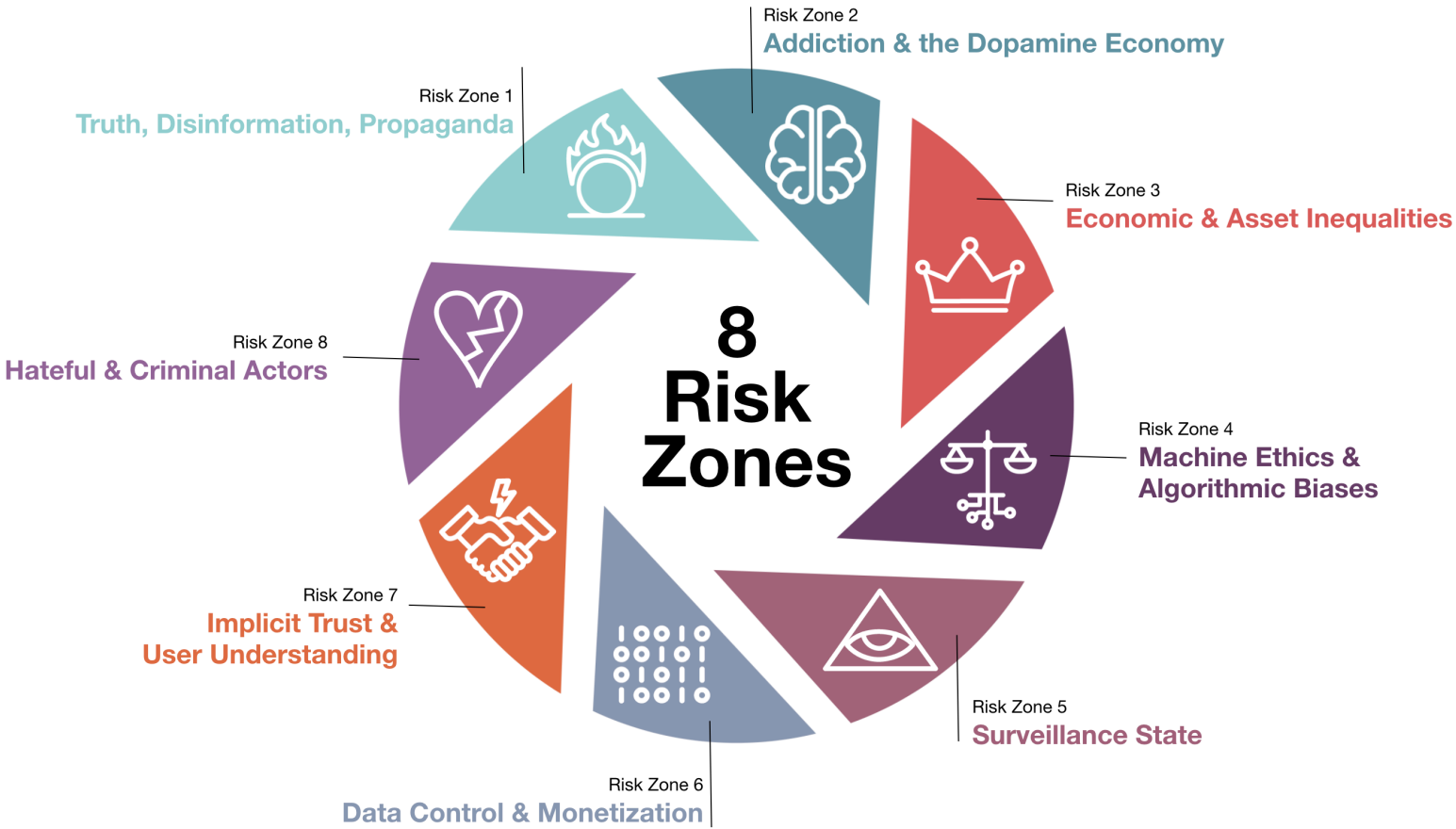
Algorithms week 5-1:
Algorithmic Accountability

Jonathan Stray
Columbia Lede Program
August 13, 2018



Institute for the Future unintended harms of technology risk zones, ethicalos.org

Algorithm Tips

Find tips for stories on algorithms

What is this?

This is a growing list of potentially newsworthy algorithms used by the U.S. government. As more decisions are influenced by algorithms, more algorithmic accountability may be needed. But where to start? Here, you can find algorithms warranting a closer look. Read about our criteria and sources.

How to get started

Search for interesting algorithms using keywords relating to facets such as agency (e.g., Dept. of Justice) or topic (e.g., health, police, etc.). Then, on our resources page, learn how to submit FOIA requests about algorithms, or find news articles and research papers about the uses and risks of algorithms.

Want to help?

There are several ways to get involved. You can submit your own tip about a government algorithm. Or, volunteer to screen the tips we receive from our users. International collaborators welcome: we want to make this a global resource. Contact us to learn how you can collaborate to help us expand the tip list.

Search terms used to find “algorithms” on .gov sites

Algorithm	Automatic ranking	Grading methodology	Rating method
Algorithmic	Automatic rating	Grading model	Rating methodology
Automated analysis	Automatic score	Numerical rating	Rating model
Automated assessment	Automatic scoring	Predictive Analytics	Scoring calculation
Automated calculation	Automatic sorting	Predictive modeling	Scoring equation
Automated filtering	Calculating matrix	Ranking calculation	Scoring formula
Automated grading	Calculating method	Ranking equation	Scoring matrix

Automated ranking	Calculating model	Ranking formula	Scoring method
Automated rating	Computation	Ranking matrix	Scoring model
Automated scoring	Computational	Ranking method	Statistical assessment
Automated simulation	Computing	Ranking methodology	Statistical methodology
Automated sorting	Grading calculation	Ranking model	Statistical model
Automatic assessment	Grading equation	Rating calculation	Statistical software
Automatic calculation	Grading formula	Rating equation	
Automatic filtering	Grading matrix	Rating formula	
Automatic grading	Grading method	Rating matrix	

Algorithm Tips: A Resource for Algorithmic Accountability in Government
Trielli, Stark, Diakopolous

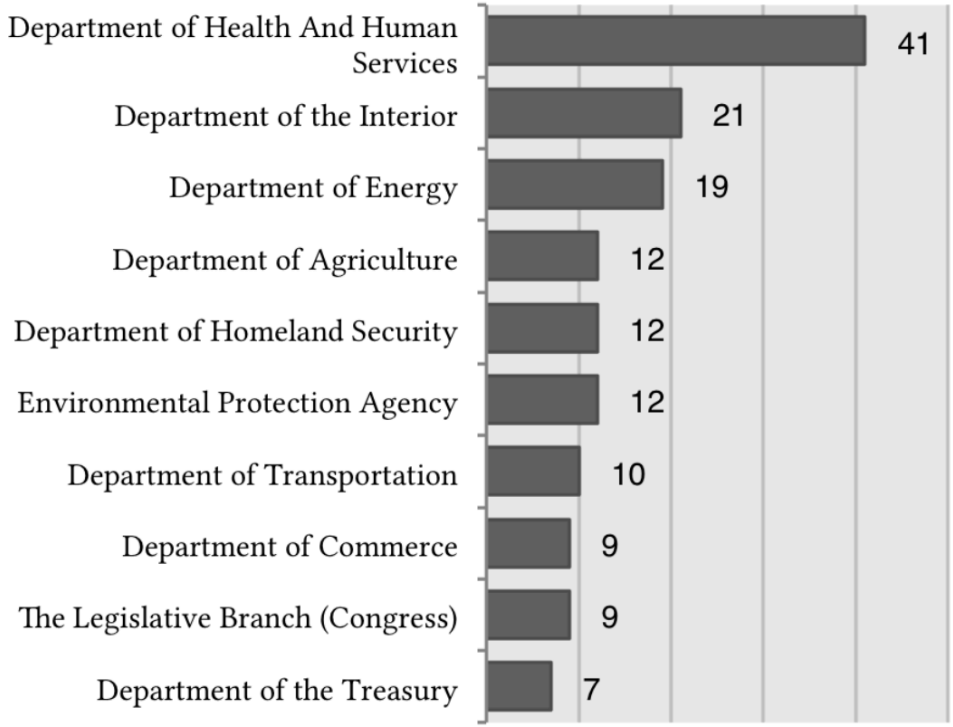
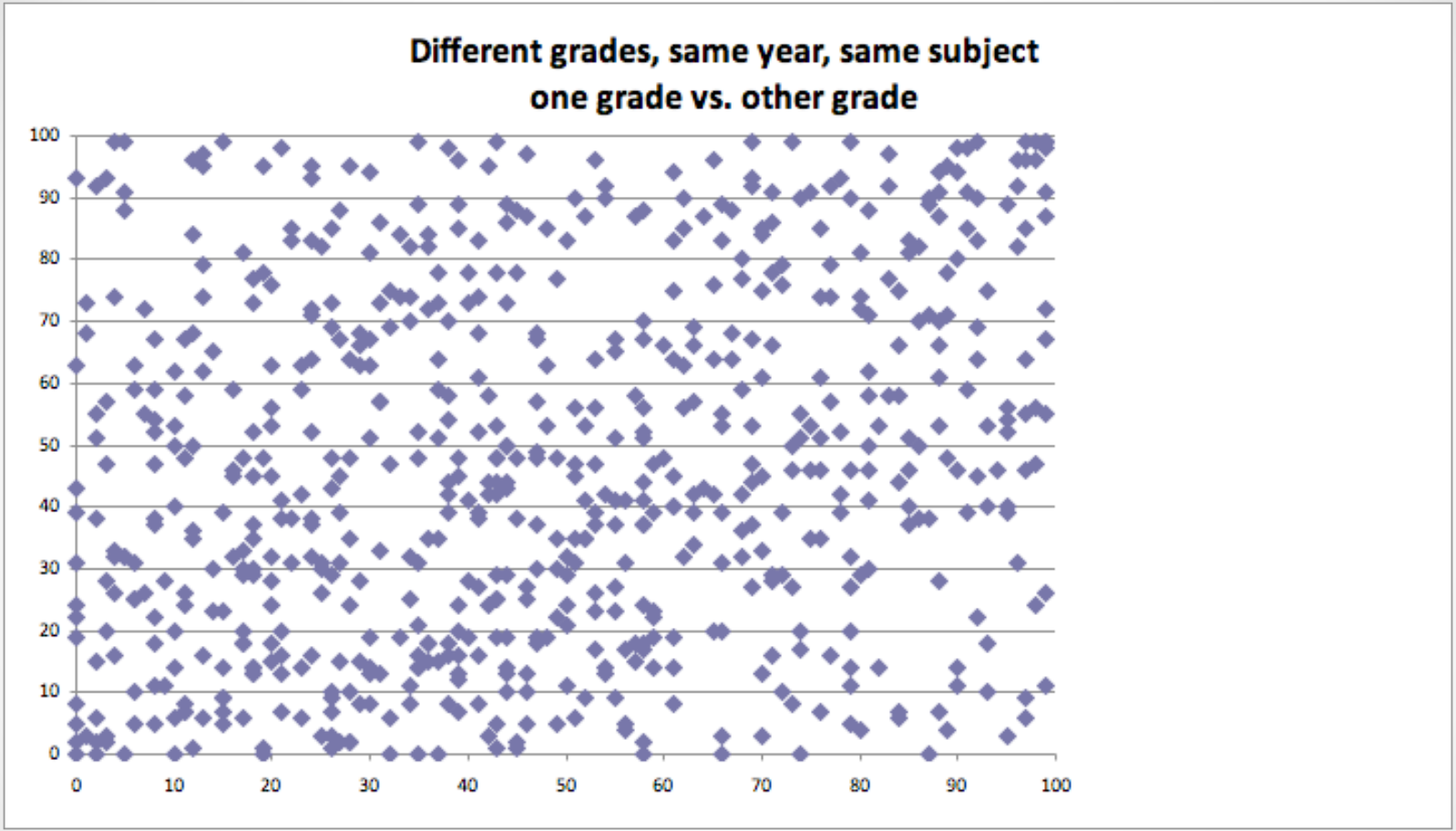


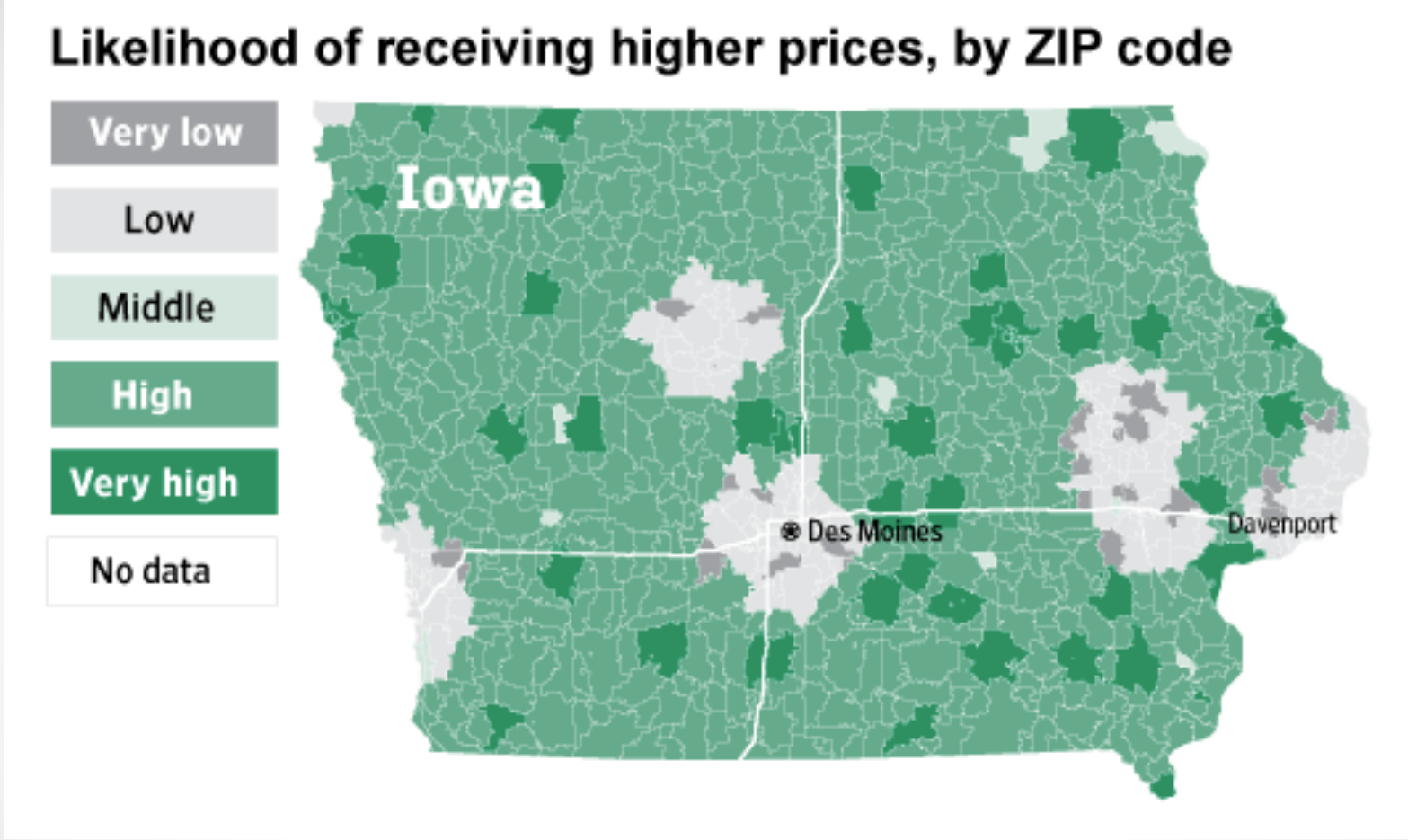
Figure 1: **Government agencies with the highest proportion of algorithms returned by the search.**

Some previous algorithmic accountability work

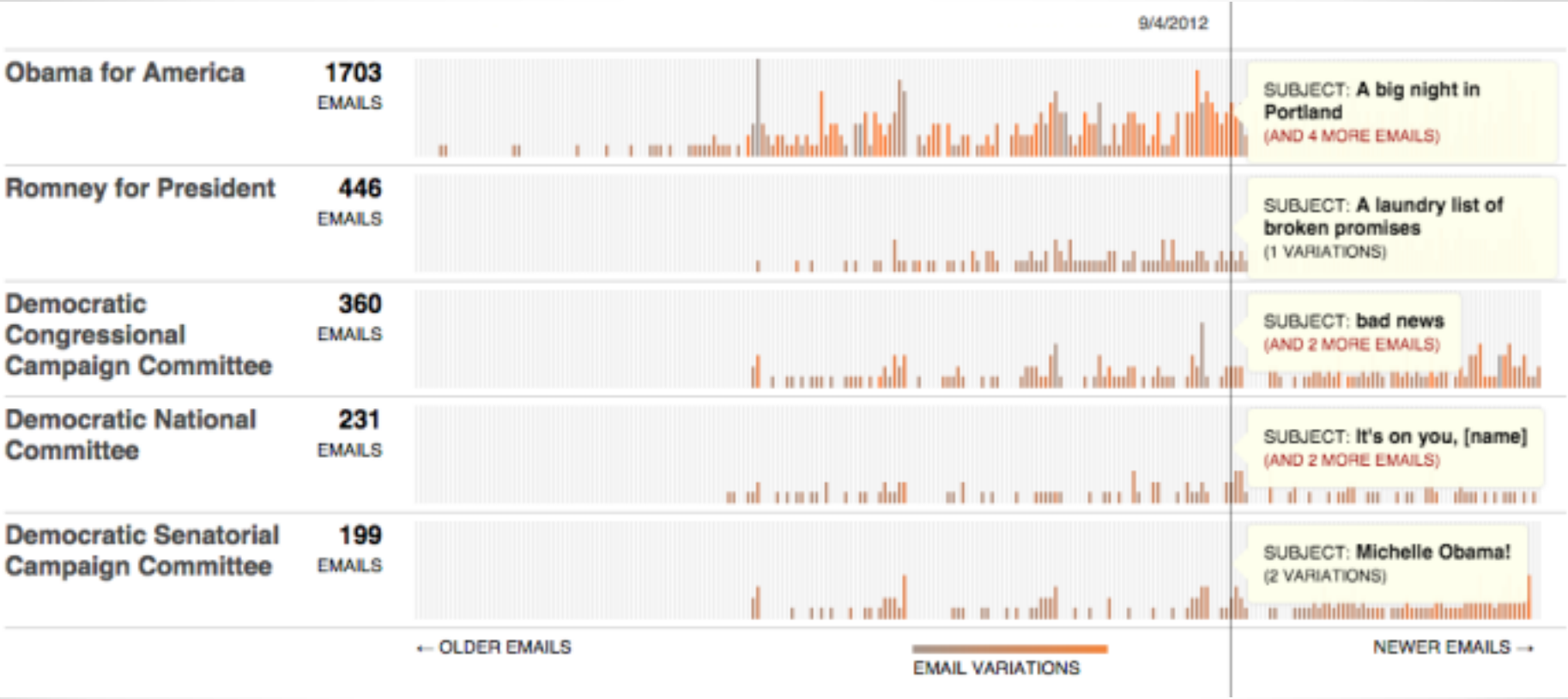
. . .



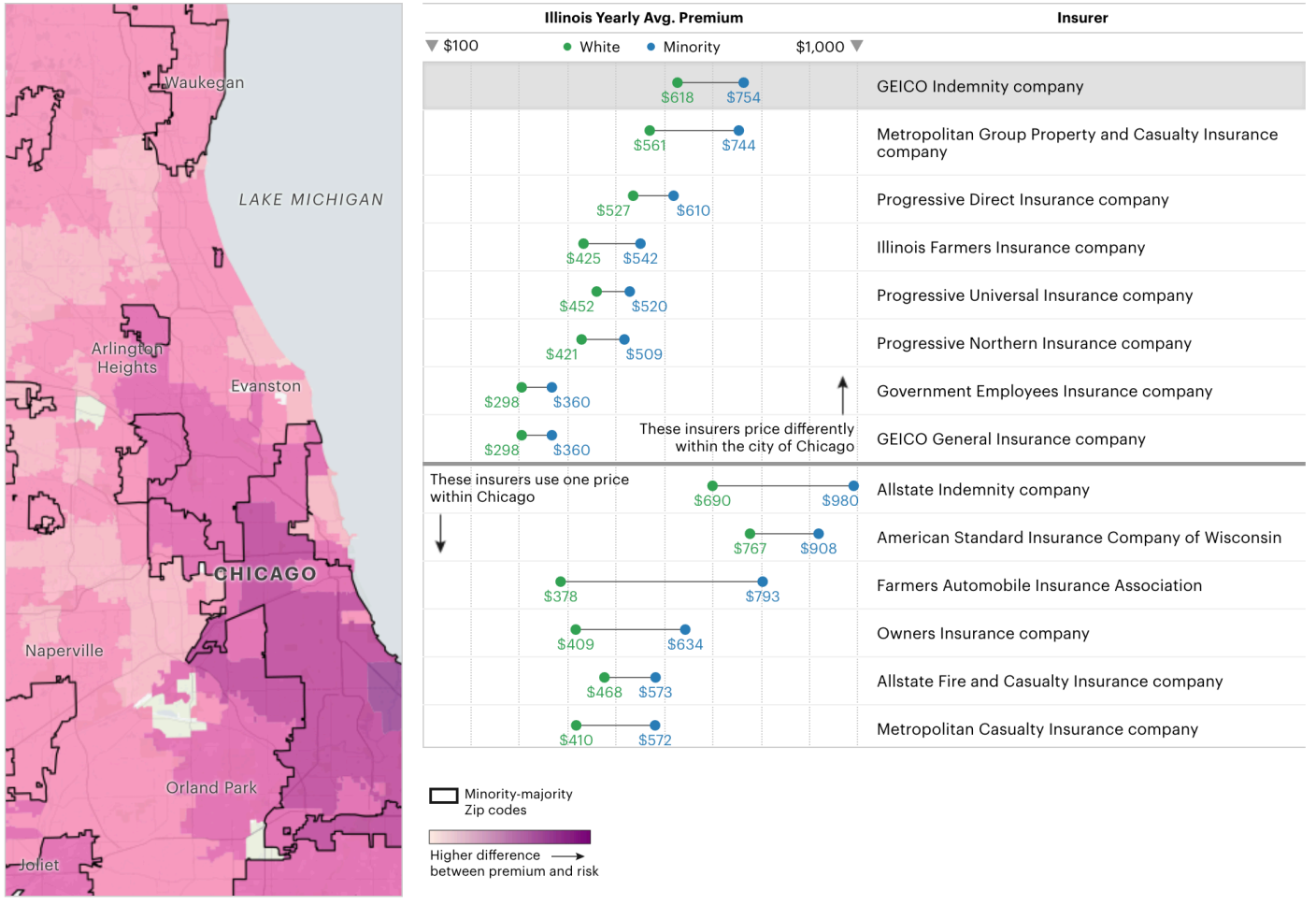
Analyzing Released NYC Value-Added Data Part 2, Gary Rubenstein



Websites Vary Prices, Deals Based on Users' Information
Valentino-Devries, Singer-Vine and Soltani, WSJ, 2012

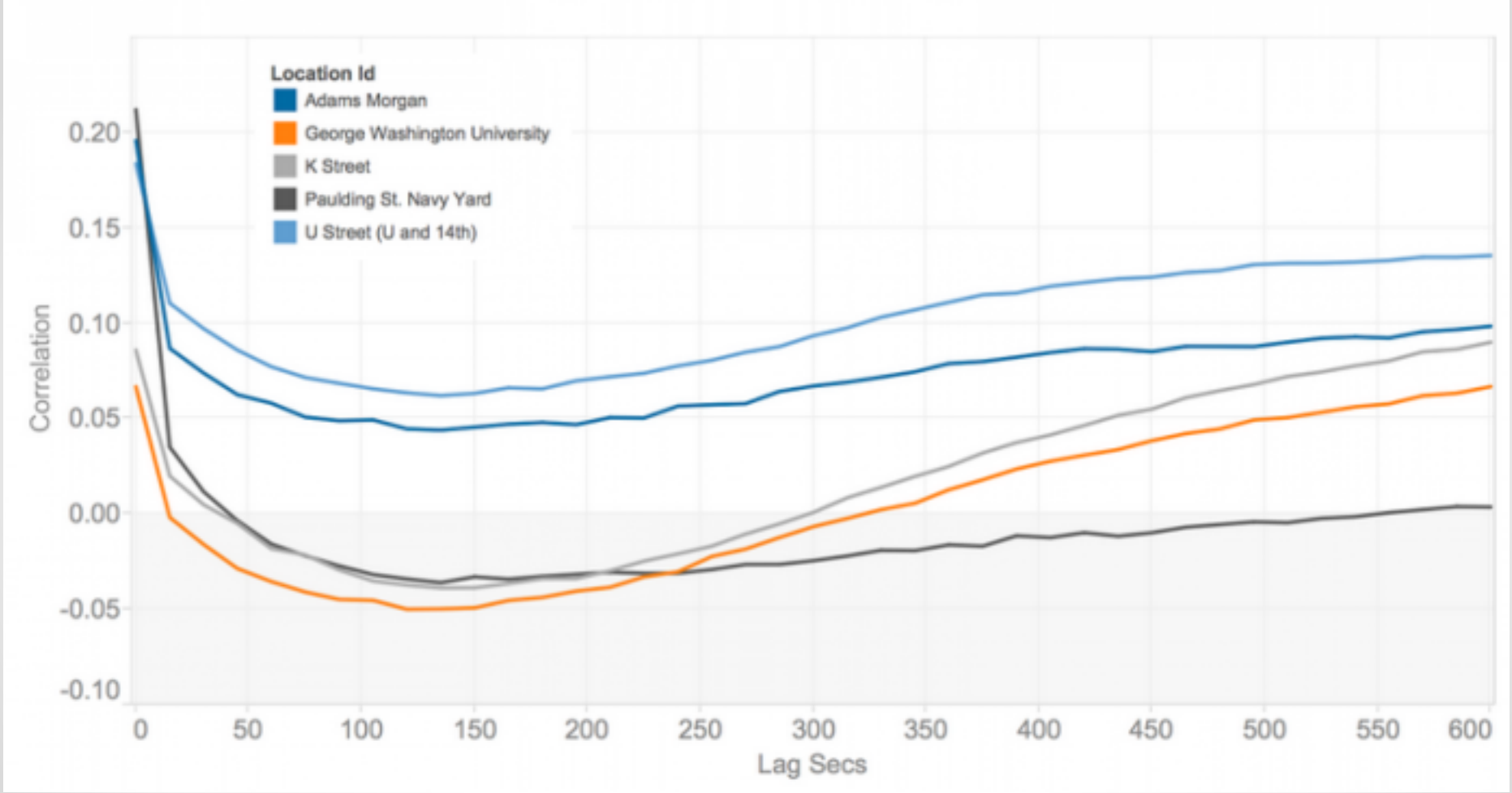


Message Machine
Jeff Larson, Al Shaw, ProPublica, 2012



Chicago Area Disparities in Car Insurance Premiums
Al Shaw, Jeff Larson, Julia Angwin, ProPublica, 2017

Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk, Angwin, Larson, Kirchner, Mattu, ProPublica, 2017



- *How Uber surge pricing really works,* Nick Diakopolous

Measuring Race and Gender Bias

. . .

Title VII of Civil Rights Act, 1964

- It shall be an unlawful employment practice for an employer -
- (1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or
- (2) to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.

Regulated Domains

Credit (Equal Credit Opportunity Act)
Education (Civil Rights Act of 1964; Education Amendments of 1972)
Employment (Civil Rights Act of 1964)
Housing (Fair Housing Act)
'Public Accommodation' (Civil Rights Act of 1964)

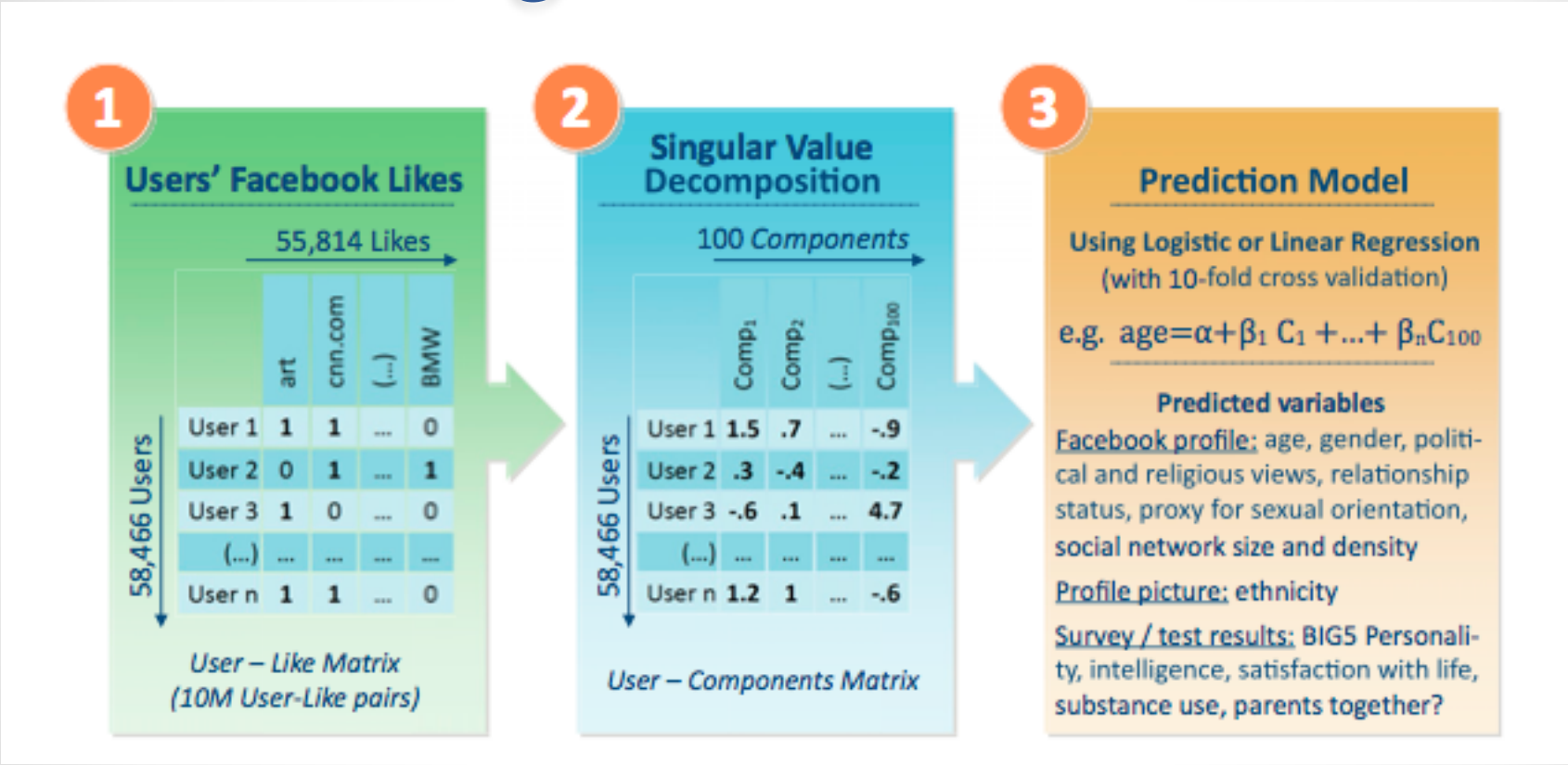
Protected Classes

Race (Civil Rights Act of 1964); **Color** (Civil Rights Act of 1964); **Sex** (Equal Pay Act of 1963; Civil Rights Act of 1964); **Religion** (Civil Rights Act of 1964); **National origin** (Civil Rights Act of 1964); **Citizenship** (Immigration Reform and Control Act); **Age** (Age Discrimination in Employment Act of 1967); **Pregnancy** (Pregnancy Discrimination Act); **Familial status** (Civil Rights Act of 1968); **Disability status** (Rehabilitation Act of 1973; Americans with Disabilities Act of 1990); **Veteran status** (Vietnam Era Veterans' Readjustment Assistance Act of 1974; Uniformed Services Employment and Reemployment Rights Act); **Genetic information** (Genetic Information Nondiscrimination Act)

Race and gender correlate with everything

. . .

Learning from Facebook likes



From Kosinski et. al., *Private traits and attributes are predictable from digital records of human behavior*

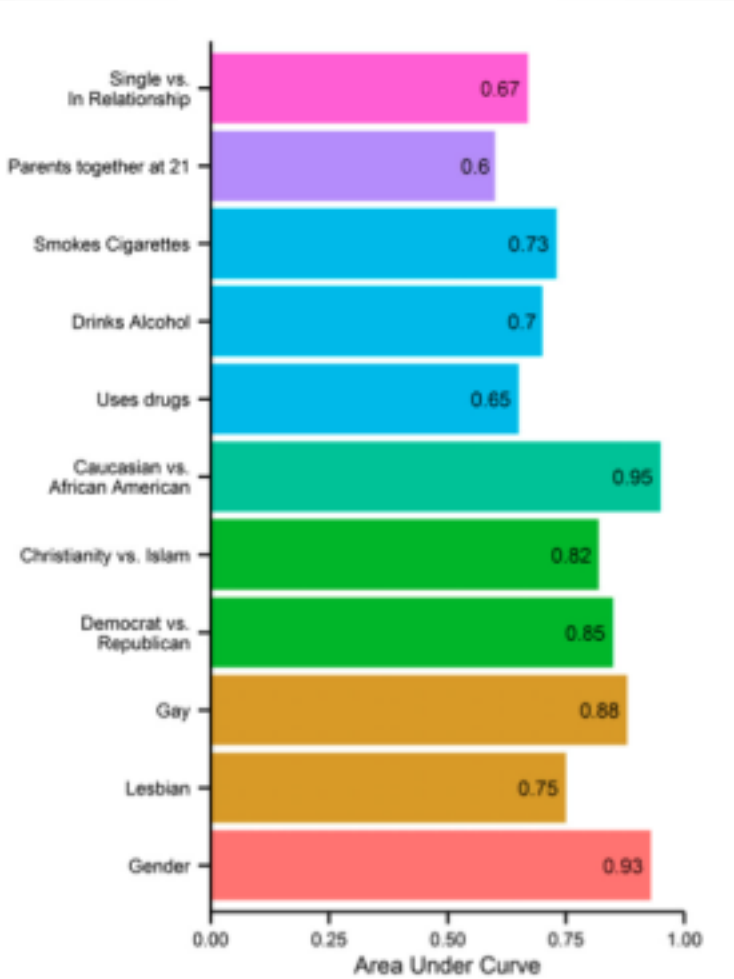


Fig. 2. Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.

Predicting gender from Twitter

Configuration	Age	Gender	Political
UserOnly	0.751	0.795	0.890
Nbr-All	0.736	0.669	0.920
Nbr-Most	0.619	0.688	0.777
Nbr-Least	0.691	0.560	0.725
Nbr-Closest	0.716	0.598	0.895
Avg-All	0.795	0.750	0.918
Avg-Most	0.739	0.749	0.885
Avg-Least	0.805	0.758	0.878
Avg-Closest	0.779	0.674	0.909
Join-All	0.764	0.799	0.932
Join-Most	0.741	0.755	0.889
Join-Least	0.782	0.774	0.873
Join-Closest	0.772	0.802	0.915

Table 1: The overall accuracy of the SVM-based classifiers on datasets constructed using different combinations of user and neighborhood data. The top row, *UserOnly*, corresponds to results obtained from feature vectors that contained only data from the user's microblog content. All other rows involve configurations that incorporated neighborhood data.

- Zamal et. al., *Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors*

Predicting race from Twitter

System	PREC	REC	F-MEAS
democrats-B1	**0.989**	0.183	0.308
democrats-B2	0.735	0.896	0.808
democrats-FULL	0.894‡	**0.936**♭	**0.915**♭
republicans-B1	**0.920**	0.114	0.203
republicans-B2	0.702	0.430	0.533
republicans-FULL	0.878‡	**0.805**♭	**0.840**♭
ethnicity-B1	**0.878**	0.421	0.569
ethnicity-B2	0.579	0.633	0.604
ethnicity-FULL	0.646‡	**0.665**♭	**0.655**♭
starbucks-B1	**0.817**	0.019	0.038
starbucks-B2	0.747	0.723	0.735
starbucks-FULL	0.762	**0.756**♭	**0.759**♭

Table 1: Overall classification results. †, ‡ and ♭ respectively indicate statistical significance at the 0.95 level with respect to B1 alone, B2 alone, and both B1 and B2.

Observational vs. Experimental

. . .

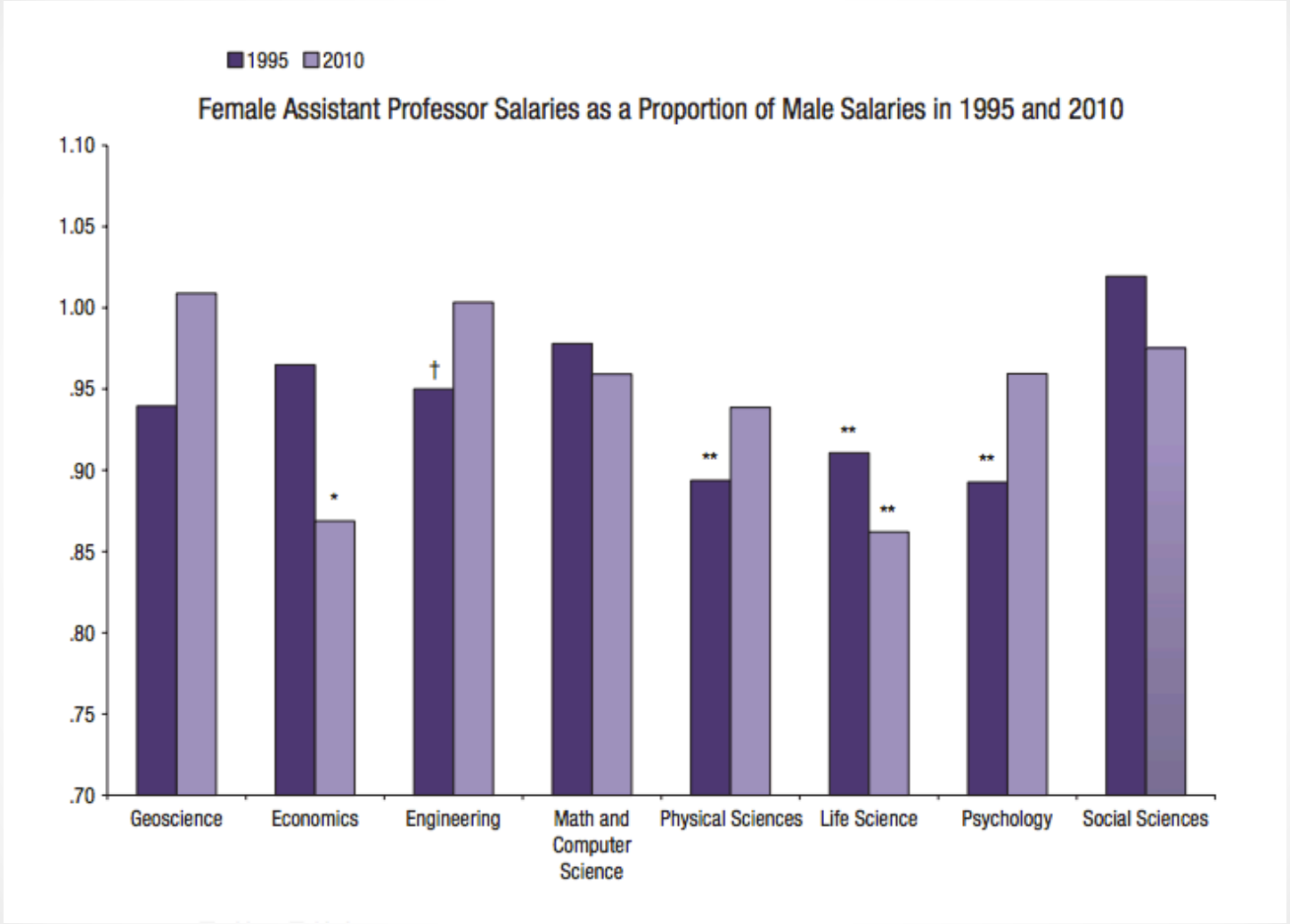

1995
2010
Female Assistant Professor Salaries as a Proportion of Male Salaries in 1995 and 2010
1.10
1.05
1.00
.95
.90
.85
.80
.75
.70
†
*
**
**
**
**
Geoscience
Economics
Engineering
Math and Computer Science
Physical Sciences
Life Science
Psychology
Social Sciences

Table 1. Percentage of Female Applicants for Tenure-Track Positions Invited to Interview and Offered Positions at 89 U.S. Research Universities

Field	Mean percentage of female applicants	Mean percentage of women invited to interview	Mean percentage of women offered position
Physics	12%	19%	20%
Biology	26%	28%	34%
Chemistry	18%	25%	29%
Civil engineering	16%	30%	32%
Electrical engineering	11%	19%	32%
Mathematics	20%	28%	32%

Note: Data shown here were drawn from Sections 3-10 and 3-13 of "Gender Differences at Critical Transitions in the Careers of Science, Engineering and Mathematics Faculty" (National Research Council, 2010).

Women in Academic science: a Changing Landscape

Ceci, et. al

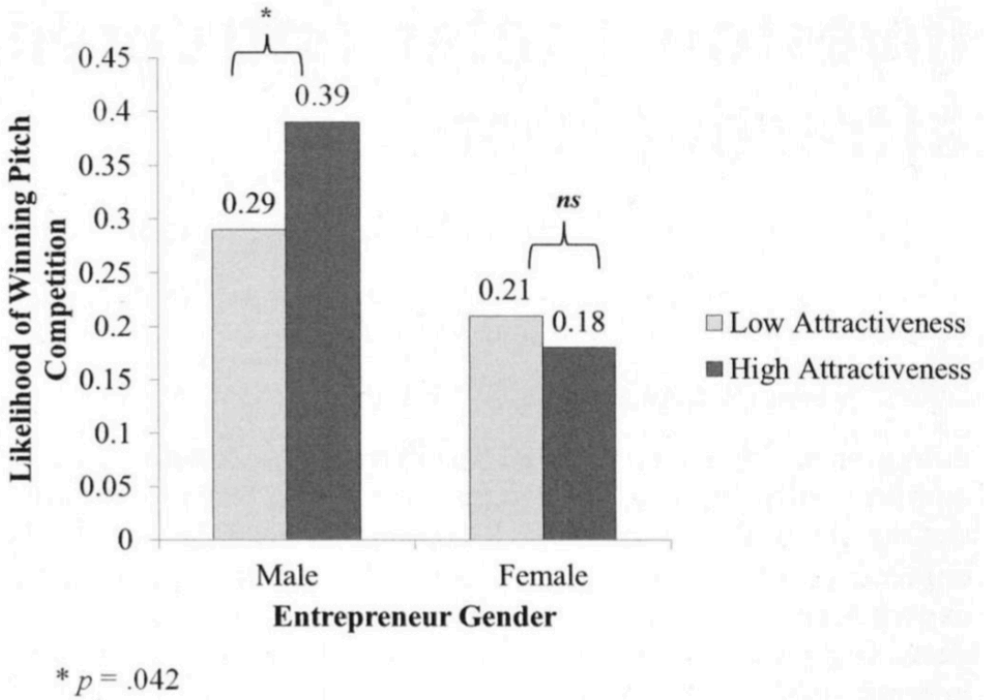


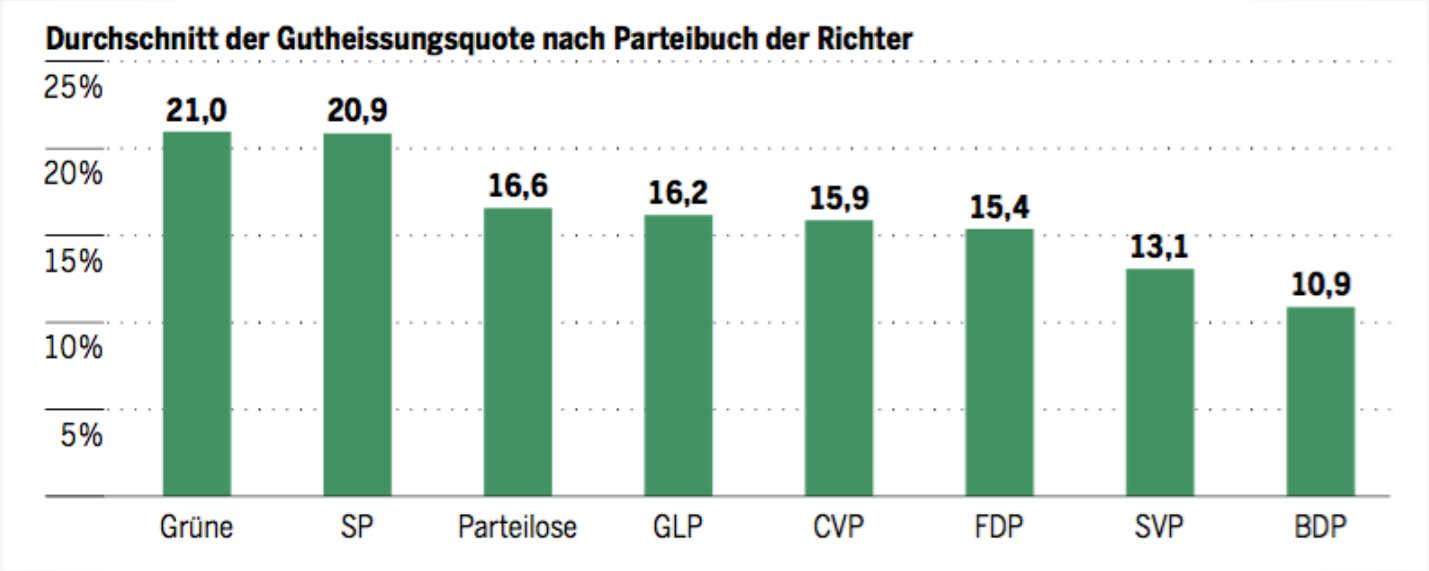
Fig. 1. The effect of entrepreneur gender and physical attractiveness on pitch success rate in a field setting (n = 90). ns, not significant.

Investors prefer entrepreneurial ventures pitched by attractive men, Brooks et. al. 2014

Measuring Bias in criminal justice

. . .

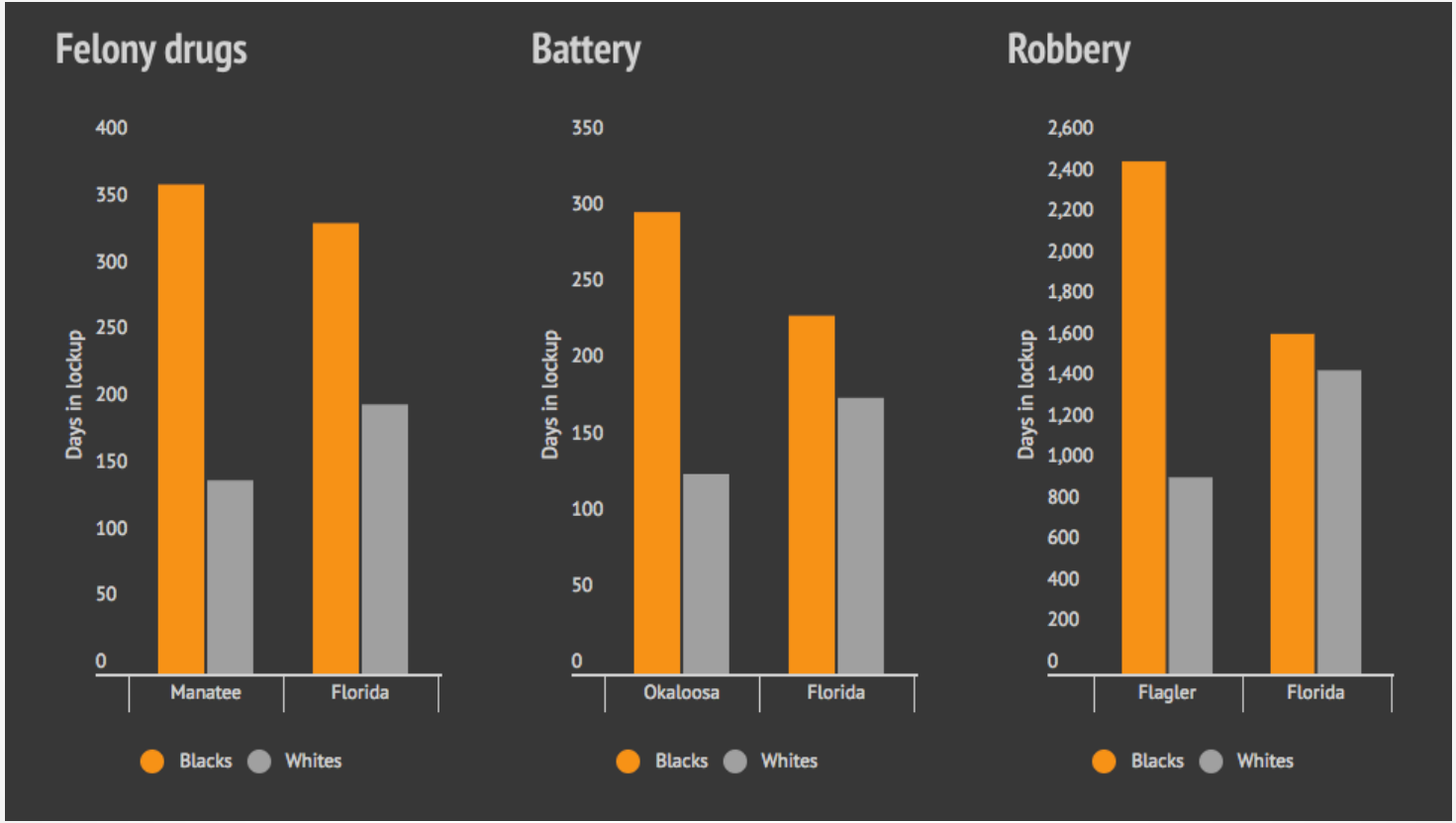
Swiss judges: a natural experiment



24 Judges of Swiss Federal Administrative court are randomly assigned to cases. They rule at different rates on migrant deportation cases. Here are their deportation rates broken down by party.

Barnaby Skinner and Simone Rau, *Tages-Anzeiger*.
https://github.com/barjacks/swiss-asylum-judges

Florida sentencing analysis adjusted for "points"



Bias on the Bench, Michael Braga, Herald Tribune

Containing 1.4 million entries, the DOC database notes the exact number of points assigned to defendants convicted of felonies. The points are based on the nature and severity of the crime committed, as well as other factors such as past criminal history, use of a weapon and whether anyone got hurt. The more points a defendant gets, the longer the minimum sentence required by law.

Florida legislators created the point system to ensure defendants committing the same crime are treated equally by judges. But that is not what happens.

…

The Herald-Tribune established this by grouping defendants who committed the same crimes according to the points they scored at sentencing. Anyone who scored from 30 to 30.9 would go into one group, while anyone who scored from 31 to 31.9 would go in another, and so on.

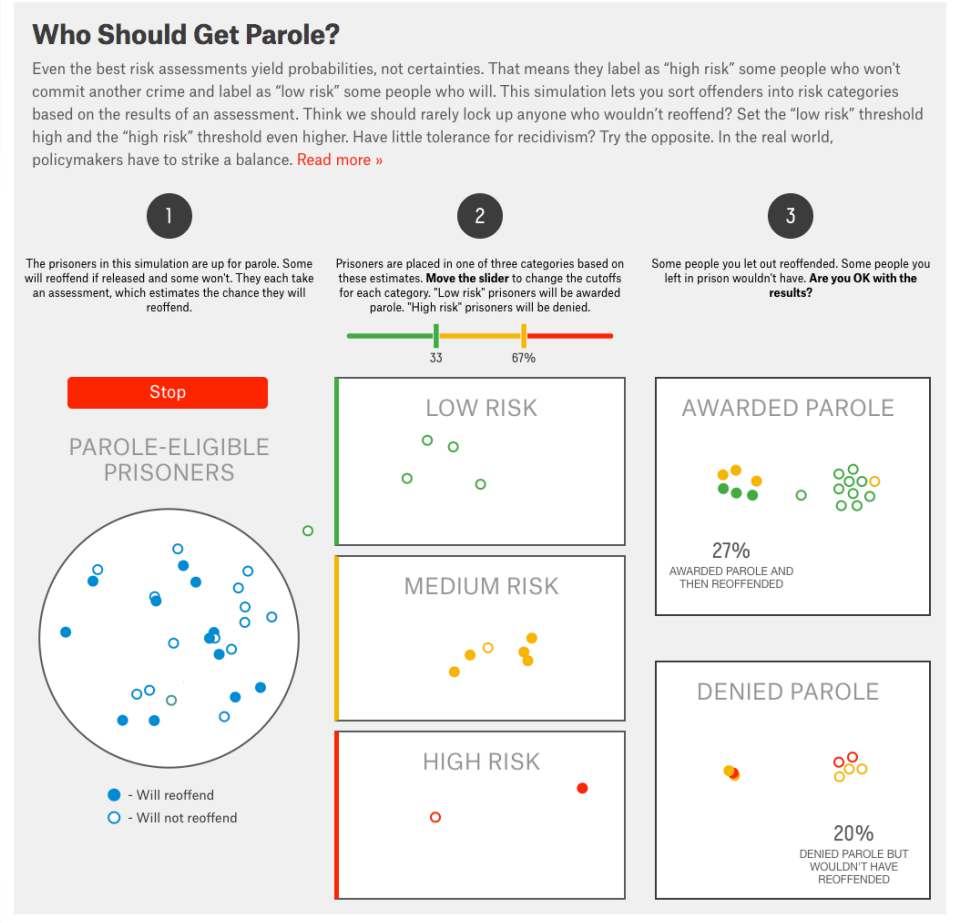
We then evaluated how judges sentenced black and white defendants within each point range, assigning a weighted average based on the sentencing gap.

If a judge wound up with a weighted average of 45 percent, it meant that judge sentenced black defendants to 45 percent more time behind bars than white defendants.

Bias on the Bench: How We Did It, Michael Braga, Herald Tribune

Unpacking ProPublica's "Machine Bias"

. . .



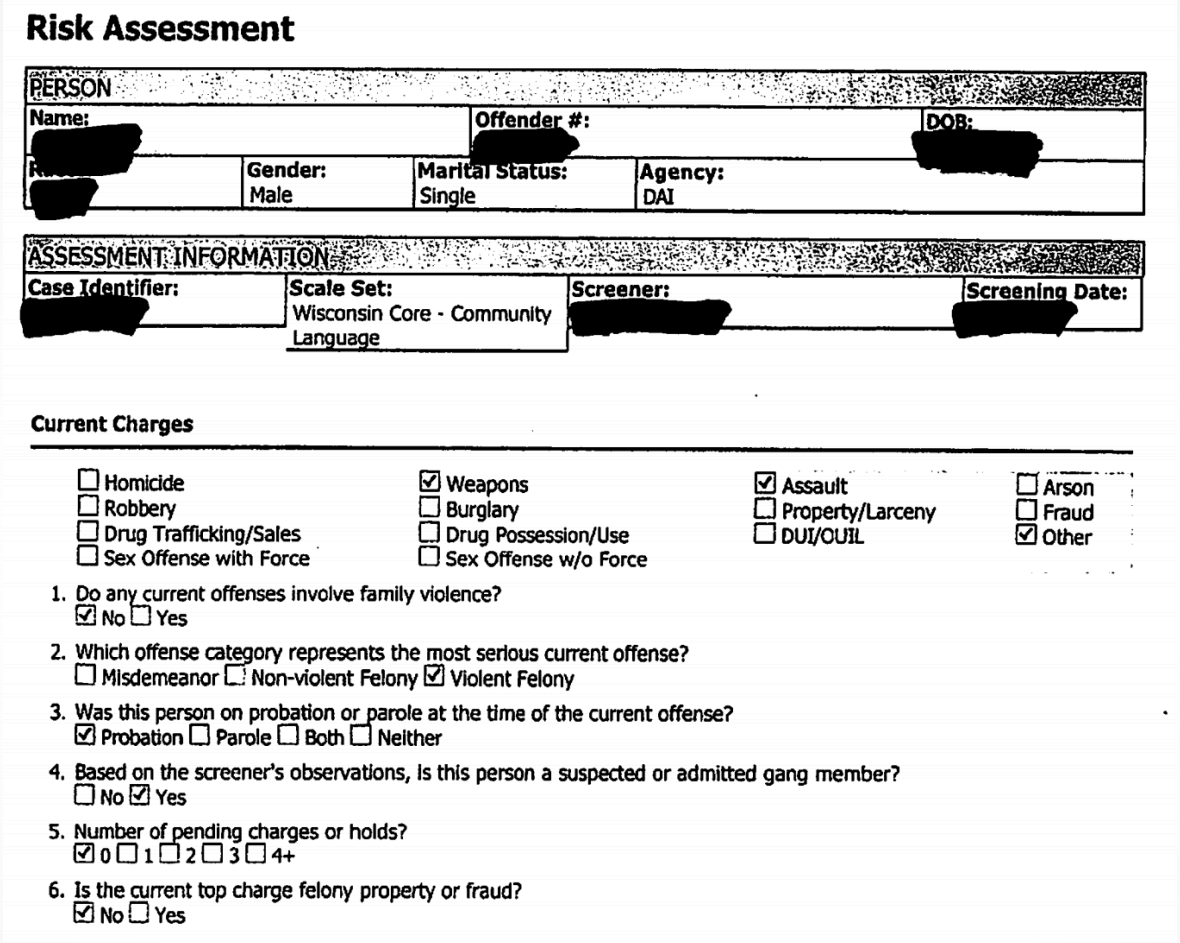
Should Prison Sentences Be Based On Crimes That Haven't Been Committed Yet?, FiveThirtyEight

All Defendants	Low	High
Survived	2681	1282
Recidivated	1216	2035
FP rate: 32.35		
FN rate: 37.40		
PPV: 0.61		
NPV: 0.69		
LR+: 1.94		
LR-: 0.55		

Black Defendants	Low	High
Survived	990	805
Recidivated	532	1369
FP rate: 44.85		
FN rate: 27.99		
PPV: 0.63		
NPV: 0.65		
LR+: 1.61		
LR-: 0.51		

White Defendants	Low	High
Survived	1139	349
Recidivated	461	505
FP rate: 23.45		
FN rate: 47.72		
PPV: 0.59		
NPV: 0.71		
LR+: 2.23		
LR-: 0.62		

How We Analyzed the COMPAS Recidivism Algorithm, ProPublica



Risk Assessment

PERSON

Name:	Offender #:	DOB:
[redacted]	[redacted]	[redacted]

[redacted]	Gender: Male	Marital Status: Single	Agency: DAI

ASSESSMENT INFORMATION

Case Identifier:	Scale Set:	Screener:	Screening Date:
[redacted]	Wisconsin Core - Community Language	[redacted]	[redacted]

Current Charges

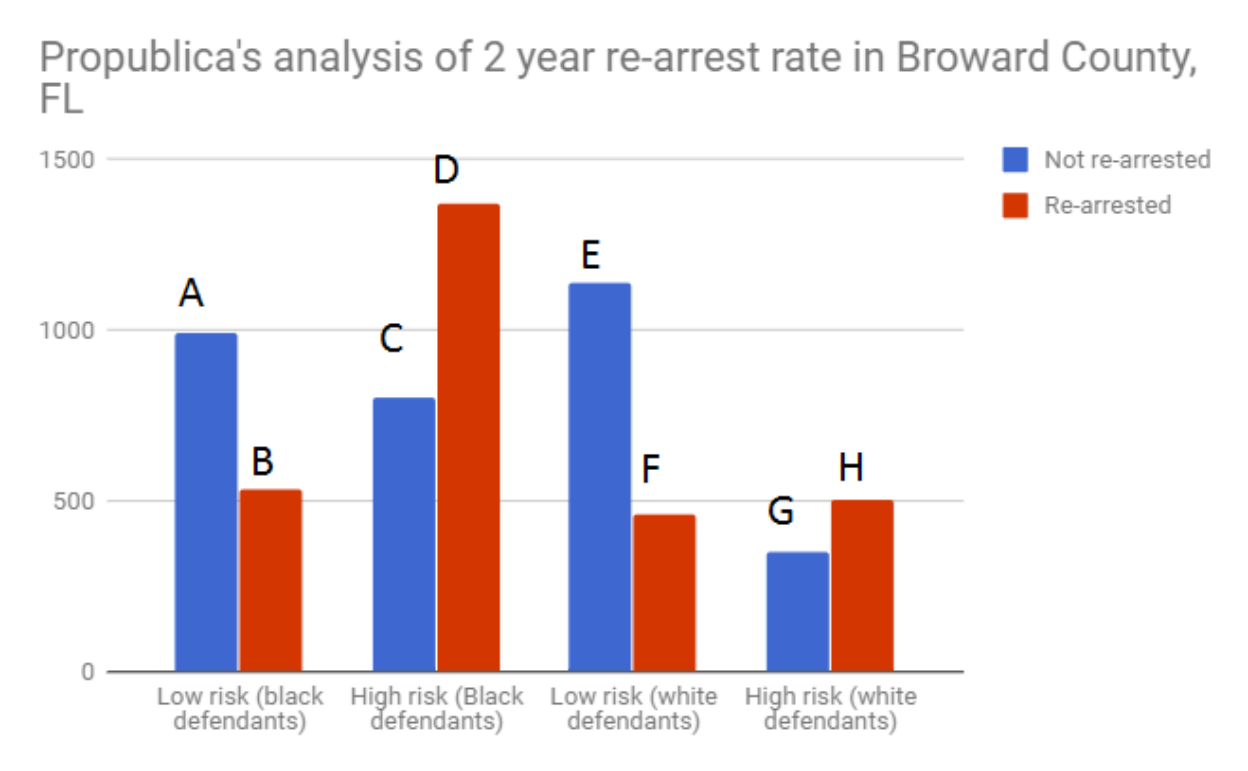
☐ Homicide	☑ Weapons	☑ Assault	☐ Arson
☐ Robbery	☐ Burglary	☐ Property/Larceny	☐ Fraud
☐ Drug Trafficking/Sales	☐ Drug Possession/Use	☐ DUI/OUIL	☑ Other
☐ Sex Offense with Force	☐ Sex Offense w/o Force		

1. Do any current offenses involve family violence?
   ☑ No ☐ Yes
2. Which offense category represents the most serious current offense?
   ☐ Misdemeanor ☐ Non-violent Felony ☑ Violent Felony
3. Was this person on probation or parole at the time of the current offense?
   ☑ Probation ☐ Parole ☐ Both ☐ Neither
4. Based on the screener's observations, is this person a suspected or admitted gang member?
   ☐ No ☑ Yes
5. Number of pending charges or holds?
   ☑ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4+
6. Is the current top charge felony property or fraud?
   ☑ No ☐ Yes

COMPAS "CORE" questionnaire, 2011.

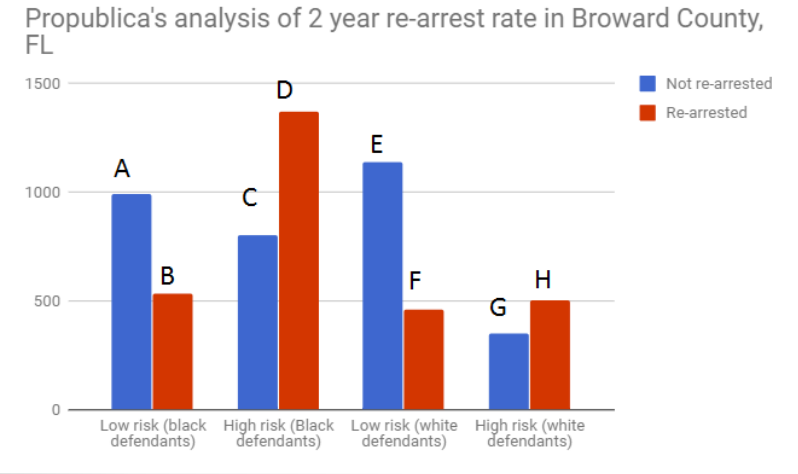
- *Includes criminal history, family history, gang involvement, drug use…*

	Low risk (black d	High risk (Black (	Low risk (white d	High risk (white defendants)
Not re-arrested	990	805	1139	349
Re-arrested	532	1369	461	505



Stephanie Wykstra, personal communication

	Low risk (black d	High risk (Black (	Low risk (white d	High risk (white defendants)
Not re-arrested	990	805	1139	349
Re-arrested	532	1369	461	505



ProPublica argument

False positive rate

P(high risk | black, no arrest) = C/(C+A) = 0.45

P(high risk | white, no arrest) = G/(G+E) = 0.23

False negative rate

P(low risk | black, arrested) = B/(B+D) = 0.28

P(low risk | white, arrested) = F/(F+H) = 0.48

Northpointe response

Positive predictive value

P(arrest | black, high risk) = D/(C+D) = 0.63

P(arrest | white, high risk) = H/(G+H) = 0.59