

---

# Defining and Designing Fair Algorithms

Sharad Goel  
Stanford University

---

---

# Risk assessment tools

Jurisdictions nationwide now use statistical algorithms to assess the likelihood a defendant will fail to appear at trial or commit a future violent crime.

---



*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

---

# The data

ProPublica analyzed 3,000 white and black defendants assigned COMPAS scores in Broward County, Florida.

[ Also determined whether these defendants recidivated. ]

---

---

**31% vs. 15%**

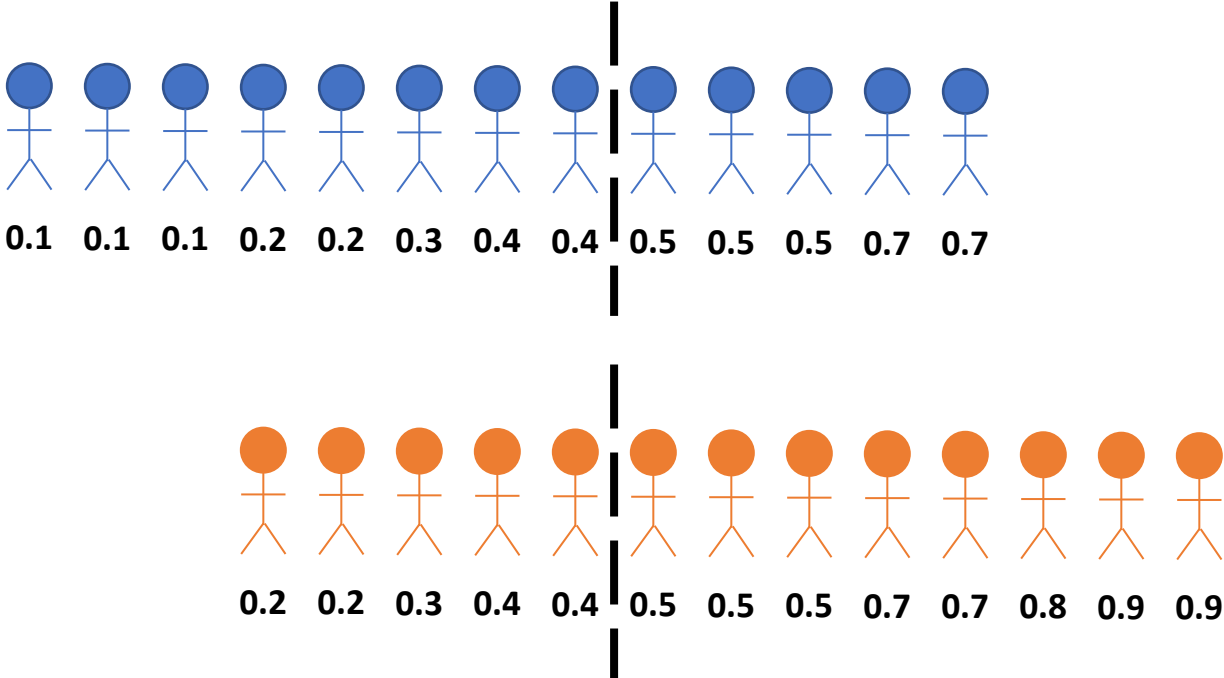
of black defendants  
who did not reoffend

of white defendants  
who did not reoffend

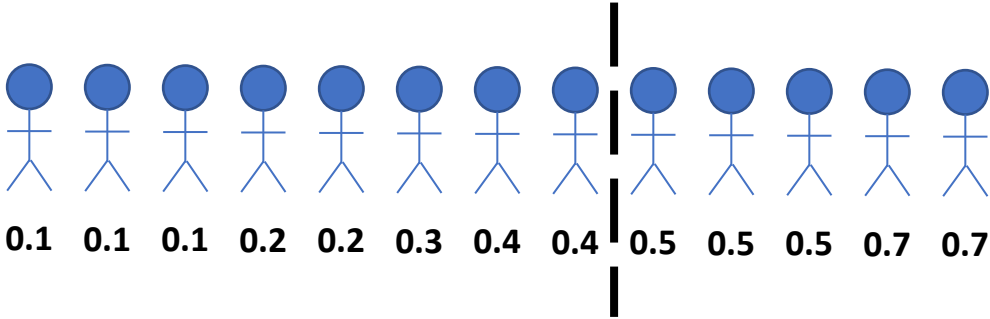
were deemed **high risk** of committing a violent crime.  
[ Higher false positive rates for black defendants. ]

---

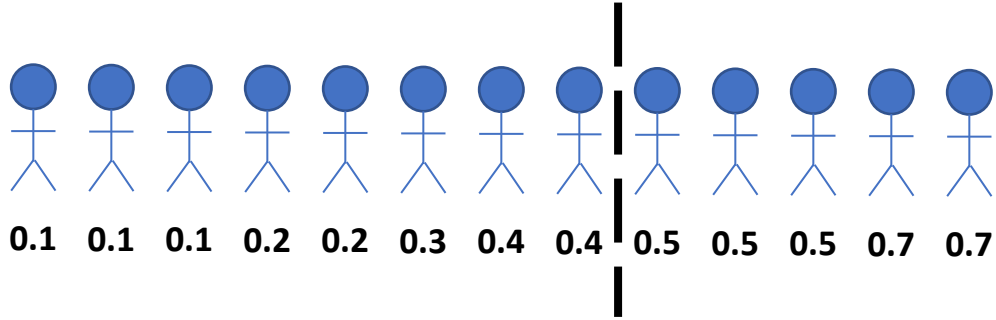
# The problem with false positive rates



# The problem with false positive rates



# The problem with false positive rates



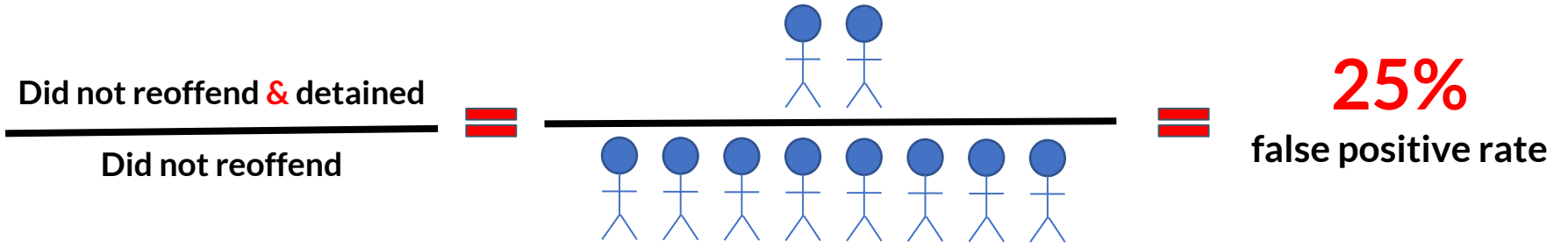
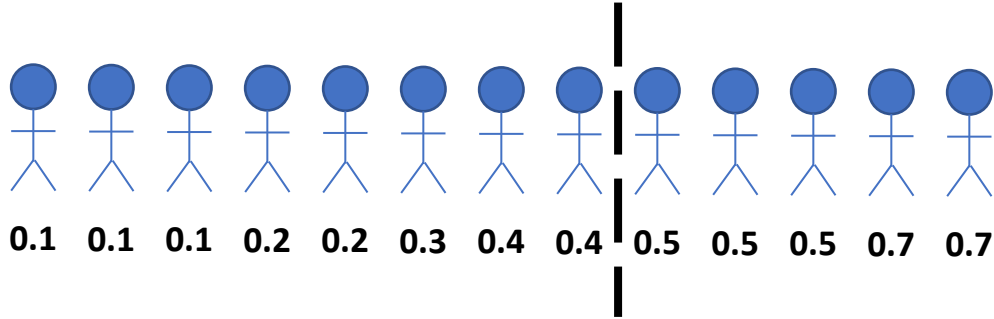
Did not reoffend & detained



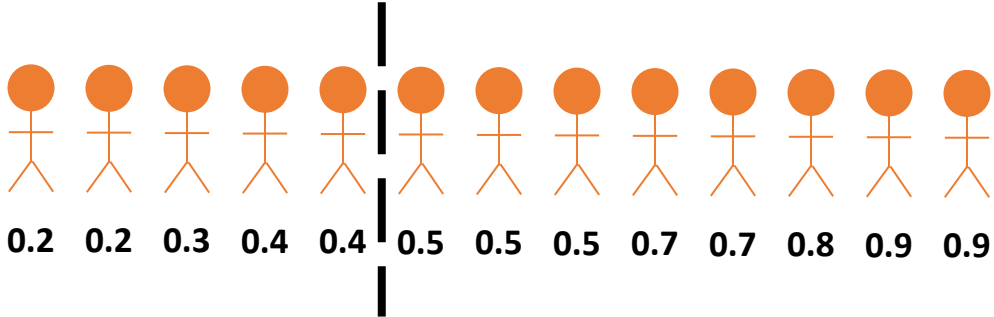
Did not reoffend



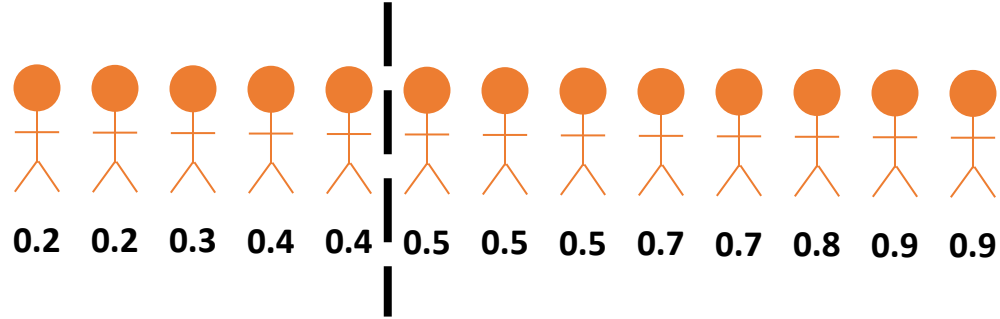
# The problem with false positive rates



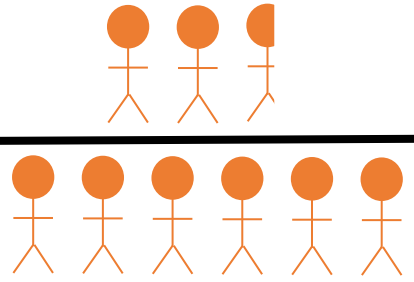
# The problem with false positive rates



# The problem with false positive rates

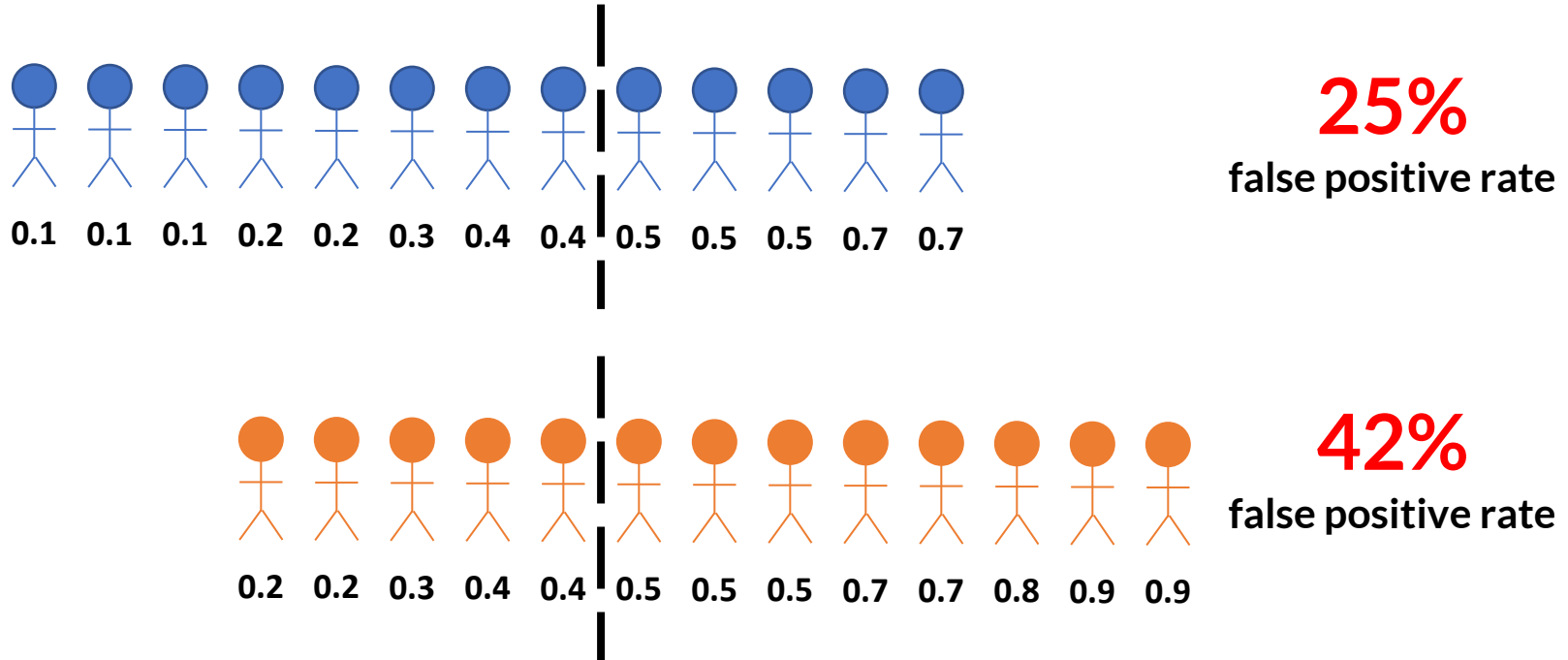


Did not reoffend & detained  
—————  
Did not reoffend

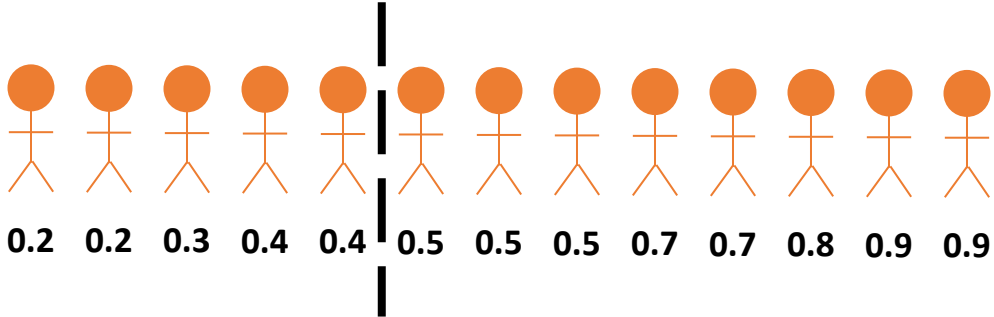


**42%**  
false positive rate

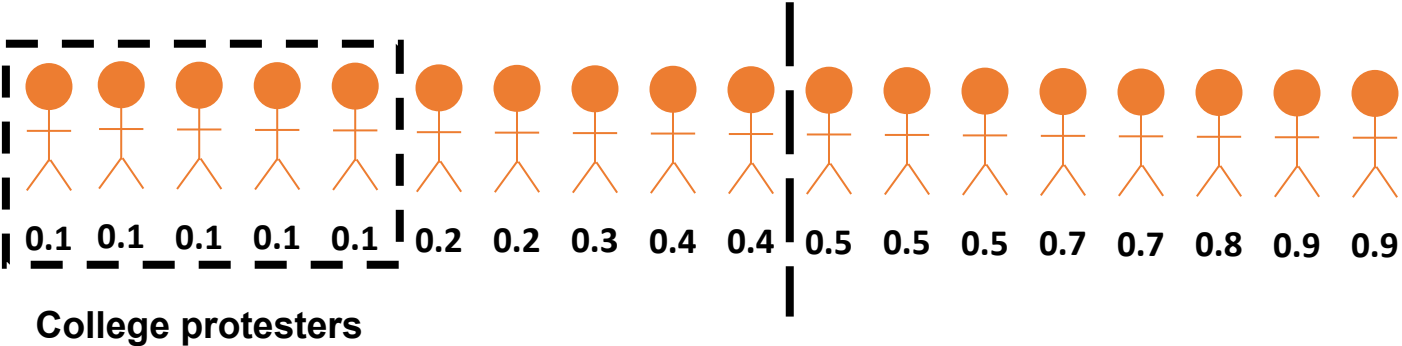
# The problem with false positive rates



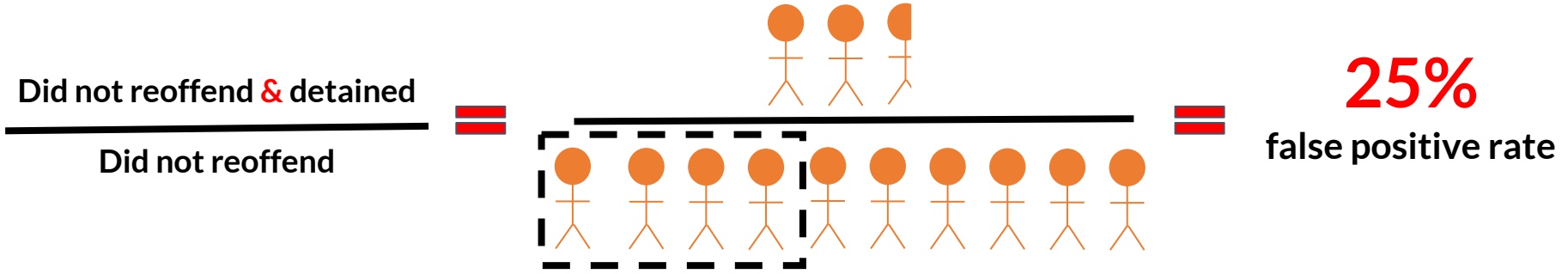
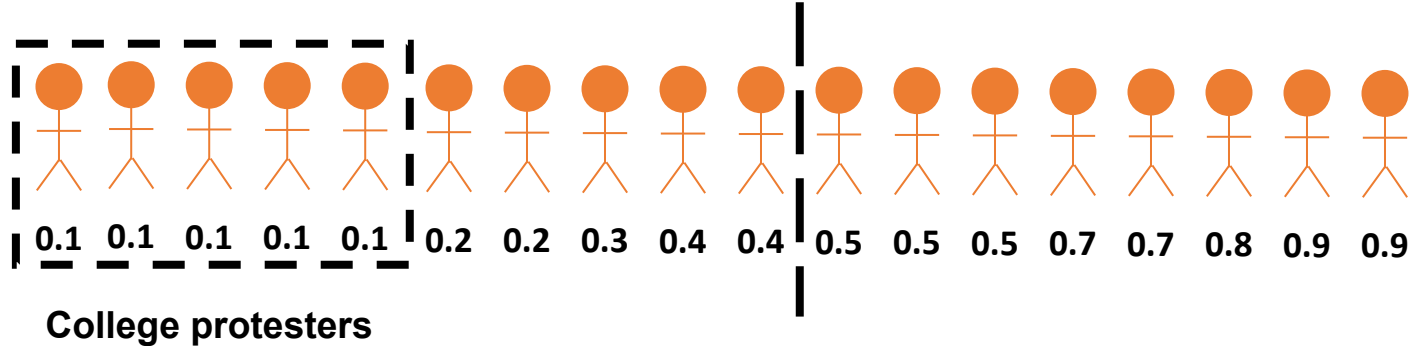
# The problem with false positive rates



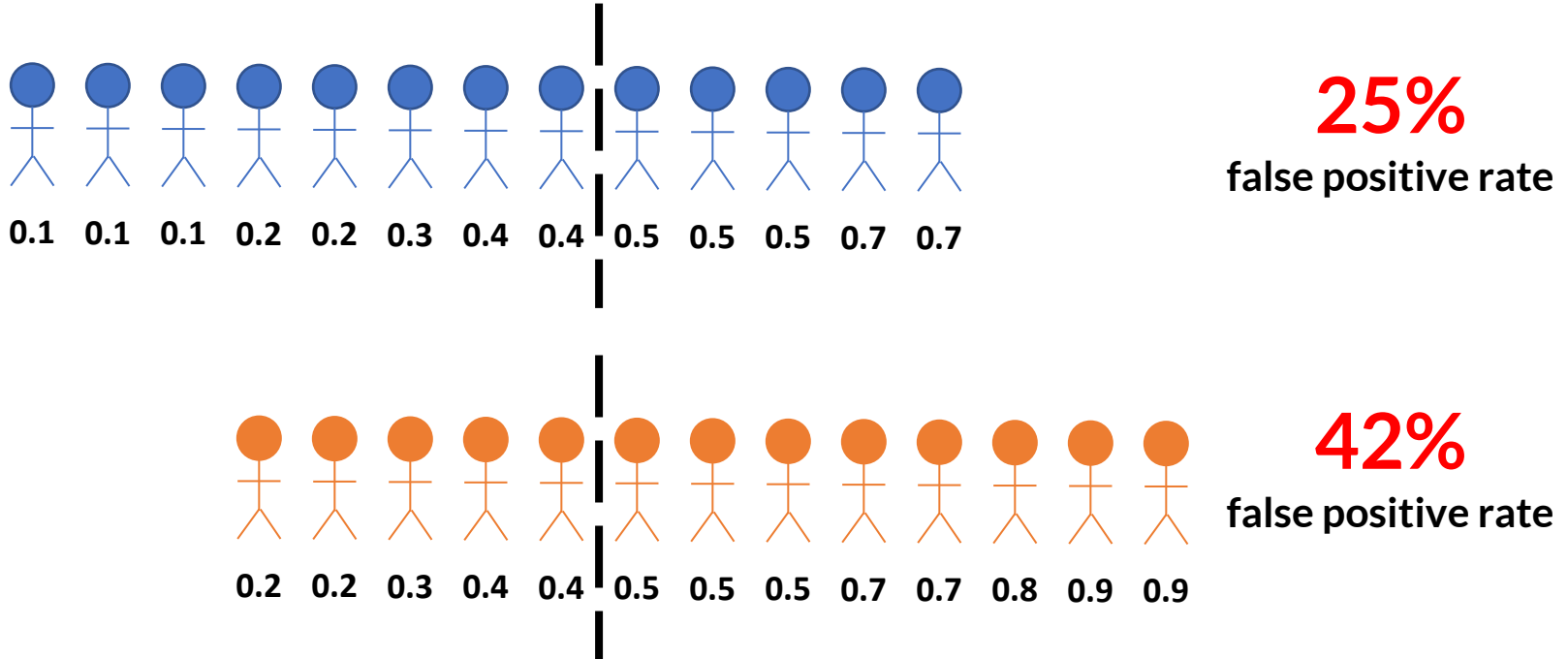
# The problem with false positive rates



# The problem with false positive rates

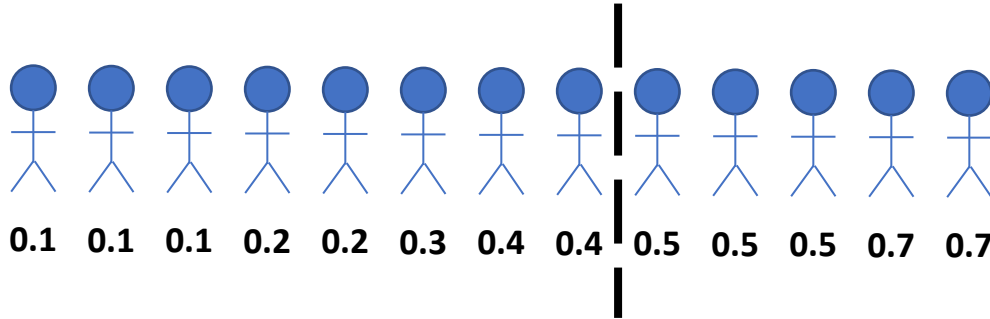


# The problem with false positive rates

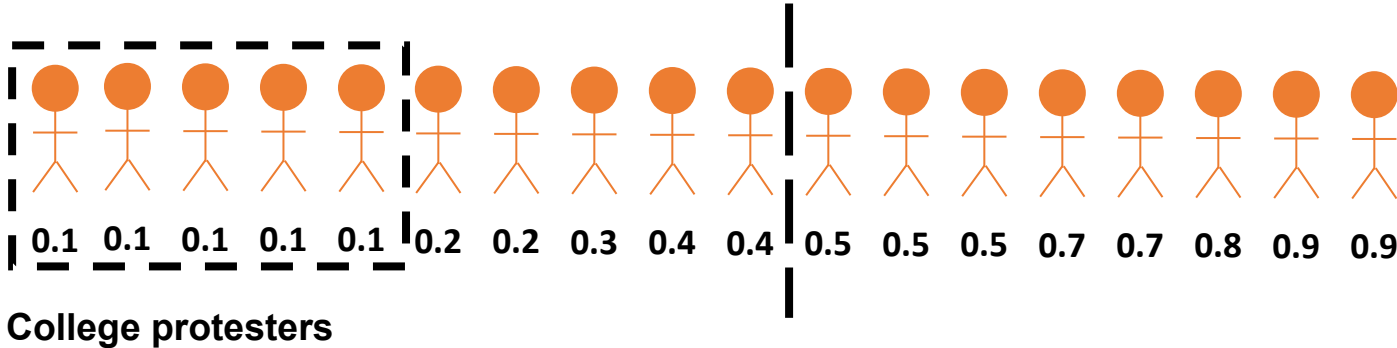




# The problem with false positive rates



**25%**  
false positive rate



**25%**  
false positive rate

---

**Are the data *biased*?**

---

---

# Biased labels

## [ Measurement error ]

St. George's Hospital in the UK developed an algorithm to sort medical school applicants. Algorithm trained to mimic past admissions decisions made by humans.

---

---

# Biased labels

## [ Measurement error ]

St. George's Hospital in the UK developed an algorithm to sort medical school applicants. Algorithm trained to mimic past admissions decisions made by humans.

But past decisions were biased against women and minorities.  
[ The algorithm codified discrimination. ]

---

---

# Biased labels

## [ Measurement error ]

Pretrial algorithms estimate the probability a defendant will be *observed / reported* committing a future crime.

Since reported crime is only a proxy for actual crime, estimates might be biased.

---

---

# What can [ should ] we do?

Algorithms are often good at synthesizing information, but we must still set effective and equitable policy.

[ Limiting money bail, providing pretrial services ]

---