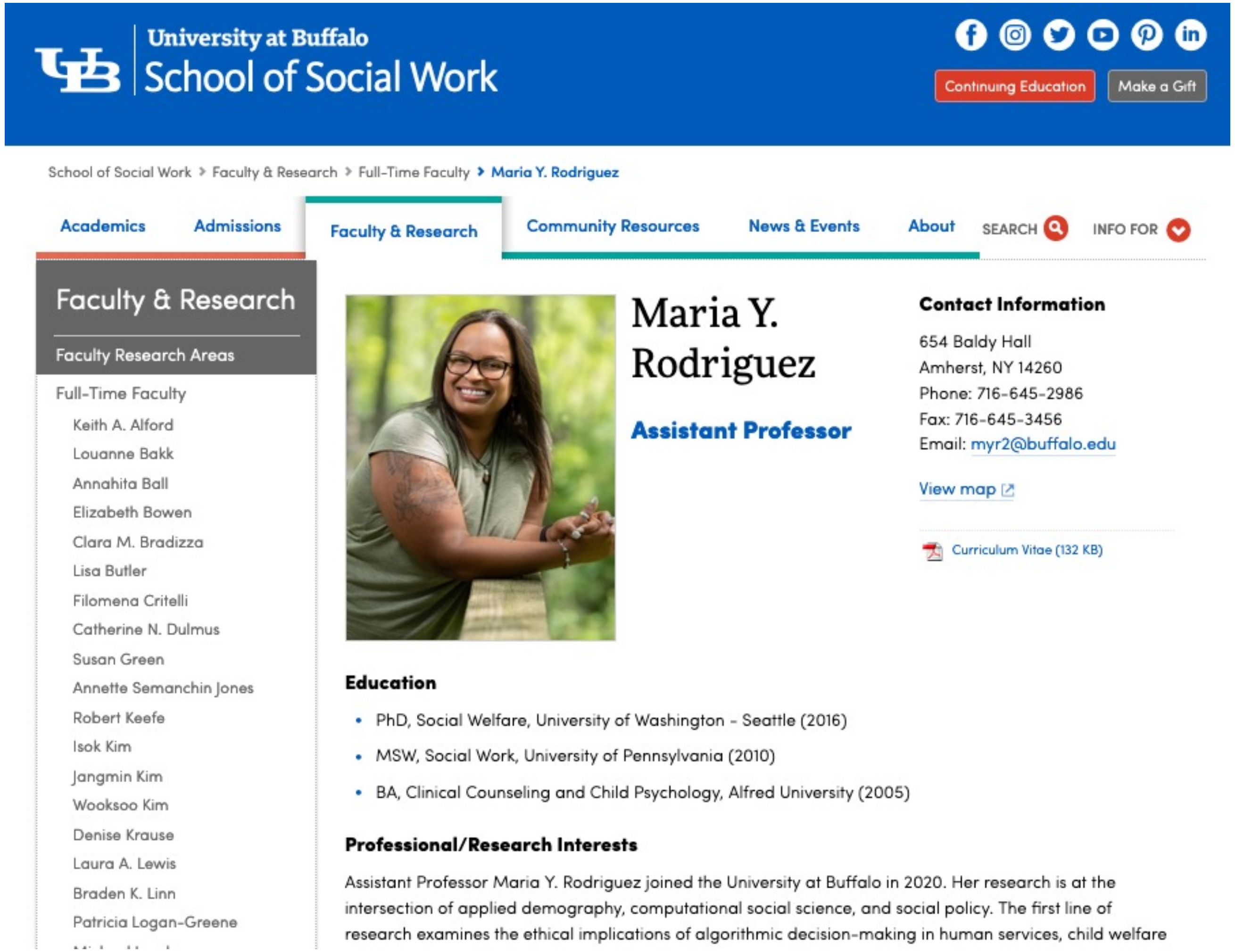
ML and Society

Mar 1, 2023

Passphrase for today: Maria Y. Rodriguez



Maria Y. Rodriguez

Assistant Professor

Contact Information

654 Baldy Hall
Amherst, NY 14260
Phone: 716-645-2986
Fax: 716-645-3456
Email: myr2@buffalo.edu

View map

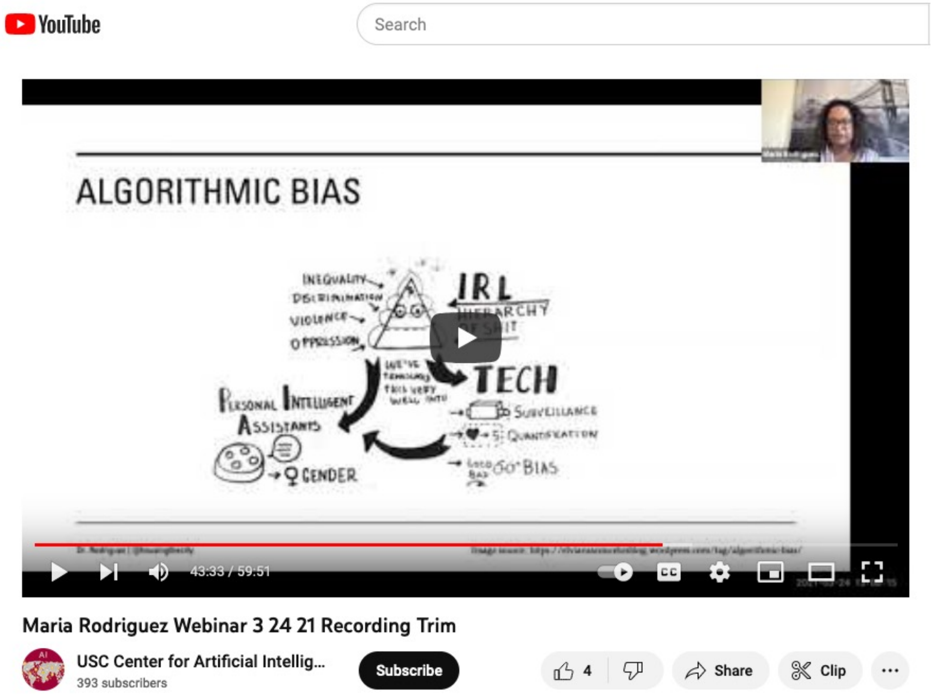
Curriculum Vitae (132 KB)

Education

- PhD, Social Welfare, University of Washington - Seattle (2016)
- MSW, Social Work, University of Pennsylvania (2010)
- BA, Clinical Counseling and Child Psychology, Alfred University (2005)

Professional/Research Interests

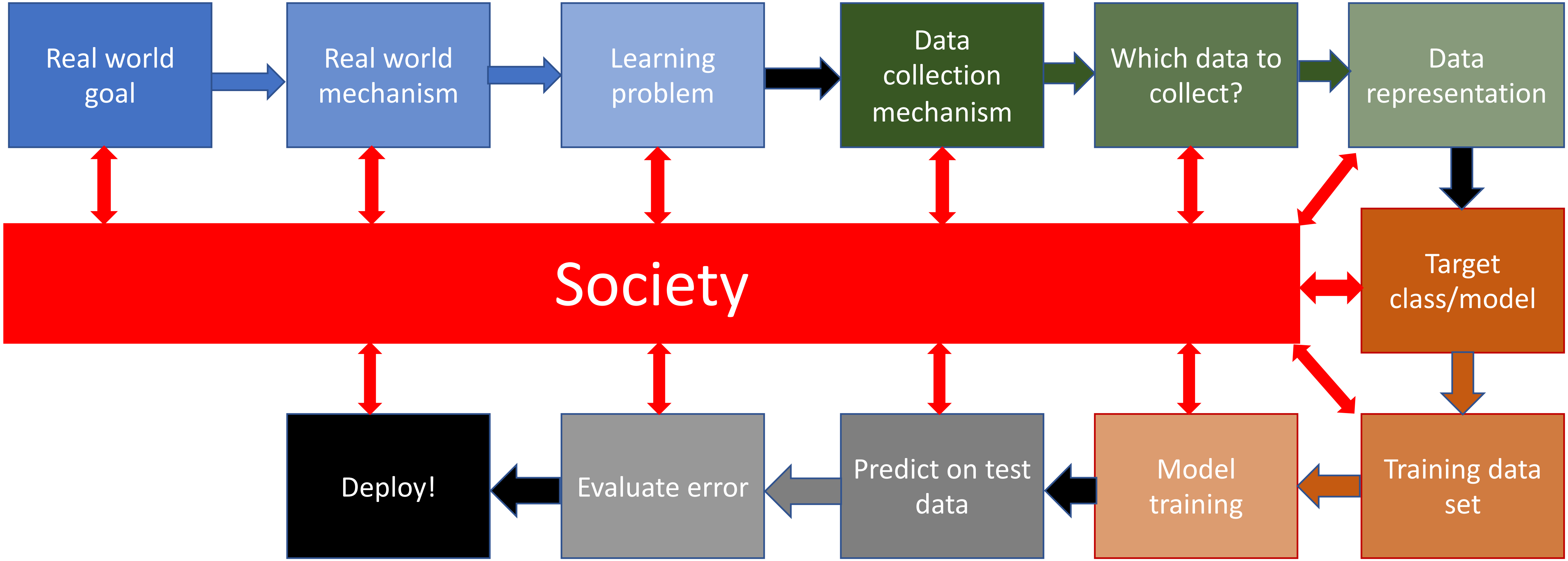
Assistant Professor Maria Y. Rodriguez joined the University at Buffalo in 2020. Her research is at the intersection of applied demography, computational social science, and social policy. The first line of research examines the ethical implications of algorithmic decision-making in human services, child welfare



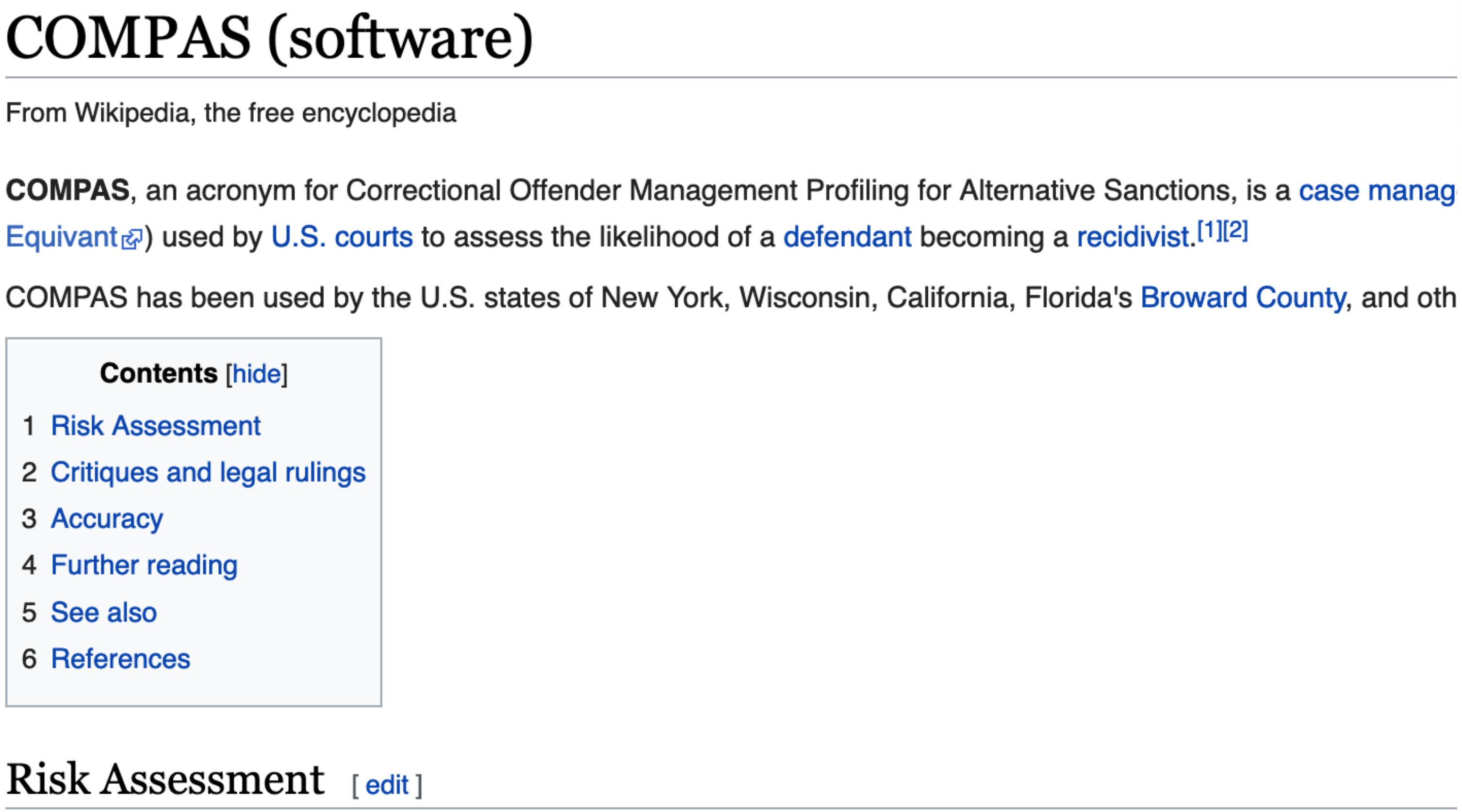
Maria Rodriguez Webinar 3 24 21 Recording Trim

USC Center for Artificial Intellig...
393 subscribers

“Done” with ML pipeline



Do you remember COMPAS?



COMPAS (software)

From Wikipedia, the free encyclopedia

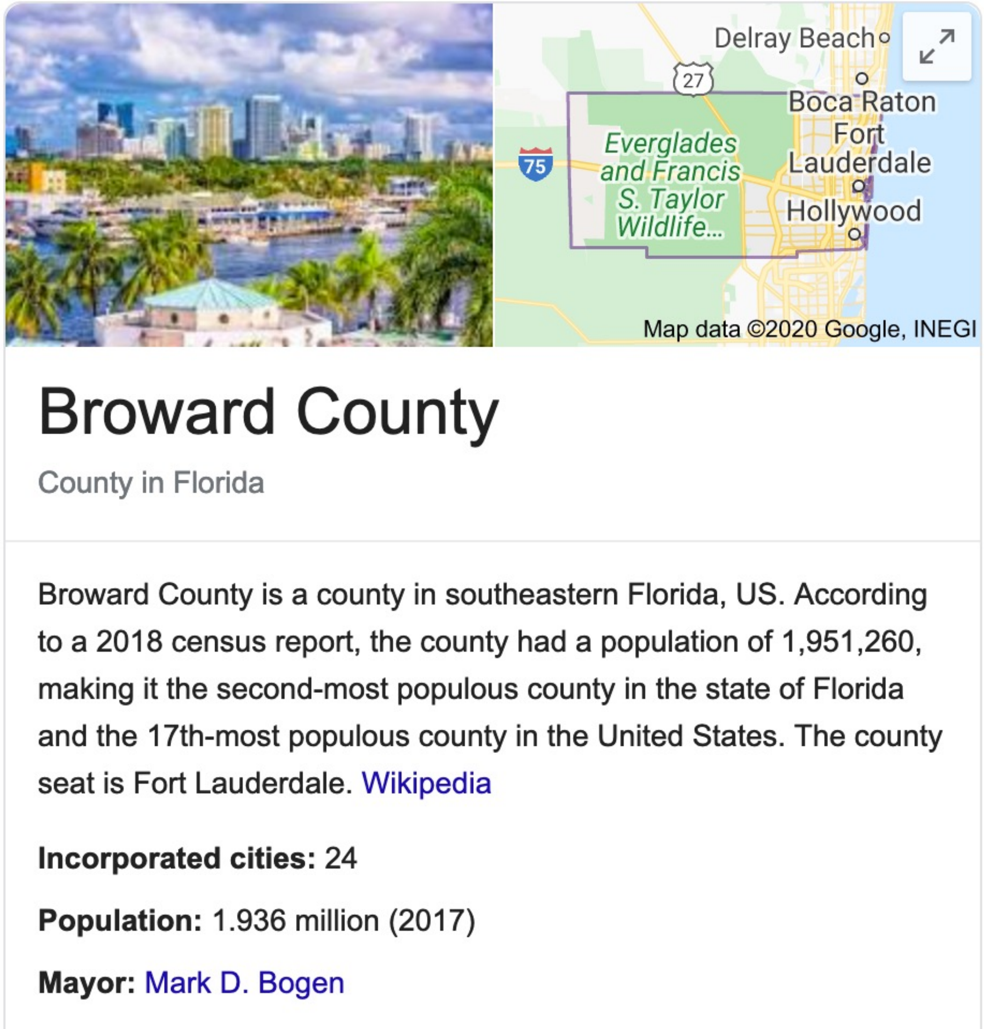
COMPAS, an acronym for Correctional Offender Management Profiling for Alternative Sanctions, is a case manag Equivant) used by U.S. courts to assess the likelihood of a defendant becoming a recidivist.[1][2]

COMPAS has been used by the U.S. states of New York, Wisconsin, California, Florida's Broward County, and oth

Contents [hide]

1 Risk Assessment
2 Critiques and legal rulings
3 Accuracy
4 Further reading
5 See also
6 References

Risk Assessment [edit]



Delray Beach
27
Boca Raton
75
Everglades and Francis S. Taylor Wildlife...
Fort Lauderdale
Hollywood
Map data ©2020 Google, INEGI

Broward County

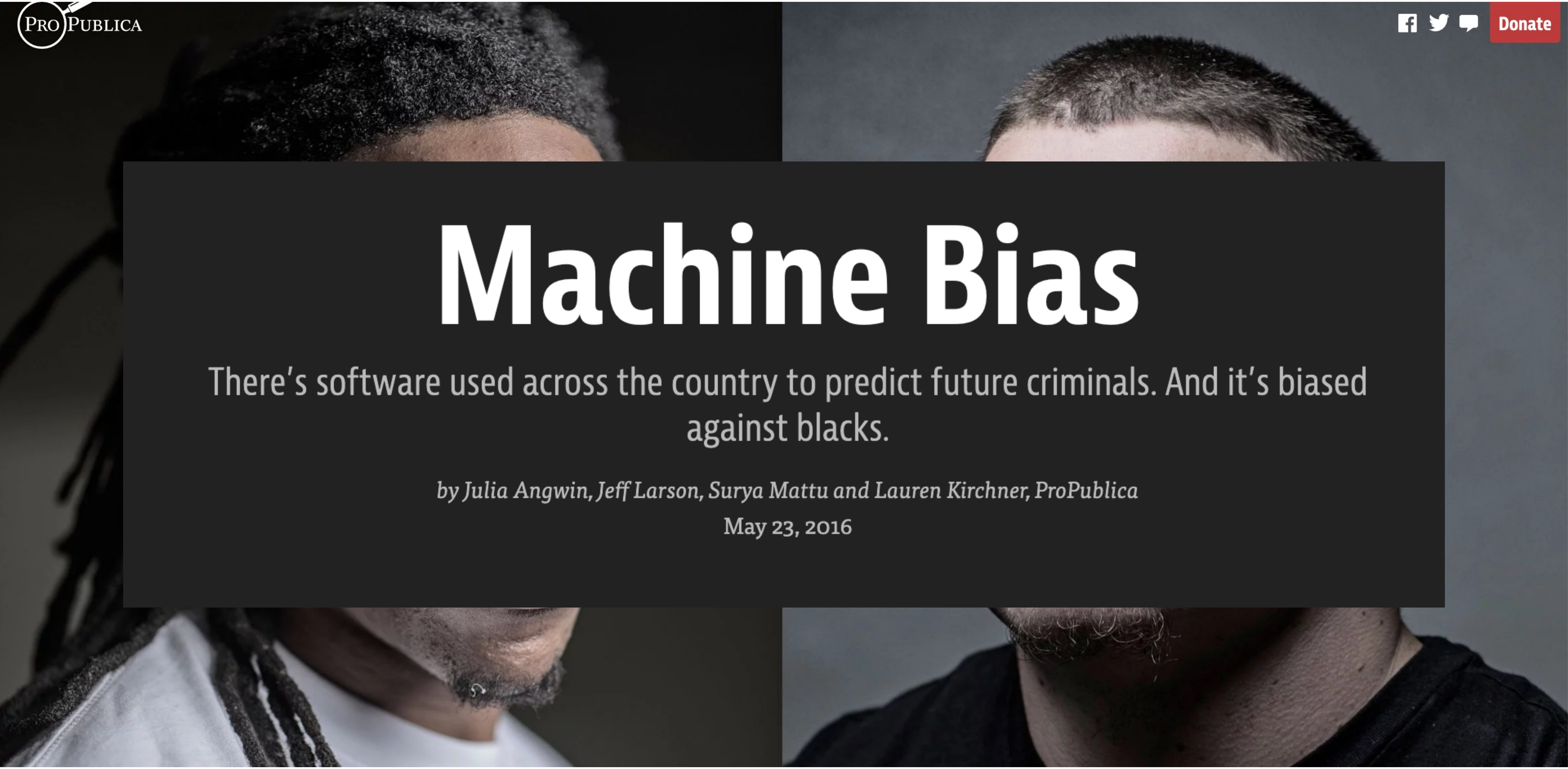
County in Florida

Broward County is a county in southeastern Florida, US. According to a 2018 census report, the county had a population of 1,951,260, making it the second-most populous county in the state of Florida and the 17th-most populous county in the United States. The county seat is Fort Lauderdale. Wikipedia

Incorporated cities: 24

Population: 1.936 million (2017)

Mayor: Mark D. Bogen



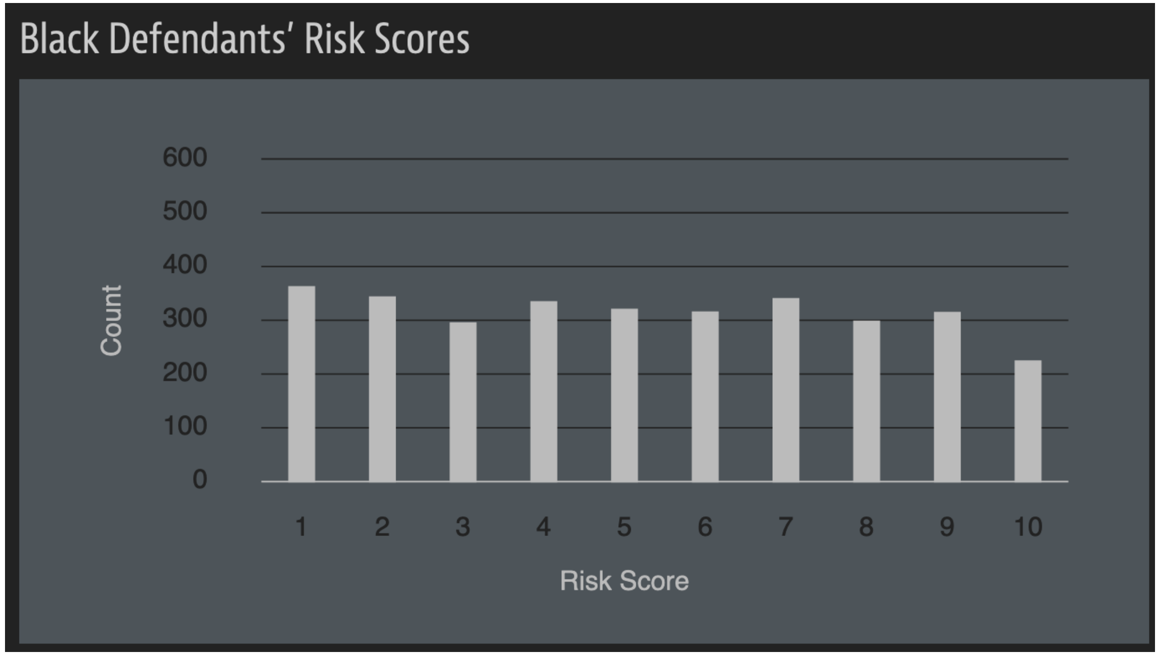
Machine Bias

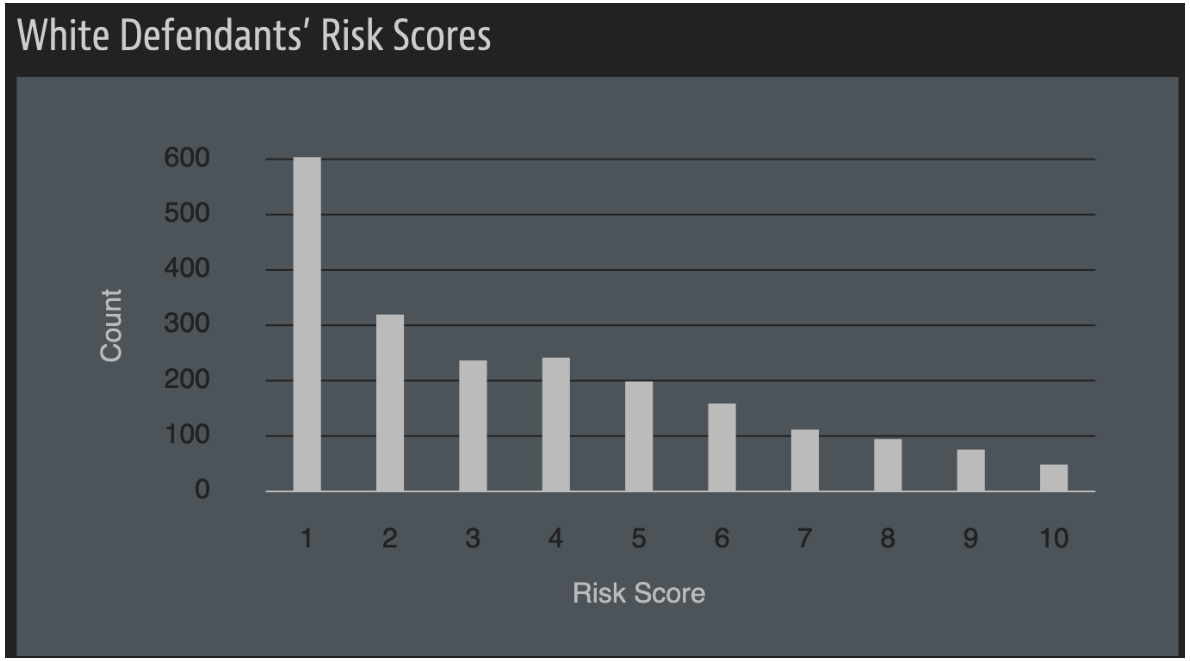
There's software used across the country to predict future criminals. And it's biased against blacks.

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

May 23, 2016

A sample of their result

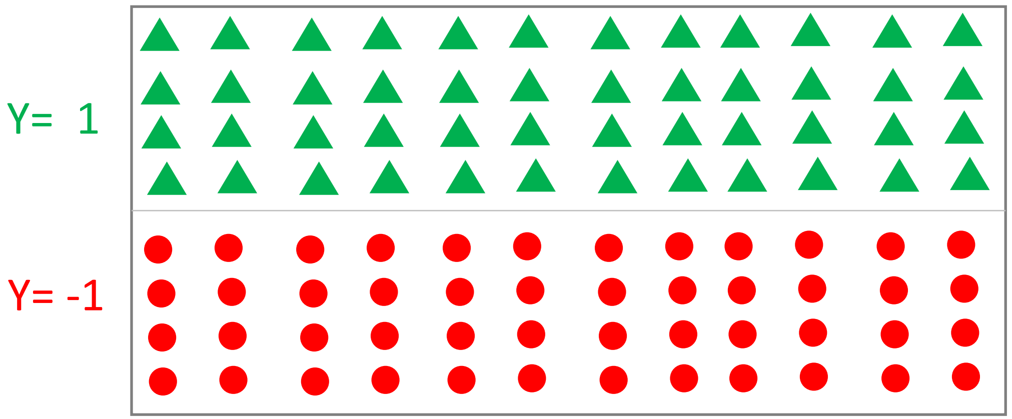




False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks."

Anthony W. Flores
California State University, Bakersfield
Kristin Bechtel
Crime and Justice Institute at CRJ
Christopher T. Lowenkamp
Administrative Office of the United States Courts
Probation and Pretrial Services Office

Going beyond # correctly classified points

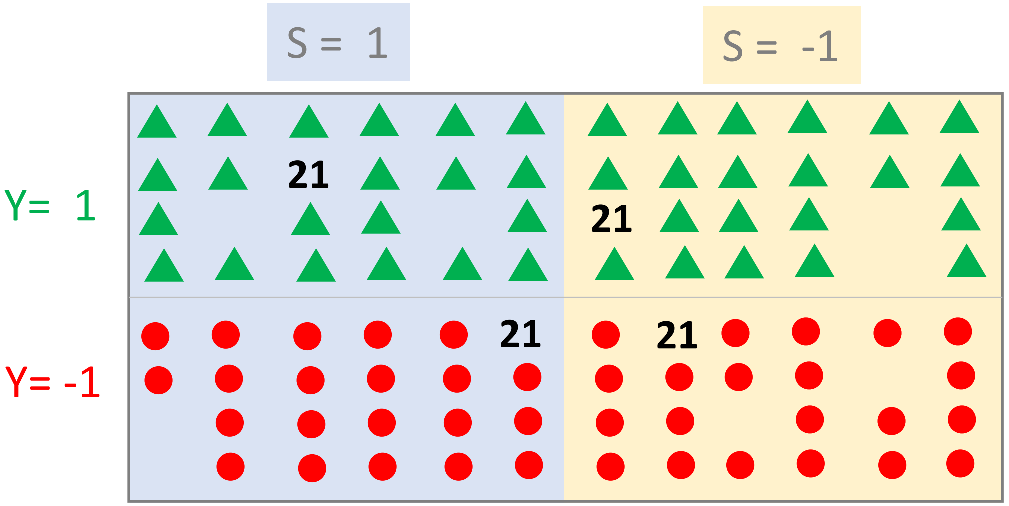


Points are NOT 2D!

Binary classifier output

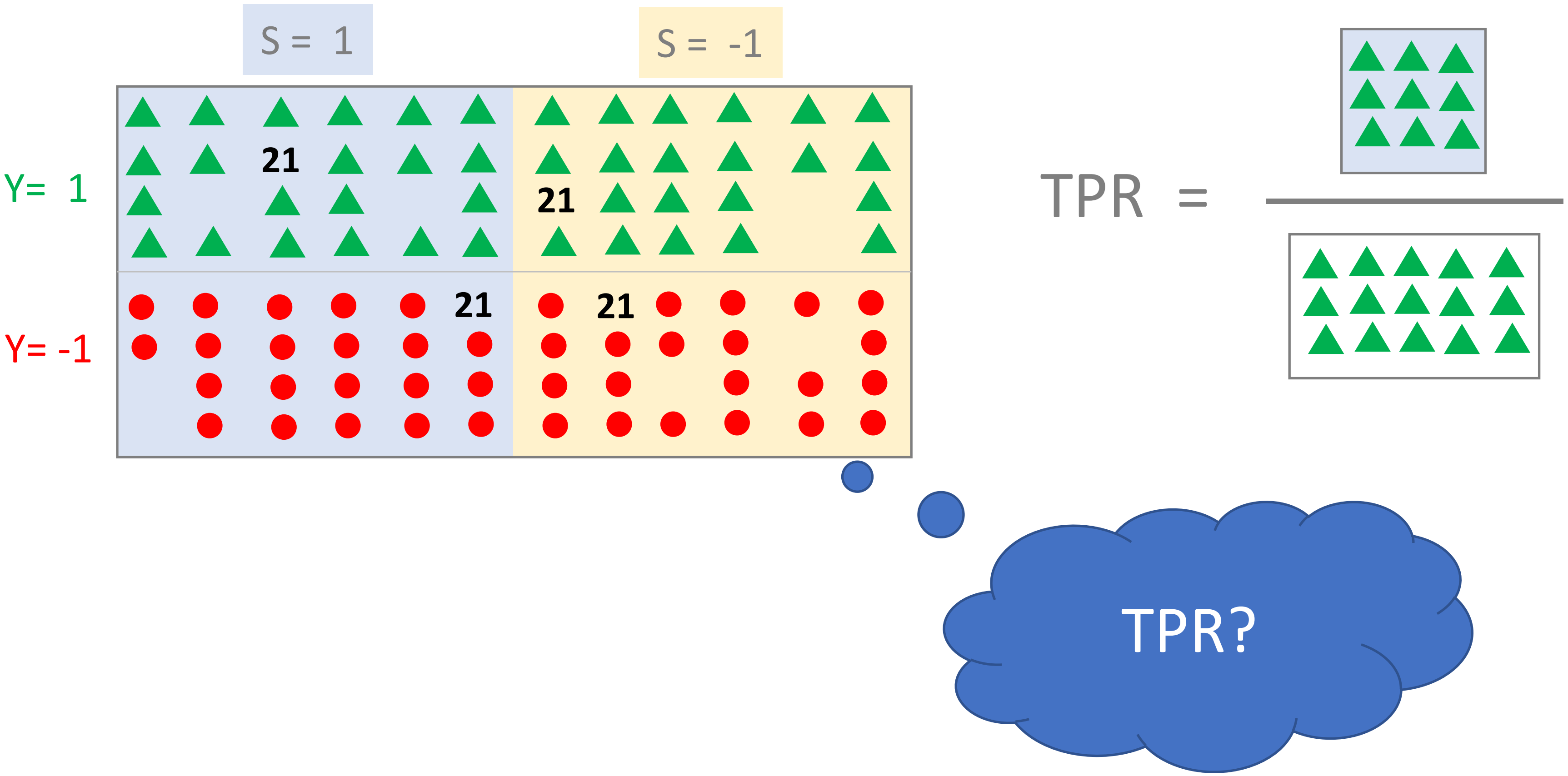
S = 1	S = -1

“Confusion matrix”

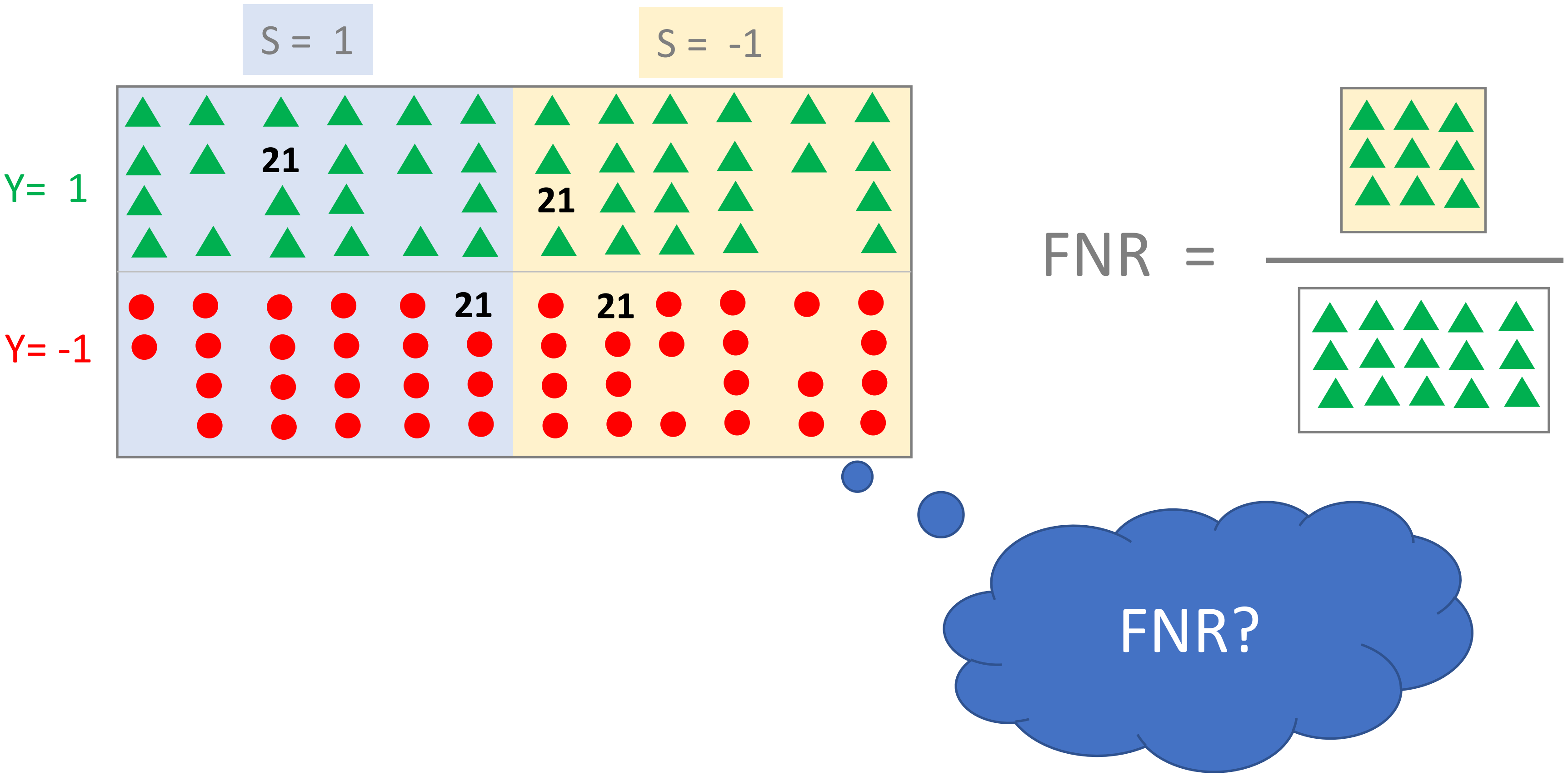


	S = 1	S = -1
Y= 1	21	21
Y= -1	21	21

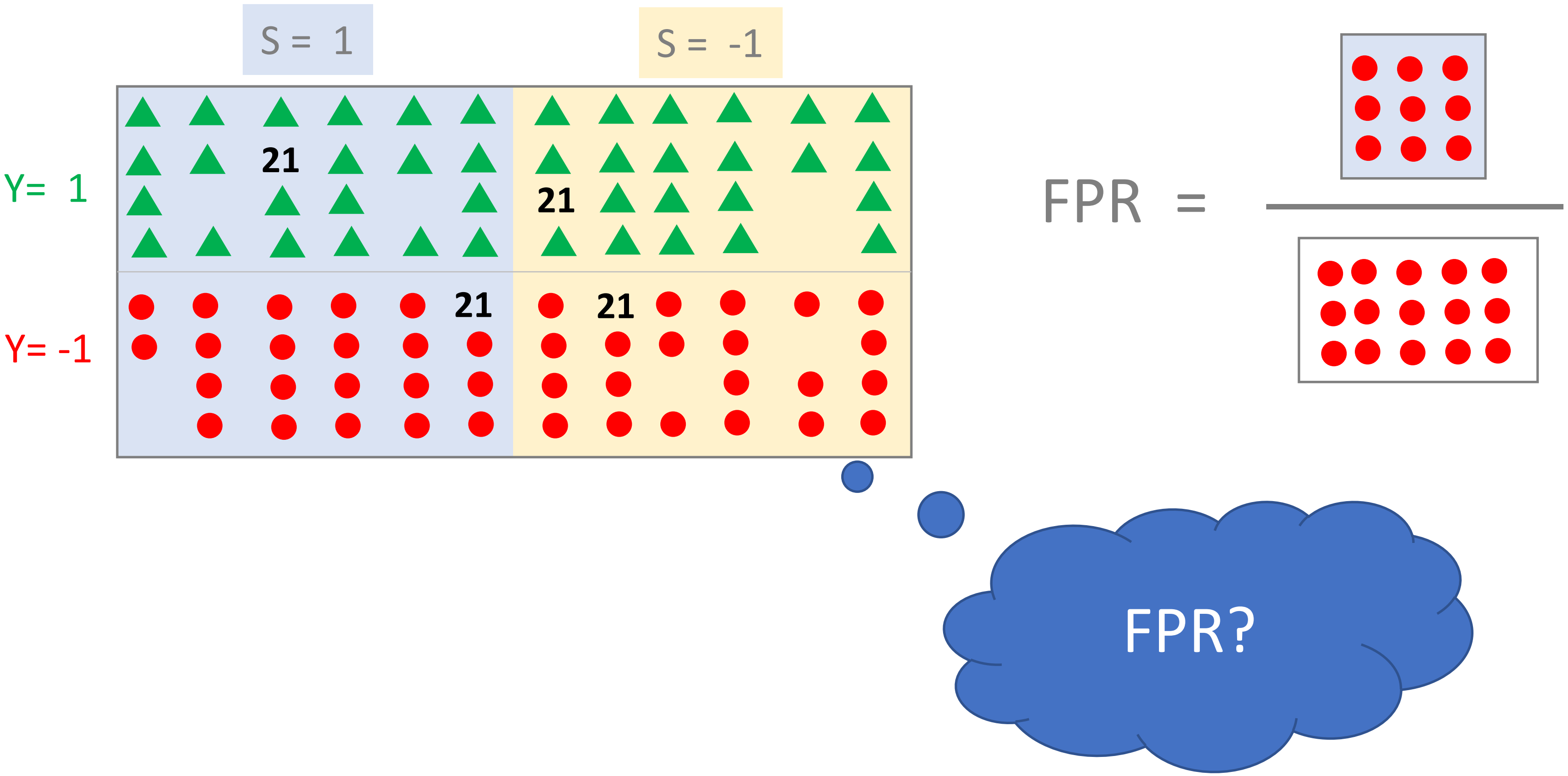
True Positive rate



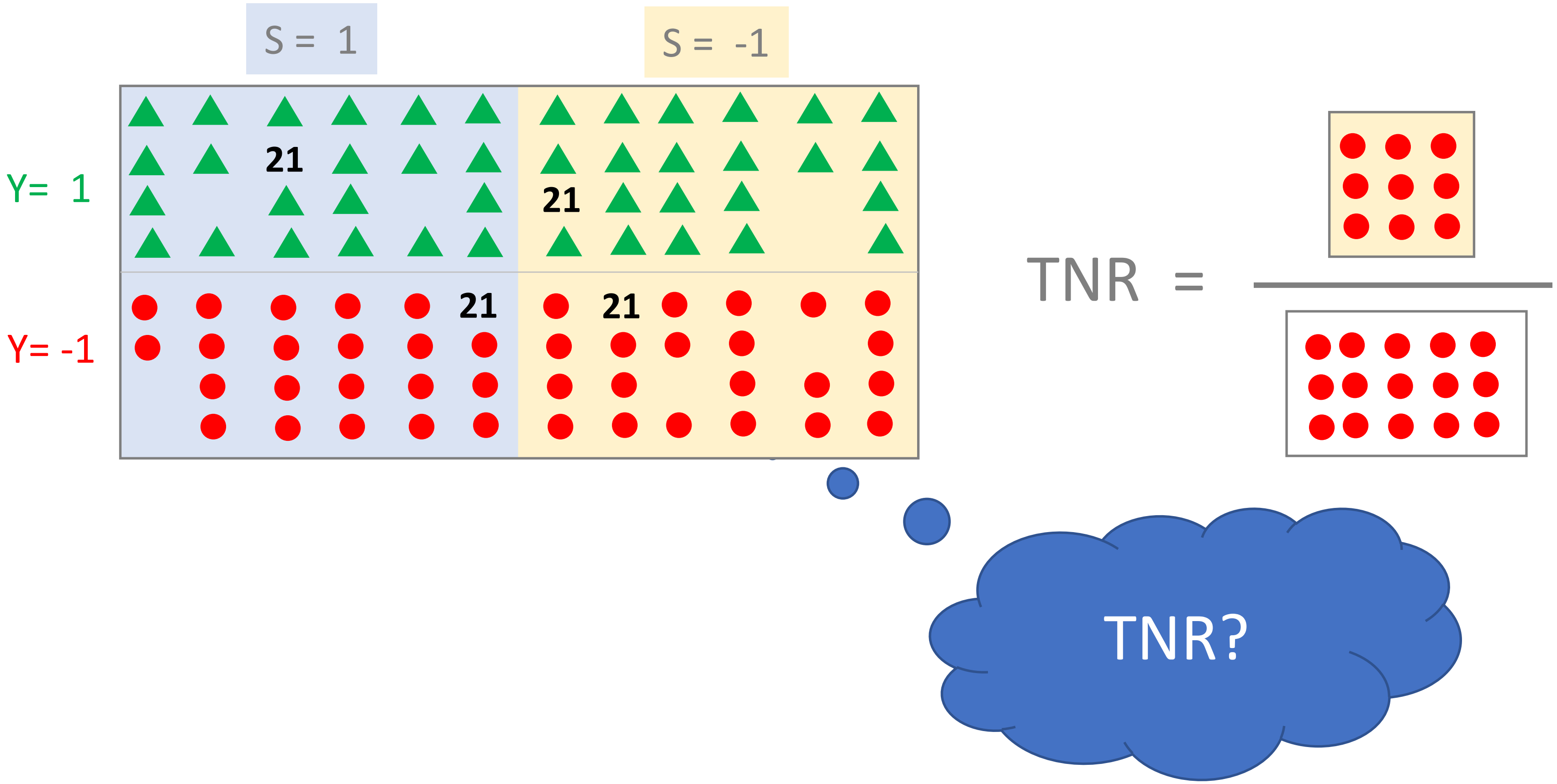
False Negative Rate (FNR)



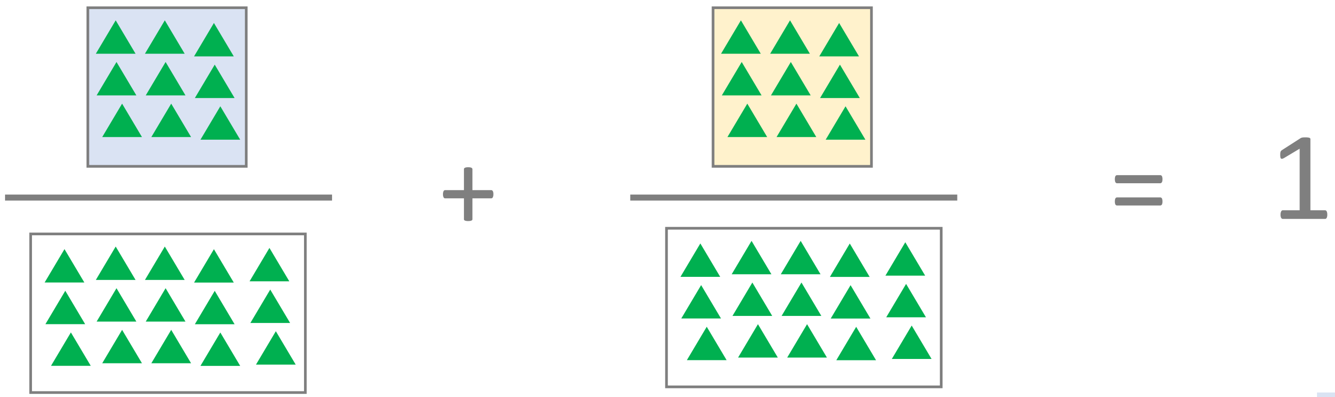
False Positive Rate (FPR)

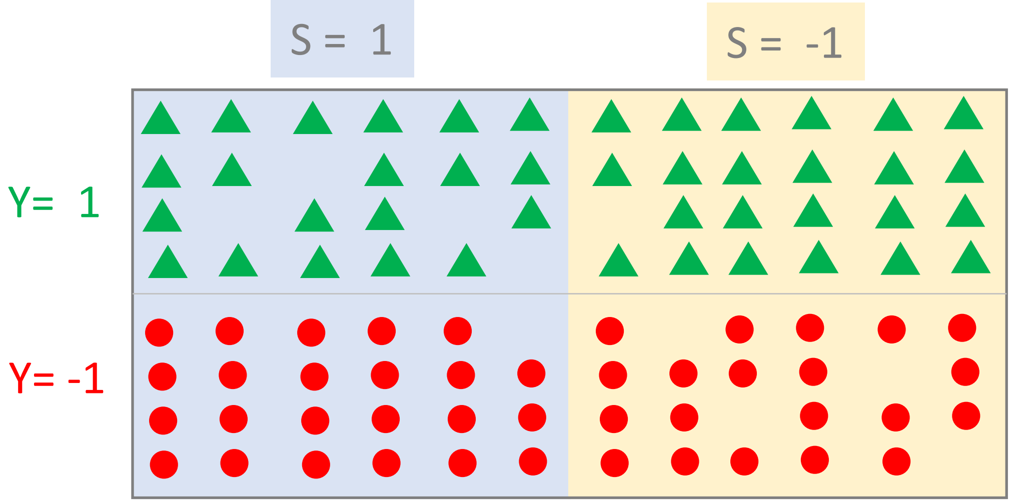


True Negative Rate (TNR)

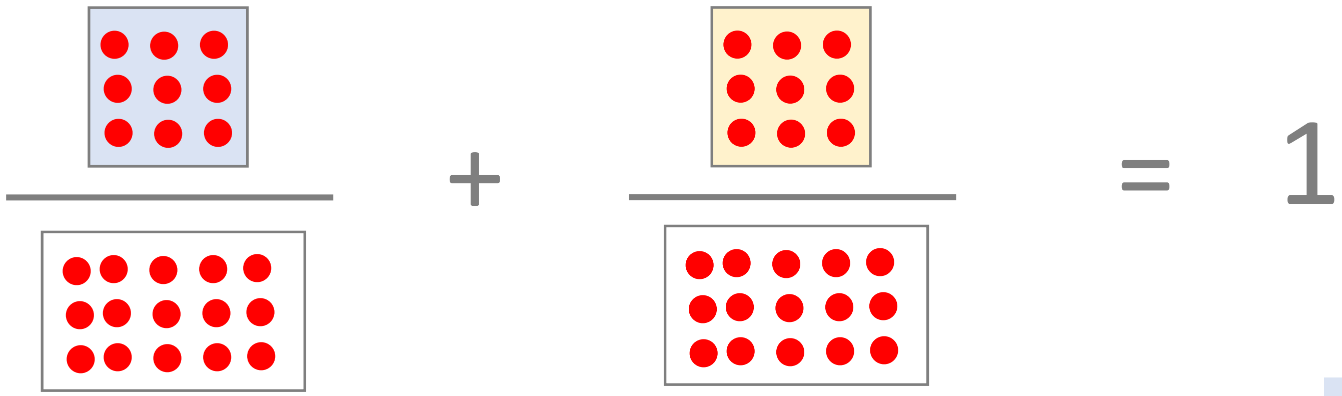


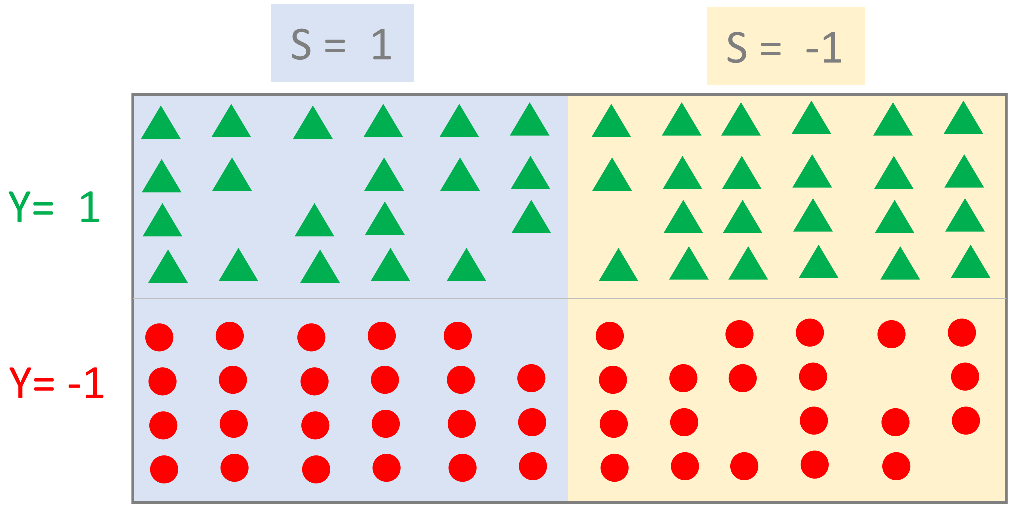
TPR + FNR = 1



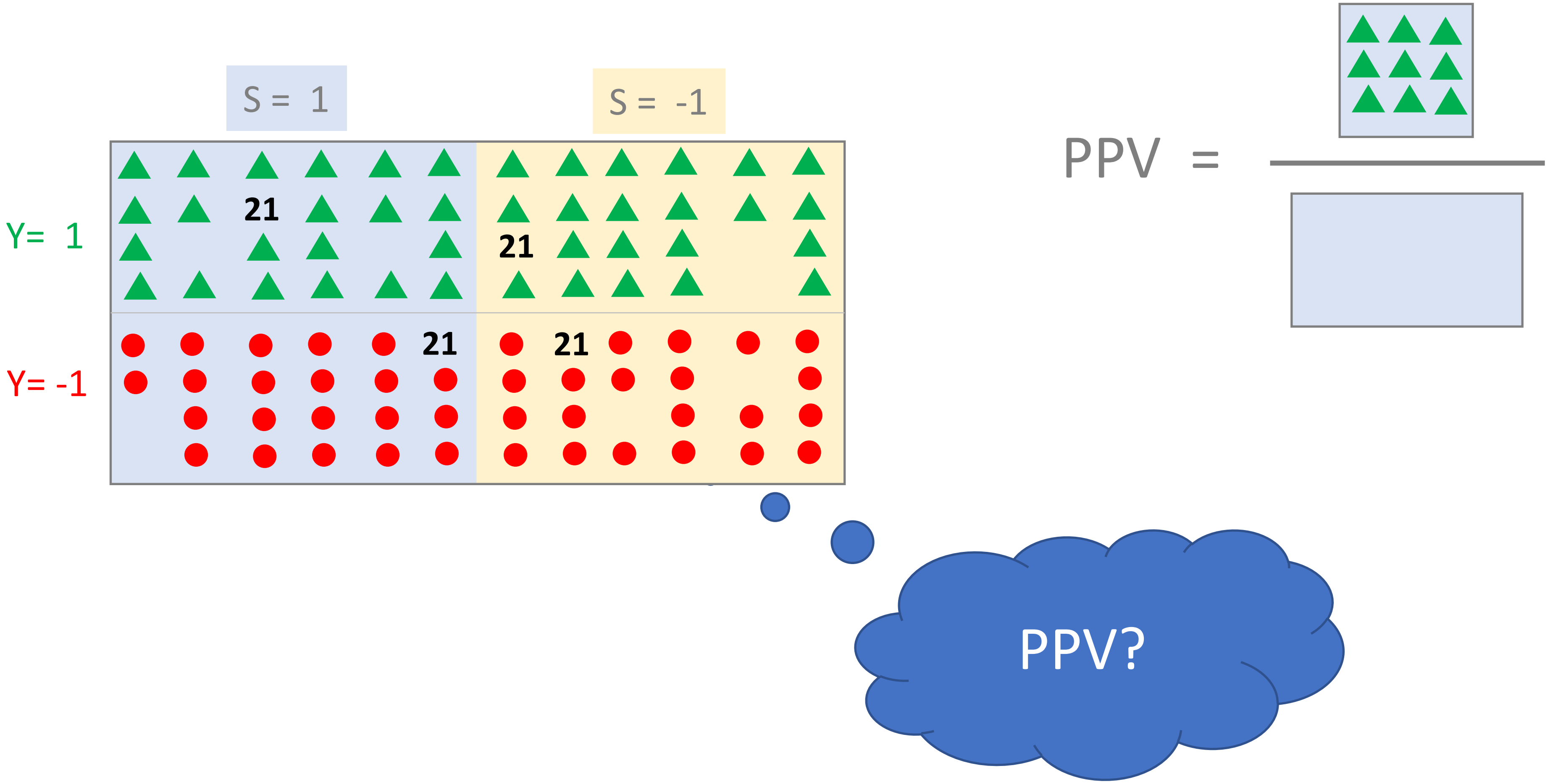


FPR + TNR = 1

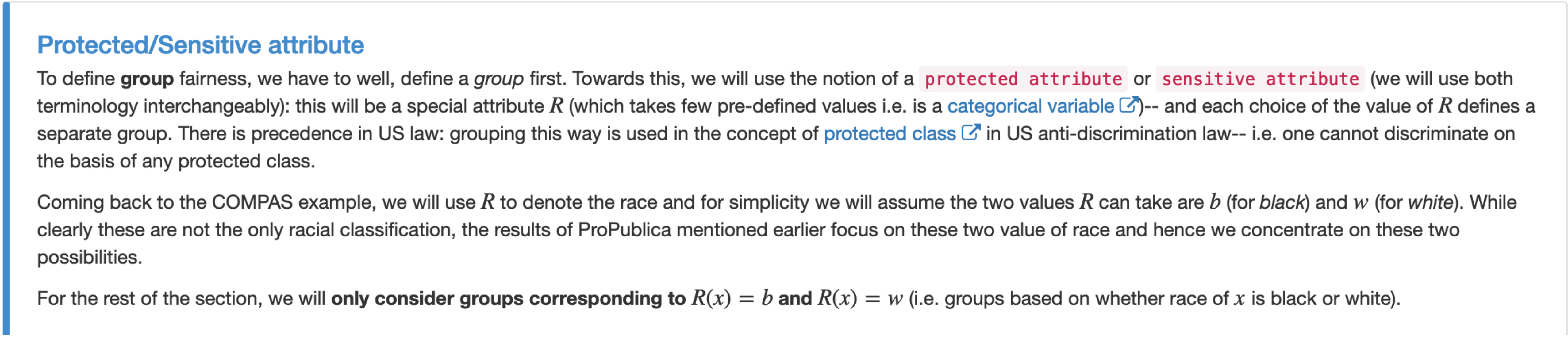




Positive Predictive Value (PPV)



Back to fairness

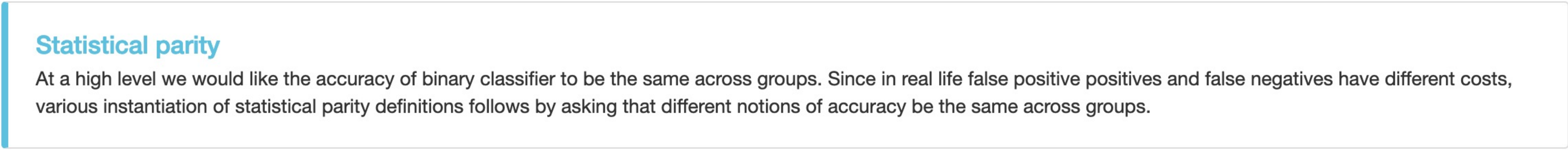


Protected/Sensitive attribute

To define **group** fairness, we have to well, define a *group* first. Towards this, we will use the notion of a `protected attribute` or `sensitive attribute` (we will use both terminology interchangeably): this will be a special attribute R (which takes few pre-defined values i.e. is a categorical variable)-- and each choice of the value of R defines a separate group. There is precedence in US law: grouping this way is used in the concept of protected class in US anti-discrimination law-- i.e. one cannot discriminate on the basis of any protected class.

Coming back to the COMPAS example, we will use R to denote the race and for simplicity we will assume the two values R can take are b (for *black*) and w (for *white*). While clearly these are not the only racial classification, the results of ProPublica mentioned earlier focus on these two value of race and hence we concentrate on these two possibilities.

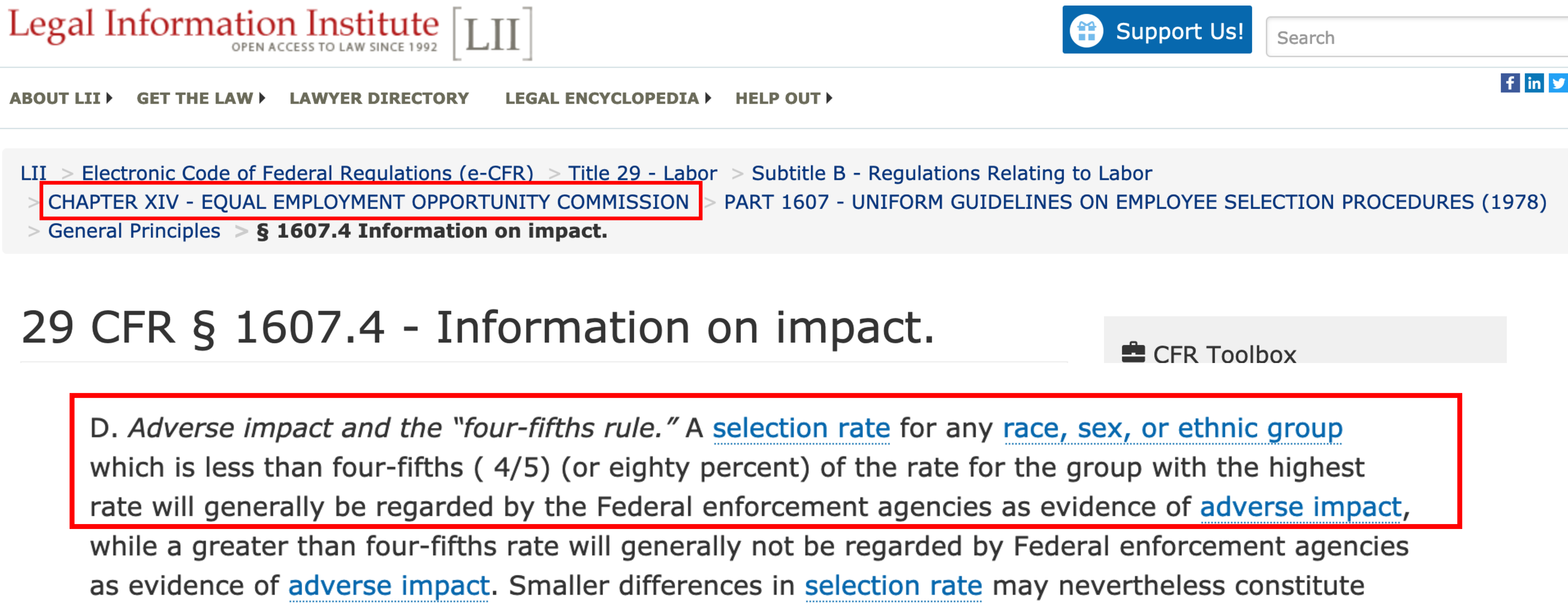
For the rest of the section, we will **only consider groups corresponding to $R(x) = b$ and $R(x) = w$** (i.e. groups based on whether race of x is black or white).



Statistical parity

At a high level we would like the accuracy of binary classifier to be the same across groups. Since in real life false positive positives and false negatives have different costs, various instantiation of statistical parity definitions follows by asking that different notions of accuracy be the same across groups.

Why statistical parity across groups?

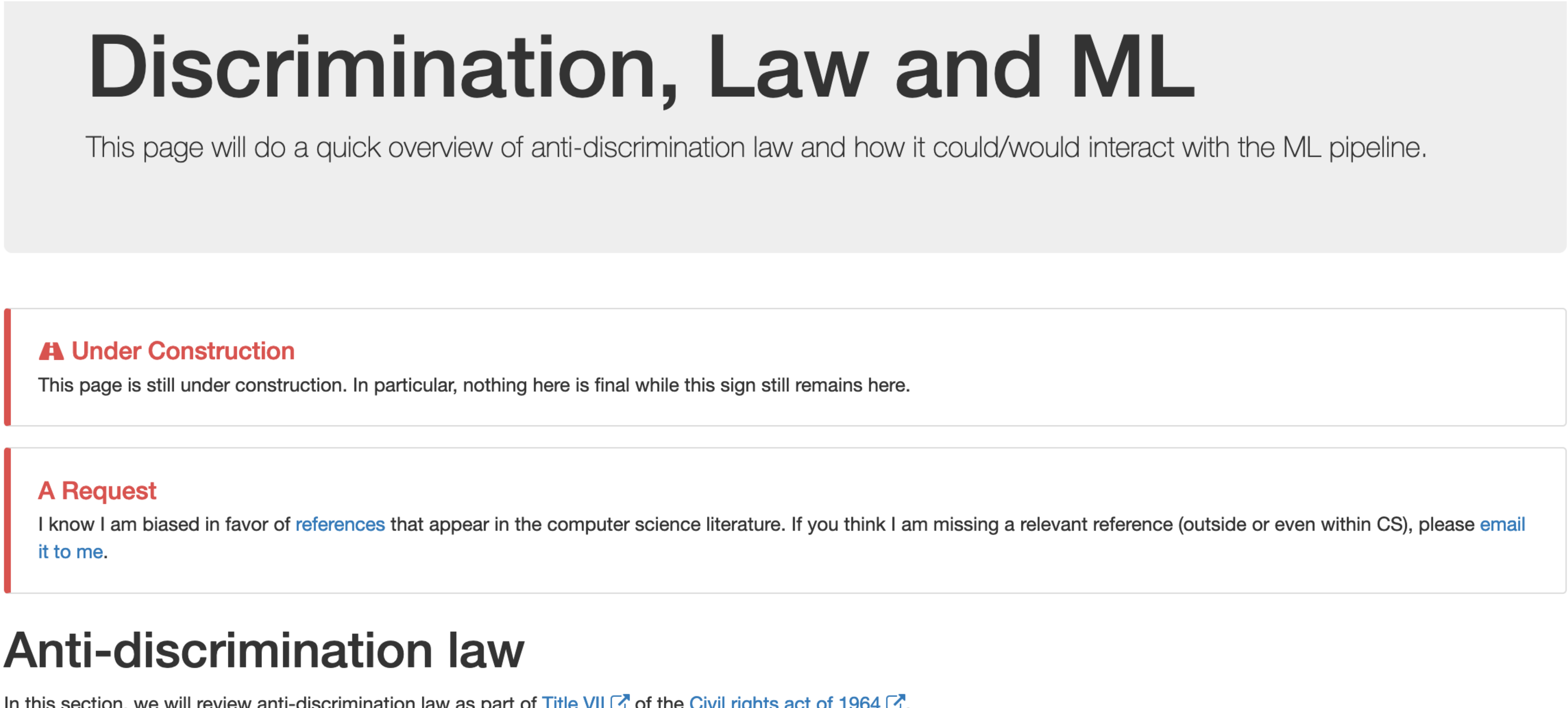


LII > Electronic Code of Federal Regulations (e-CFR) > Title 29 - Labor > Subtitle B - Regulations Relating to Labor > CHAPTER XIV - EQUAL EMPLOYMENT OPPORTUNITY COMMISSION > PART 1607 - UNIFORM GUIDELINES ON EMPLOYEE SELECTION PROCEDURES (1978) > General Principles > **§ 1607.4 Information on impact.**

29 CFR § 1607.4 - Information on impact.

D. *Adverse impact and the "four-fifths rule."* A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute

Notes on ML and law



Discrimination, Law and ML

This page will do a quick overview of anti-discrimination law and how it could/would interact with the ML pipeline.

Under Construction

This page is still under construction. In particular, nothing here is final while this sign still remains here.

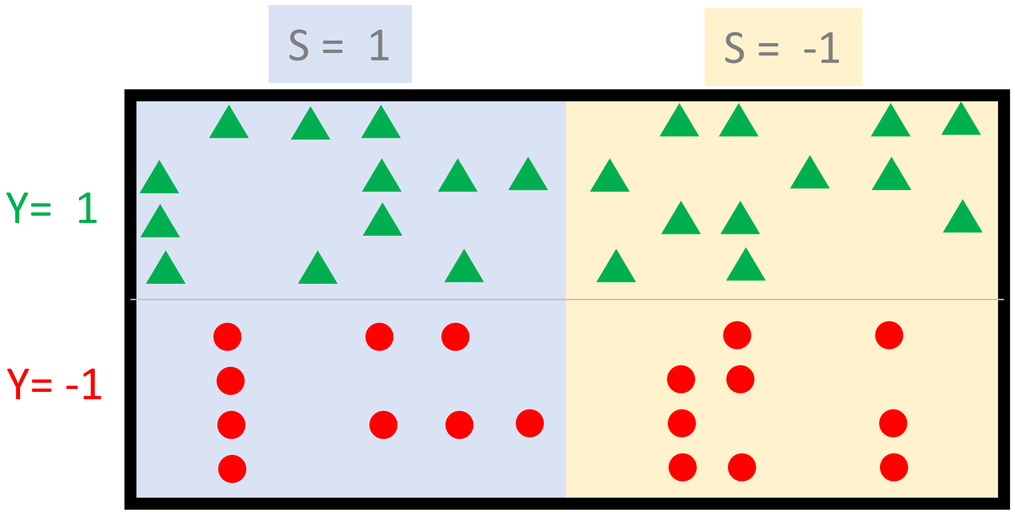
A Request

I know I am biased in favor of references that appear in the computer science literature. If you think I am missing a relevant reference (outside or even within CS), please email it to me.

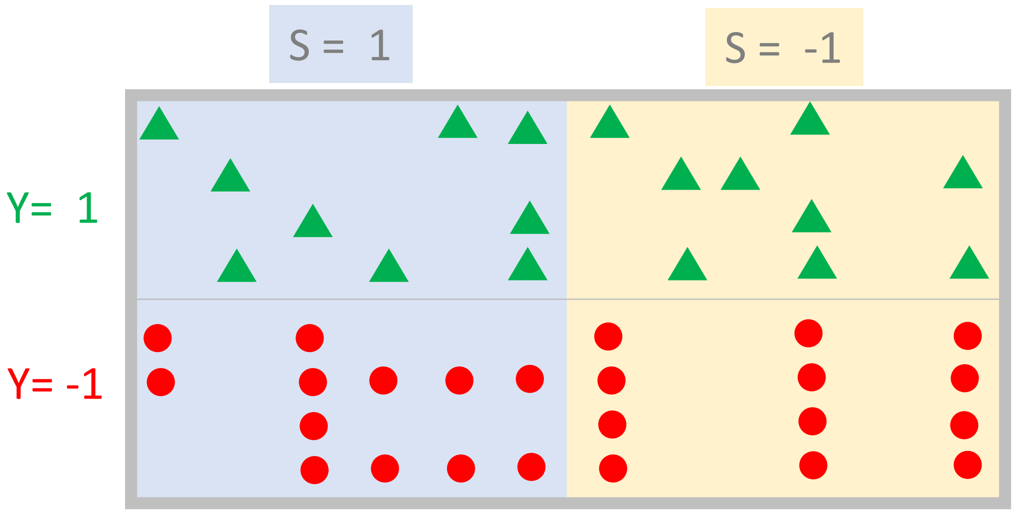
Anti-discrimination law

In this section, we will review anti-discrimination law as part of Title VII of the Civil rights act of 1964.

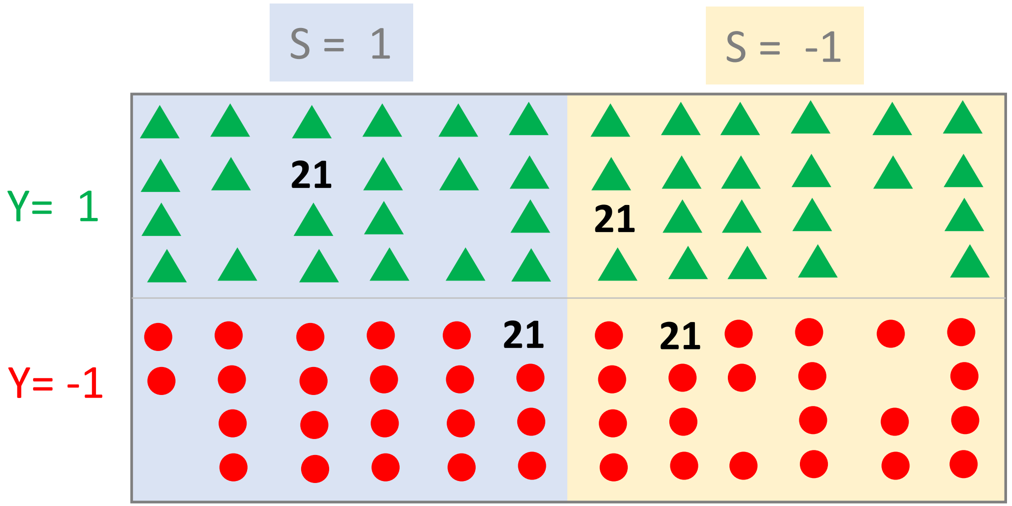
Rates for groups



b

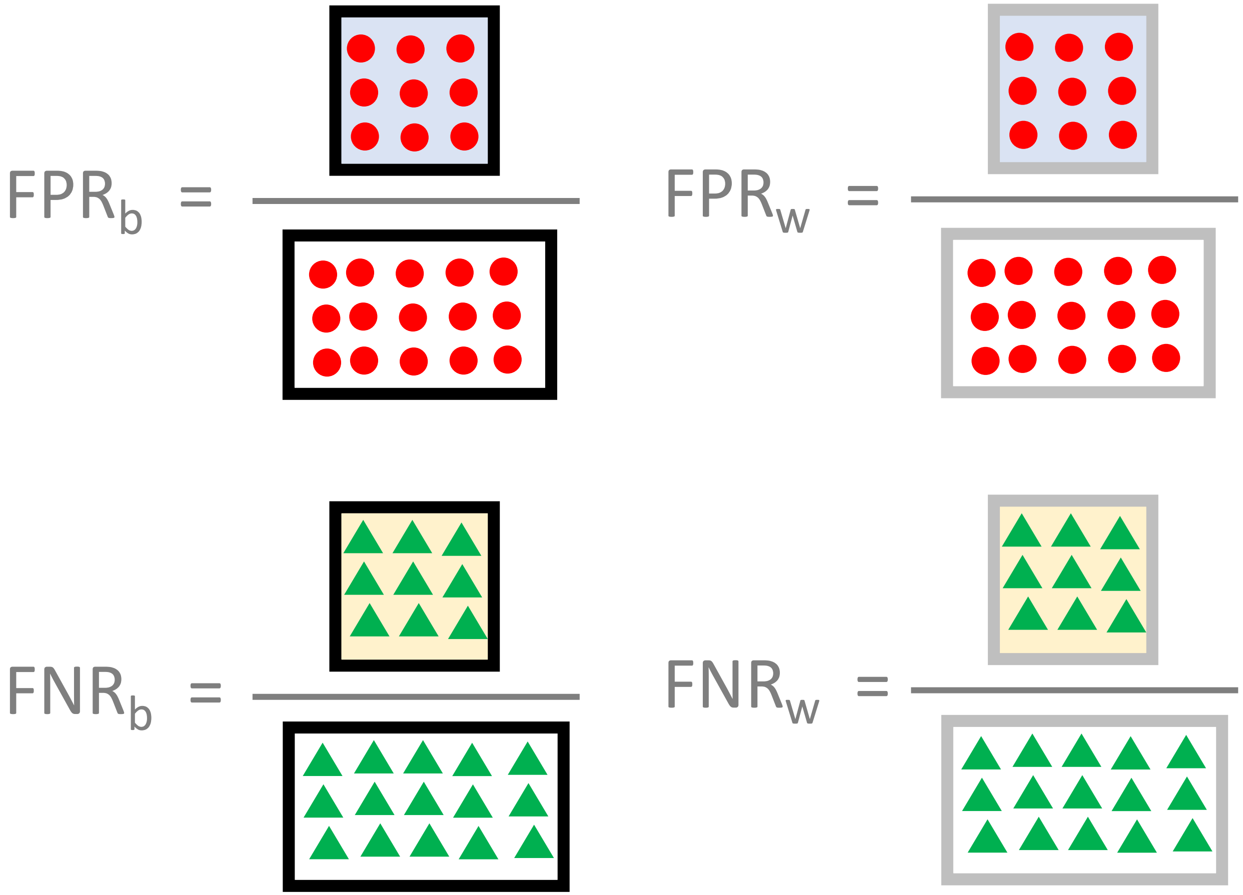


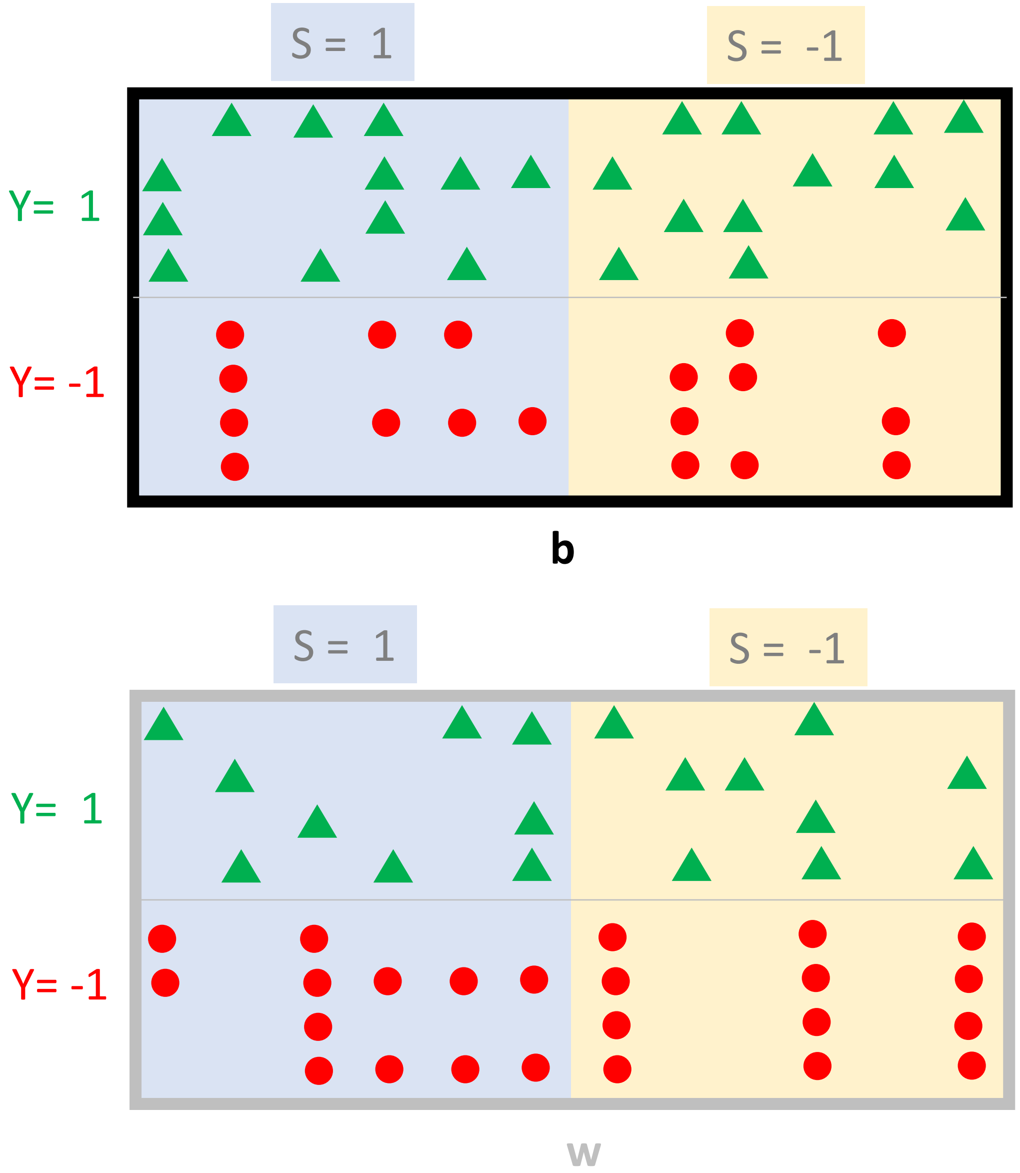
w



Calculate the rates

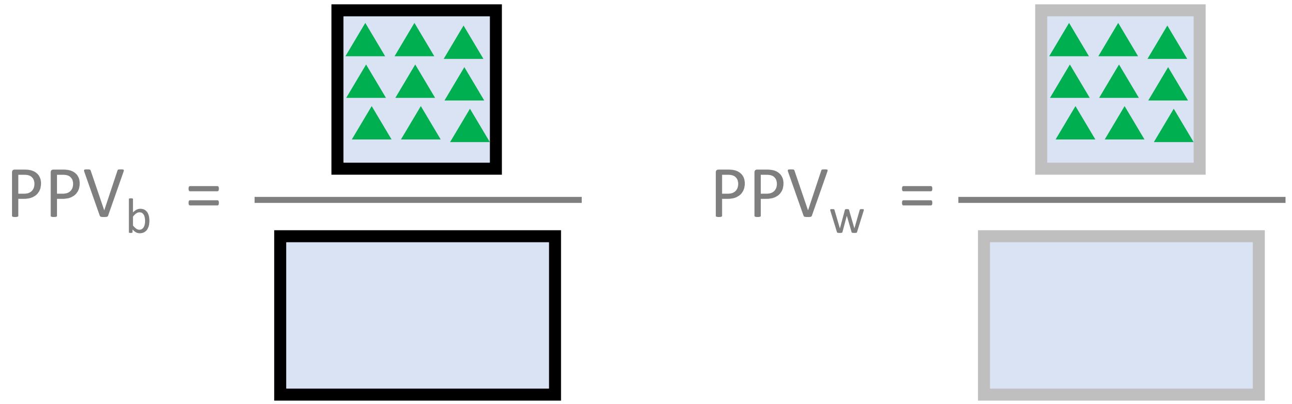
FPR and FNR for groups

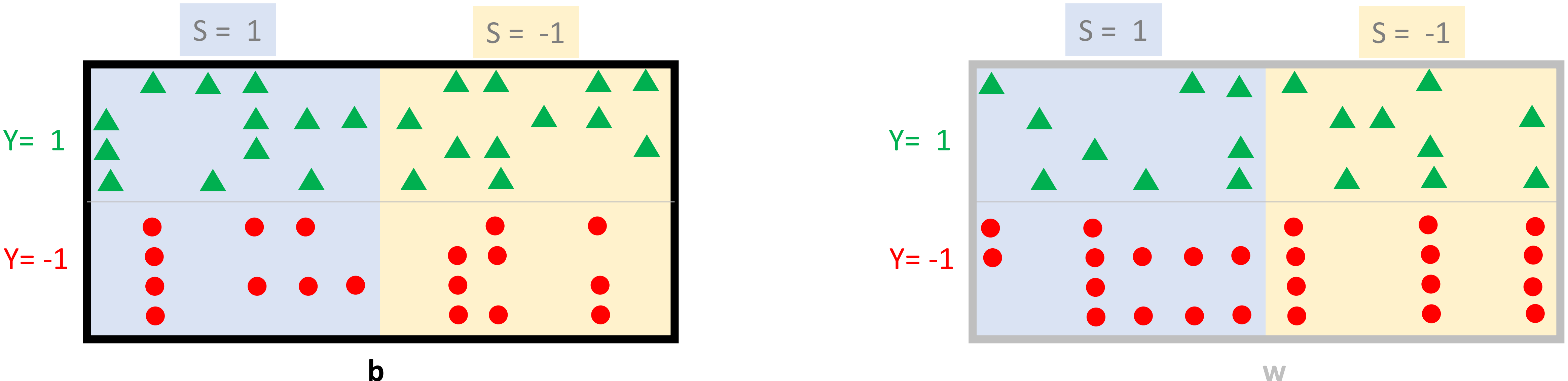




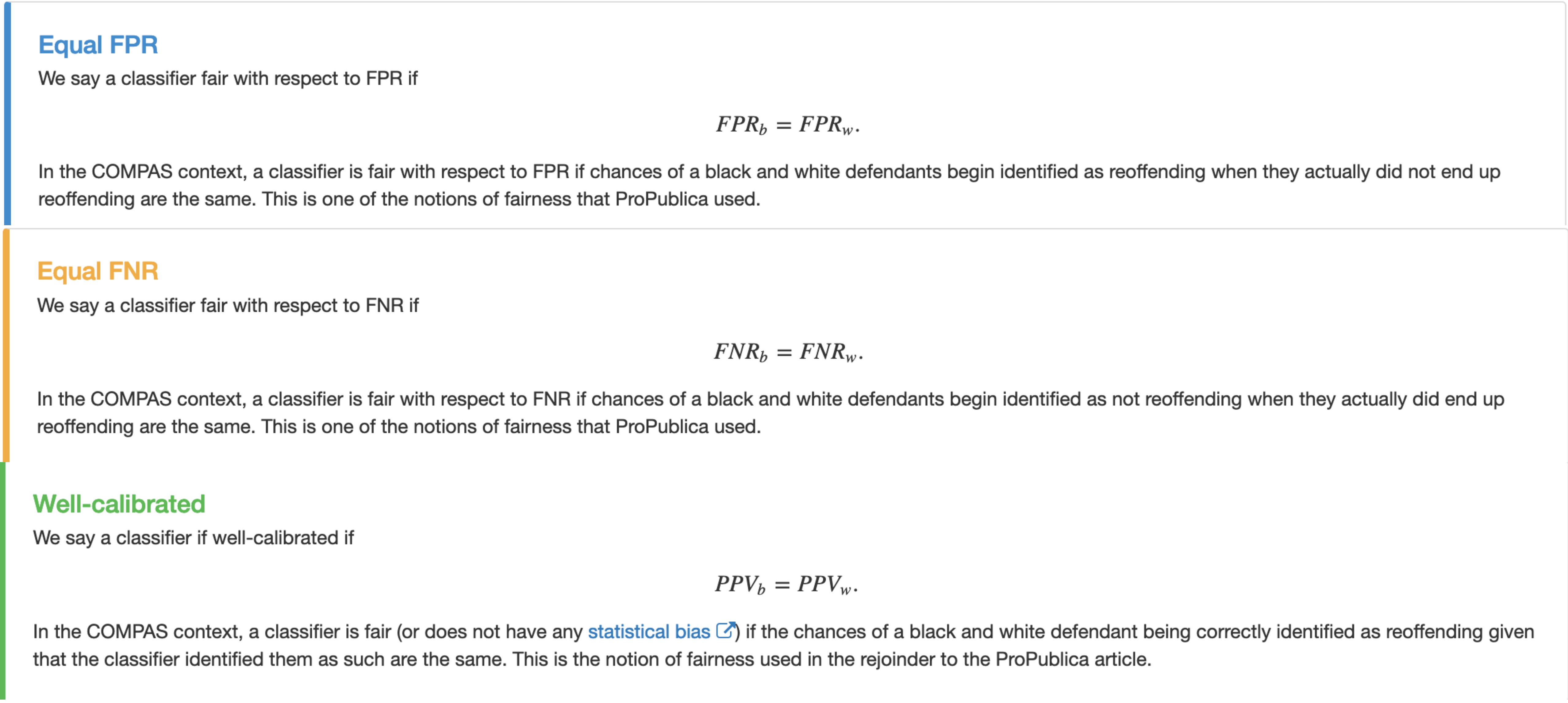
PPV for groups

Calculate the values





Finally, the formal fairness definitions



Equal FPR

We say a classifier fair with respect to FPR if

$$FPR_b = FPR_w.$$

In the COMPAS context, a classifier is fair with respect to FPR if chances of a black and white defendants begin identified as reoffending when they actually did not end up reoffending are the same. This is one of the notions of fairness that ProPublica used.

Equal FNR

We say a classifier fair with respect to FNR if

$$FNR_b = FNR_w.$$

In the COMPAS context, a classifier is fair with respect to FNR if chances of a black and white defendants begin identified as not reoffending when they actually did end up reoffending are the same. This is one of the notions of fairness that ProPublica used.

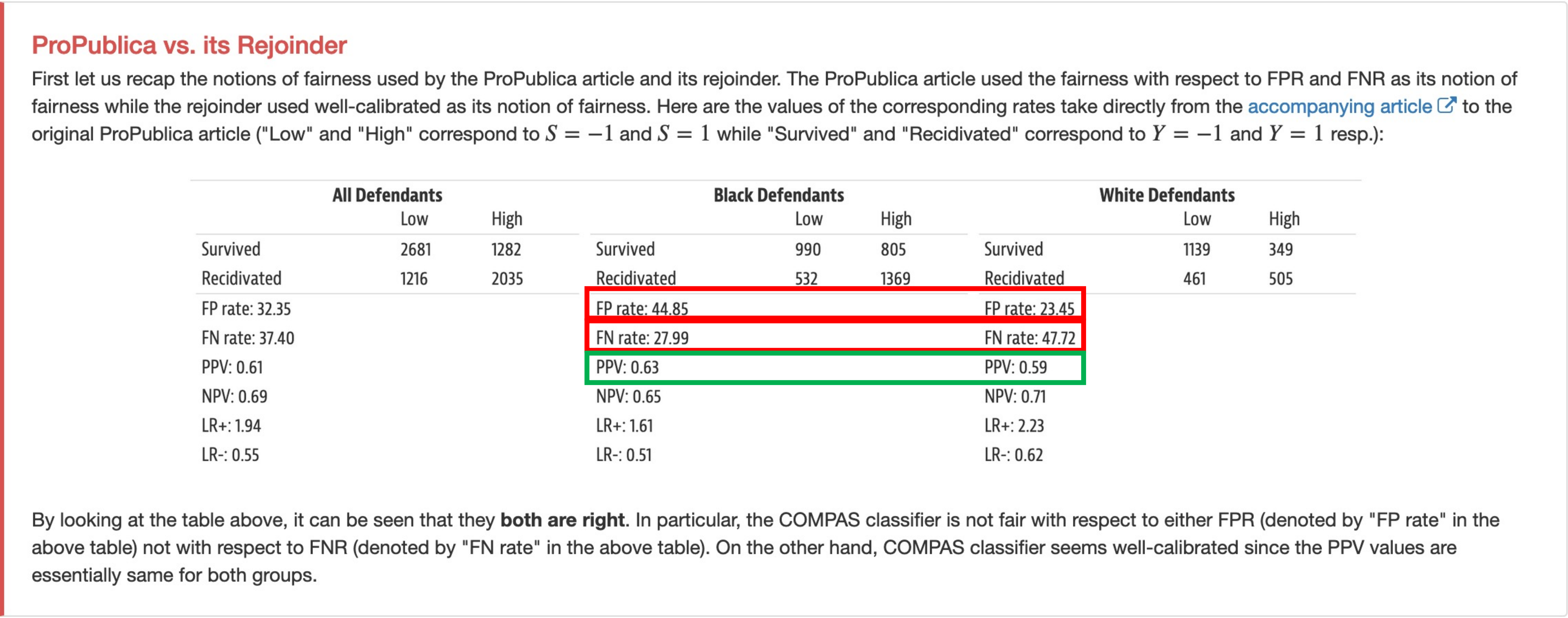
Well-calibrated

We say a classifier if well-calibrated if

$$PPV_b = PPV_w.$$

In the COMPAS context, a classifier is fair (or does not have any statistical bias) if the chances of a black and white defendant being correctly identified as reoffending given that the classifier identified them as such are the same. This is the notion of fairness used in the rejoinder to the ProPublica article.

Connecting back to COMPAS story



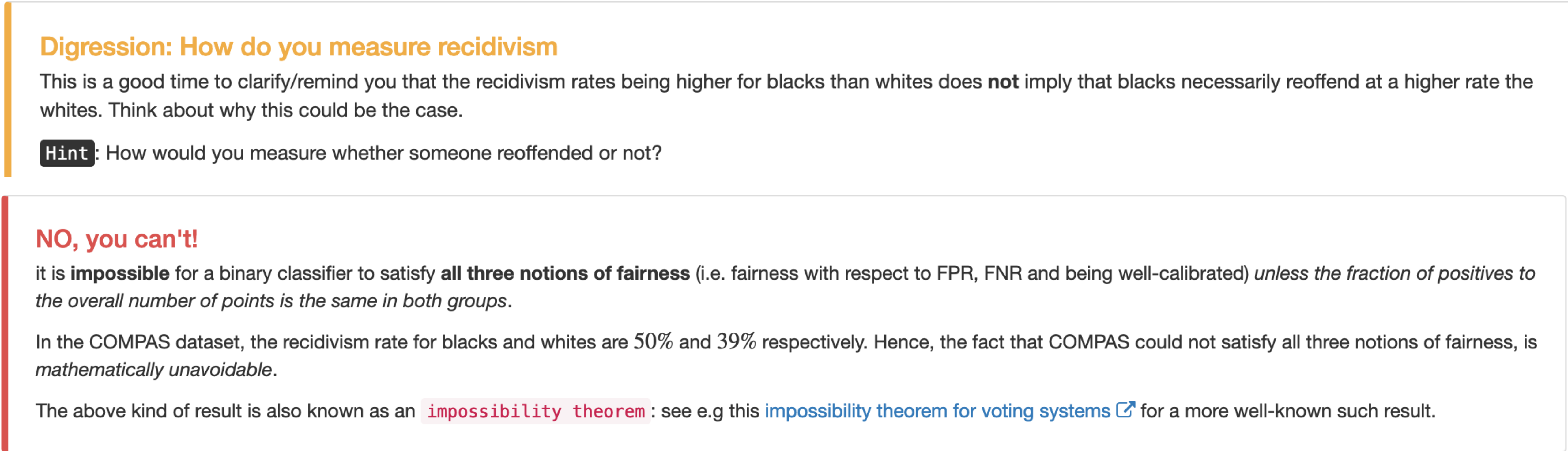
ProPublica vs. its Rejoinder

First let us recap the notions of fairness used by the ProPublica article and its rejoinder. The ProPublica article used the fairness with respect to FPR and FNR as its notion of fairness while the rejoinder used well-calibrated as its notion of fairness. Here are the values of the corresponding rates take directly from the accompanying article to the original ProPublica article ("Low" and "High" correspond to $S = -1$ and $S = 1$ while "Survived" and "Recidivated" correspond to $Y = -1$ and $Y = 1$ resp.):

All Defendants			Black Defendants			White Defendants		
	Low	High		Low	High		Low	High
Survived	2681	1282	Survived	990	805	Survived	1139	349
Recidivated	1216	2035	Recidivated	532	1369	Recidivated	461	505
FP rate: 32.35			FP rate: 44.85			FP rate: 23.45		
FN rate: 37.40			FN rate: 27.99			FN rate: 47.72		
PPV: 0.61			PPV: 0.63			PPV: 0.59		
NPV: 0.69			NPV: 0.65			NPV: 0.71		
LR+: 1.94			LR+: 1.61			LR+: 2.23		
LR-: 0.55			LR-: 0.51			LR-: 0.62		

By looking at the table above, it can be seen that they **both are right**. In particular, the COMPAS classifier is not fair with respect to either FPR (denoted by "FP rate" in the above table) not with respect to FNR (denoted by "FN rate" in the above table). On the other hand, COMPAS classifier seems well-calibrated since the PPV values are essentially same for both groups.

Perhaps COMPAS can be improved?



Digression: How do you measure recidivism

This is a good time to clarify/remind you that the recidivism rates being higher for blacks than whites does **not** imply that blacks necessarily reoffend at a higher rate the whites. Think about why this could be the case.

`Hint`: How would you measure whether someone reoffended or not?

NO, you can't!

it is **impossible** for a binary classifier to satisfy **all three notions of fairness** (i.e. fairness with respect to FPR, FNR and being well-calibrated) *unless the fraction of positives to the overall number of points is the same in both groups*.

In the COMPAS dataset, the recidivism rate for blacks and whites are 50% and 39% respectively. Hence, the fact that COMPAS could not satisfy all three notions of fairness, is *mathematically unavoidable*.

The above kind of result is also known as an `impossibility theorem`: see e.g this impossibility theorem for voting systems for a more well-known such result.

Discussion 3

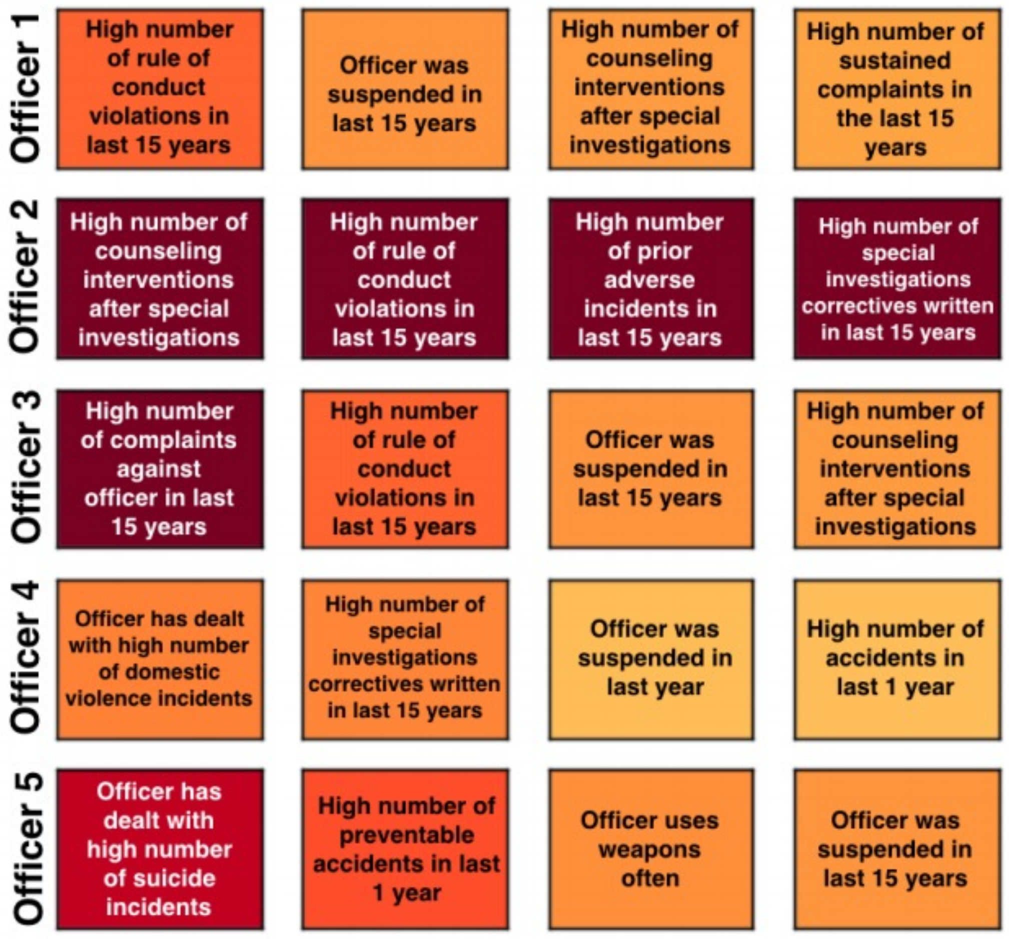
Kenneth (Kenny) Joseph

University at Buffalo
Department of Computer Science
and Engineering
School of Engineering and Applied Sciences

Kenny’s summary

- This seemed to be a difficult week for folks. We were happy to see that

What is "AI for Social Good"?



Officer 1	High number of rule of conduct violations in last 15 years	Officer was suspended in last 15 years	High number of counseling interventions after special investigations	High number of sustained complaints in the last 15 years
Officer 2	High number of counseling interventions after special investigations	High number of rule of conduct violations in last 15 years	High number of prior adverse incidents in last 15 years	High number of special investigations correctives written in last 15 years
Officer 3	High number of complaints against officer in last 15 years	High number of rule of conduct violations in last 15 years	Officer was suspended in last 15 years	High number of counseling interventions after special investigations
Officer 4	Officer has dealt with high number of domestic violence incidents	High number of special investigations correctives written in last 15 years	Officer was suspended in last year	High number of accidents in last 1 year
Officer 5	Officer has dealt with high number of suicide incidents	High number of preventable accidents in last 1 year	Officer uses weapons often	Officer was suspended in last 15 years

Figure 1: An illustration of five at-risk officers that will go on to have an adverse incident and their risk factors. The darker the red, the stronger the importance of that feature.

What is “AI for Social Good”?

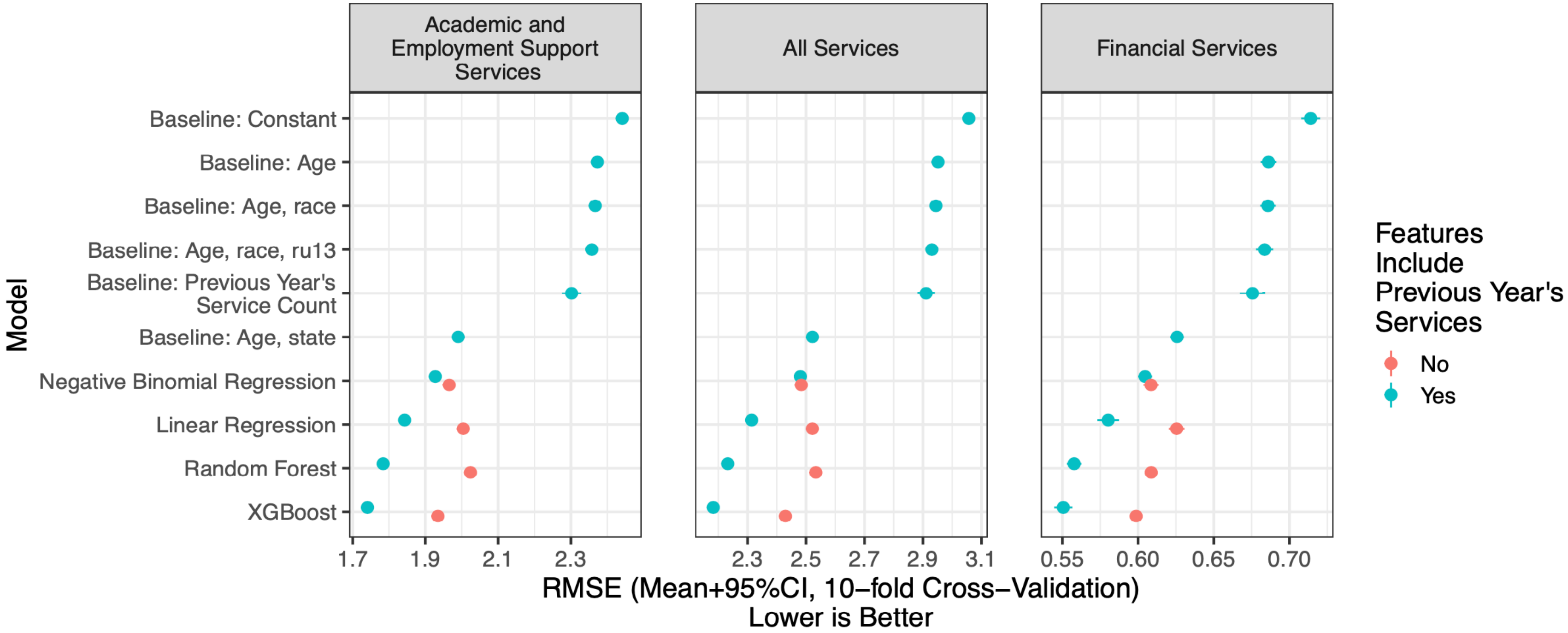
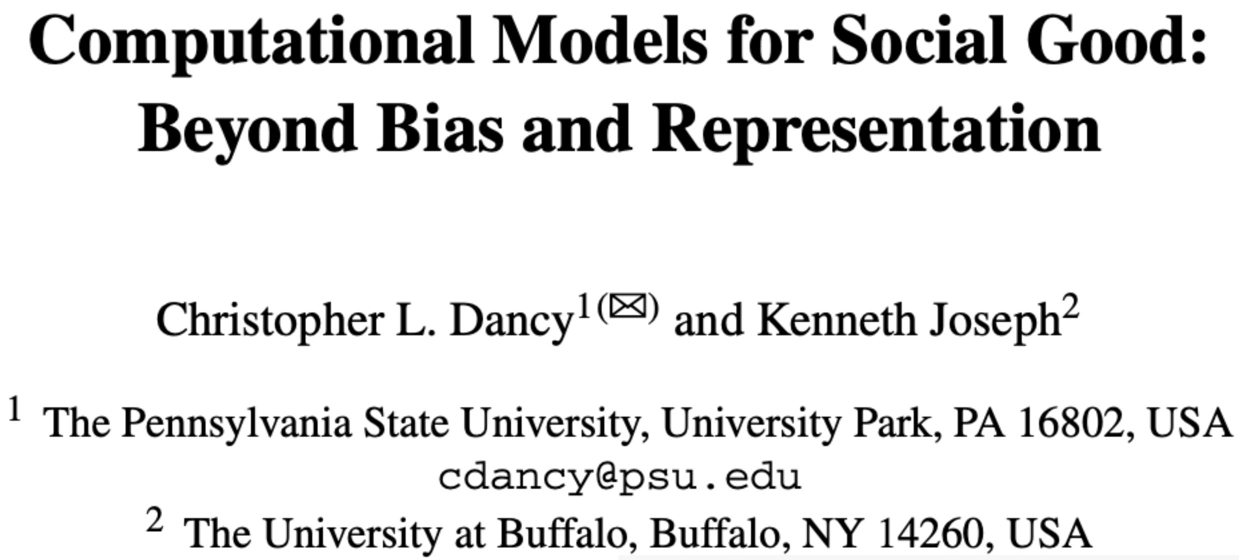


Figure 3: Results of our predictive experiment, using Root Mean Squared Error (RMSE; y-axis) as the outcome of interest. Each row represents a different prediction model, and each of the three sub-plots shows results for the three different dependent variables we analyzed, respectively.

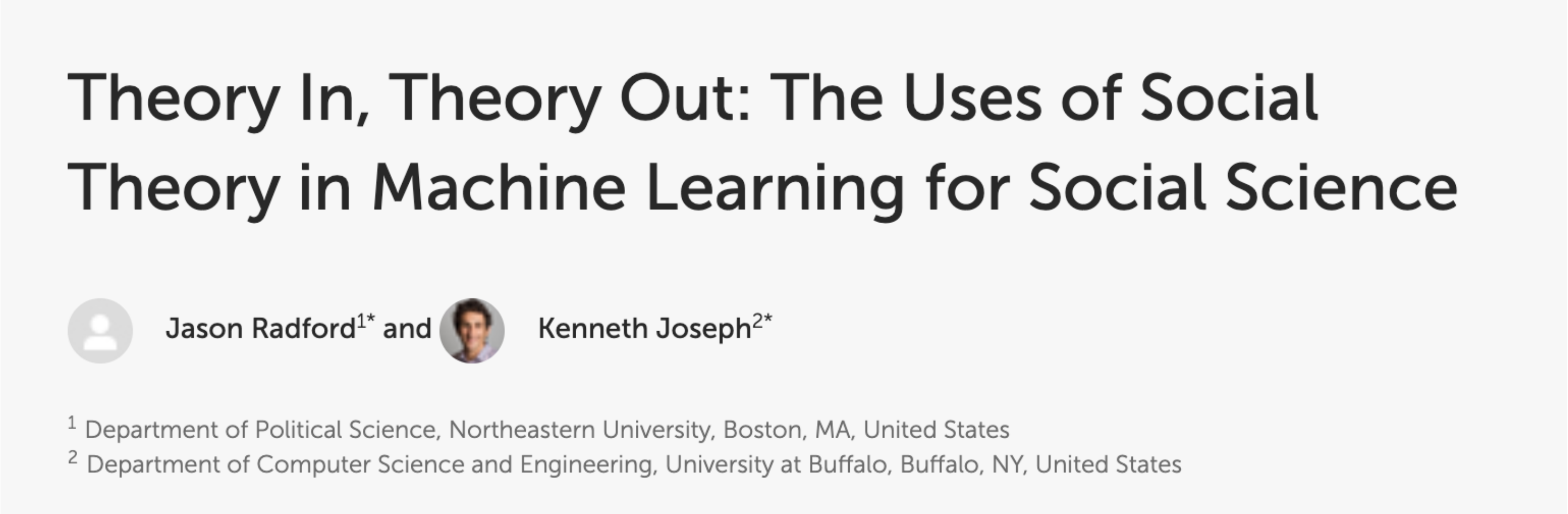
What isn’t ”AI for Social Good”



Computational Models for Social Good: Beyond Bias and Representation

Christopher L. Dancy[1](✉) and Kenneth Joseph[2]

[1] The Pennsylvania State University, University Park, PA 16802, USA
cdancy@psu.edu
[2] The University at Buffalo, Buffalo, NY 14260, USA



Theory In, Theory Out: The Uses of Social Theory in Machine Learning for Social Science

Jason Radford[1*] and Kenneth Joseph[2*]

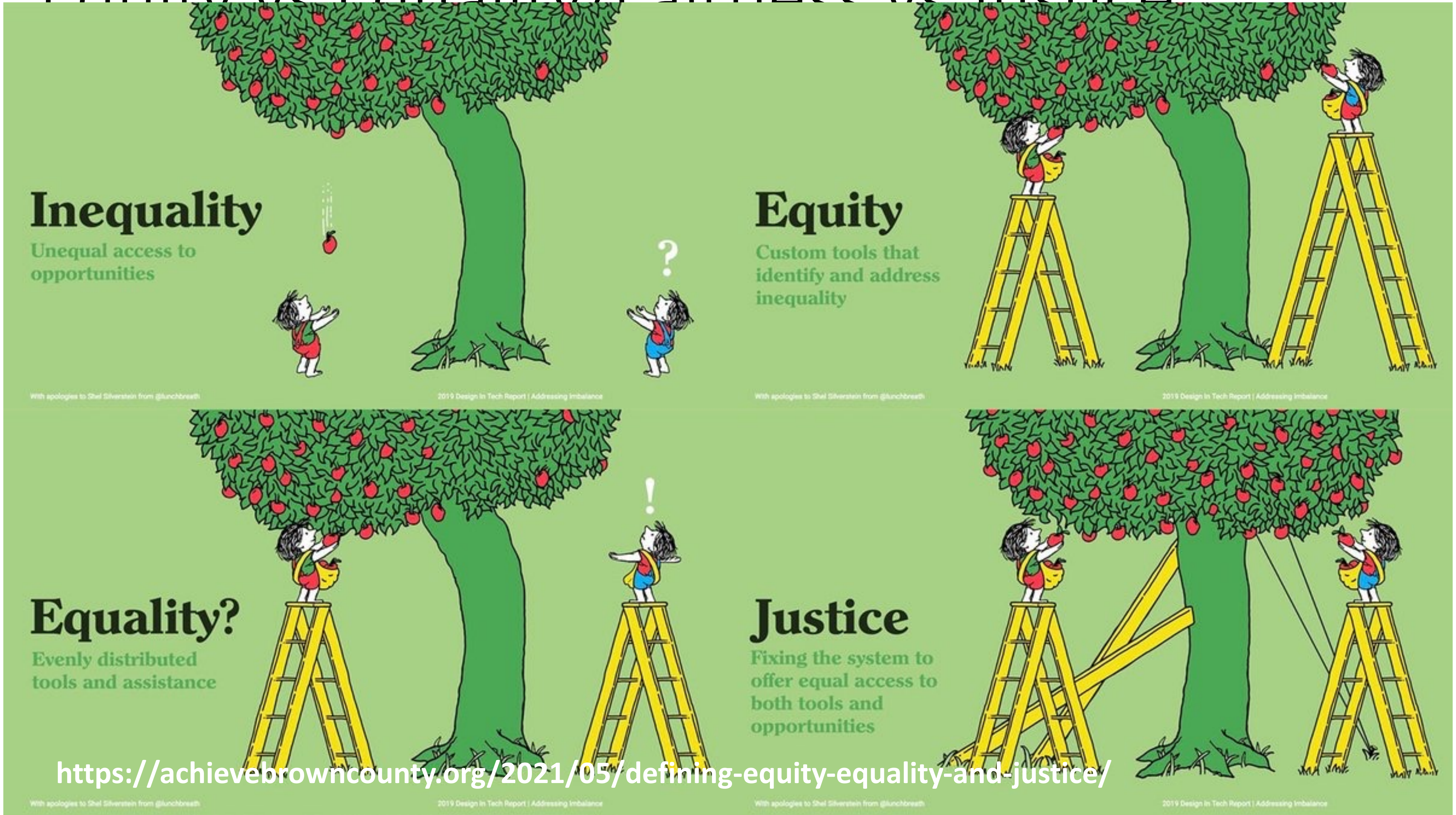
[1] Department of Political Science, Northeastern University, Boston, MA, United States
[2] Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, United States

What is redlining?

- In short, the process by which white families, governments, and private companies effectively “barred Black home buyers from qualifying for secure mortgages from many mainstream banks.” (https://www.nytimes.com/2021/08/17/realestate/what-is-redlining.html)

- This was followed by the tearing apart of Black communities established in redlined areas once established as white families wanted pathways back into the city (by creating highways that ran through the center of them)

Equity vs Equality/Fairness vs Justice


Inequality
Unequal access to opportunities
?
With apologies to Shel Silverstein from @lunchbreath
2019 Design In Tech Report | Addressing Imbalance
Equity
Custom tools that identify and address inequality
With apologies to Shel Silverstein from @lunchbreath
2019 Design In Tech Report | Addressing Imbalance
Equality?
Evenly distributed tools and assistance
!
With apologies to Shel Silverstein from @lunchbreath
2019 Design In Tech Report | Addressing Imbalance
Justice
Fixing the system to offer equal access to both tools and opportunities
With apologies to Shel Silverstein from @lunchbreath
2019 Design In Tech Report | Addressing Imbalance

https://achievebrowncounty.org/2021/05/defining-equity-equality-and-justice/

Round 1

- Were the orphan trains really that bad? (Several folks)

Round 2

- Is it feasible to do tech with mostly non-tech people? (Gopi)
- Why isn't this always the case? (Herman)

Round 3

- Why can't we just live together peacefully and co-exist happily? (Hitesh)
- What chance do we have for Zuckerberg and Dr. Rodriguez to ever co-exist happily? (Chaithanya)

Round 4

- What is the limit of the willingness of the advantaged group to change? (Alex)
- What do we have to sacrifice for equity? For equality?
- What are **you willing** to sacrifice for equity?

Other things

- What percent of people need to benefit before we start building? (Aishwarya)
- What is justice? (Alex)
- What is justice, vs. fairness, and which is preferrable? (Joe)
- Who gets to access the "minor privleges of AI" e.g. robot vaccums, not picking up dog poop (Dhiraj)?
- I'm not a [Black American]. What can I really do?
- Will AI cost jobs? (Several)
- What is AI if we are not simply training on past data?
- "To know something fully, you need to experience it" (Hitesh)... what does this mean for us?