS

Issues JEL More Content Submit Purchase About All The Quarterly Jou

QUARTERLY JOURNAL OF ECONOMICS

Volume 133, Issue 1
February 2018

Article Contents

Abstract
I. Introduction
II. Data and Context
III. Empirical Strategy
IV. Judge Decisions and Machine Predictions
V. Are Judges Really Making Mistakes?
VI. Understanding Judge Misprediction
VII. Conclusion
Supplementary Material
Footnotes
References
Supplementary data

< Previous Next >

Human Decisions and Machine Predictions*

Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan

The Quarterly Journal of Economics, Volume 133, Issue 1, February 2018, Pages 237–293,
https://doi-org.sire.ub.edu/10.1093/qje/qjx032
Published: 26 August 2017

PDF Split View Cite Permissions Share

Abstract

Can machine learning improve human decision making? Bail decisions provide a good test case. Millions of times each year, judges make jail-or-release decisions that hinge on a prediction of what a defendant would do if released. The concreteness of the prediction task combined with the volume of data available makes this a promising machine-learning application. Yet comparing the algorithm to judges proves complicated. First, the available data are generated by prior judge decisions. We only observe crime outcomes for released defendants, not for those judges detained. This makes it hard to evaluate counterfactual decision rules based on algorithmic predictions. Second, judges may have a broader set of preferences than the variable the algorithm predicts; for instance, judges may care specifically about violent crimes or about racial inequities. We deal with these problems using different econometric strategies, such as quasi-random assignment of cases to judges. Even accounting for these concerns, our results suggest potentially large welfare gains: one policy simulation shows crime reductions up to 24.7% with no change in jailing rates, or jailing rate reductions up to 41.9% with no increase in crime rates. Moreover, all categories of crime, including violent crimes, show reductions; these gains can be achieved while simultaneously reducing racial disparities. These results suggest that while machine learning can be valuable, realizing this value requires integrating these tools into an economic framework: being clear about the link between predictions and decisions; specifying the scope of payoff functions; and constructing unbiased decision counterfactuals.

Bail is a set of pre-trial restrictions that are imposed on a suspect to ensure that they will not hamper the judicial process.

Bail is the conditional release of a defendant with the promise to appear in court when required.

Motivation

Which police departments should the feds investigate?

Arrests per 100 residents (2019) and police killings per 100,00 residents (2013-20) by race alongside disparities between those numbers for the police departments with the 37 largest jurisdictions in the U.S.

	ARRESTS/100			KILLINGS/100K		
POLICE DEPARTMENT	WHITE	BLACK	DIS.	WHITE	BLACK	DIS.
Albuquerque, NM	4.0	10.4	2.6	5.5	19.5	3.6
Austin, TX	2.5	9.4	3.8	4.0	7.2	1.8
Baltimore, MD	1.9	5.5	2.9	2.4	7.6	3.2
Boston, MA	0.8	2.4	2.9	0.3	5.8	17.6
Charlotte-Mecklenburg, NC	1.0	4.8	4.8	1.0	3.7	3.7
Chicago, IL*	1.7	6.8	4.1	0.3	7.4	22.1
Columbus, OH	1.0	2.5	2.6	2.5	12.7	5.1
Dallas, TX	2.0	5.0	2.5	3.1	5.1	1.6
Denver, CO	3.6	11.0	3.1	3.0	8.0	2.7
Detroit, MI	1.1	2.0	1.8	1.4	2.5	1.7
El Paso, TX	2.6	5.2	2.0	5.6	8.7	1.6
Fort Worth, TX	1.8	4.1	2.3	1.8	5.7	3.2
Fresno, CA**	5.6	11.2	2.0	3.5	2.7	0.8
Honolulu, HI	2.2	5.0	2.2	2.2	0.0	0.0
Houston, TX	1.1	3.5	3.2	1.6	7.5	4.7
Indianapolis, IN	2.8	6.1	2.2	2.1	7.5	3.5
Jacksonville, FL*	2.4	6.1	2.6	4.2	8.6	2.1
Las Vegas Metro, NV	3.9	13.2	3.4	3.3	5.9	1.8
Los Angeles, CA	1.8	4.4	2.4	1.8	8.2	4.6
Louisville Metro, KY	4.3	10.0	2.3	2.5	9.1	3.7
Memphis, TN	2.3	6.3	2.7	2.4	3.6	1.5
Mesa, AZ	3.1	12.9	4.2	4.6	0.0	0.0
Milwaukee, WI	1.2	4.4	3.8	1.0	7.0	7.3
Nashville Metropolitan, TN	2.7	6.5	2.4	0.8	3.8	4.7
New York, NY*	2.0	5.5	2.7	0.4	2.9	7.9
Oklahoma City, OK	2.1	6.3	3.0	5.0	27.2	5.5
Philadelphia, PA**	2.3	4.6	2.0	0.5	3.9	7.0
Phoenix, AZ	3.5	10.6	3.0	7.2	15.2	2.1
Portland, OR	3.0	12.8	4.3	2.9	11.1	3.9
Sacramento, CA	3.0	8.3	2.8	3.7	9.3	2.5
San Antonio, TX	2.5	9.3	3.7	3.0	10.5	3.5
San Diego, CA	2.8	8.7	3.2	2.0	3.5	1.7
San Francisco, CA	2.0	11.9	5.8	1.4	11.5	8.1
San Jose, CA	2.8	6.7	2.4	2.6	3.4	1.3
Seattle, WA	1.1	7.0	6.1	2.2	12.4	5.7
Tucson, AZ	7.2	20.2	2.8	4.2	7.9	1.9
Washington, D.C.*	0.9	6.4	7.3	0.4	5.4	13.4

*The departments serving Chicago, Jacksonville, New York and Washington, D.C., do not report their arrests disaggregated by race to the FBI, but release their data independently. We excluded arrests for traffic violations to make their data comparable to that released by the FBI.

**Data from the Fresno and Philadelphia police departments are from 2018.

SOURCES: MAPPING POLICE VIOLENCE, FBI UNIFORM CRIME REPORT, U.S. CENSUS BUREAU

Motivation

How numbers that appear equitable can obscure bias

Let's say a police officer is patrolling the street, looking for people with contraband. The officer sees 100 people, some of whom have contraband on their person. Say the crowd is evenly split between Black and white people.

FiveThirtyEight

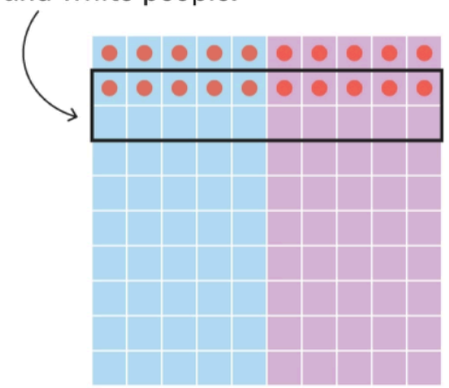
Motivation

How numbers that appear equitable can obscure bias

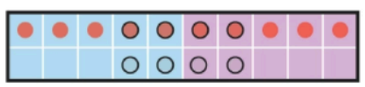
Let's say a police officer is patrolling the street, looking for people with contraband. The officer sees 100 people, some of whom have ● contraband on their person. Say the crowd is evenly split between Black and white people.

SCENARIO 1

The police officer stops **20** people, pulling aside equal numbers of Black and white people.



Of the **20** people stopped, the officer uses ○ force against **8** of them.



The police officer used force against stopped white people and stopped Black people at the same rate: **40%**.

But that's not the only scenario that can lead to that **40%** number.

FiveThirtyEight

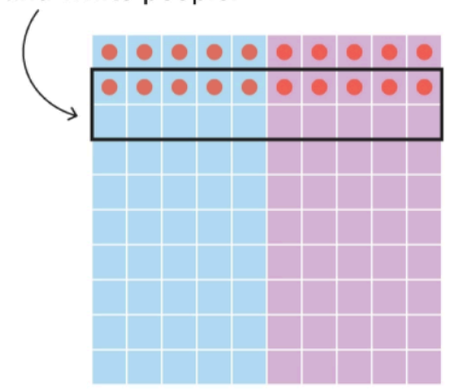
Motivation

How numbers that appear equitable can obscure bias

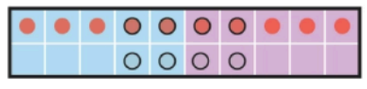
Let's say a police officer is patrolling the street, looking for people with contraband. The officer sees 100 people, some of whom have ● contraband on their person. Say the crowd is evenly split between Black and white people.

SCENARIO 1

The police officer stops **20** people, pulling aside equal numbers of Black and white people.



Of the **20** people stopped, the officer uses ○ force against **8** of them.

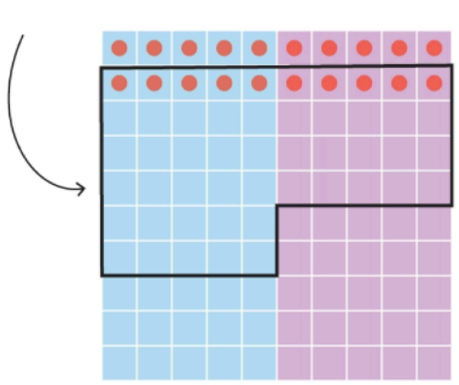


The police officer used force against stopped white people and stopped Black people at the same rate: **40%**.

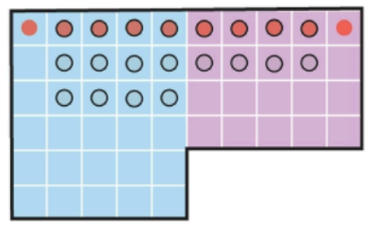
But that's not the only scenario that can lead to that **40%** number.

SCENARIO 2

This time, of the **100** people the officer sees, he stops **50**. But this time he is biased in whom he pulls aside.



The officer uses force against **20** people this time.



This time, like last time, the police officer used force against stopped white people and stopped Black people at the same rate: **40%**.

FiveThirtyEight

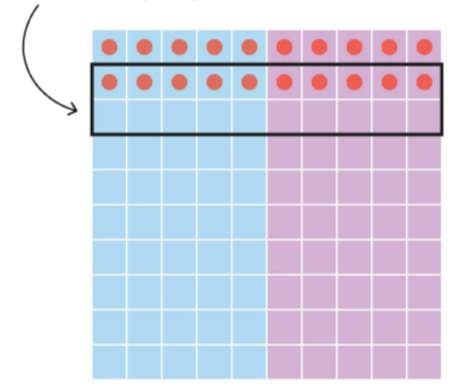
Motivation

How numbers that appear equitable can obscure bias

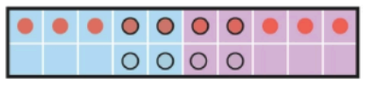
Let's say a police officer is patrolling the street, looking for people with contraband. The officer sees 100 people, some of whom have ● contraband on their person. Say the crowd is evenly split between Black and white people.

SCENARIO 1

The police officer stops **20** people, pulling aside equal numbers of Black and white people.



Of the **20** people stopped, the officer uses ○ force against **8** of them.

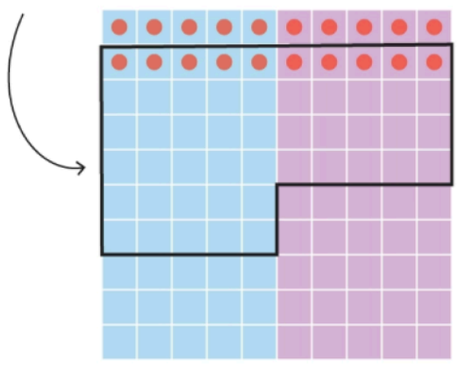


The police officer used force against stopped white people and stopped Black people at the same rate: **40%**.

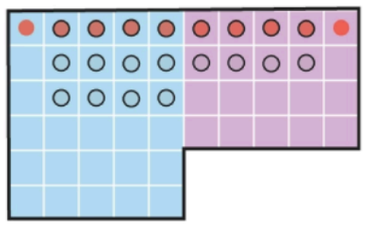
But that's not the only scenario that can lead to that **40%** number.

SCENARIO 2

This time, of the **100** people the officer sees, he stops **50**. But this time he is biased in whom he pulls aside.



The officer uses force against **20** people this time.



This time, like last time, the police officer used force against stopped white people and stopped Black people at the same rate: **40%**.

ANALYSIS

Things might appear equal, but in the second scenario, more Black people were stopped by the police than white people.

While use of force among **stopped** people is equal, use of force among all **observed** people is not:

$\frac{12}{50}$ = **24%** of Black people have force used against them

$\frac{8}{50}$ = **16%** of white people have force used against them

FiveThirtyEight

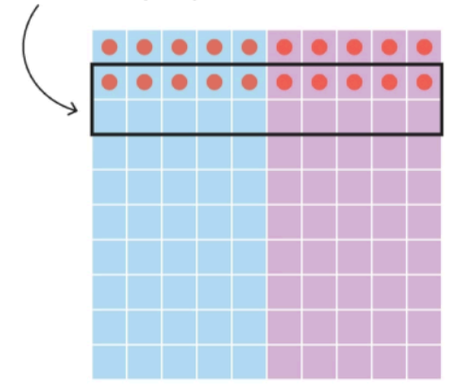
Motivation

How numbers that appear equitable can obscure bias

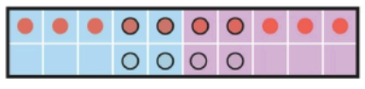
Let's say a police officer is patrolling the street, looking for people with contraband. The officer sees 100 people, some of whom have ● contraband on their person. Say the crowd is evenly split between Black and white people.

SCENARIO 1

The police officer stops **20** people, pulling aside equal numbers of Black and white people.



Of the **20** people stopped, the officer uses ○ force against **8** of them.

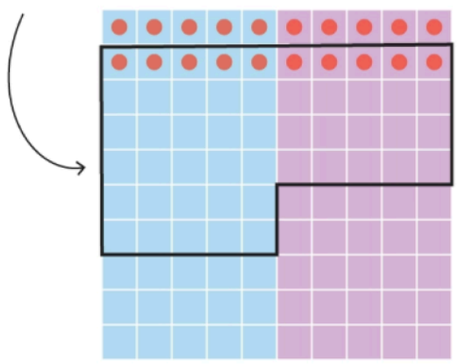


The police officer used force against stopped white people and stopped Black people at the same rate: **40%**.

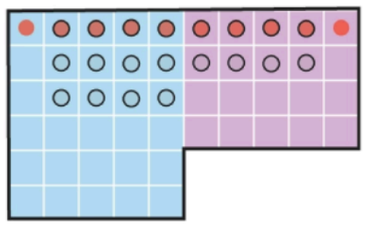
But that's not the only scenario that can lead to that **40%** number.

SCENARIO 2

This time, of the **100** people the officer sees, he stops **50**. But this time he is biased in whom he pulls aside.



The officer uses force against **20** people this time.



This time, like last time, the police officer used force against stopped white people and stopped Black people at the same rate: **40%**.

ANALYSIS

Things might appear equal, but in the second scenario, more Black people were stopped by the police than white people.

While use of force among **stopped** people is equal, use of force among all **observed** people is not:

$\frac{12}{50}$ = **24%** of Black people have force used against them

$\frac{8}{50}$ = **16%** of white people have force used against them

CONCLUSION

This is why knowing how often police use force against people they've stopped is **not enough information** to know whether use of force is racially biased. In real life, we don't have data on everyone who was observed but not stopped, but we need that to know whether use of force is biased overall.

FiveThirtyEight

Motivation

How can we evaluate in a sound way the rule "*Women and children first..*" in the case of the Titanic? Were low class passengers discriminated?

Journal of Economic Perspectives—Volume 28, Number 2—Spring 2014—Pages 3–28

Big Data: New Tricks for Econometrics[*]

Hal R. Varian

Computers are now involved in many economic transactions and can capture data associated with these transactions, which can then be manipulated and analyzed. Conventional statistical and econometric techniques such as regression often work well, but there are issues unique to big datasets that may require different tools.

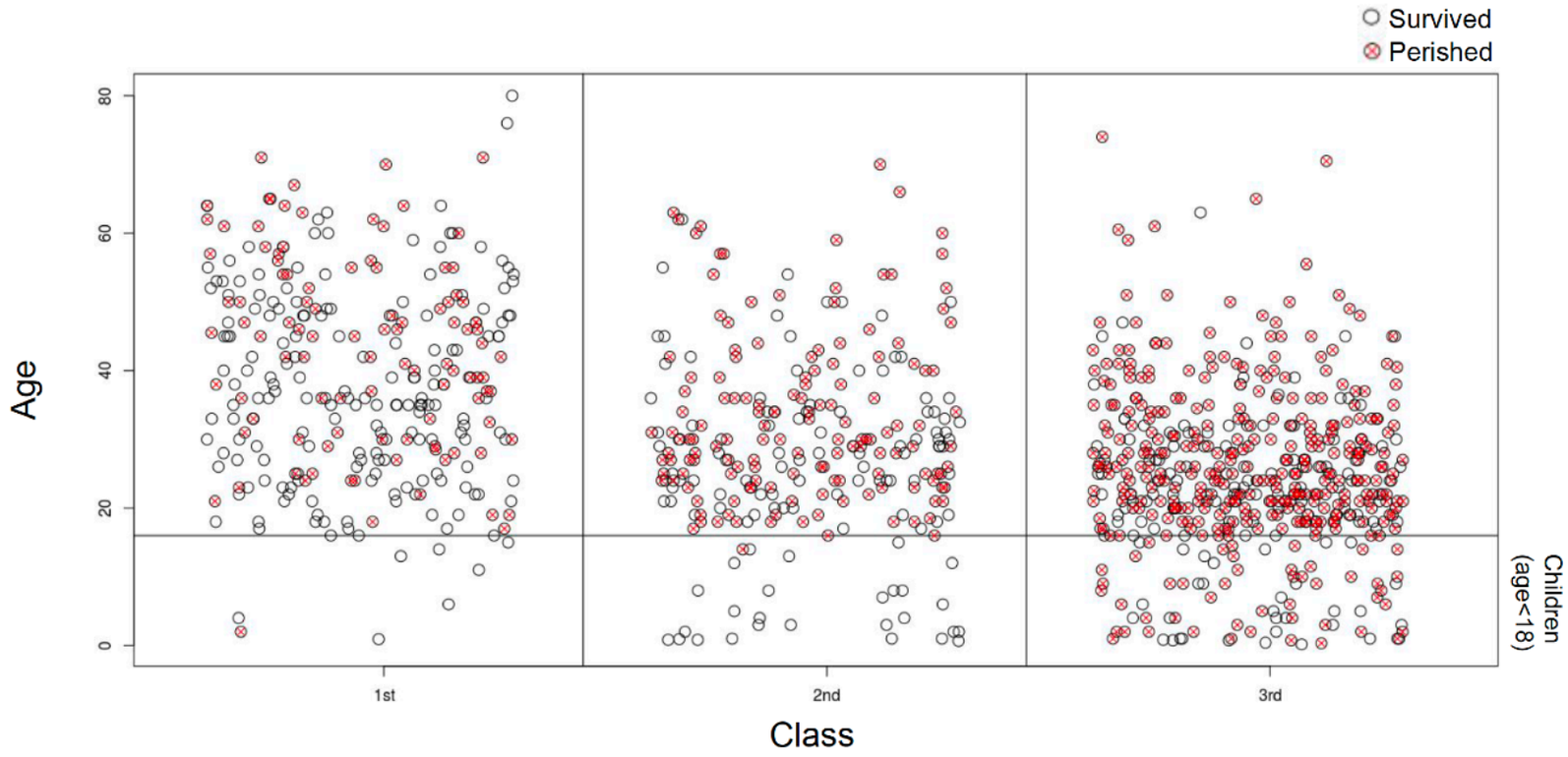
First, the sheer size of the data involved may require more powerful data manipulation tools. Second, we may have more potential predictors than appropriate for estimation, so we need to do some kind of variable selection. Third, large datasets may allow for more flexible relationships than simple linear models. Machine learning techniques such as decision trees, support vector machines, neural nets, deep learning, and so on may allow for more effective ways to model complex relationships.

In this essay, I will describe a few of these tools for manipulating and analyzing big data. I believe that these methods have a lot to offer and should be more widely known and used by economists. In fact, my standard advice to graduate students these days is go to the computer science department and take a class in machine learning. There have been very fruitful collaborations between computer scientists and statisticians in the last decade or so, and I expect collaborations between computer scientists and econometricians will also be productive in the future.

■ *Hal Varian is Chief Economist, Google Inc., Mountain View, California, and Emeritus Professor of Economics, University of California, Berkeley, California. His email address is hal@ischool.berkeley.edu.*

[*] To access the Appendix and disclosure statements, visit http://dx.doi.org/10.1257/jep.28.2.3

doi=10.1257/jep.28.2.3



Spoiler:
The rule that was applied to the Titanic class was "**Women and children first... particularly if they were traveling first class**"

Motivation

discriminate

verb

UK /dɪˈskrɪm.ɪ.neɪt/ **US** /dɪˈskrɪm.ə.neɪt/

Cambridge Dictionary
Make your words meaningful

discriminate *verb* (TREAT DIFFERENTLY)

C1 [I]

to treat a person or particular group of people differently, especially in a worse way from the way in which you treat other people, because of their skin colour, sex, sexuality, etc.:

C2 [I + adv/prep] formal

to be able to recognize the difference between people or things:

Motivation

Under the most advanced law systems, everyone is protected from **unlawful behavior** (discrimination) when the cause of this behavior is that they **have** or **are perceived to have** a "*protected characteristic*" or are associated with someone who has a protected characteristic:

- Age
- Disability
- Gender
- Civil state
- Pregnancy and maternity
- Race
- Religion and belief
- Sex
- Sexual orientation

Motivation

There are several types of discrimination:

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

Disparate treatment/Direct discrimination:
Treatment depends on class membership

Disparate impact or indirect discrimination:
Outcome depends on class membership

Motivation

An employer does not interview a job applicant because of the applicant's ethnic background

A hair salon owner has a policy of not employing stylists who cover their hair, believing it is important for them to exhibit their flamboyant haircuts.

An employer offers flexible working to all staff. Requests are supposed to be considered based on business need. A manager allows a man's request to work flexibly to train for a qualification but does not allow another man's request to work flexibly to care for his disabled child.

An employer dismisses a worker because she has had three months' sick leave. The employer is aware that the worker has multiple sclerosis and most of her sick leave is disability-related.

An employer has a policy that designated car parking spaces are only offered to senior managers. A worker who is not a manager, but has a mobility impairment is not given a designated car parking space.

A builder addresses abusive and hostile remarks to a customer because of her race after their business relationship has ended.

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

Motivation

1

An employer does not interview a job applicant because of the applicant's ethnic background

A hair salon owner has a policy of not employing stylists who cover their hair, believing it is important for them to exhibit their flamboyant haircuts.

An employer offers flexible working to all staff. Requests are supposed to be considered based on business need. A manager allows a man's request to work flexibly to train for a qualification but does not allow another man's request to work flexibly to care for his disabled child.

An employer dismisses a worker because she has had three months' sick leave. The employer is aware that the worker has multiple sclerosis and most of her sick leave is disability-related.

An employer has a policy that designated car parking spaces are only offered to senior managers. A worker who is not a manager, but has a mobility impairment is not given a designated car parking space.

A builder addresses abusive and hostile remarks to a customer because of her race after their business relationship has ended.

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

Motivation

1

An employer does not interview a job applicant because of the applicant's ethnic background

2

A hair salon owner has a policy of not employing stylists who cover their hair, believing it is important for them to exhibit their flamboyant haircuts.

An employer offers flexible working to all staff. Requests are supposed to be considered based on business need. A manager allows a man's request to work flexibly to train for a qualification but does not allow another man's request to work flexibly to care for his disabled child.

An employer dismisses a worker because she has had three months' sick leave. The employer is aware that the worker has multiple sclerosis and most of her sick leave is disability-related.

An employer has a policy that designated car parking spaces are only offered to senior managers. A worker who is not a manager, but has a mobility impairment is not given a designated car parking space.

A builder addresses abusive and hostile remarks to a customer because of her race after their business relationship has ended.

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

Motivation

1. An employer does not interview a job applicant because of the applicant's ethnic background

2. A hair salon owner has a policy of not employing stylists who cover their hair, believing it is important for them to exhibit their flamboyant haircuts.

4. An employer offers flexible working to all staff. Requests are supposed to be considered based on business need. A manager allows a man's request to work flexibly to train for a qualification but does not allow another man's request to work flexibly to care for his disabled child.

An employer dismisses a worker because she has had three months' sick leave. The employer is aware that the worker has multiple sclerosis and most of her sick leave is disability-related.

An employer has a policy that designated car parking spaces are only offered to senior managers. A worker who is not a manager, but has a mobility impairment is not given a designated car parking space.

A builder addresses abusive and hostile remarks to a customer because of her race after their business relationship has ended.

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

Motivation

1. An employer does not interview a job applicant because of the applicant's ethnic background

2. A hair salon owner has a policy of not employing stylists who cover their hair, believing it is important for them to exhibit their flamboyant haircuts.

3. An employer dismisses a worker because she has had three months' sick leave. The employer is aware that the worker has multiple sclerosis and most of her sick leave is disability-related.

4. An employer offers flexible working to all staff. Requests are supposed to be considered based on business need. A manager allows a man's request to work flexibly to train for a qualification but does not allow another man's request to work flexibly to care for his disabled child.

An employer has a policy that designated car parking spaces are only offered to senior managers. A worker who is not a manager, but has a mobility impairment is not given a designated car parking space.

A builder addresses abusive and hostile remarks to a customer because of her race after their business relationship has ended.

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

Motivation

1

An employer does not interview a job applicant because of the applicant's ethnic background

2

A hair salon owner has a policy of not employing stylists who cover their hair, believing it is important for them to exhibit their flamboyant haircuts.

4

An employer offers flexible working to all staff. Requests are supposed to be considered based on business need. A manager allows a man's request to work flexibly to train for a qualification but does not allow another man's request to work flexibly to care for his disabled child.

3

An employer dismisses a worker because she has had three months' sick leave. The employer is aware that the worker has multiple sclerosis and most of her sick leave is disability-related.

5

An employer has a policy that designated car parking spaces are only offered to senior managers. A worker who is not a manager, but has a mobility impairment is not given a designated car parking space.

A builder addresses abusive and hostile remarks to a customer because of her race after their business relationship has ended.

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

Motivation

1. An employer does not interview a job applicant because of the applicant's ethnic background

2. A hair salon owner has a policy of not employing stylists who cover their hair, believing it is important for them to exhibit their flamboyant haircuts.

3. An employer dismisses a worker because she has had three months' sick leave. The employer is aware that the worker has multiple sclerosis and most of her sick leave is disability-related.

4. An employer offers flexible working to all staff. Requests are supposed to be considered based on business need. A manager allows a man's request to work flexibly to train for a qualification but does not allow another man's request to work flexibly to care for his disabled child.

5. An employer has a policy that designated car parking spaces are only offered to senior managers. A worker who is not a manager, but has a mobility impairment is not given a designated car parking space.

6. A builder addresses abusive and hostile remarks to a customer because of her race after their business relationship has ended.

1. **Direct discrimination** . This means treating someone less favourably than someone else because of a protected characteristic.
2. **Direct discrimination by perception**. This means treating one person less favourably than someone else, because you incorrectly think they have a protected characteristic.
3. **Discrimination arising from disability**. This means treating a disabled person unfavourably because of something connected with their disability when this cannot be objectively justified.
4. **Direct discrimination by association.** This means treating someone less favourably than another person because they are associated with a person who has a protected characteristic.
5. **Failing to make reasonable adjustments.** To do this for disabled people is also a form of discrimination.
6. **Harassment.** Harassment is unwanted behaviour related to a protected characteristic which has the purpose or effect of violating someone's dignity or which creates a hostile, degrading, humiliating or offensive environment.

https://www.equalityhumanrights.com/sites/default/files/ea_legal_definitions_0.pdf

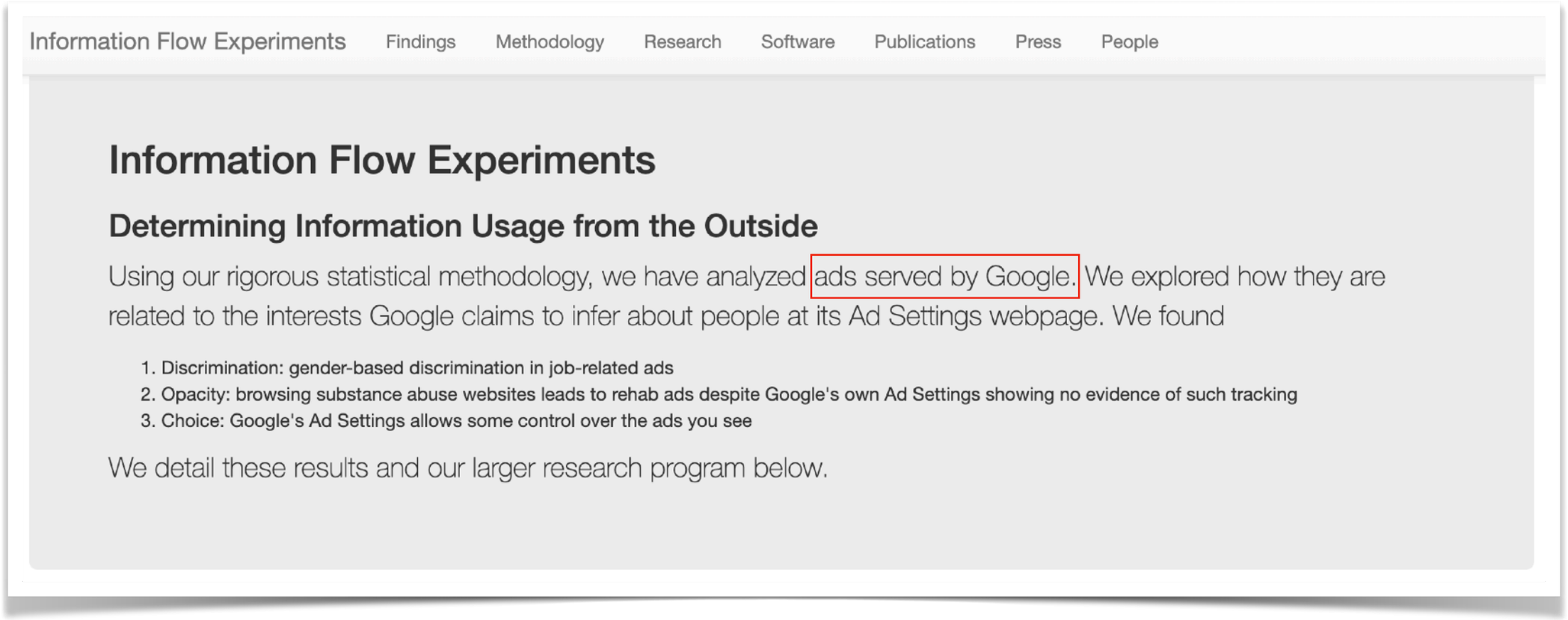
Motivation

Algorithmic discrimination scenarios:

- Access to employment
- Access to education
- Access to government/companies benefits
- Access to penitentiary alternatives
- Etc.

Anti-discrimination legislation typically seeks **equal access** to employment, working conditions, education, social protection, goods, and services.

Motivation



https://www.cs.cmu.edu/~mtschant/ife/

Motivation

Over hundreds of browsers, we randomly edited the profile to be either “female” or “male” and visited job-related websites. We found that the “male” instances were much more likely to receive ads promoting high paying jobs than the “female” instances.

Top ads for identifying the female group

Ad Title	Ad URL	Times shown to Females	Times shown to Males
Jobs (Hiring Now)	www.jobsinyourarea.co	45	8
4Runner Parts Service	www.westernpatoyotaservice.com	36	5
Criminal Justice Program	www3.mc3.edu/Criminal+Justice	29	1
Goodwill - Hiring	goodwill.careerboutique.com	121	39
UMUC Cyber Training	www.umuc.edu/cybersecuritytraining	38	30

Top ads for identifying the male group

Ad Title	Ad URL	Times shown to Females	Times shown to Males
$200k+ Jobs - Execs Only	careerchange.com	311	1816
Find Next $200k+ Job	careerchange.com	7	36
Become a Youth Counselor	www.youthcounseling.degreeleap.com	0	310
CDL-A OTR Trucking Jobs	www.tadrivers.com/OTRJobs	0	8
Free Resume Templates	resume-templates.resume-now.com	8	10

Learning objectives

Given a large database of historical **decision records**, find, measure and mitigate discriminatory situations and practices.