Stanford University

CS 329P : Practical Machine Learning (2021 Fall)

Lecture 14 - Fairness

Qingqing Huang, Mu Li, Alex Smola

https://c.d2l.ai/stanford-cs329p

Outline

- **Examples**
- **Law**
- **Algorithmic Fairness**
  - Evaluating estimators
  - Fairness criteria
  - Impossibility results
- **In Practice**
  - Human evaluation
  - Poison needle in the haystack
  - Decisions, risk and estimates

Simson’s Paradox

UC Berkeley Student Admission Rates (1973)

	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	**44%**	4,321	**35%**

Per Department

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	***825***	62%	108	**82%**
B	585	63%	***560***	63%	25	**68%**
C	918	35%	325	**37%**	***593***	34%
D	792	34%	417	33%	375	**35%**
E	584	25%	191	**28%**	***393***	24%
F	714	6%	373	6%	341	**7%**
Total	4526	39%	2691	45%	1835	30%

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

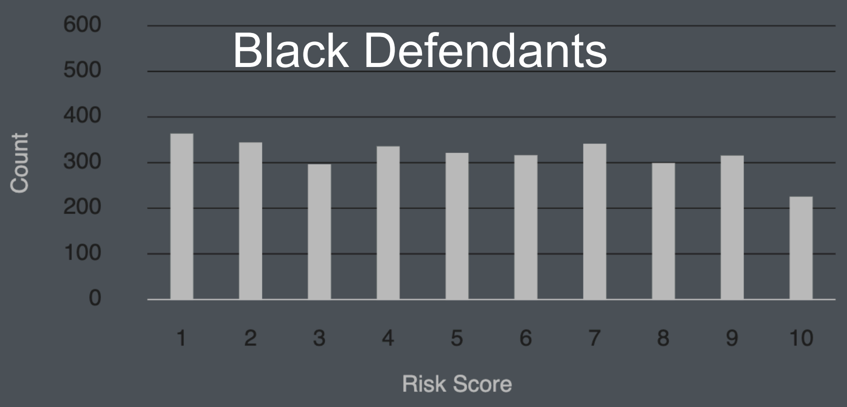
Machine Bias

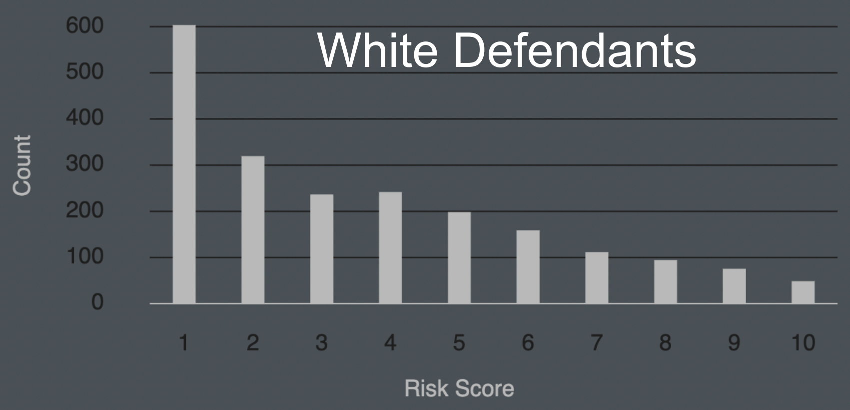
There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

COMPAS Study (Pro Publica, 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.)

Bias in Online Advertising?

Lambrecht & Tucker (ftc.gov)

Key idea - show ad and check for male/female bias

STEM Careers
Information about STEM Careers

Table: Raw Data Reported as an Average per Country

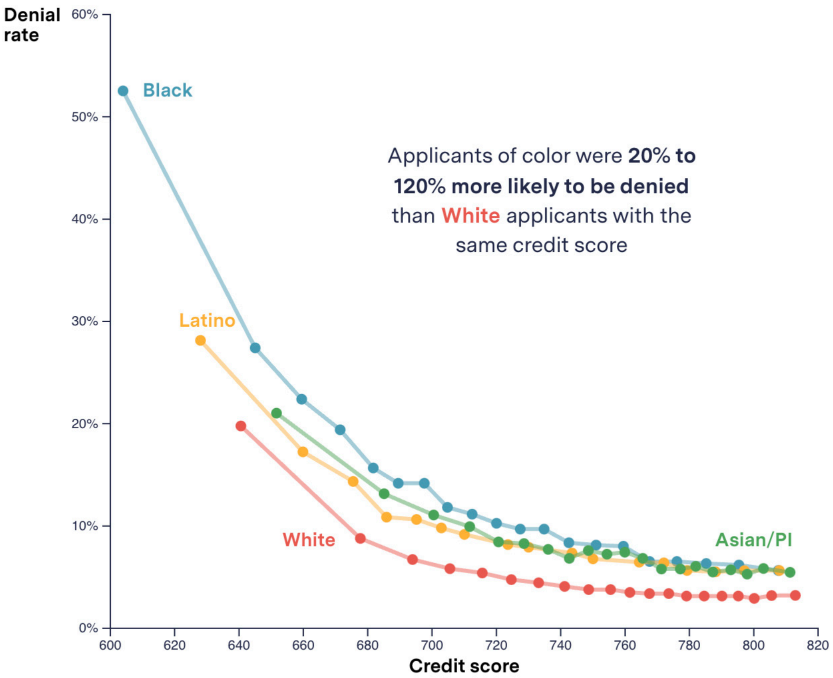
Age Group	Male Impr.	Female Impr.	Male Clicks	Female Clicks
Age18-24	3909	3401	6	6
Age25-34	3471	2597	5	4
Age35-44	2159	1485	3	3
Age45-54	1611	1177	2	2
Age55-64	1097	924	2	2
Age 65+	1007	808	2	2

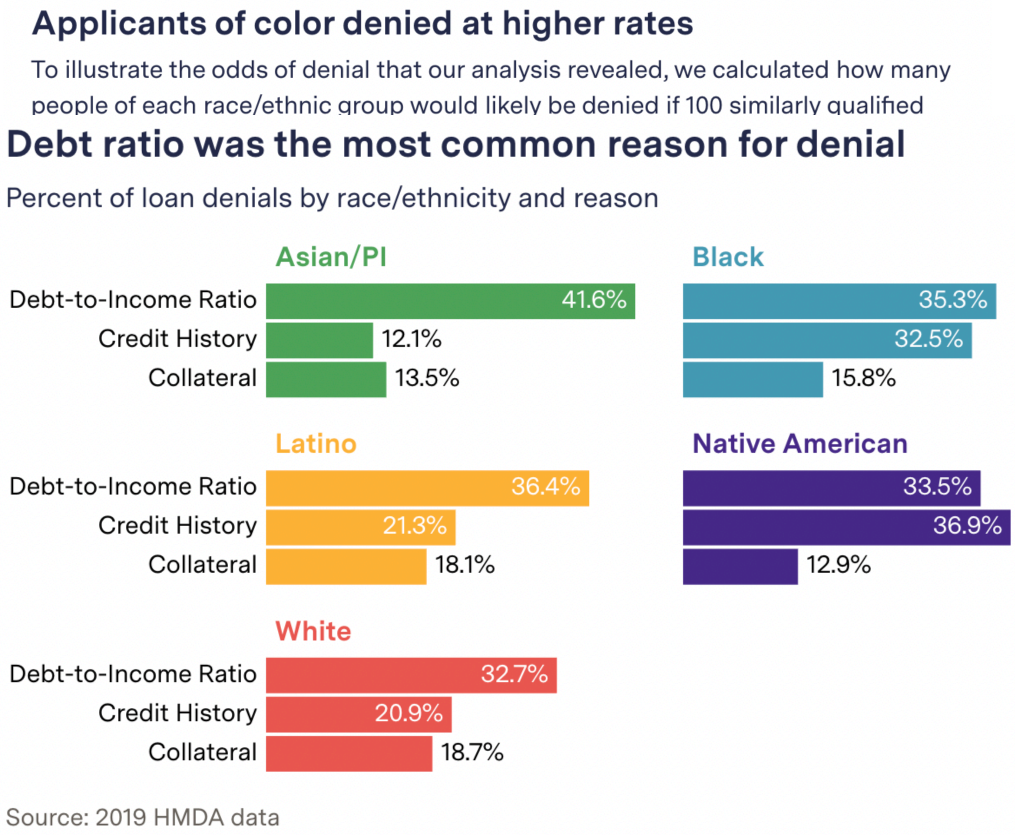
Really?!?

Really, this paper doesn't need any complex analysis

Bias in Lending?

Martinez & Kirchner, 2021



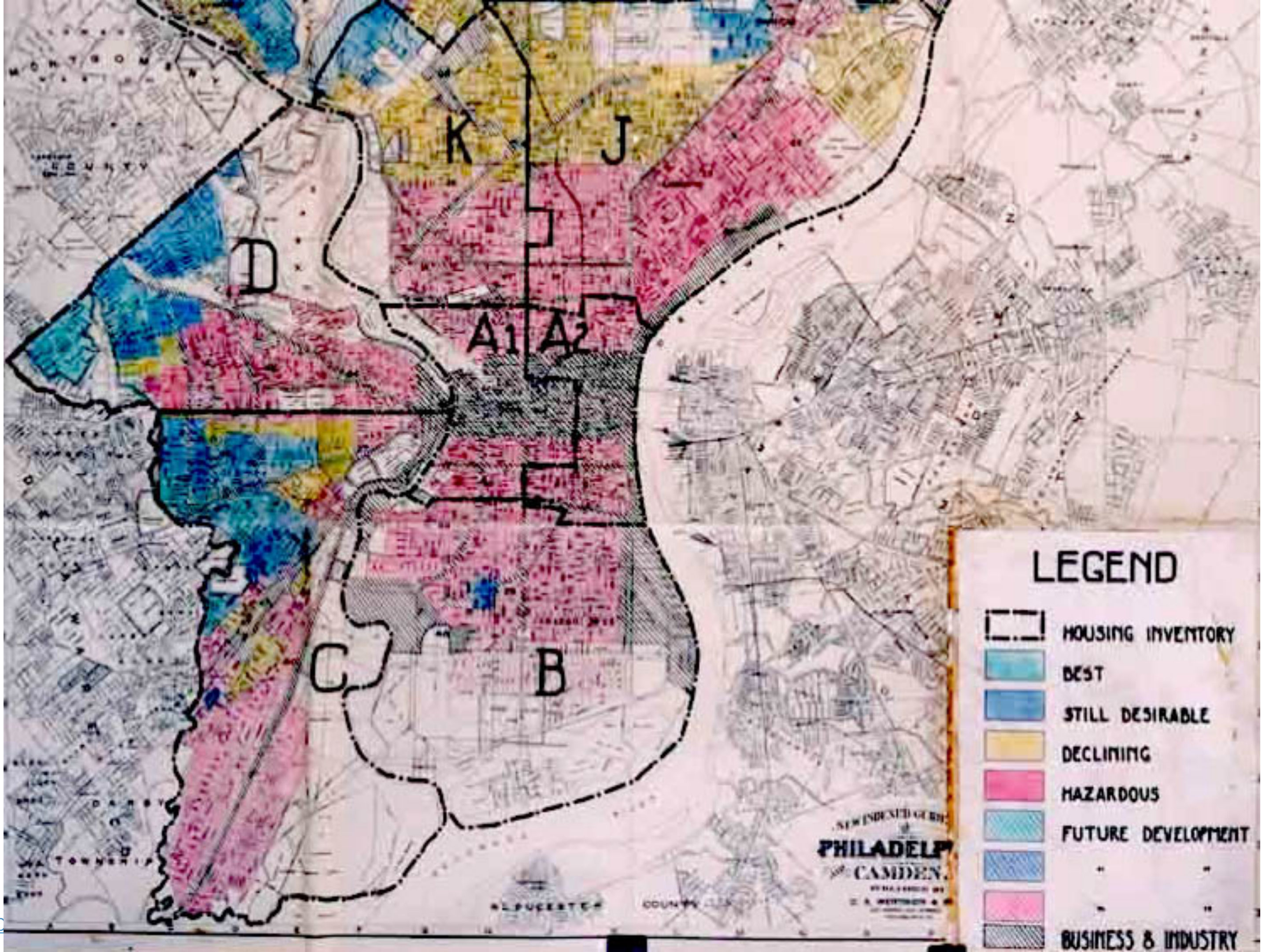


https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms

Redlining

Location based credit approval in the 1930s

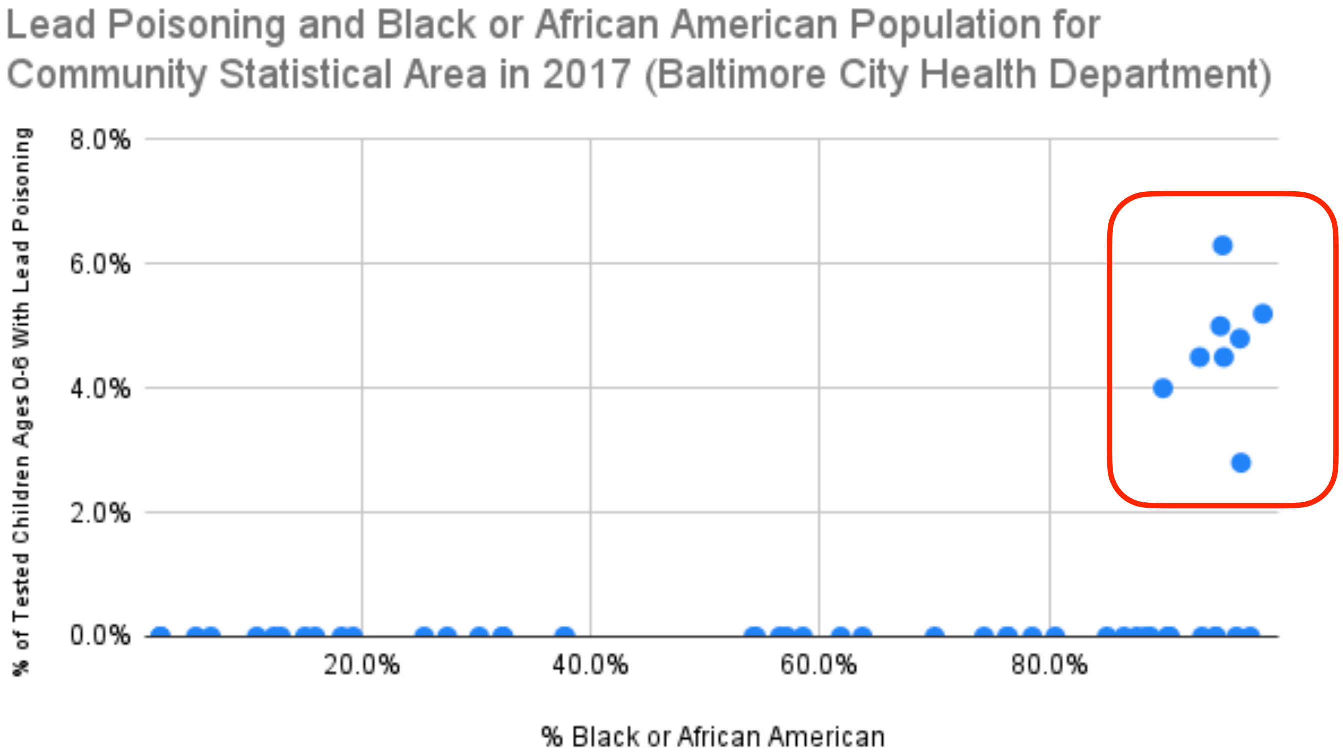
Discrimination against minorities



Redlining effects 90 years later

Location based credit approval in the 1930s

Discrimination against minorities



Bias and lack of fairness
can harm people for a century!

We owe it to everyone to be mindful!

Reading Material

- Corbett Davis & Goel - ICML 2019 tutorial https://policylab.stanford.edu/projects/defining-and-designing-fair-algorithms.html
- Hutchinson & Mitchell, 2018 (50 years fairness review) https://arxiv.org/abs/1811.10104
- Fazelpour & Lipton, 2020 (Non-ideal perspective) https://arxiv.org/abs/2001.09773
- Wachter, Mittelstadt & Russell, 2020 (Why fairness cannot be automated) https://arxiv.org/abs/2005.05906