Ethics and Data Science

Katie Shelef, PhD.
Co-instructor for Ethics in Data Science, Fall 2017

Introduction to Data Science
Spring 2018

THE UNIVERSITY OF UTAH

It is fair to use machines to judge humans

Strongly Agree — Neutral or Unsure — Strongly Disagree

Hiring

College admission

Loan

Data scientists have no responsibility for the outcome of their analyses

Strongly Agree

Neutral or Unsure

Strongly Disagree

Hiring

College admission

Loan

THE UNIVERSITY OF UTAH

Data scientists have no responsibility for the outcome of their analyses

"Even if AI is evil, most developers don't think it's the fault of the programmers. Fifty-eight percent say that ethics are the responsibility of upper management, 23 percent the inventor of the unethical idea, and just 20 percent think that they're the responsibility of the developer who actually wrote the code."

- Peter Bright, ars Technica, 3/13/18

https://arstechnica.com/gadgets/2018/03/developers-love-trendy-new-languages-but-earn-more-with-functional-programming/

Algorithms can't be unfair!

"Data is truth"

"Math is not racist"

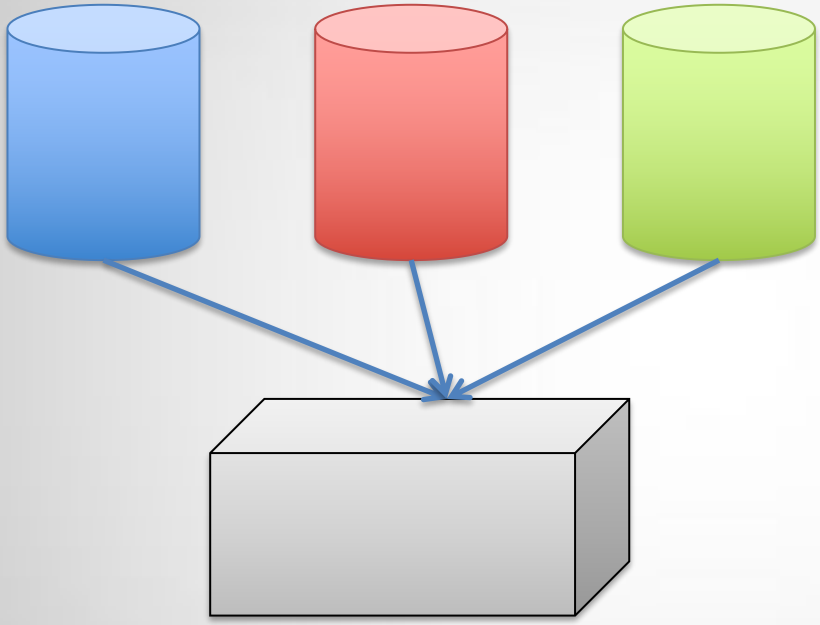
"This is not a CS/Data Science problem"

The age of automated decision-making

- Dealing with scale

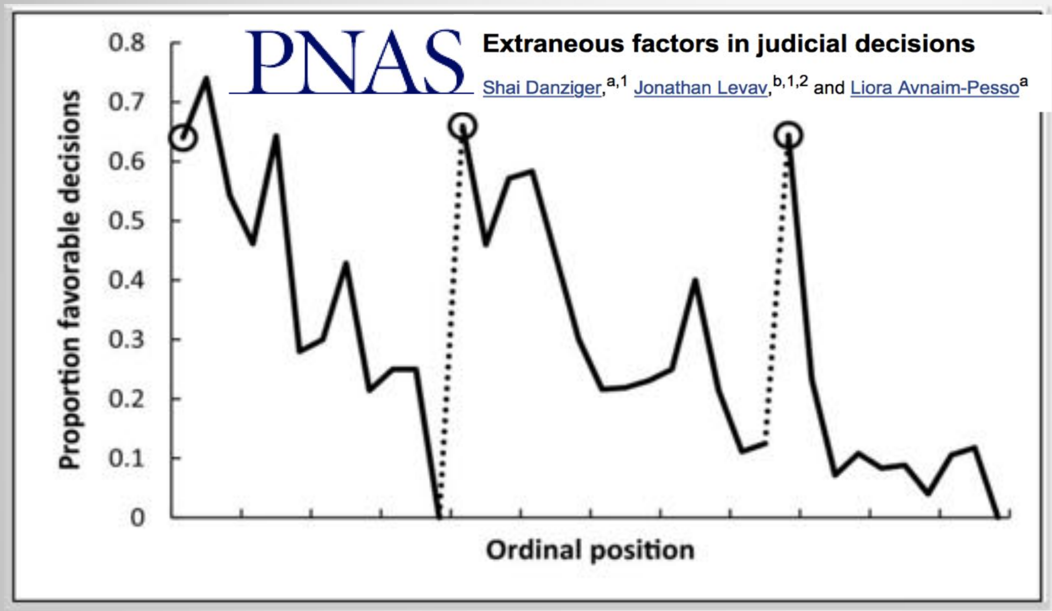
THE UNIVERSITY OF UTAH

The age of automated decision-making



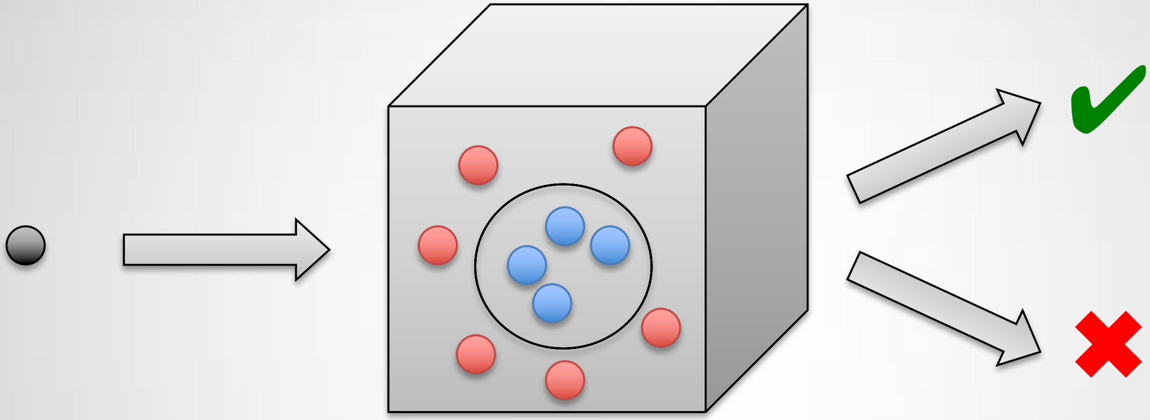
- Dealing with scale
- Dealing with complexity

The age of automated decision-making



- Dealing with scale
- Dealing with complexity
- Dealing with objectivity

Where does bias come from?



Data Sources

Algorithm design

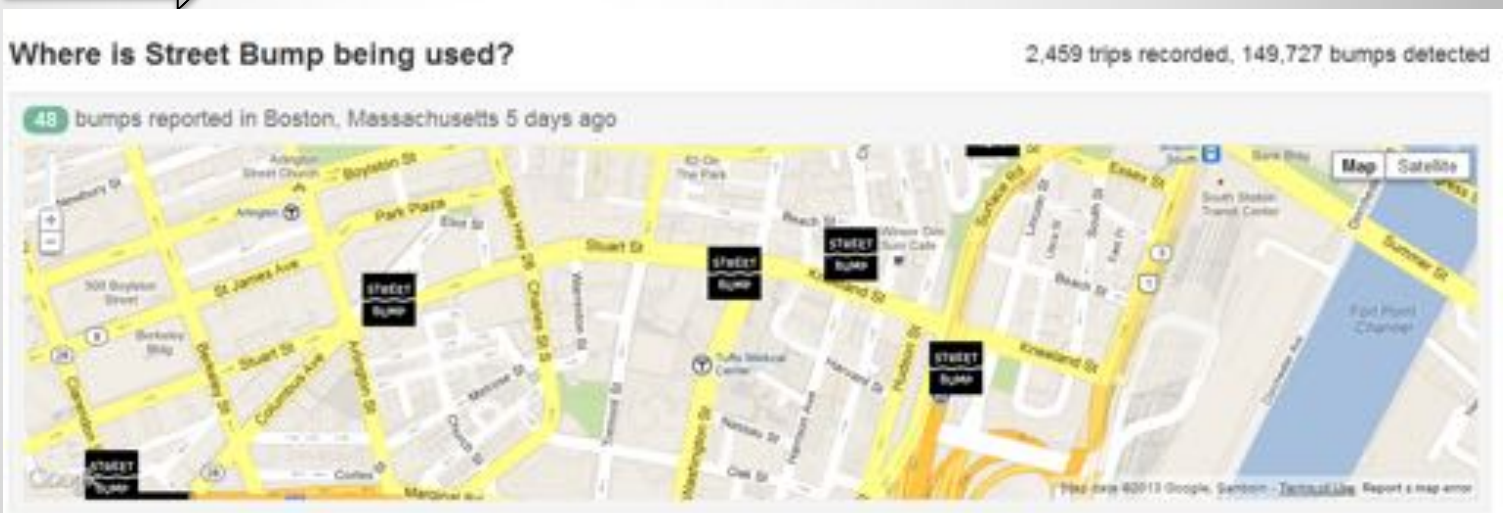
Feedback & Prediction

Where does bias come from?

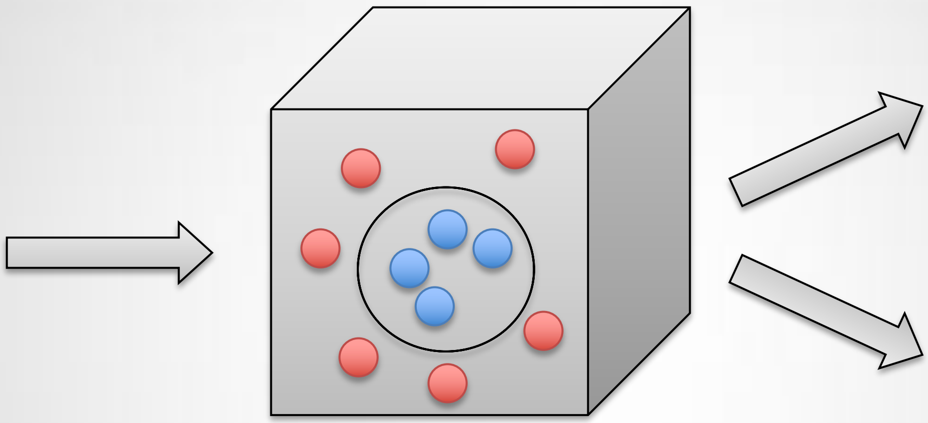
Many Cars Tone Deaf To Women's Voices

Female voices pose a bigger challenge for voice-activated technology than men's voices

Data Sources



Where does bias come from?



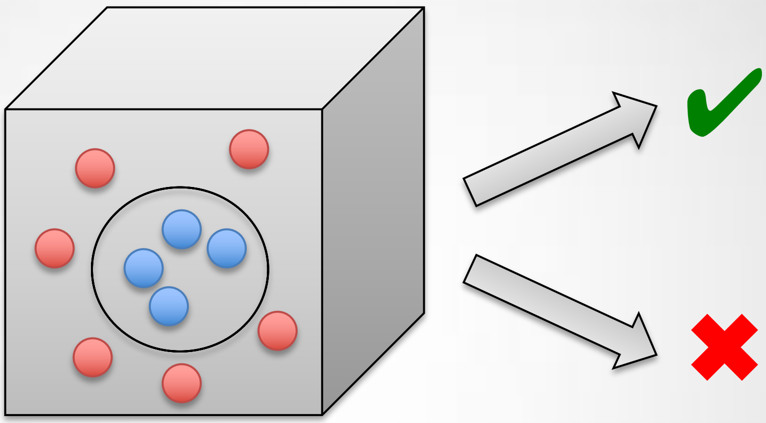
Algorithm
design

Where does bias come from?

Discrimination in Online Ad Delivery

Latanya Sweeney
Harvard University

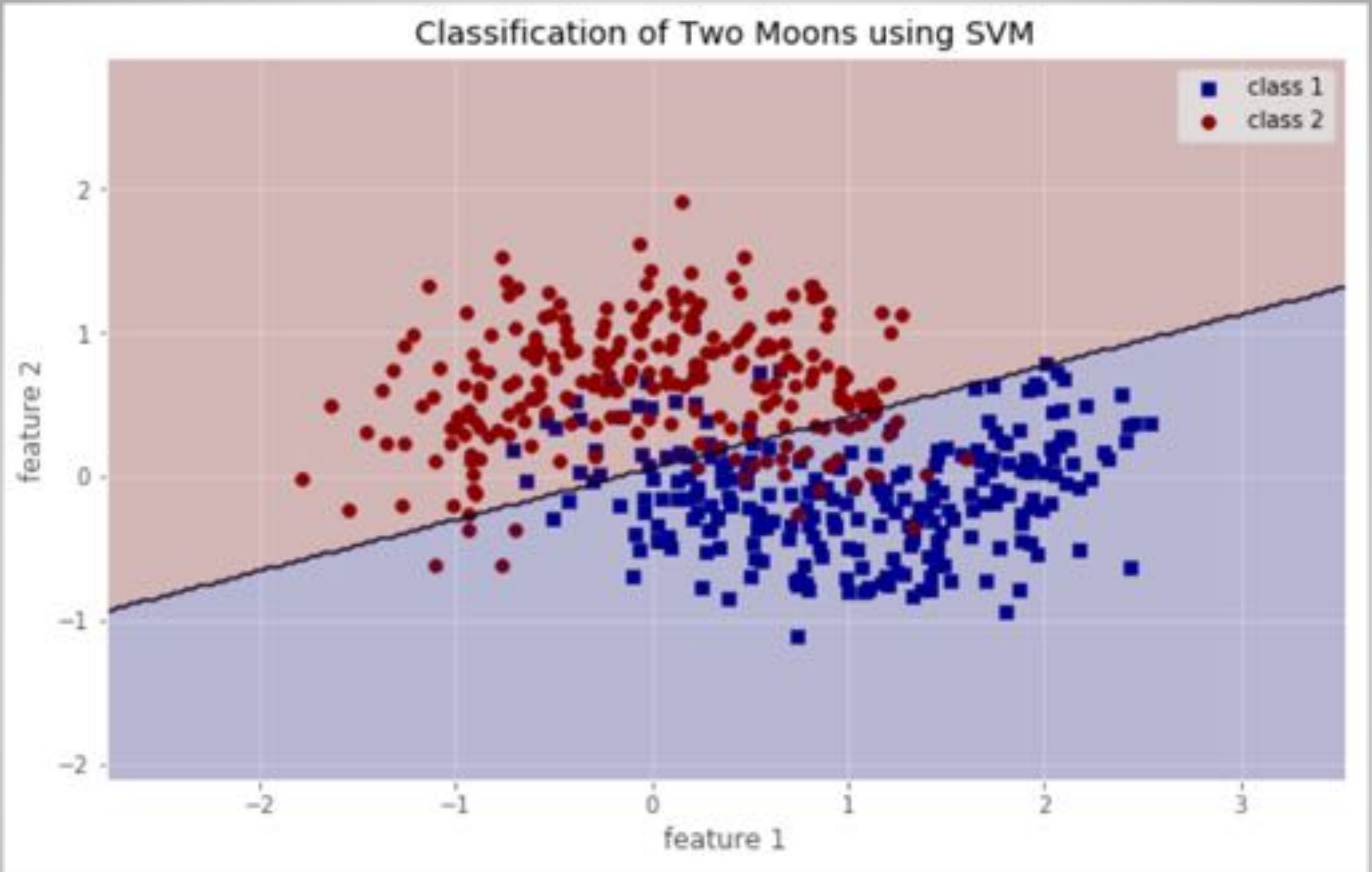
Where does bias come from?



(Exclusive) Crime-prediction tool PredPol amplifies racially biased policing, study shows

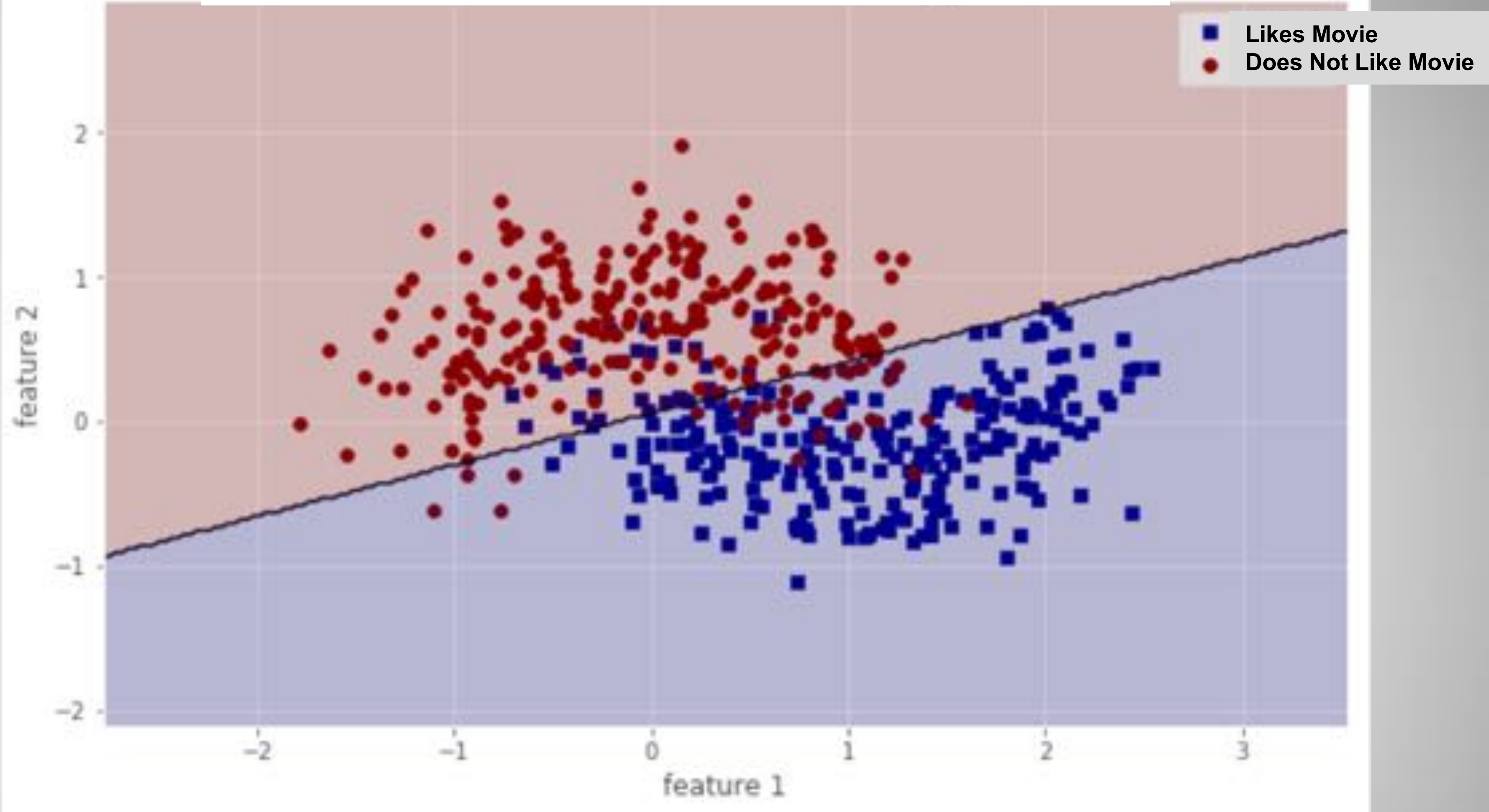
Feedback
&
Prediction

THE UNIVERSITY OF UTAH



	Predicted Class 1	Predicted Class 2
Actual Class 1	213	37
Actual Class 2	33	217

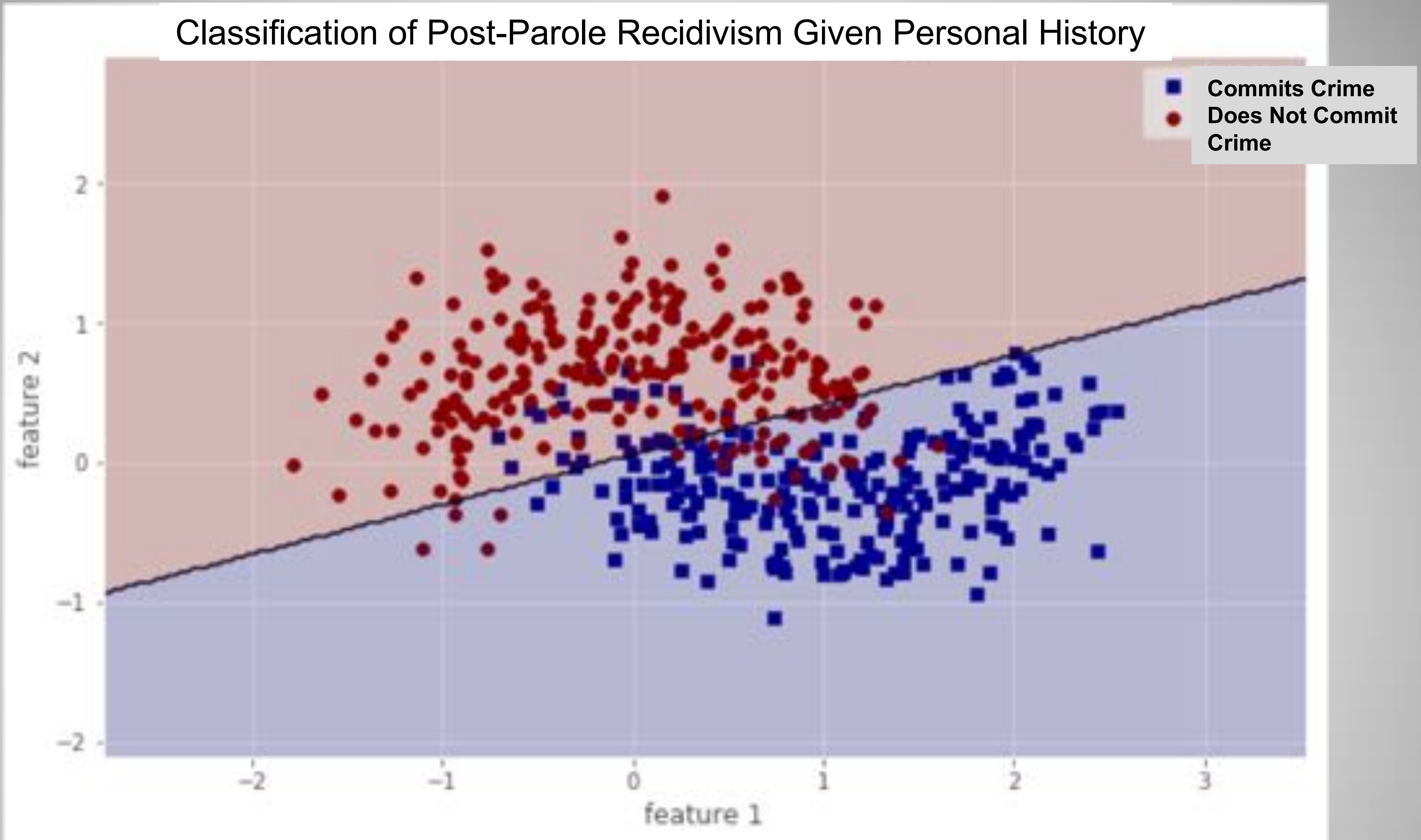
Classification of Movie Preference Given Past Preferences



	Predicted Likes Movie	Predicted Does Not Like
Actual Likes Movie	213	37
Actual Does Not Like	33	217

False positive: Annoyed customers?

False negative: Lost business



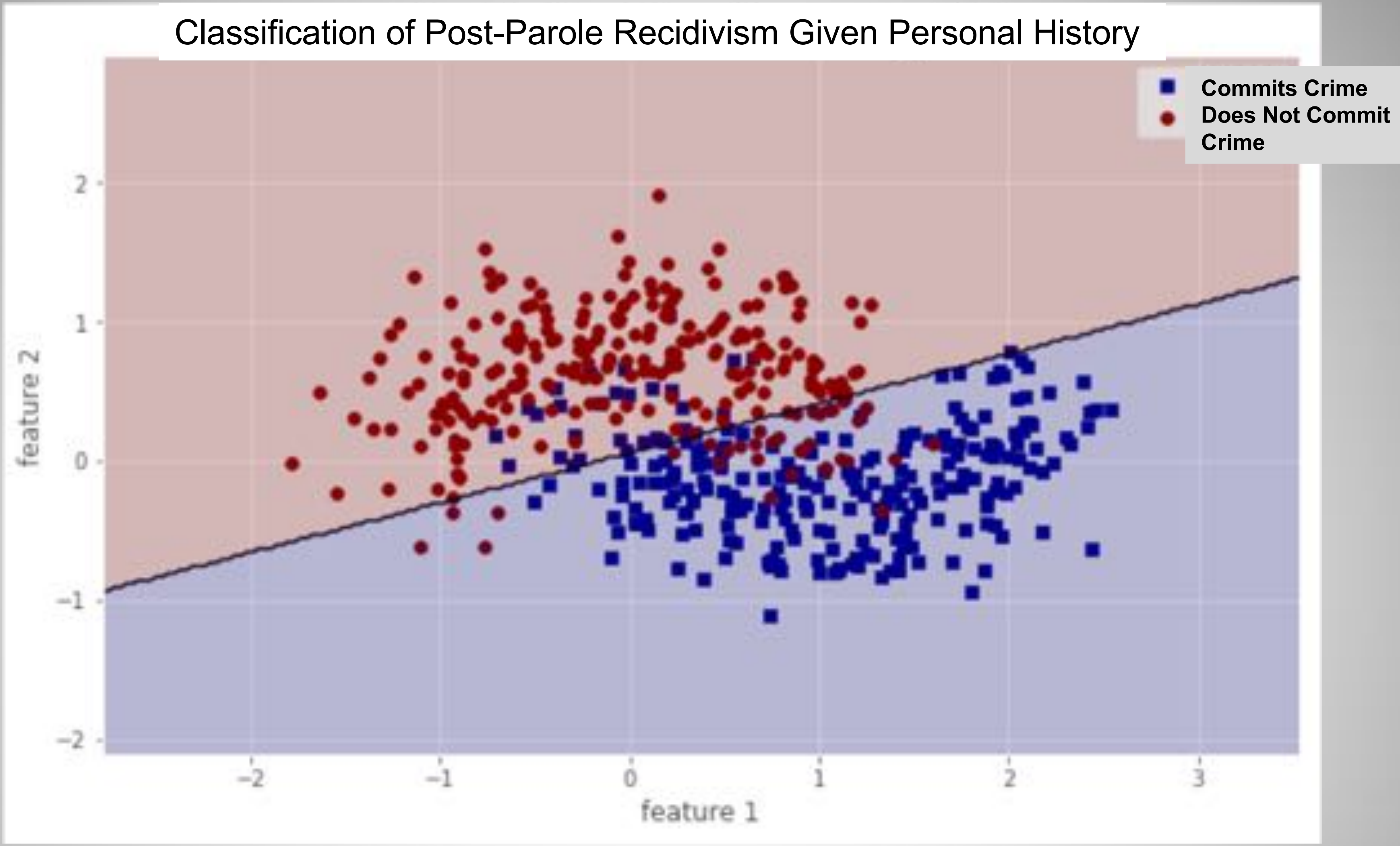
	Predicted Commits Crime	Predicted Does Not Commit
Actual Commits Crime	213	37
Actual Does Not Commit	33	217

THE UNIVERSITY OF UTAH

False positive: Punish Innocent People

cience

False negative: More Victims



	Predicted Commits Crime	Predicted Does Not Commit
Actual Commits Crime	213	37
Actual Does Not Commit	33 **Racial Bias?**	217

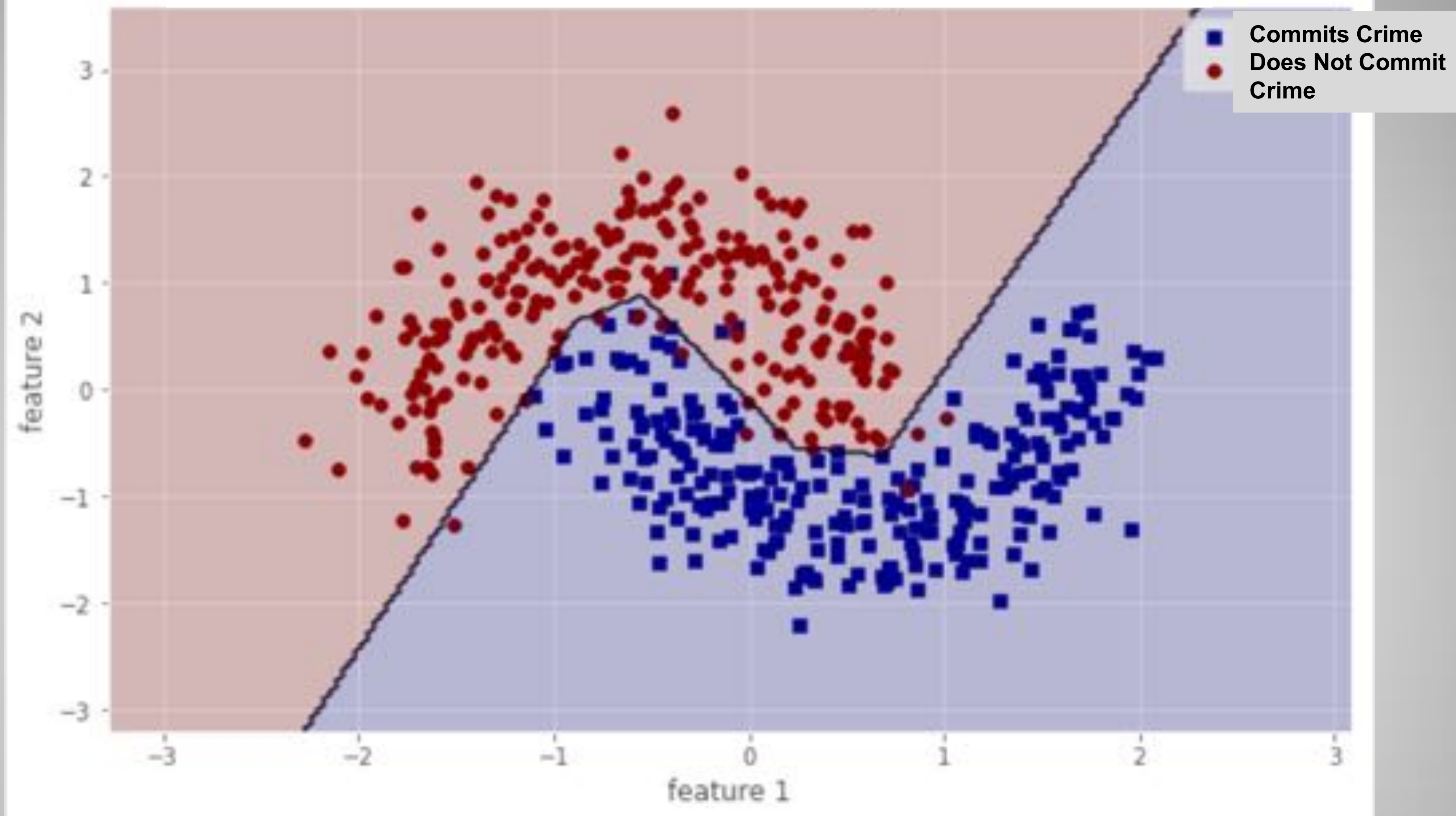
False positive: Punish Innocent People

False negative: More Victims

“ I did then what I knew how to do. Now that I know better, I do better.”

- Maya Angelou

Classification of Post-Parole Recidivism Given Personal History



	Predicted Commits Crime	Predicted Does Not Commit
Actual Commits Crime	237	13
Actual Does Not Commit	4	246

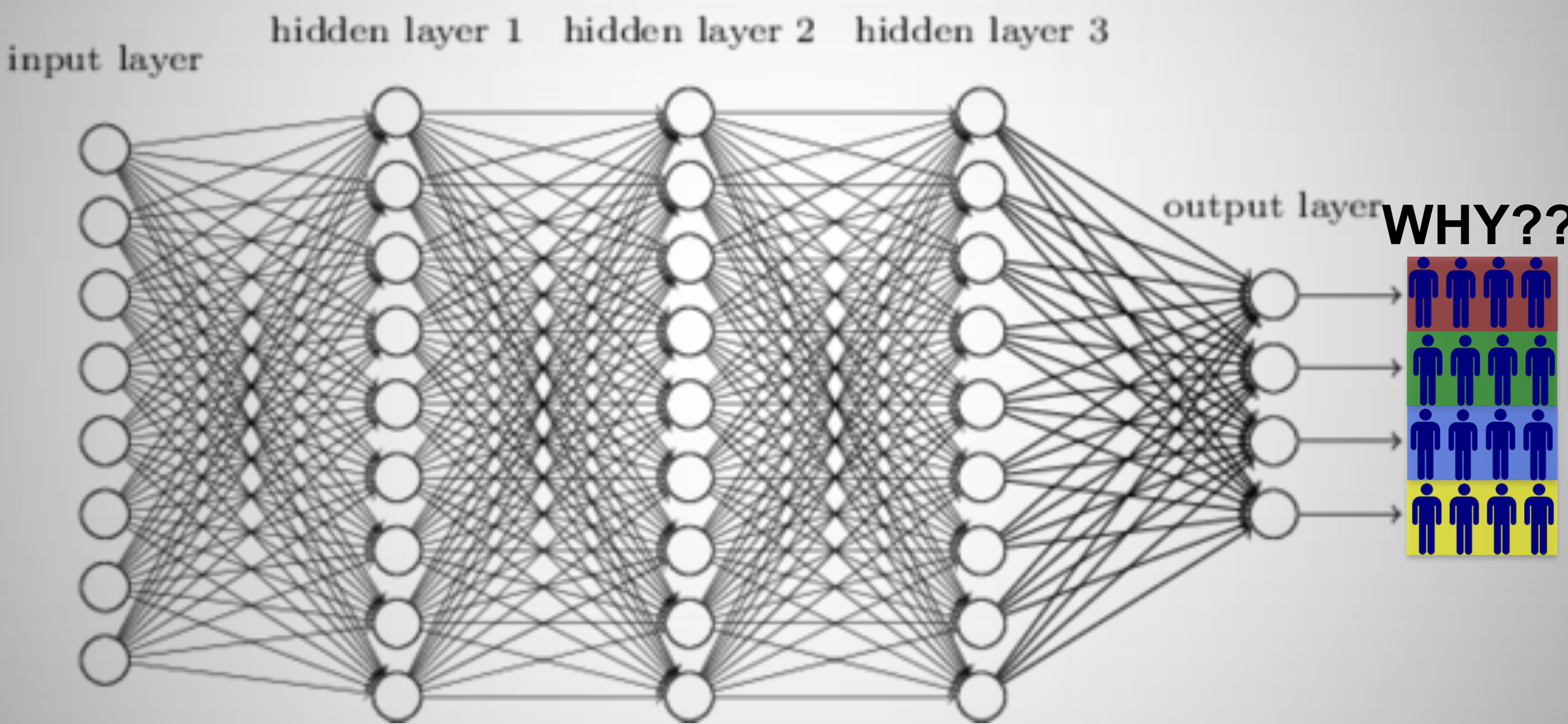
THE UNIVERSITY OF UTAH

False positive: Punish Innocent People

cience

False negative: More Victims

Make sure algorithms are *interpretable* and *transparent*



Because my neural network said so!!

Make sure we can query black box algorithms

In six major same-day delivery cities, however, the service area excludes predominantly black ZIP codes to varying degrees, according to a Bloomberg analysis that compared Amazon same-day delivery areas with U.S. Census Bureau data.

http://www.bloomberg.com/graphics/2016-amazon-same-day/

Make sure we can query black box algorithms

Complement analysis with more interpretable model

Make sure we can query black box algorithms

Model auditing: Analysis of the influence of features used to train the model

- Features independent and identically distributed
- Co-variance among features
- Joint interactions – features only important when they interact with other features

Knowledge and Information Systems
January 2018, Volume 54, Issue 1, pp 95–122 | Cite as

Auditing black-box models for indirect influence

Authors | Authors and affiliations

Philip Adler, Casey Falk, Sorelle A. Friedler, Tionney Nix, Gabriel Rybeck, Carlos Scheidegger, Brandon Smith, Suresh Venkatasubramanian

Relationships to Data

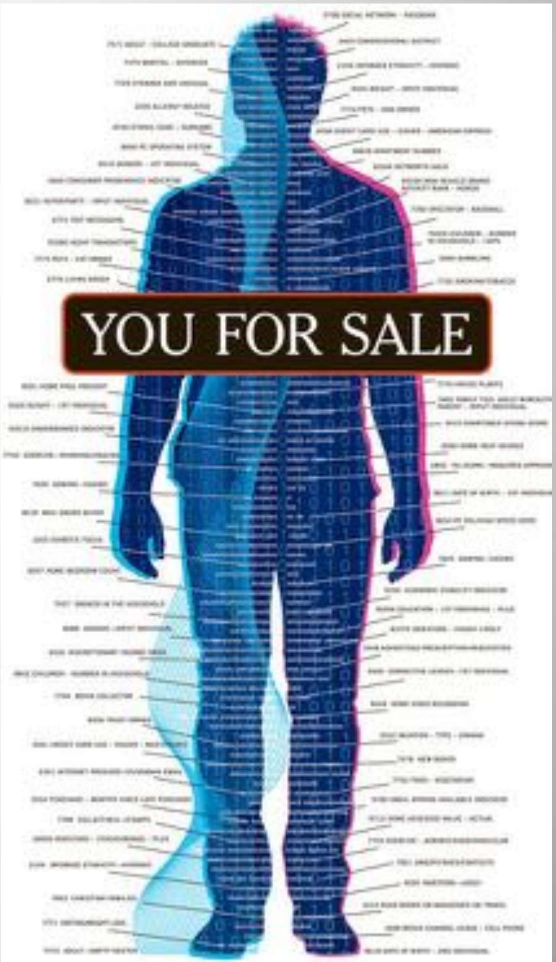
Relationships to Data

Commodity – Data Mined and Monetized

- Data Brokers (e.g. Acxiom)
- Cambridge Analytica

acxiom®

Cambridge Analytica



Relationships to Data

Personal Property
- Privacy
- Informed Consent

Relationships to Data

Public Resource

- NIH All of Us: Collect genetic data of over 1,000,000 people for new insights into disease treatment
- Eviction Lab: Publically available nationwide eviction database (evictionlab.org)

Data is People

Human experimentation in medical research

- JM Sims and the birth of gynecology, 1845-1849

- Medical experimentation by Nazi regime,1941-1945

- Tuskegee Syphilis experiment, 1932-1972

Belmont Report (1979) – Ethical Principles

- Respect for persons
  - Treat individuals as autonomous agents
  - Protect those with diminished autonomy
- Beneficence
  - Minimize possible harms
  - Justify possible benefits
- Justice
  - Equitable distribution of the burdens and benefits of research

Belmont Report (1979) – Application

- Informed Consent: Information, Comprehension, and Voluntariness
- Assessment of Risks and Benefits
  - Probability
  - Magnitude
- Selection of Subjects
  - Fair procedures and outcomes
  - Address social, racial, sexual and cultural biases
  - Protection of vulnerable populations

(Physical) Data is People

Institutional Review Boards

- Review and monitor biomedical research involving human subjects
- Protects the rights and welfare of human subjects

Data is People

Human experimentation by tech companies

PNAS

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory and Jeffrey T. Hancock

PNAS June 17, 2014. 111 (24) 8788-8790; published ahead of print June 2, 2014.
https://doi.org/10.1073/pnas.1320040111

facebook

Cornell University

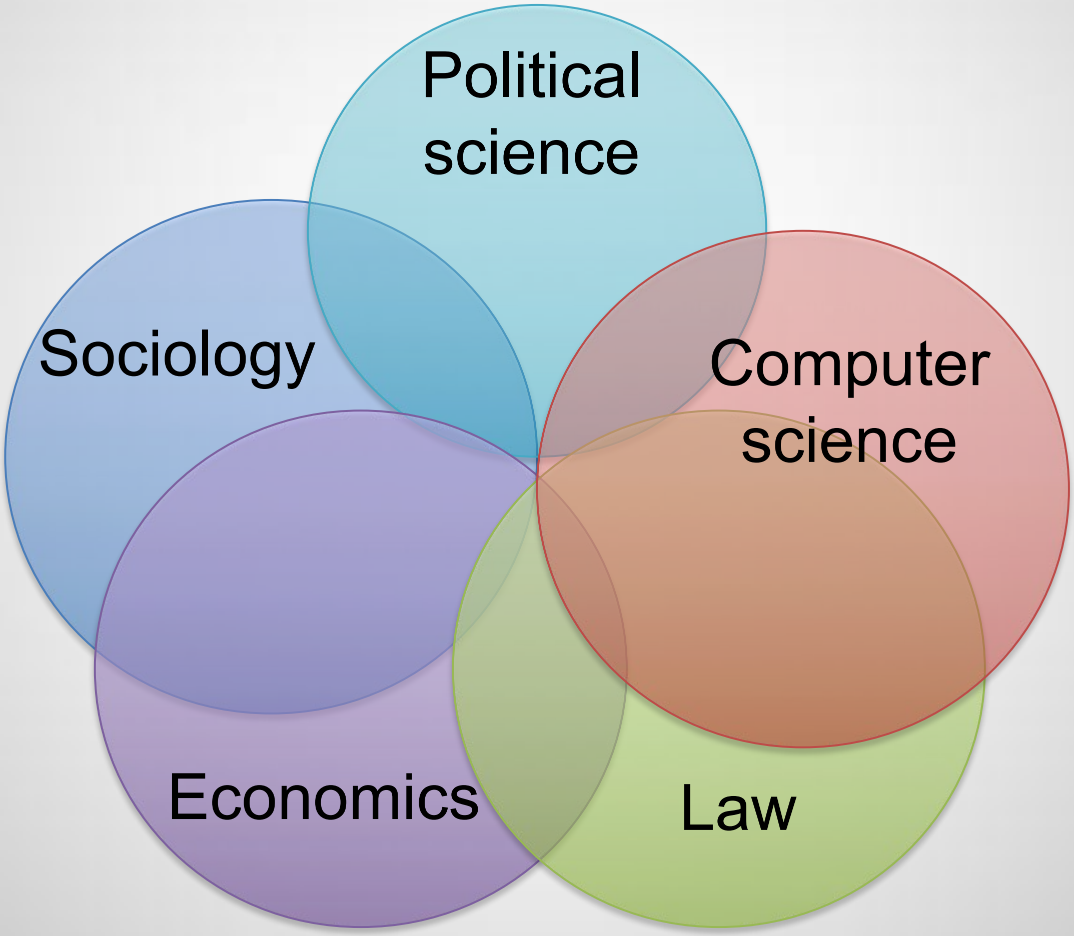
THE UNIVERSITY OF UTAH

(Digital) Data is People

Consumer Subject Review Boards (Ryan Calo, UW)

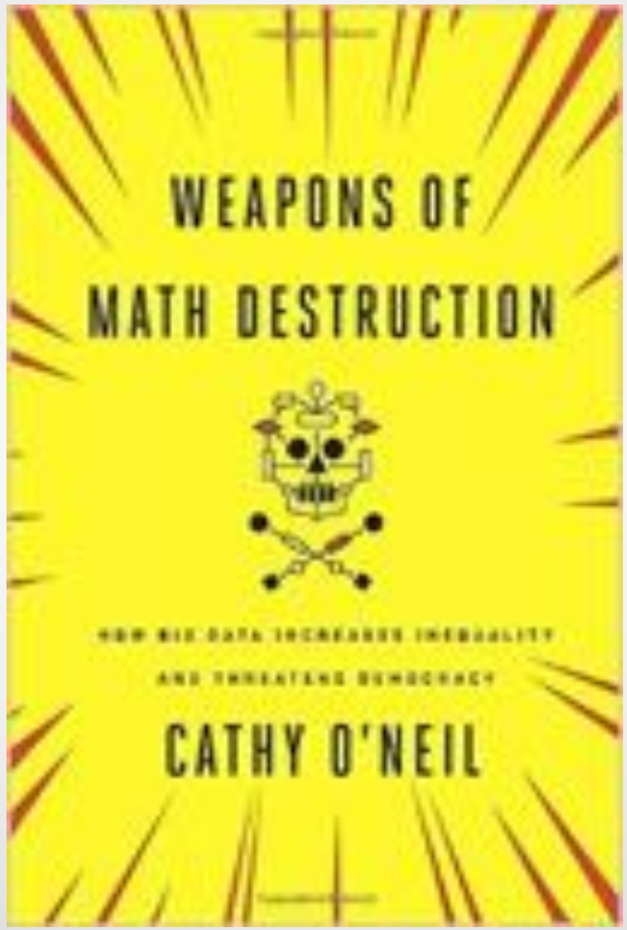
- Ensure informed consent and limit data experimentation when informed consent impossible
- Perform risk/benefit analysis for non-contextual data uses
- Conduct disparate impact analysis: Biases, burdens, and the use of sensitive attributes

The questions



THE UNIVERSITY OF UTAH

Weapons of Math Destruction



THE UNIVERSITY OF UTAH

Ethical considerations of Final Projects

- Where in the process of your analysis were ethical decisions made? What were they?
- Stakeholder analysis
  - Who are the different “personas” relevant to your project?
  - What are some incentives that may align or compete among these groups?
- Successive iterations of analysis: What would you change? Why?