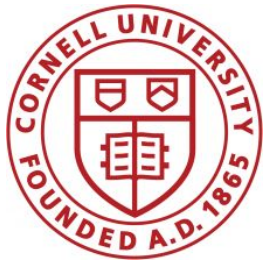


DSFA

Spring 2018

Lecture 39

Privacy

Announcements

- Piazza post about final exam coming soon.
- Final exam review session **tentatively** scheduled for Sunday, May 13, 3 pm, Gates G01.
- Recitations are done. None this week.

Question

Do you believe that Facebook is good or bad for society?

A. Excellent
B. Good
C. Neutral
D. Bad
E. Awful

Question

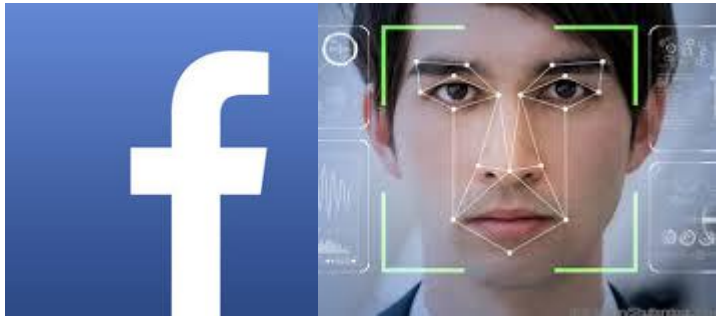
Do you believe that facial recognition technology is good or bad for society?

A. Excellent
B. Good
C. Neutral
D. Bad
E. Awful

Question

Do you believe that Facebook using facial recognition technology is good or bad for society?

A. Excellent
B. Good
C. Neutral
D. Bad
E. Awful



Privacy

Q: What are ways that Facebook (or similar apps) could violate your privacy?

Q: What *is* privacy?

Privacy definitions

- The right to be let alone
  - Samuel Warren and Justice Louis Brandeis
  - Fourth Amendment
- Fair information practices
  - US Federal Trade Commission
  - Notice, consent [opt-in, opt-out], access
- Contextual integrity
  - Prof. Helen Nissenbaum, CS 5436 “Privacy in the Digital Age”
  - Appropriate flows of information in a context

Why care about privacy?

- Reputation management
- Maintaining social boundaries
- Limits on government power
- Law enforcement
- Freedom of thought, speech, politics
- Opportunity for second chances

Data and Privacy

Automated License Plate Readers

(Demo)

Voter Databases

(Demo)

Linking

(Demo)

Identification in the real world

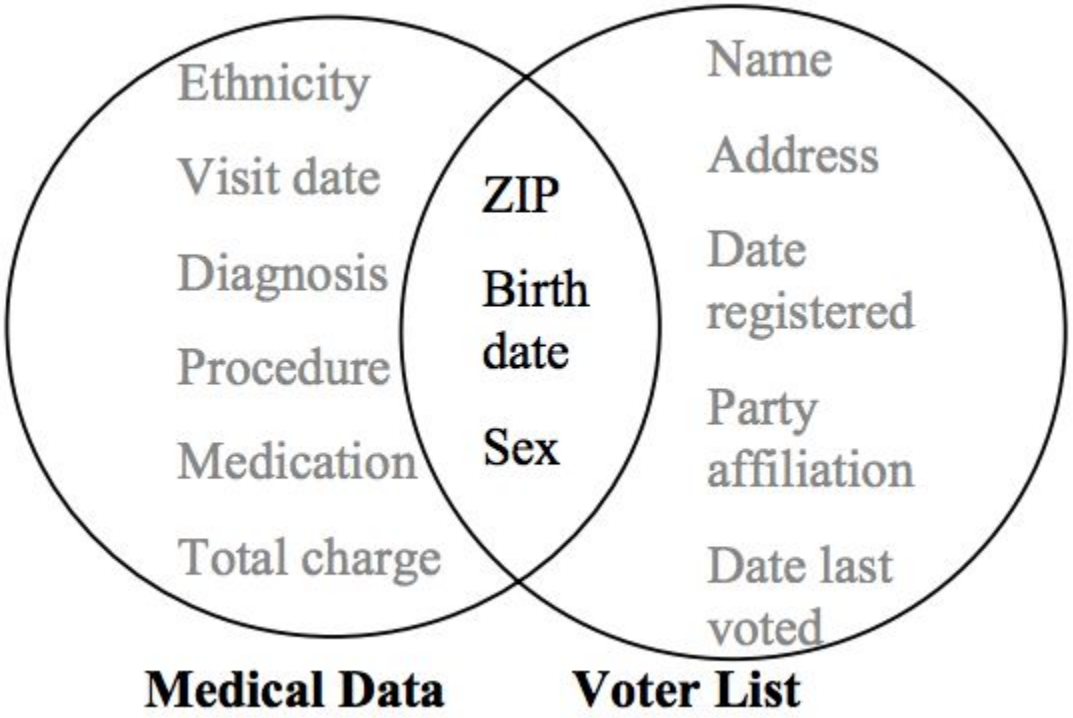
87% of US population
uniquely identifiable
from 5-digit ZIP, sex, date of birth

Prof. Latanya Sweeney (Harvard)
Former Chief Technologist at US FTC

Discovered in 2000 from 1990 census

Linking in the real world

"Anonymized" health insurance database



Massachusetts voter database

Sweeney (2002) re-identified then-governor of MA this way

Linking in the real world

Narayanan and Shmatikov (2006): Use public IMDb movie ratings to link against and de-anonymize Netflix “anonymized” movie ratings

Database privacy

- K-anonymity: each individual is indistinguishable from K-1 other individuals in database [Sweeney 2002]
- How to guarantee?
  - Suppress some data
  - Generalize some data
- Tradeoff:
  - Improve privacy
  - Decrease *utility* of data for analysis

(Demo)

Guidelines for data privacy

- Seek consent
- Select minimal identity
- Limit storage
- Avoid linking