Week 9 A03

Reading Discussion

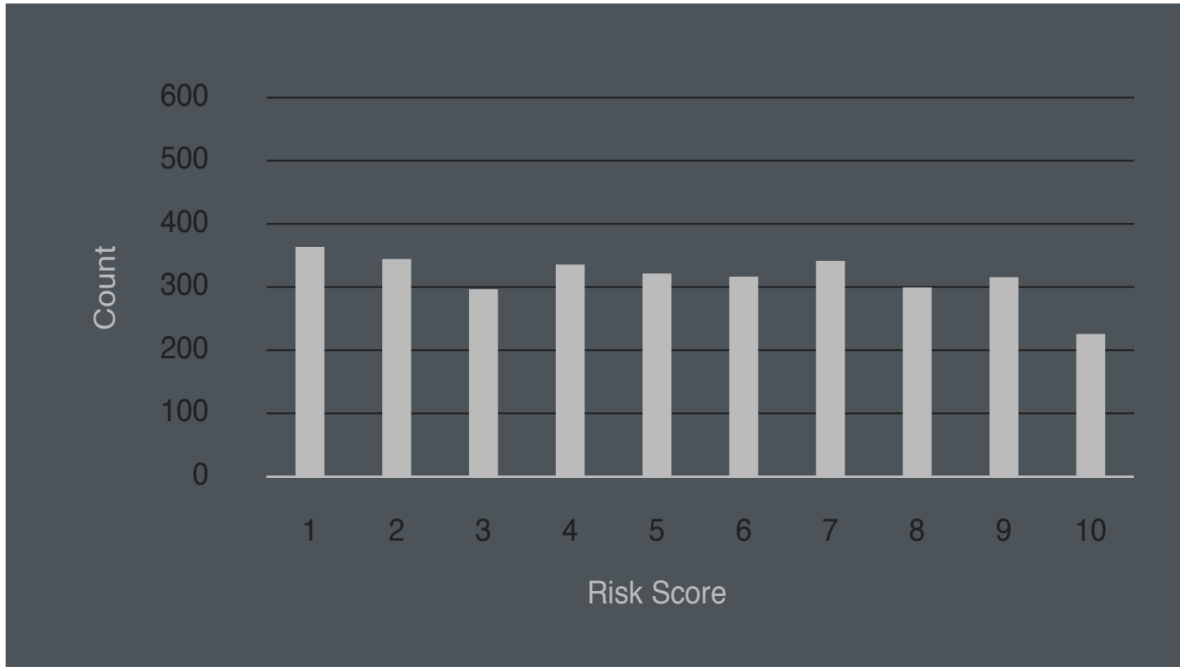
Machine Bias

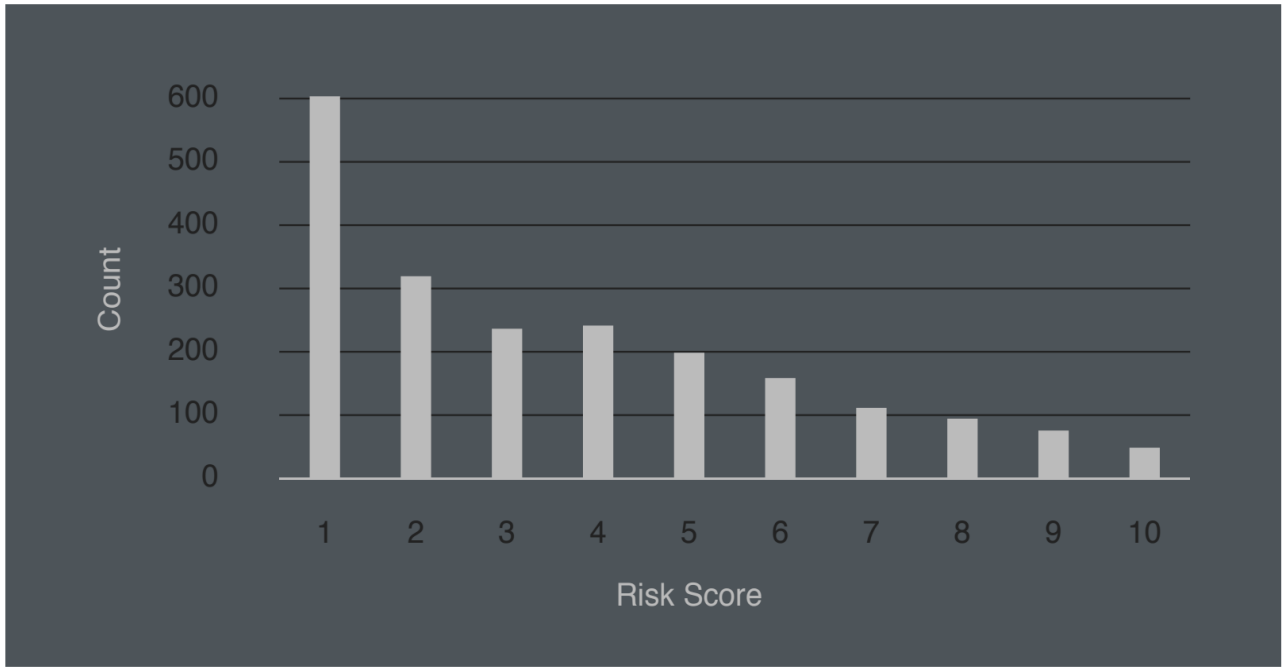
by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

What is machine bias?

1) Machine bias is the effect of an incorrect assumption in a machine learning (ML) model that's caused by overestimating or underestimating the importance of a particular parameter.
2) In the article, risk assessment algorithms have been found to be biased against black defendants, leading to harsher sentences.

Evidence of machine bias





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.)***

Why machine bias exists?

1) Because human perceptions cannot be free from bias, and the source of the data is derived from the objective human bias.

2) Because, in the article, risk assessment algorithms are often trained on historical data that reflects existing racial disparities in the criminal justice system (page 2).

The limitations of the risk assessments

1) Rely on historical data that may not accurately reflect an individual's current circumstances or potential for rehabilitation.

2) Defendants rarely have an opportunity to challenge their assessments.

3) The calculations used to generate risk scores are often not transparent.

How to address machine bias in criminal justice?

1) Improving the transparency and accountability of risk assessment algorithms by requiring that they be subject to independent audits and evaluations.
2) Ensuring the data used to train these algorithms is representative and free from bias.
3) Taking into account a broader range of factors beyond just criminal history, e.g., more representative data.