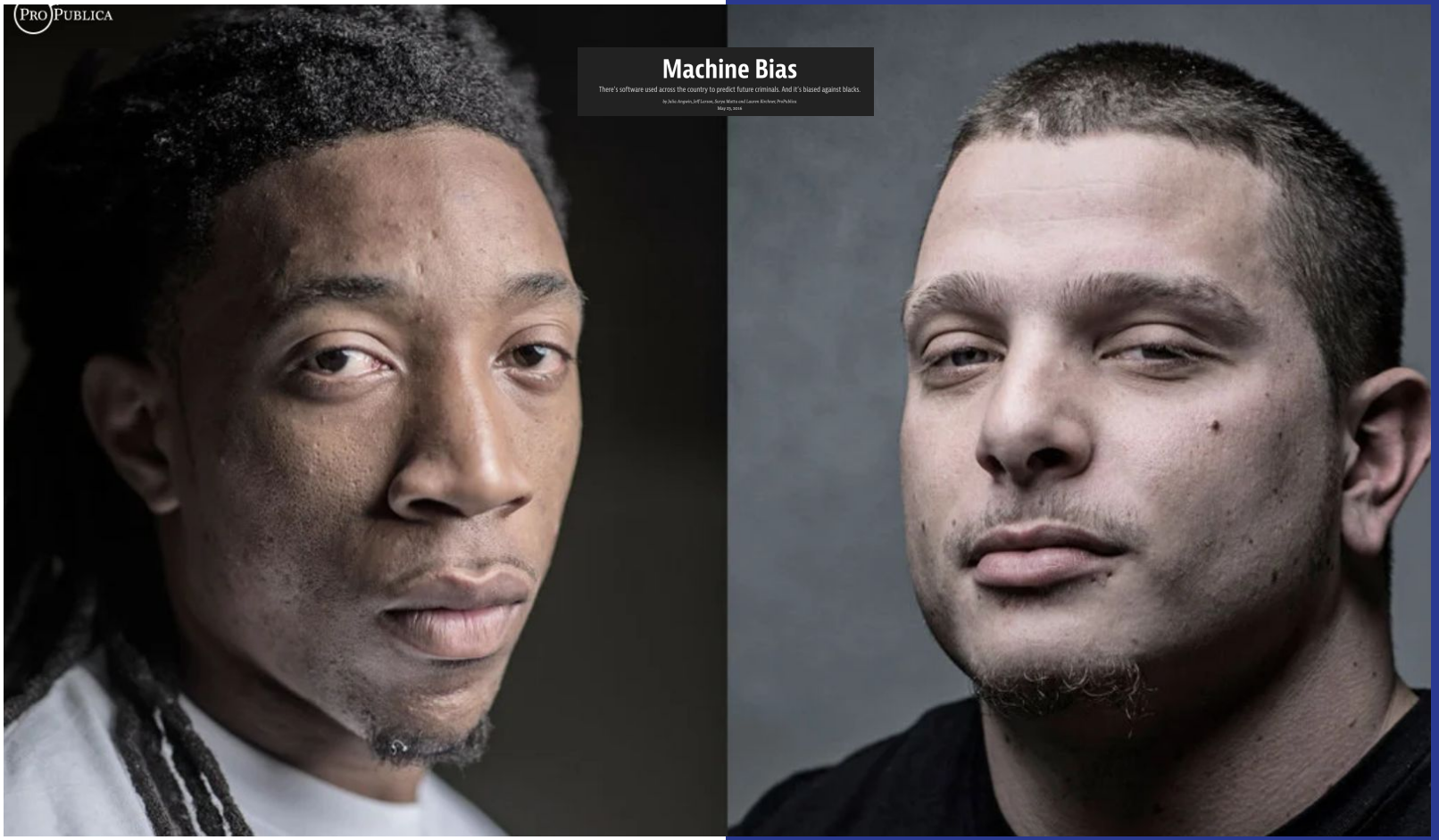
Machine Bias

There’s software used across the country to predict future criminals. And it’s biased against blacks.

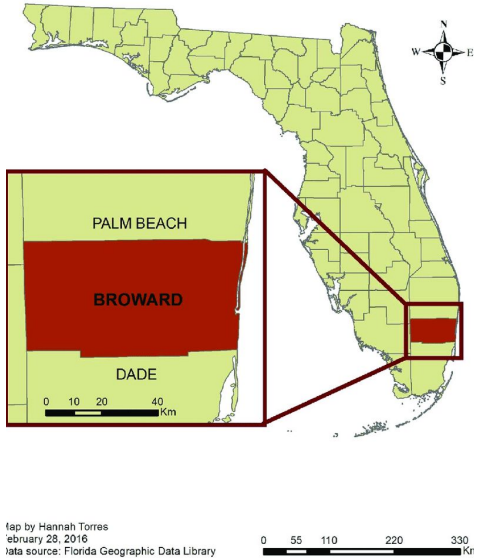
Propublica Article: Source

Presentation by Group 2: Isabel, Sherry, Lee



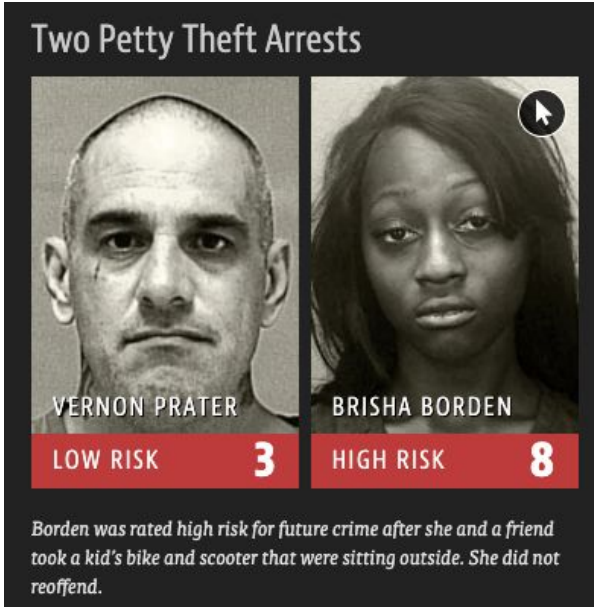
ProPublica Article: Source

BROWARD COUNTY, FL
POPULATION: 1,952,778



Broward County Stats: Here.

MACHINE BIAS



Methodology: Here and Here.

RISK ASSESSMENT. RECIDIVISM. REHABILITATION.

SIMULATING RISK

The Marshall Project: Source

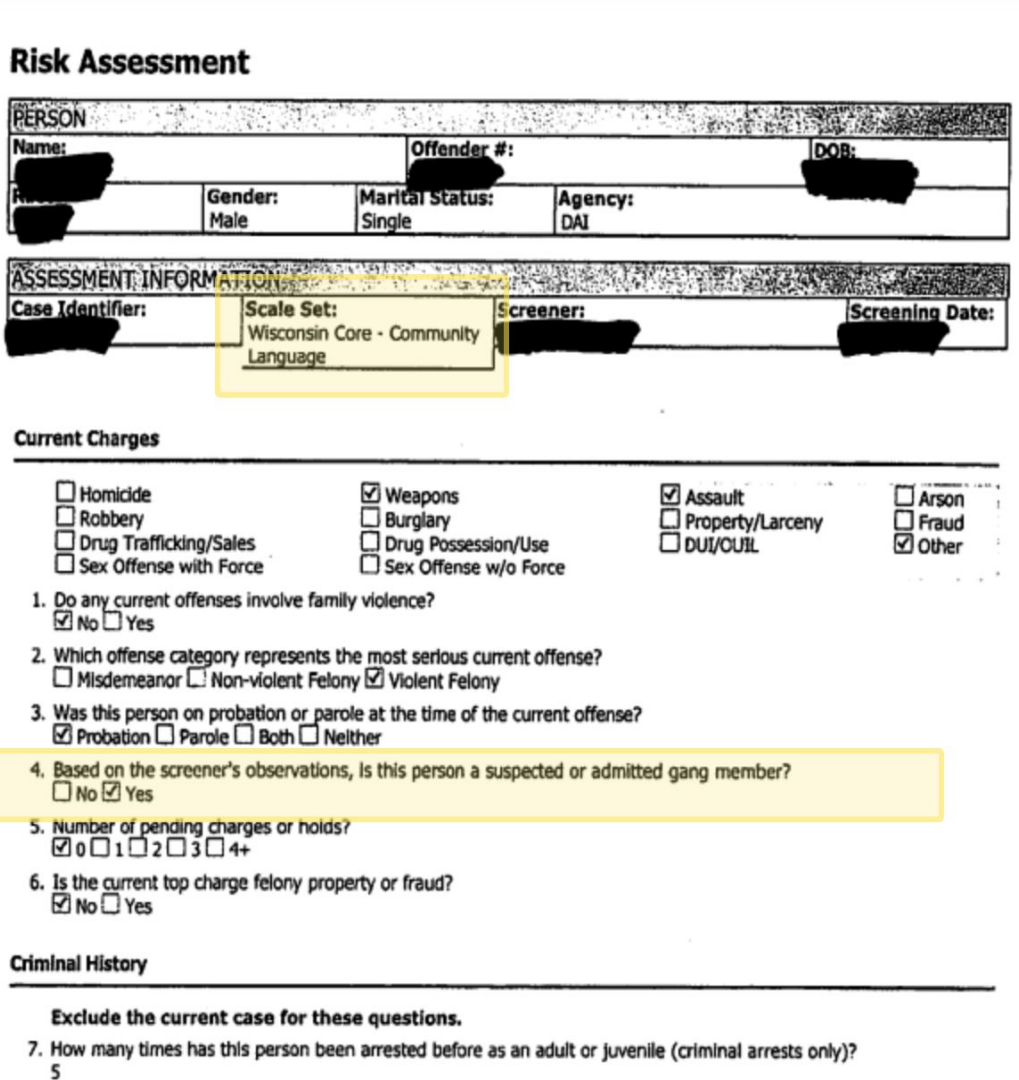
COMPAS - RISK ASSESSMENT

CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS

137 Questions

Gang Membership
Parental Separation
Friends Arrested
Residential Stability
Neighborhood Crime
School Suspensions
Money
Boredom
Anger
Criminal Thinking

Northpointe Sample: Source



Risk Assessment

PERSON

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

Gender:	Marital Status:	Agency:
Male	Single	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/CUIL	☑ 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

Criminal History

Exclude the current case for these questions.

7. How many times has this person been arrested before as an adult or juvenile (criminal arrests only)?
   5

p. 30

4.1.1 Pretrial Release Risk

The Pretrial Release Risk Scale was developed through a pretrial release outcomes study conducted in a large sample of felony defendants assessed with COMPAS in Kent County, Michigan Pretrial Services (Dieterich, 2010). The Pretrial Release Risk Scale was constructed to predict failure to appear (FTA) and new felony arrest among defendants on pretrial release.

Prior pretrial risk assessment research has consistently identified a set of factors that are predictive of pretrial failure. The most common risk factors include current charges, pending charges, prior arrest history, previous pretrial failure, residential stability, employment status, community ties, and substance abuse (VanNostrand, 2003). We selected items from the COMPAS assessment and included them as candidates for risk model development on the basis of this prior research.

One purpose of pretrial release risk assessment is to sort a pretrial caseload into low-, moderate-, and high-risk groups based on the likelihood of failure to appear in court or commit a new crime pending trial. Use of the risk assessment tool by pretrial services agencies should result in consistent and equitable decisions regarding release and conditions of release. The use of objective risk assessment tools is recommended by the National Association of Pretrial Services Agencies (2004). The risk assessment tool should be empirically derived and have demonstrated predictive validity in the jurisdiction in which it is deployed. The factors that enter into the risk assessment score should be consistent with applicable state statutes.[1] These and other guiding principles for pretrial risk assessment are outlined in Pretrial Services Legal and Evidence-based Practices (VanNostrand, 2007).

4.1.2 General Recidivism

How recidivism is defined ◀ PREVIOUS NEXT ▶

assessment date. The outcome used for the original scale construction was a new misdemeanor or felony offense within two years of the COMPAS administration date.

For most of our analysis of COMPAS risk scores, we defined recidivism as a new arrest within two years. We based this decision on Northpointe's practitioners guide, which says that its recidivism score is meant to predict "a new misdemeanor or felony offense within two years of the COMPAS administration date."

scores, as a low score would suggest there is little risk of general recidivism. It is important to note that the risk scores are generally taken from static information and that current level of needs, e.g. substance abuse or other issues can very much influence a person's likelihood of acting out or recidivating. In a later discussion the concept of Low risk/High needs will be covered.

General recidivism refers to a broad range of potential acts, therefore, versatility is an element for consideration. The COMPAS Typologies document delineates the typologies that have been discovered through research at Northpointe. One trait that lends itself to recidivism is versatility.

[1]For example in New York a pretrial risk assessment instrument cannot be based on age, gender, or marital status (Division of Probation and Correctional Alternatives, 2007).

RECIDIVISM

“Only 20 percent of the people predicted to commit violent crimes actually went on to do so.”

Propublica Article: Source

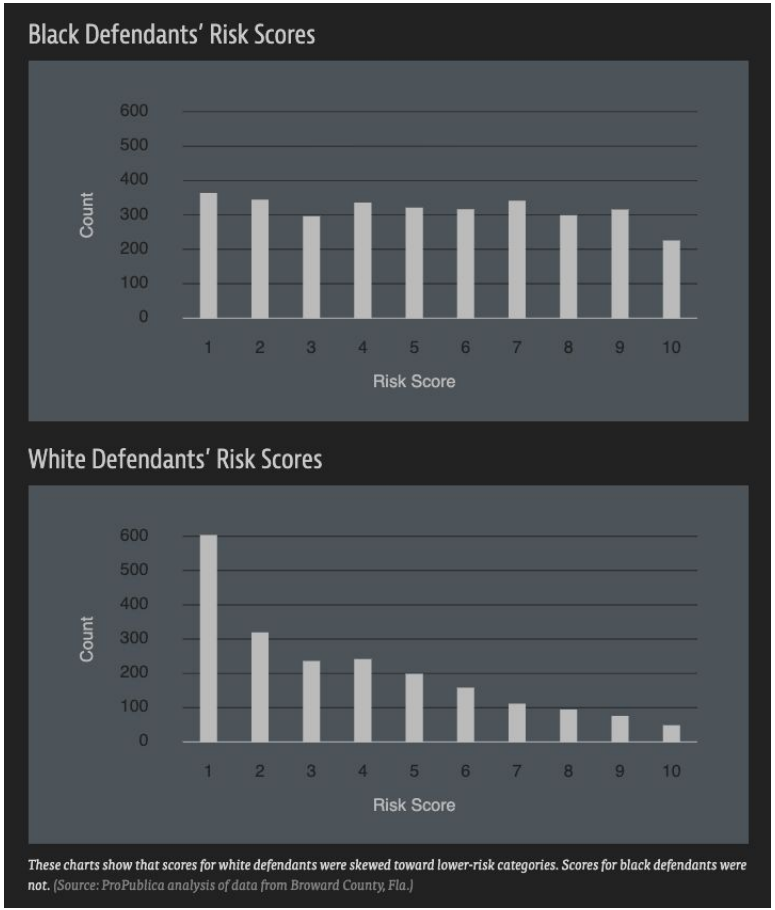
Wrongly labeling black defendants at almost twice the rate as white defendants.

White defendants were mislabeled as low risk more often than black defendants.

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.)



These charts show that scores for white defendants were skewed toward lower-risk categories. Scores for black defendants were not. (Source: ProPublica analysis of data from Broward County, Fla.)

Methods / Python Notebook: Source

SENTENCING & REHABILITATION

A BLACK BOX

"Risk assessments should be impermissible unless both parties get to see all the data that go into them. It should be an open, full-court adversarial proceeding."

- Christopher Slobogin, director of the criminal justice program at Vanderbilt Law School.

SYSTEMIC PATTERNS

CODED BIAS: A DOCUMENTARY

Datasets and models are biased by power structures of the population from which they are derived

nature

NEWS · 24 OCTOBER 2019 · UPDATE 26 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it.

Heidi Ledford

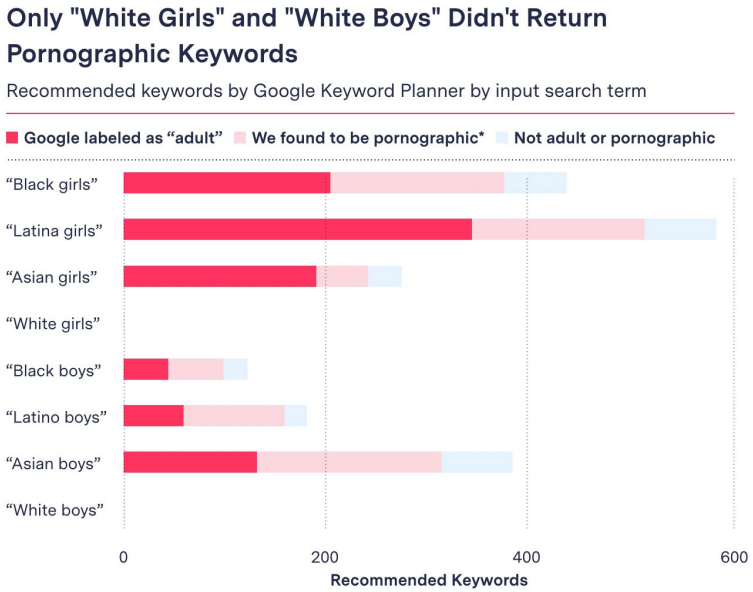
HEALTH

AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind

Machine learning has the potential to save thousands of people from skin cancer each year—while putting others at greater risk.

ANGELA LASHBROOK AUGUST 16, 2018

Source: https://www.nature.com/articles/d41586-019-03228-6 | https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/



Biased text datasets

One day GPT-2, an earlier publicly available version of the automated language generation model developed by the research organization OpenAI, started talking to me openly about "white rights." Given simple prompts like "a white man is" or "a Black woman is," the text the model generated would launch into discussions of "white Aryan nations" and "foreign and non-white invaders."

Biased image datasets

Using **symptoms** of a problem (*high-rise building for white-collar or run-down buildings for violent crimes*) **as predictors** creates **highly precise** but **fundamentally biased** tools.

Source: https://whitecollar.thenewinquiry.com/#dr5regh

IS THIS SOFTWARE STILL USED TODAY?

This article came out in 2016, is similar software still used today?

Yes. Tools like COMPAS and COMPAS itself are in widespread use across the criminal justice system.

Different Tools Used Today:

- Predictive Policing tools (PredPol, HunchLab) give estimates of where crime is most likely to occur
- Pattern Recognition tools (Patternizr) match similar crimes to crimes detectives are currently investigating
- Facial Recognition tools detect possible suspects from video footage
- Predictive Models detect high risk individuals

Source:
https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

How is COMPAS used today?

- COMPAS is still used and still criticized.
- Common concerns with COMPAS and Risk Assessment Instruments (RAIs) as a whole are their lack of individualization, absence of transparency under secret claims, possibility of bias, and questions of their true impact
- Loomis vs. Wisconsin, a 2016 Wisconsin Supreme Court case addressed these concerns particularly against COMPAS; results being that risk scores can be considered by judges during sentencing, but there must be warnings about the tool's limitations given with the scores
- COMPAS is still one of the most widely used RAIs today

Source: https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

QUESTIONS

- How do we as engaged citizens help to address these problems of algorithmic bias when the algorithms are hidden and privatized?
- What role should the creators of the algorithm play after they have sold the software?
- How do we ensure that governments and corporations are accounting for systemic anti-black bias in their models, especially as “bureaucracy by AI” becomes more popular?
- What is the right mix of machine and human decision-making when designing a predictive system?
- How does one “cross-examine” an algorithm?
- How is similar software used in other spheres (not criminal justice)? What biases/general problems are present?