

Data and privacy

SYSTEM ERROR



WHERE BIG TECH
WENT WRONG
AND HOW WE CAN
REBOOT

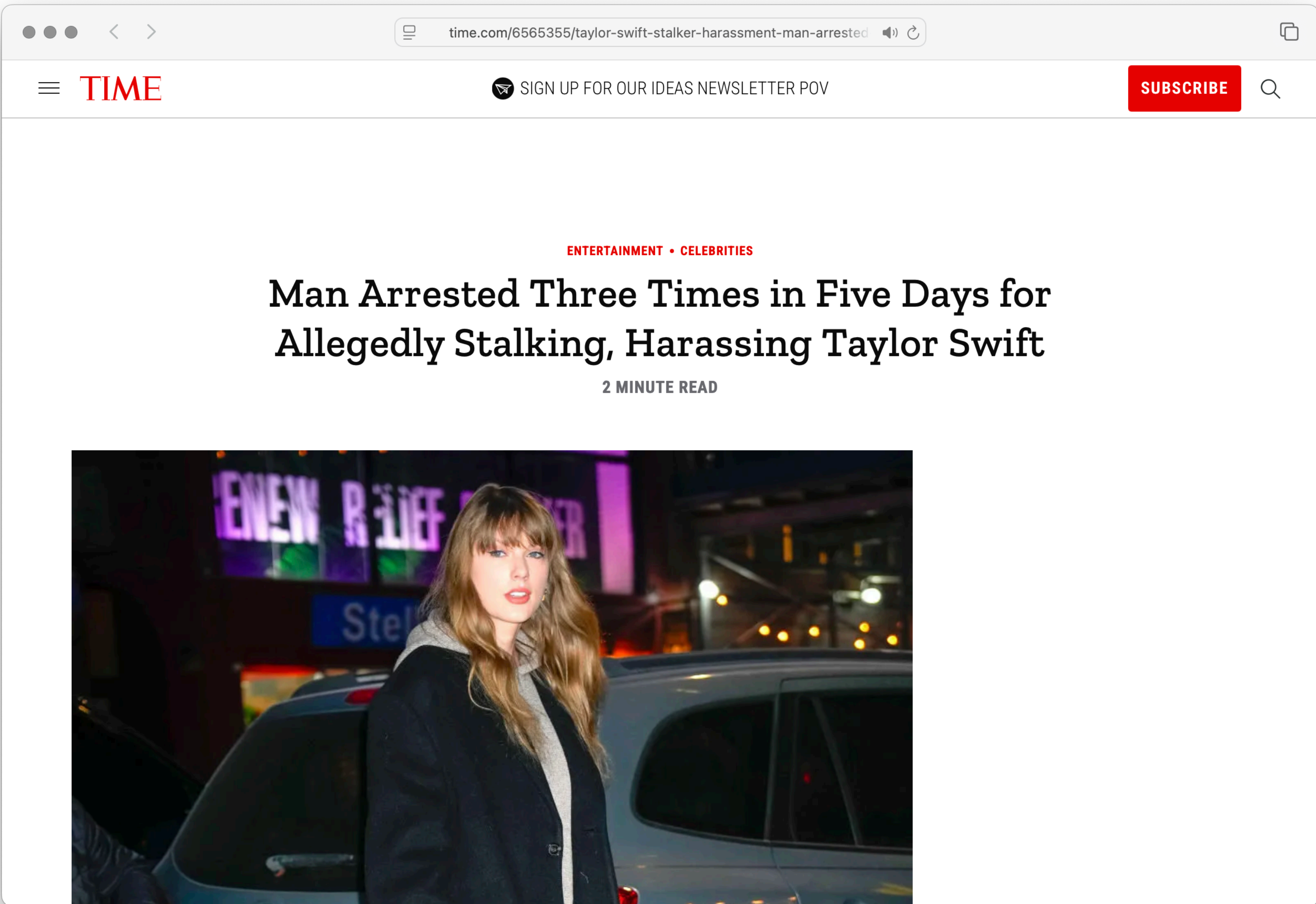
ROB
REICH

MEHRAN
SAHAMI

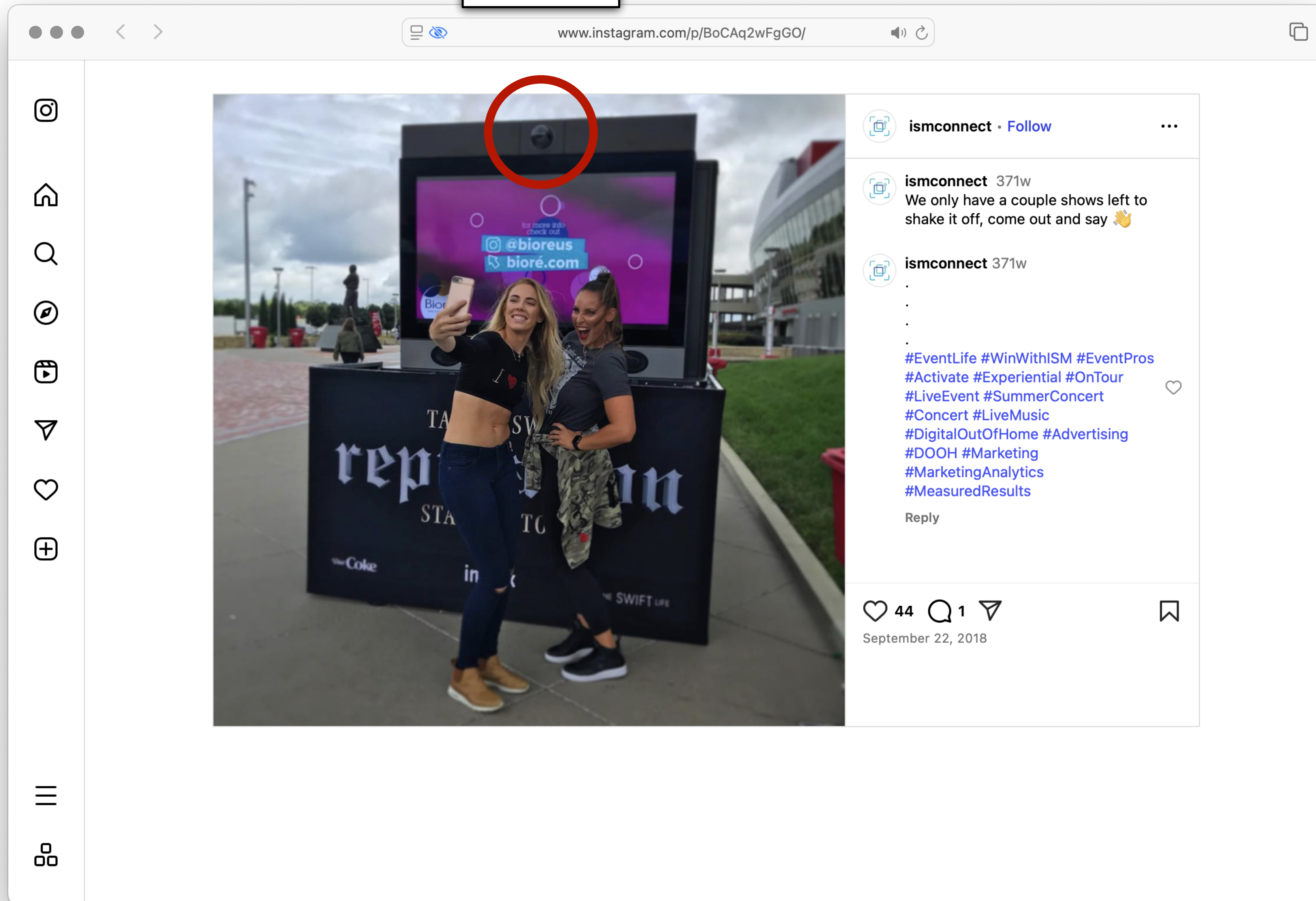
JEREMY M.
WEINSTEIN

Case study:
Taylor Swift





Camera



“The idea of protecting Taylor Swift from known stalkers makes sense. But *greater safety is something we all want*, not just celebrities who have their own private security teams. Cities around the world are using a mix of linked cameras, aerial surveillance, and facial recognition technologies to make the apprehension of criminals more likely and deter future criminality.”

Weinstein, Sahami, and Reich, *System Error: Where Big Tech Went Wrong and How We Can Reboot*, 2021

An aerial night photograph of Baltimore, Maryland, showing the harbor, city skyline, and waterfront developments. The water reflects the city lights, and several boats are visible in the harbor. The sky is filled with soft, colorful clouds from the sunset or sunrise.

Case study: Baltimore

www.economist.com/united-states/2017/06/29/crime-and-despair

The Economist

SubscribeEnterpriseLog inMenu

Weekly editionThe world in briefWar in the Middle EastWar in UkraineUnited StatesThe world economyBusinessArtificial intelligenceGames


United States

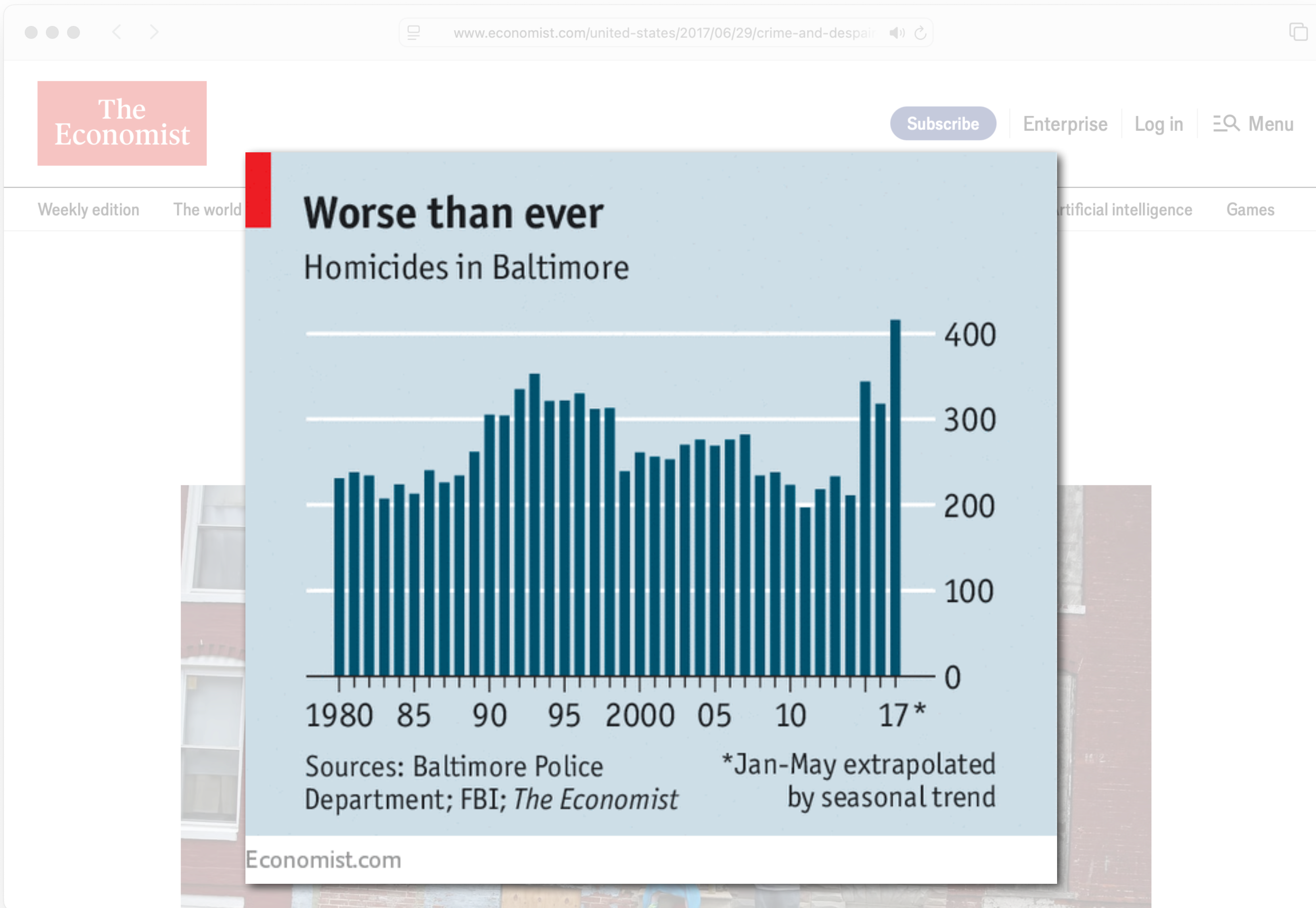
| An exceptionally murderous city

Crime and despair in Baltimore

As America gets safer, Maryland's biggest city does not

Share





<

>

www.cnn.com/2016/08/09/us/baltimore-justice-department-report

CNN

US


Crime + Justice

• Watch

Subscribe


Sign in


Racial bias pervasive among Baltimore police, DOJ says





By [Emanuella Grinberg](#), CNN


🕒 7 min read · Updated 9:35 PM EDT, Wed August 10, 2016



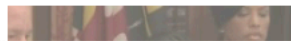










🗨 Video Ad Feedback

 DOJ: African-American

 Video shows cop plant

 Baltimore police

www.rollingstone.com/culture/culture-features/eyes-over-baltimor

RollingStone

MUSICPOLITICSTV & MOVIESCULTURERS RECOMMENDS


Got A Tip?

CULTURE

EYES OVER BALTIMORE: HOW POLICE
USE MILITARY TECHNOLOGY TO
SECRETLY TRACK YOU

"They view people as enemy combatants," says activist, as cops
adopt surveillance, tracking, facial recognition programs designed
for war zones

By BENJAMIN POWERS
JANUARY 6, 2017



THE LATEST

THE MORE THE MERRIER

Bad Bunny, Karol G, CA7RIEL
and Paco Amoroso, Fuerza
Regida to Perform at 2025 Latin
Grammys

16 MINUTES AGO

EXCLUSIVE

Bob Dylan Accepts Honorary
Doctorate From Berklee College
of Music

52 MINUTES AGO

SIR BECKS

Arise, Sir Goldenballs: David
Beckham Is Officially a Knight

58 MINUTES AGO

SHOW MUST GO ON

Ariana Grande ‘Beyond
Devastated’ to Miss Brazil
‘Wicked: For Good’ Premiere
Due to Flight Delays

2 HOURS AGO

www.aclu.org/press-releases/federal-appeals-court-rules-baltimo

ACLU

About

Issues

Our Work

News

Take Action

Shop

Give

PRESS RELEASES

Federal Appeals Court Rules Baltimore Aerial Surveillance Program is Unconstitutional

Case: Leaders of a Beautiful Struggle v. Baltimore Police Department

June 24, 2021 2:15 pm

Media Contact

media@aclu.org

(212) 549-2666

125 Broad Street

18th Floor

New York, NY 10004

United States

BALTIMORE — The Fourth Circuit Court of Appeals, sitting en banc,ruled today that the Baltimore Police Department’s (BPD) aerial surveillance program, which put the daytime movements of virtually all Baltimore residents under surveillance for 12 hours a day over six months, is unconstitutional. The decision comes in a lawsuit filed by a group of Black activist leaders in Baltimore, with the support of the American Civil Liberties Union and ACLU of Maryland, requesting that the court temporarily block the BPD from deploying and conducting a six-month trial of the aerial surveillance program. Although the Fourth Circuit heard the plaintiffs’ appeal after the six-month trial of surveillance flights had already come to an end, the BPD continued to possess unlawfully acquired data, and the court’s decision will mean that the BPD will be prohibited from accessing data collected through

Once the government has information about your activities, can you trust it to only be used to prevent and solve crimes?

Case study: White House protestor

TRUMP LIES

INVESTIGATE
LEAVE NO STONE
UNTURNED

THIS

SPECIAL

“[One of our students] said that her mother had won a prize for her work as a schoolteacher and was invited to the White House with other teachers for an award ceremony. Upon presenting her credentials at the White House gate, she had been denied entry. When she asked why, she learned that her face had come up in a database of people who had participated in public protests against President Trump, and protestors were not welcome.”

Weinstein, Sahami, and Reich, *System Error: Where Big Tech Went Wrong and How We Can Reboot*, 2021



“If I’m not doing anything
wrong, I don’t need to
hide anything from other
people”

“She had done nothing wrong. She had nothing to hide. Her face had been recorded while she was behaving lawfully in public... It’s easy to see that with surveillance comes a potential loss of essential liberties in a democratic society, such as a freedom to protest and freedom of expression.”

Weinstein, Sahami, and Reich, *System Error: Where Big Tech Went Wrong and How We Can Reboot*, 2021

Case study: Anonymization and privacy

Latanya Sweeney is a computer scientist and privacy researcher at Harvard University.

In 1997, Sweeney demonstrated that anonymized datasets aren't necessarily anonymous!



Photograph by Kayana-Szmczak

To anonymize data, we remove *personally identifiable information* (PII) – anything that can be traced back to a person, like names or social security numbers.

In the 1990s, the state of Massachusetts – through the Group Insurance Commission (GIC) – released anonymized hospital records for use in research.

The records included fields like

ZIP code,

Sex,

Date of birth,

Diagnosis, procedures, and prescriptions.

To anonymize them, they removed all names and other direct identifiers.

At the same time, voter registration data was publicly available, including

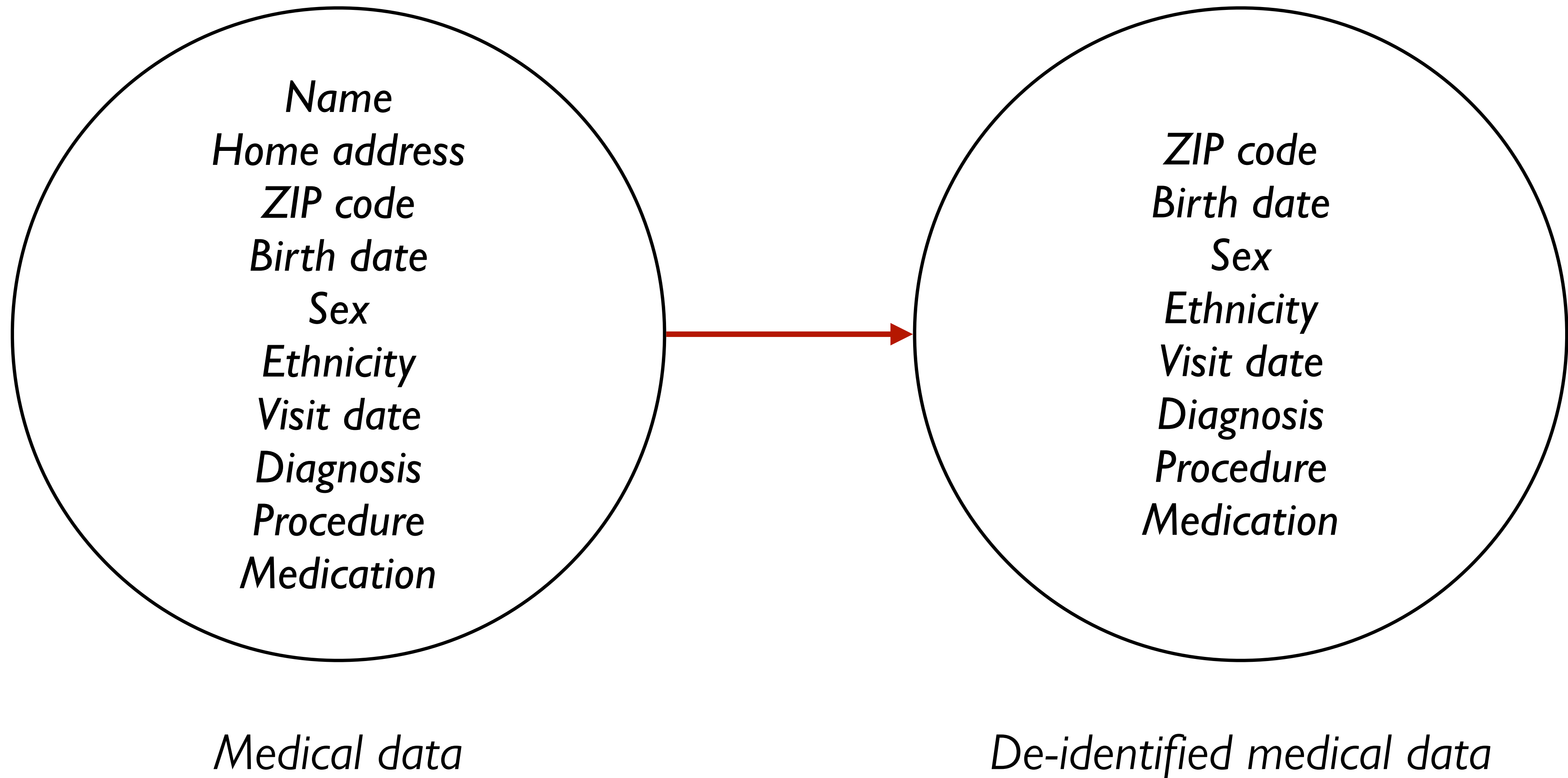
Name,

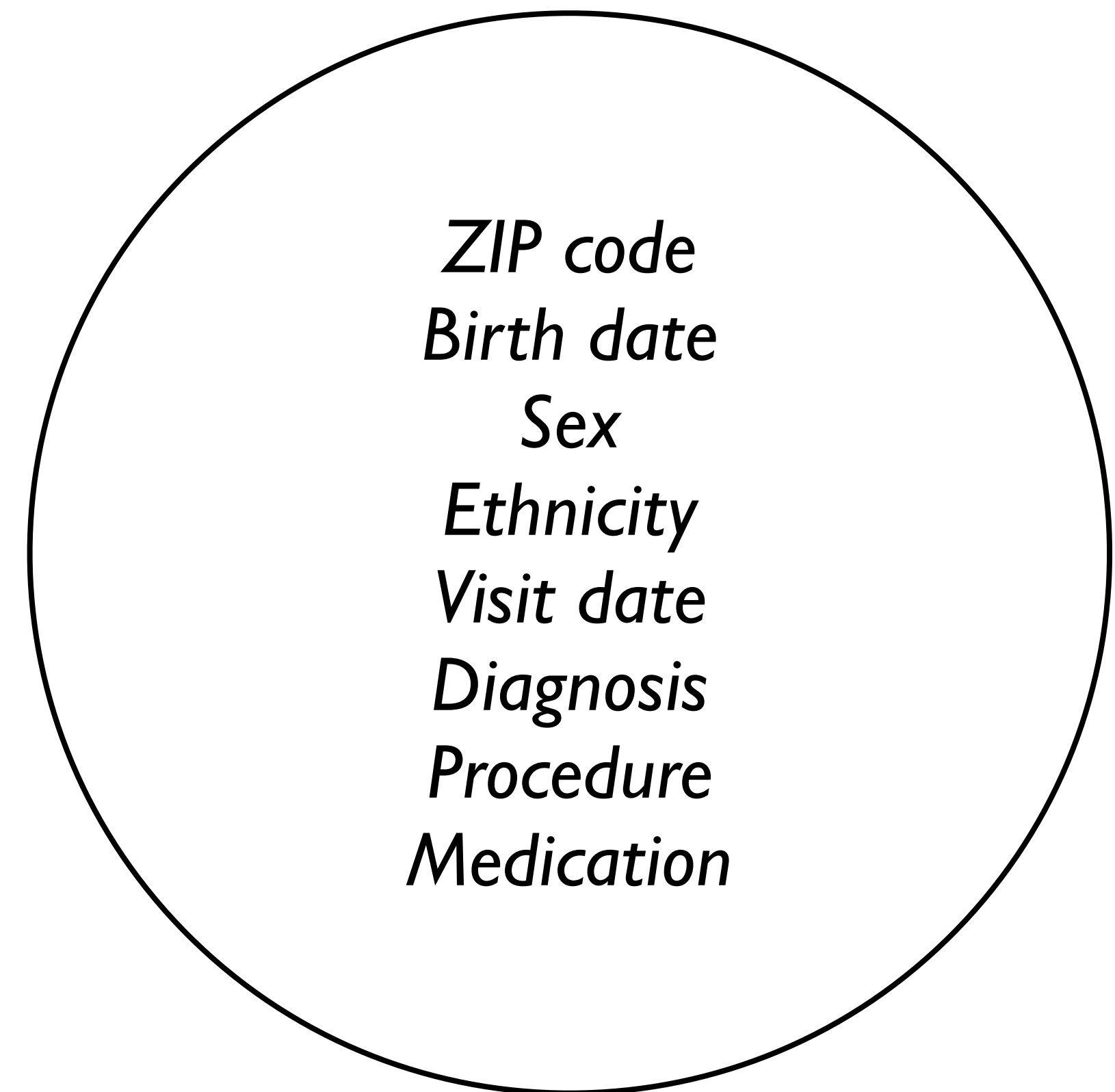
Address,

Gender, and

Date of birth.

Sweeney purchased the voter registration list for \$20.





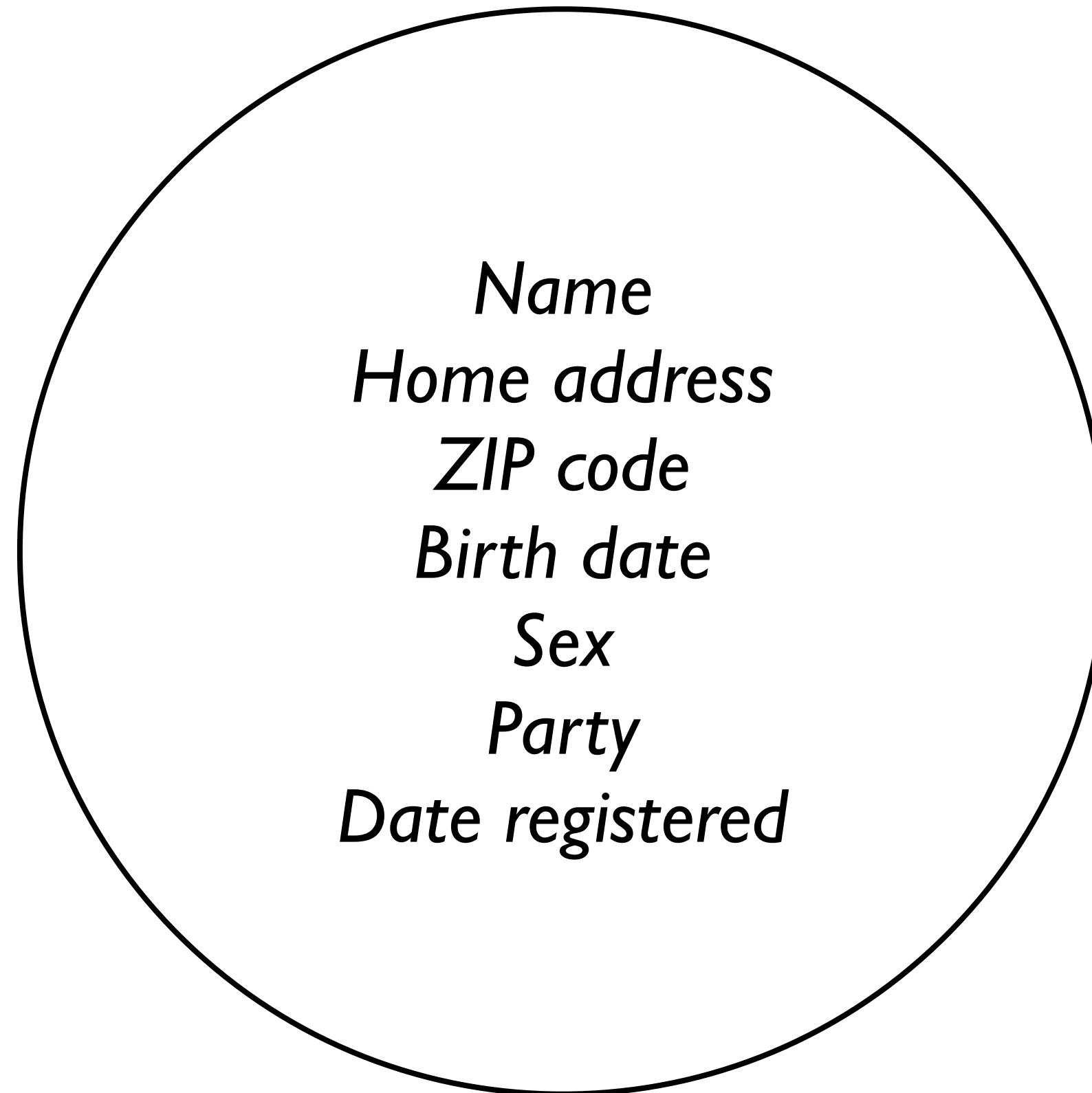
De-identified medical data

Public!

*ZIP code
Birth date
Sex
Ethnicity
Visit date
Diagnosis
Procedure
Medication*

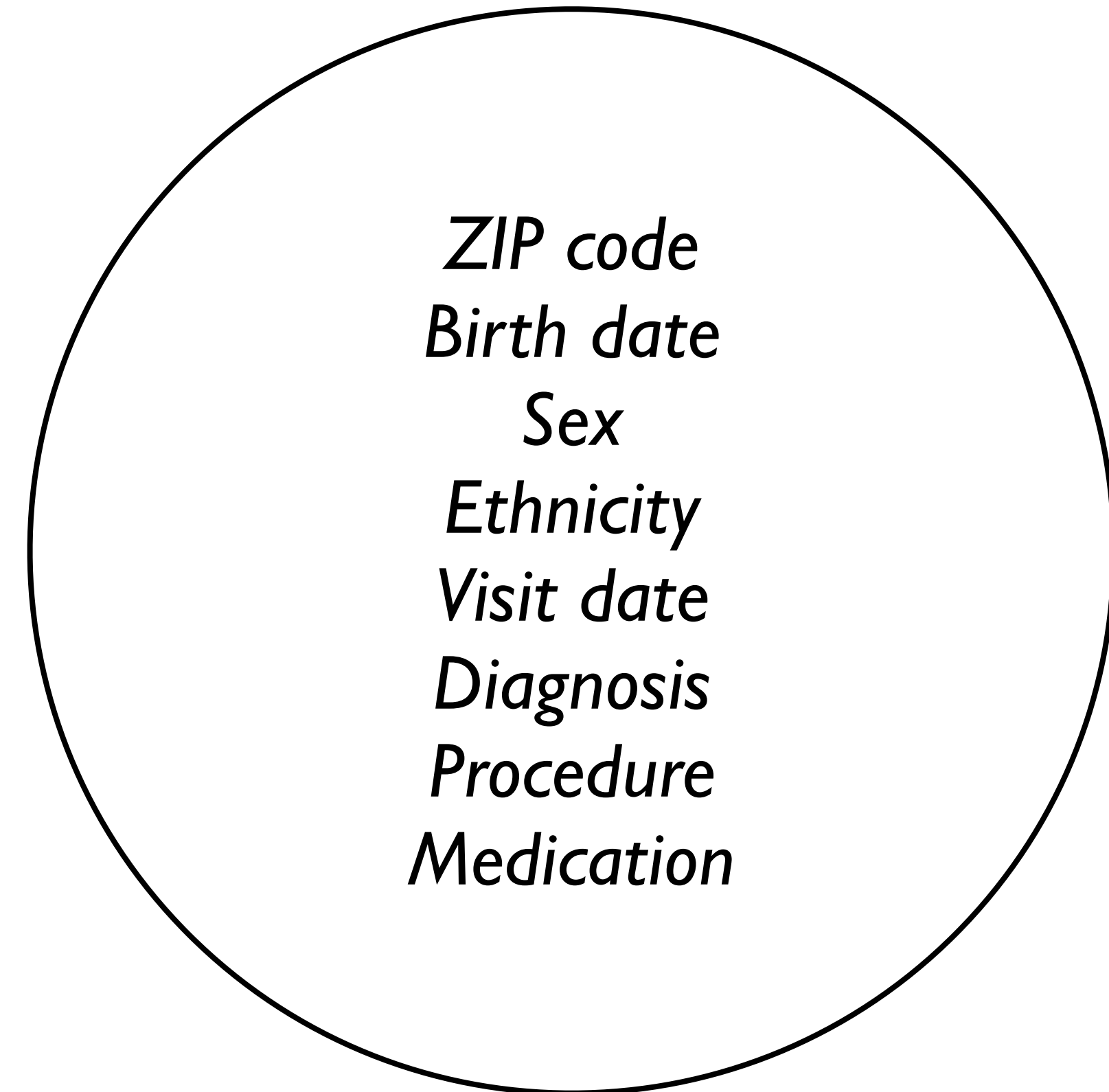
De-identified medical data

Bought for \$20!



Voter registration data

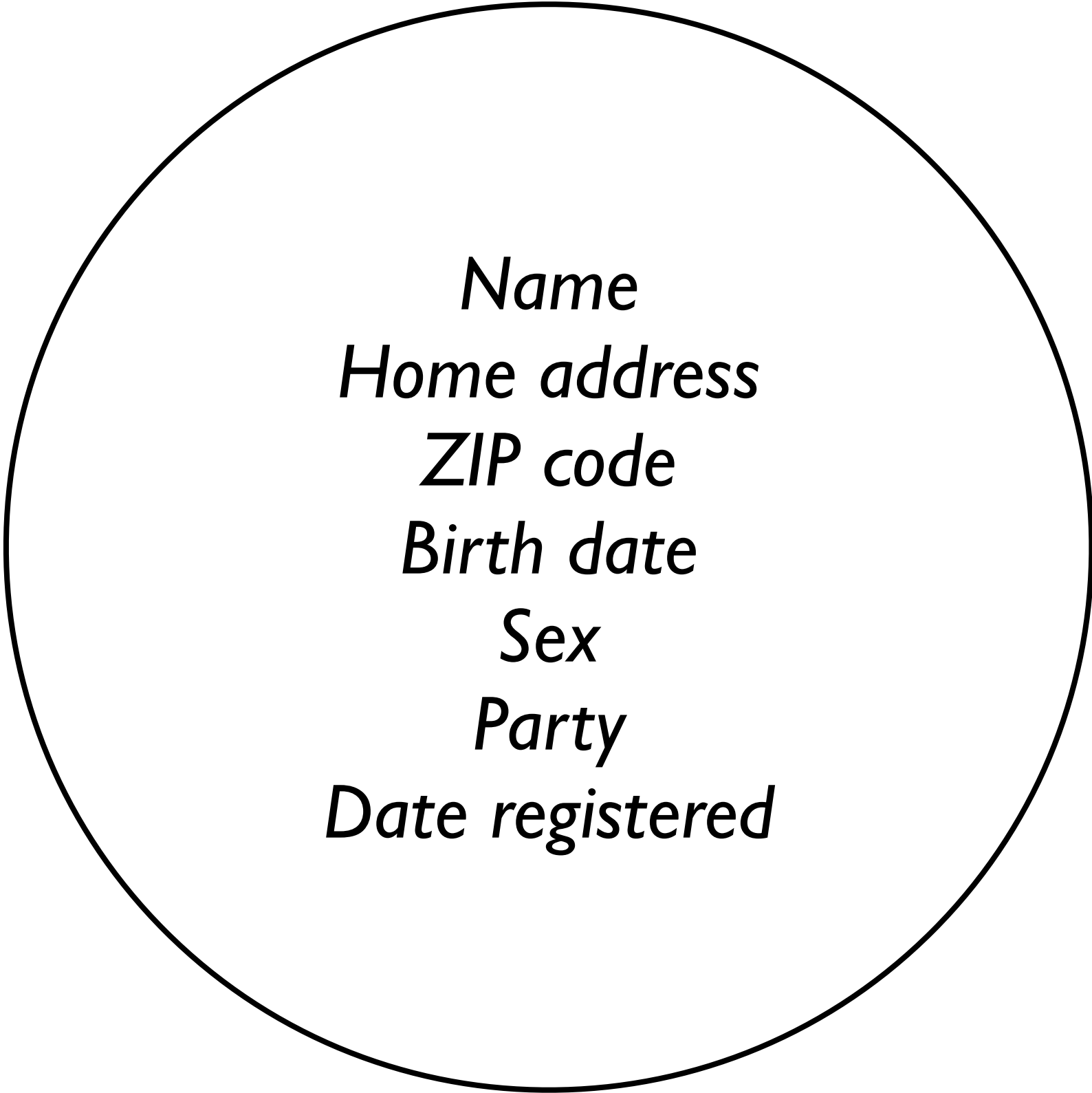
Public!



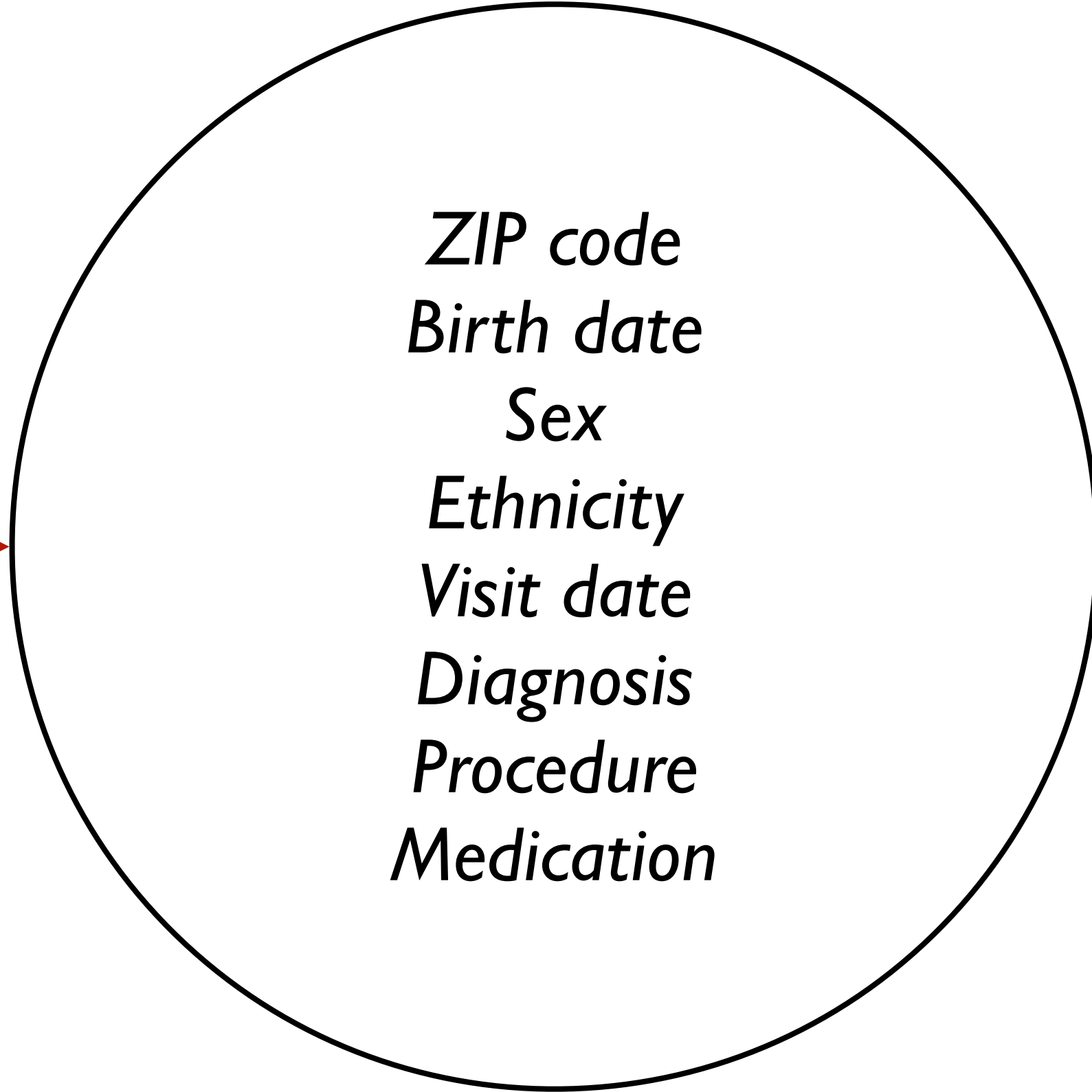
De-identified medical data

Bought for \$20!

Public!



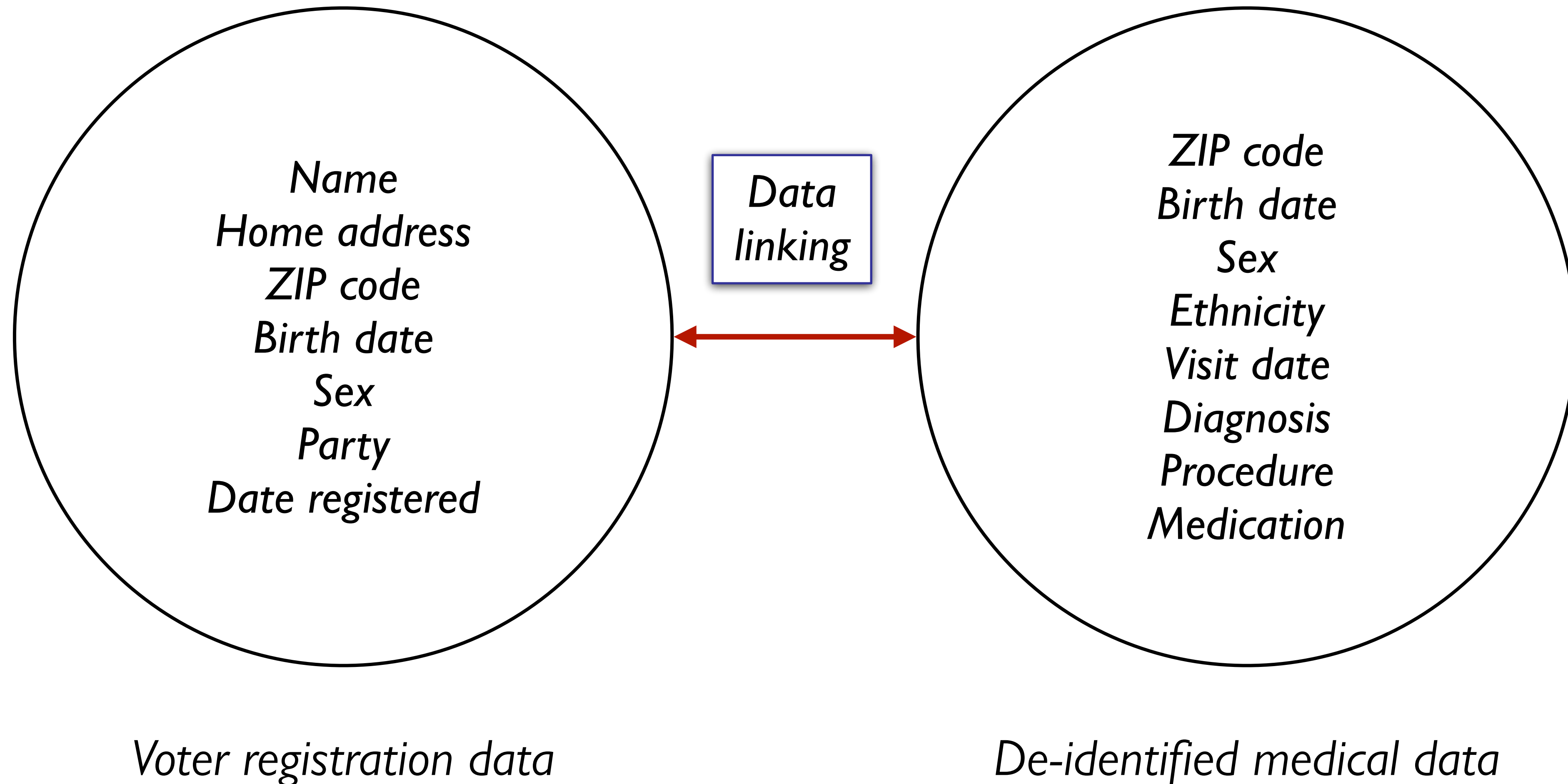
*Data
linking*

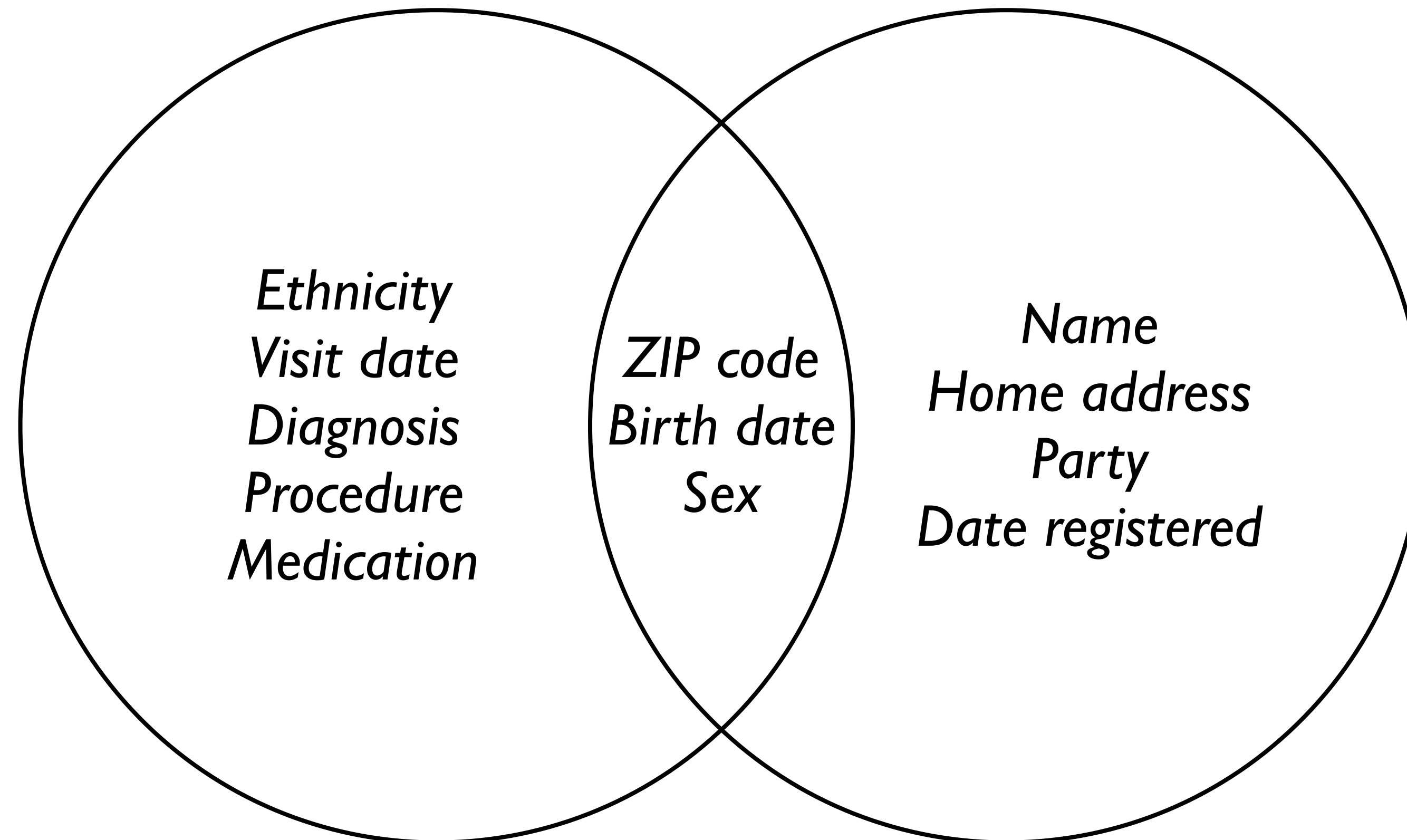


Voter registration data

De-identified medical data

```
records.join("Birth Year", voters)
```





Re-*identified medical data*

Notebook: Sweeney linking study

Risk of combining data sources



Safe

Risk of combining data sources



Safe

+



Safe

Risk of combining data sources



Safe

+



Safe

=



Danger

Bill Weld was Governor of Massachusetts at the time.

He lived in Cambridge, MA

His medical records were included in the anonymized GIC dataset

According to the Cambridge voter list:

6 people shared his birthdate

Only 3 of them were men

Only 1 man lived in his (5-digit) ZIP code.

That 1 person was uniquely identifiable: Governor Weld.



Photograph by Brian Snyder / Reuters

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 was designed to protect the privacy of medical records.

It allows “anonymized” data sharing by removing direct identifiers – but *is that enough?*

Sweeney showed that even HIPAA-compliant data could be vulnerable:

87% of Americans could be uniquely identified using just

ZIP code

Birth date

Gender

These quasi-identifiers are not considered protected under HIPAA

Anonymized \neq anonymous!

What are the potential harms from disclosure of personal information?

It might be helpful to think about different categories of harms:

Political

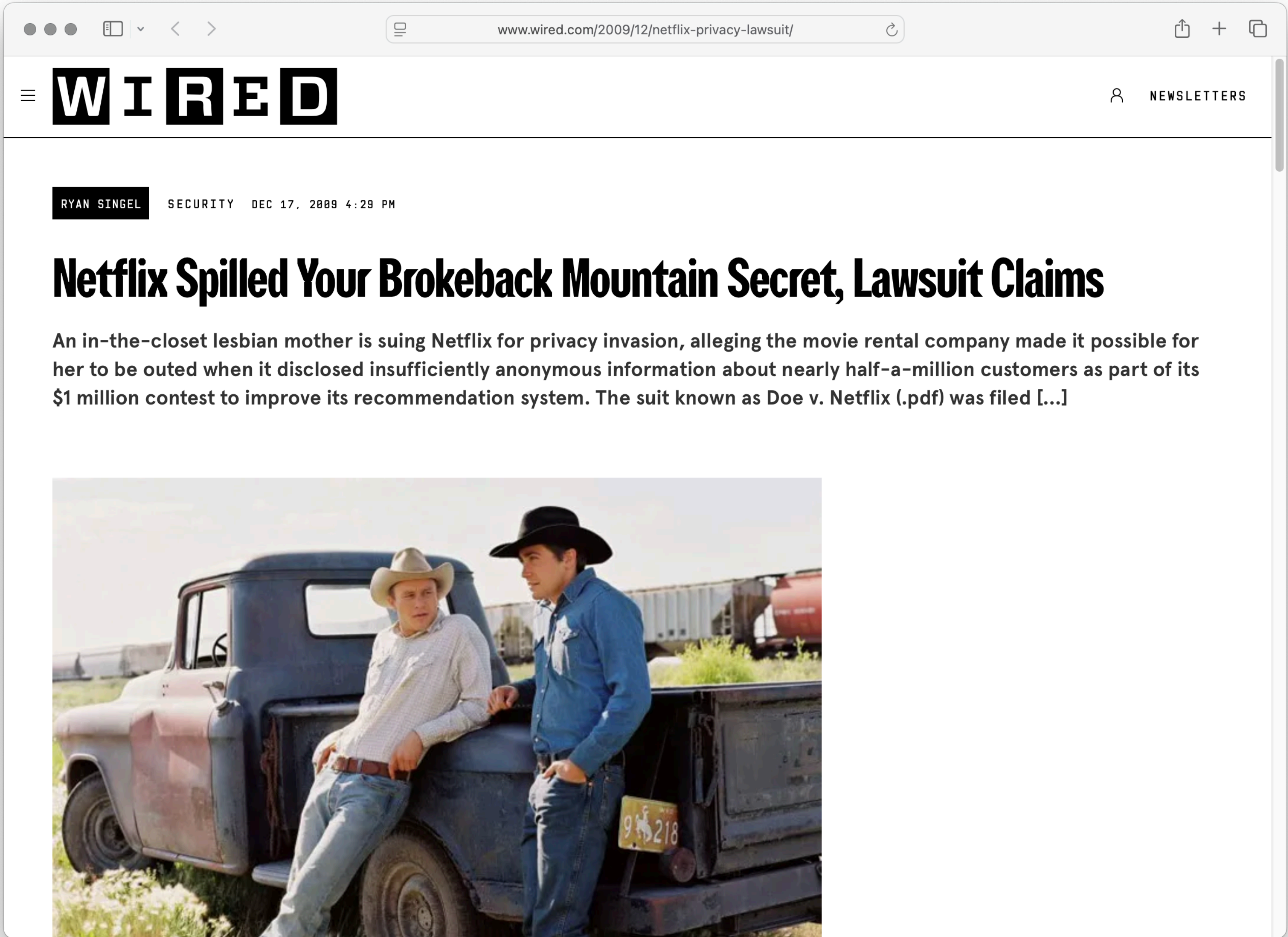
Economic

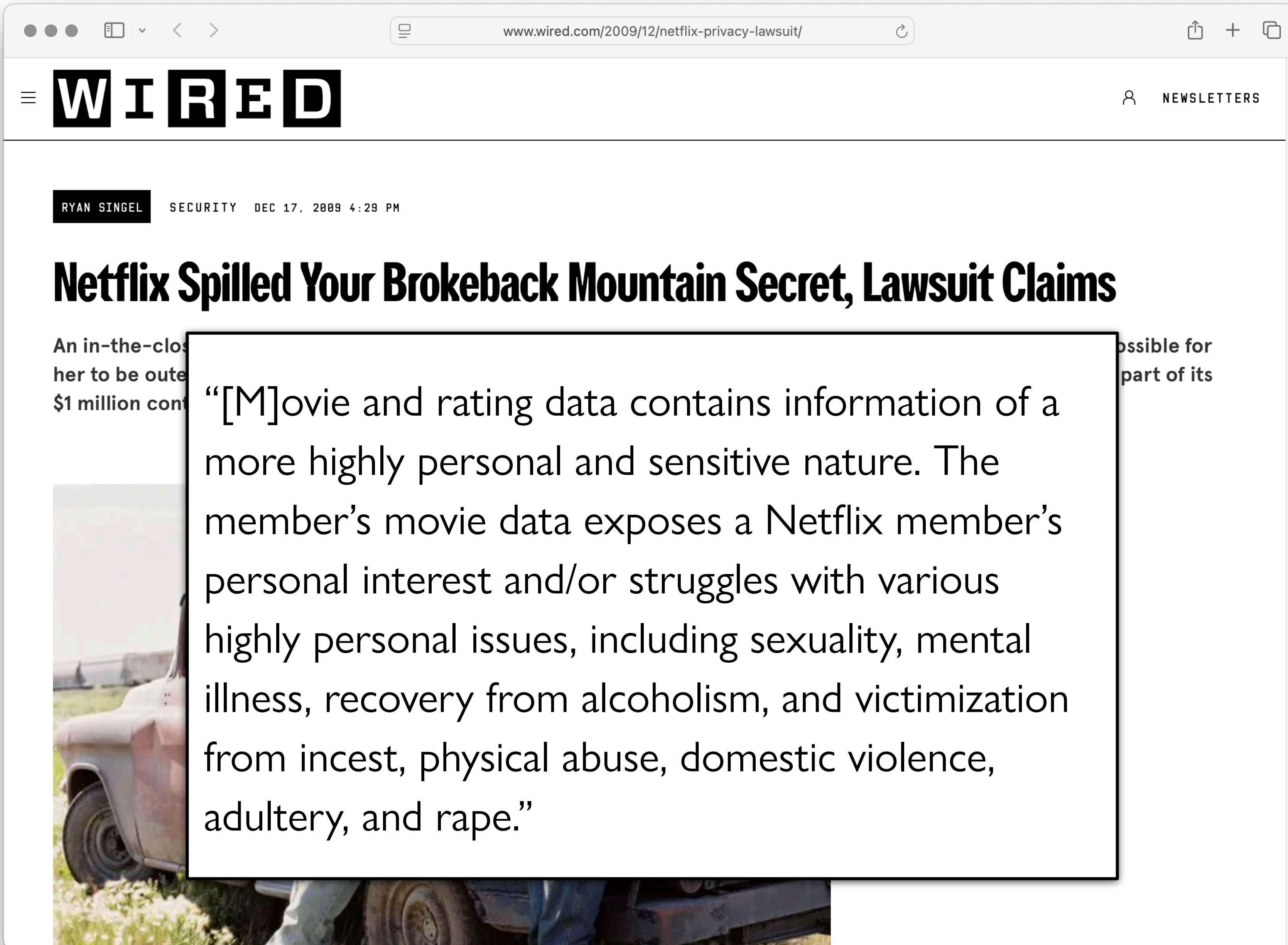
Social

Psychological

Criminal

Collective





RYAN SINGEL

SECURITY DEC 17, 2009 4:29 PM

Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims

An in-the-closet
her to be outed
\$1 million con

“[M]ovie and rating data contains information of a more highly personal and sensitive nature. The member’s movie data exposes a Netflix member’s personal interest and/or struggles with various highly personal issues, including sexuality, mental illness, recovery from alcoholism, and victimization from incest, physical abuse, domestic violence, adultery, and rape.”

ossible for
part of its

Who's who?



Big Bird



Oscar



Ernie



Bert



Grover



Cookie Monster

From Facebook:

Big Bird is the most popular

Oscar's only friend is Bert

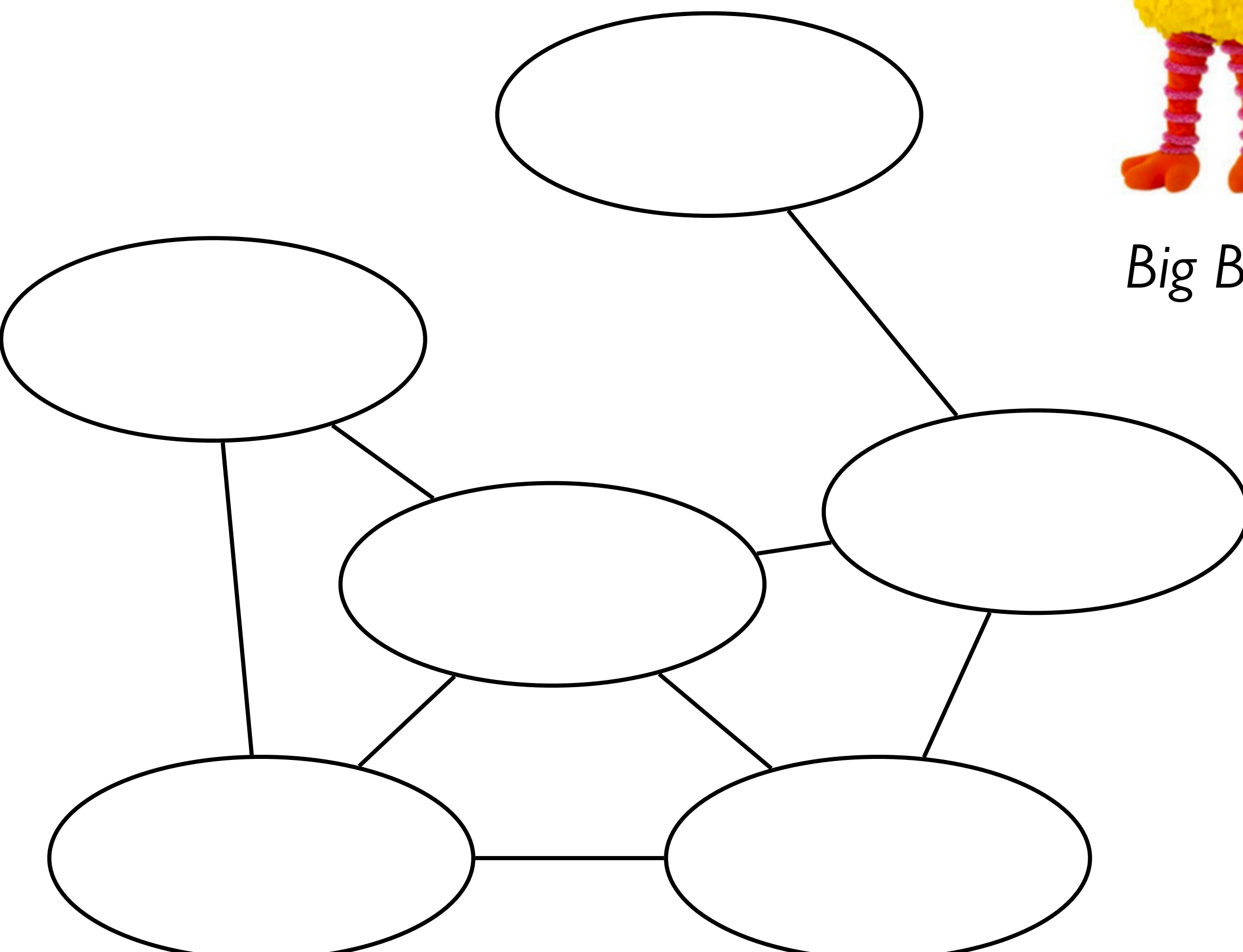
From rate-your-beverage.com:

Bernie, Bert, and Oscar only drink tea

Big Bird, Grover, and Cookie Monster only drink coffee

Ernie is friends with three coffee drinkers

"Anonymous" Social Network Graph



1 Under what conditions are we willing to trade privacy for something else?

2 Who should be responsible for ensuring privacy? How would this be enforced?

1 Under what conditions are we willing to trade privacy for something else?

Personal safety?

National security?

Research and innovation?

Convenience?

2 Who should be responsible for ensuring privacy? How would this be enforced?

Government regulatory agencies?

Individual consumers?

Free market?

Private companies?

Data scientists, engineers, and product managers creating the products?

Your data has economic value!

Data (privacy) is not removed from history or politics

A handful of major events in data privacy in Europe and the US:

- 1930s German census workers gather data on residents' nationality, native language, religion. Data later used to implement the Holocaust.
- 2001 9/11 terrorist attacks and the PATRIOT Act.
- 2013 Edward Snowden leaks classified information about the NSA
- 2018 General Data Protection Regulation (GDPR) passed by the European Union
- 2018 Cambridge Analytica scandal of using Facebook data to build voter profiles
- 2018 California Consumer Privacy Act (CCPA) passes
- 2024/5 EU AI Act, State AI Laws (quantify risk, data use, discrimination)

More extensive history

How should we move forward as a society?

How should we move forward as a society?

Some possibilities:

Companies forced to abide by *data minimization*

“Digital trust intermediary”

Government regulatory agency with more teeth

Informed citizenry

