Privacy & Fairness in Data Science

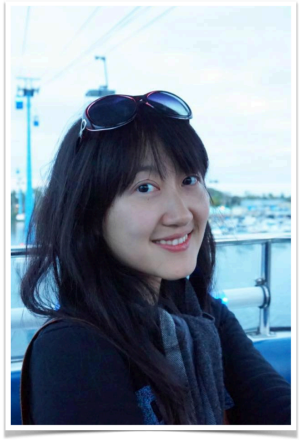
CompSci 590.01 Fall 2018

DUKE
COMPUTER SCIENCE

Instructor

Ashwin Machanavajjhala:
big-data, small devices and individual privacy

Co-Instructors



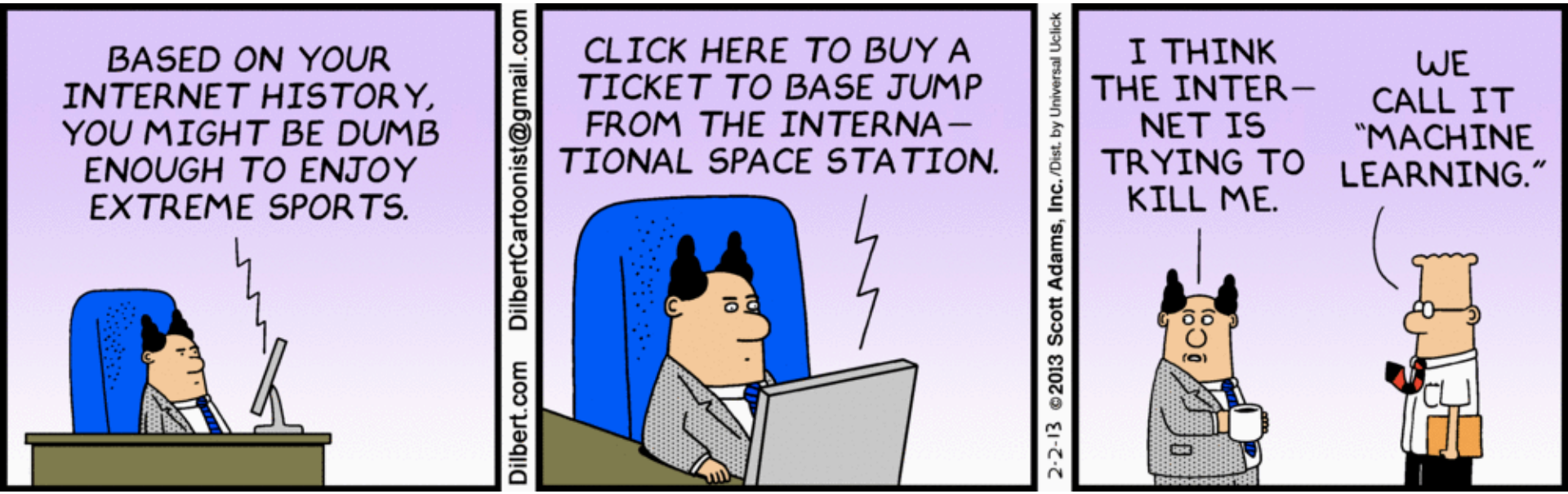
Xi He:
Privacy in data analytics

Brandon Fain:
Algorithmic Fairness

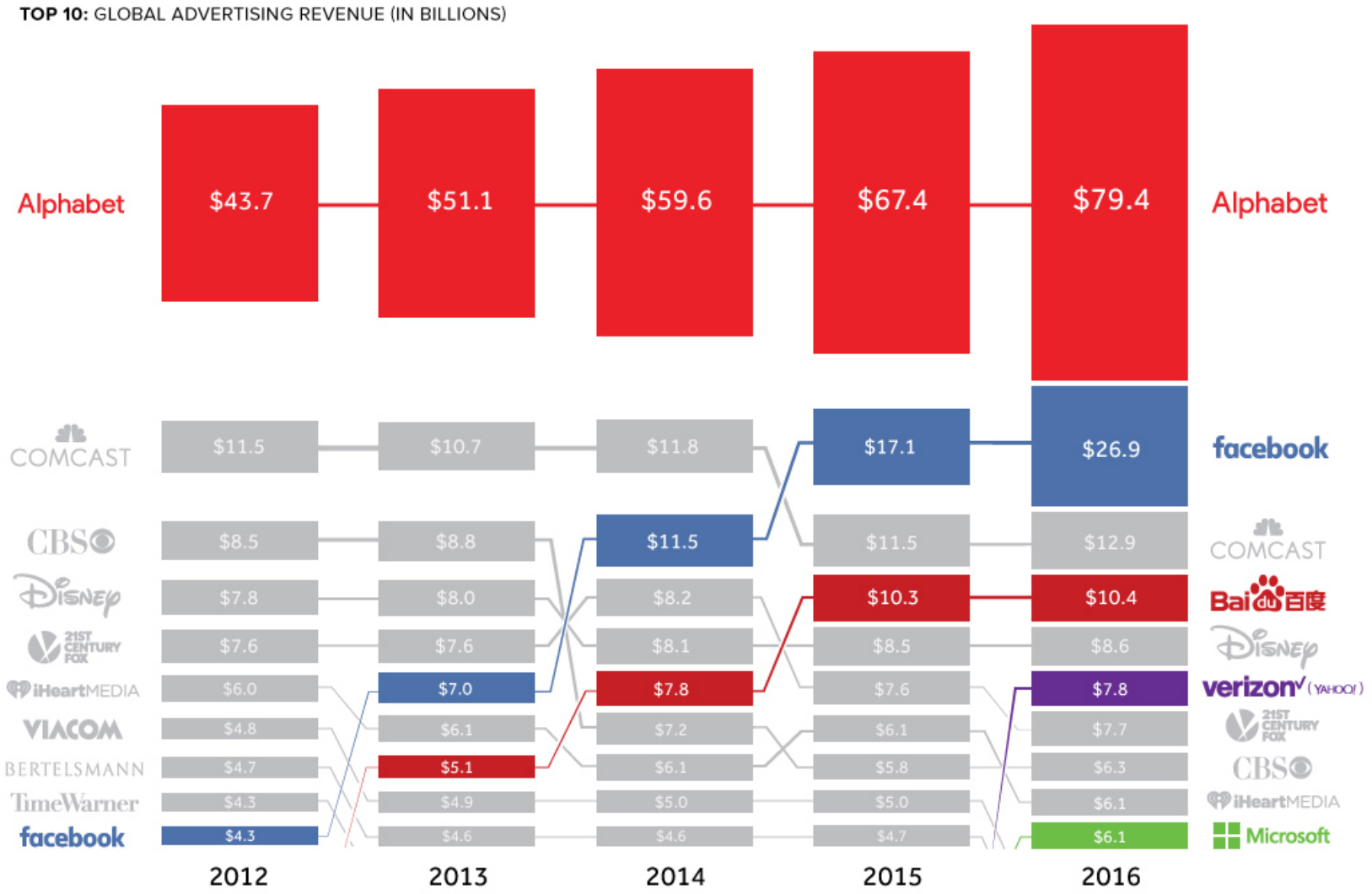
Tell me …

… why do you want to do this course?

Personalization …



Online Advertising

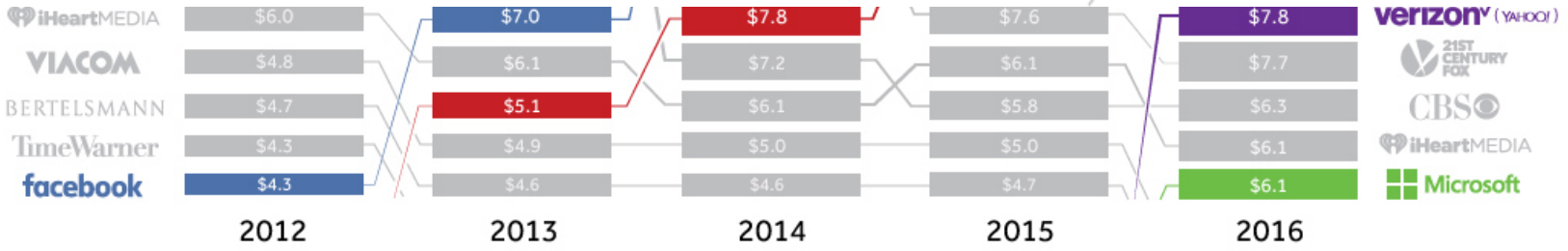


SOURCE: Bloomberg, Zenith Media

Online Advertising

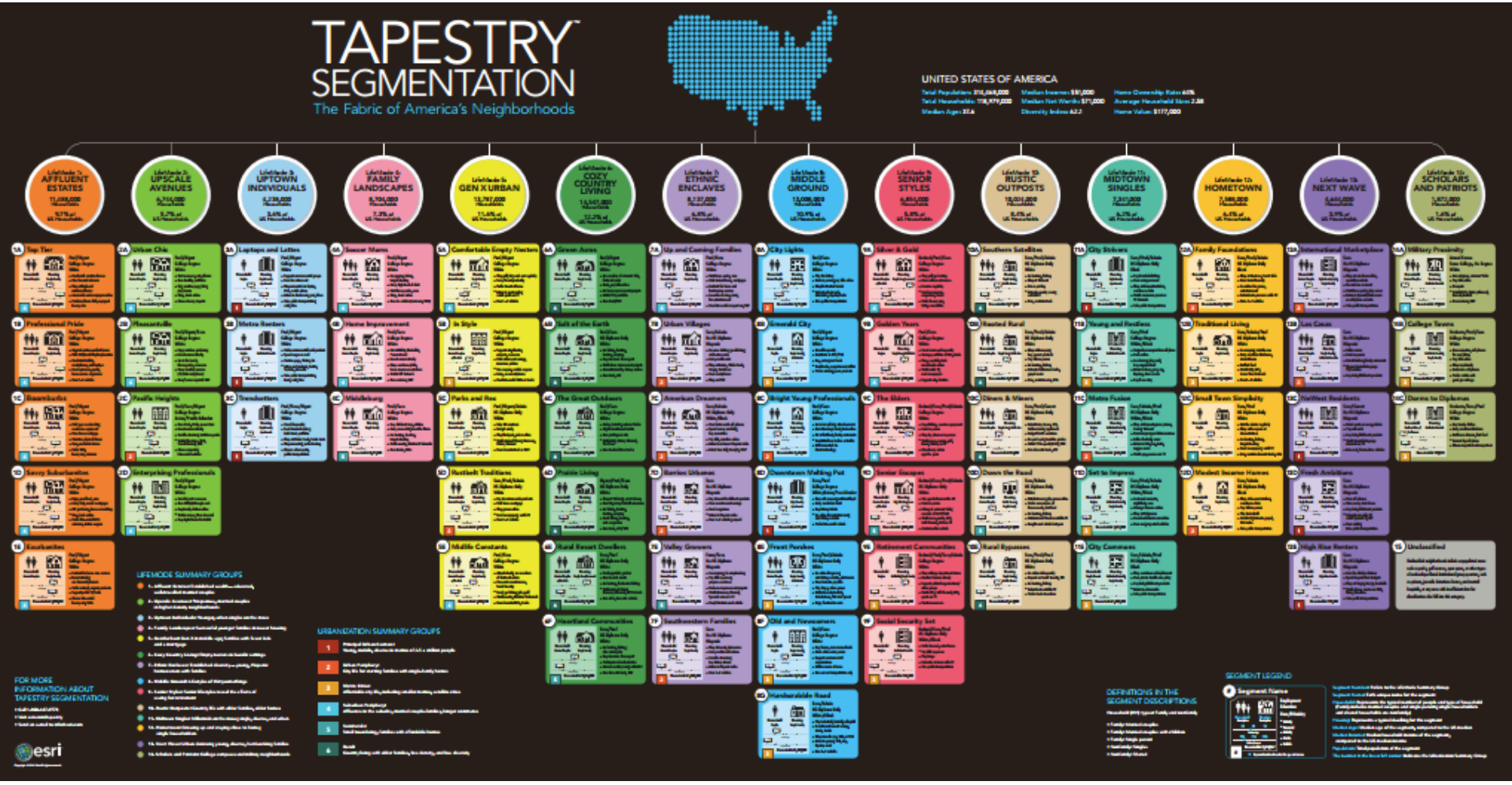
Ad-Supported Internet Brings Over $1 Trillion To The U.S. Economy, Representing 6 Percent Of Country's Total GDP, According To IAB Study Led By Harvard Business School Professor

03.15.17

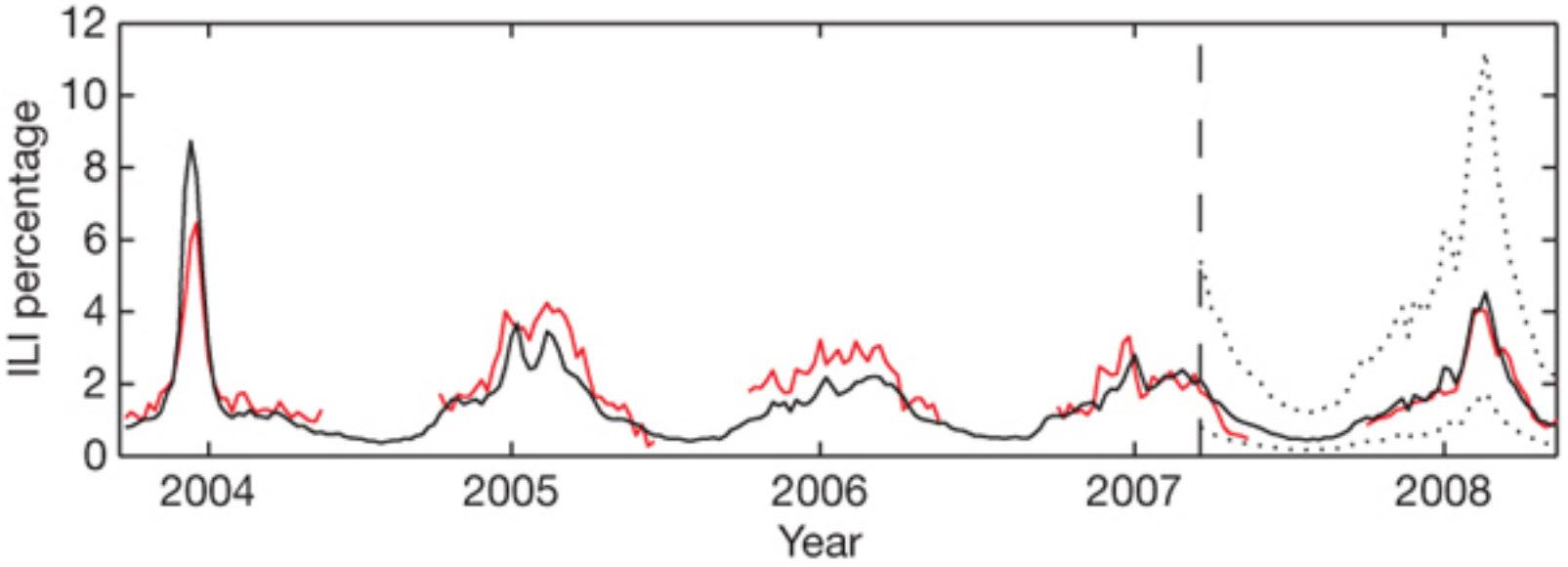


SOURCE: Bloomberg, Zenith Media

visualcapitalist.com


TAPESTRY SEGMENTATION
The Fabric of America's Neighborhoods
UNITED STATES OF AMERICA
LIFEMODE SUMMARY GROUPS
URBANIZATION SUMMARY GROUPS
FOR MORE INFORMATION ABOUT TAPESTRY SEGMENTATION
DEFINITIONS IN THE SEGMENT DESCRIPTIONS
SEGMENT LEGEND
esri

Health



Red: official numbers from Center for Disease Control and Prevention; weekly
Black: based on Google search logs; daily (potentially instantaneously)

Detecting influenza epidemics using search engine query data

http://www.nature.com/nature/journal/v457/n7232/full/nature07634.html

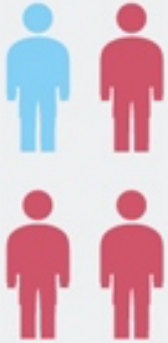
IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. ABILIFY (aripiprazole)
Schizophrenia

2. NEXIUM (esomeprazole)
Heartburn

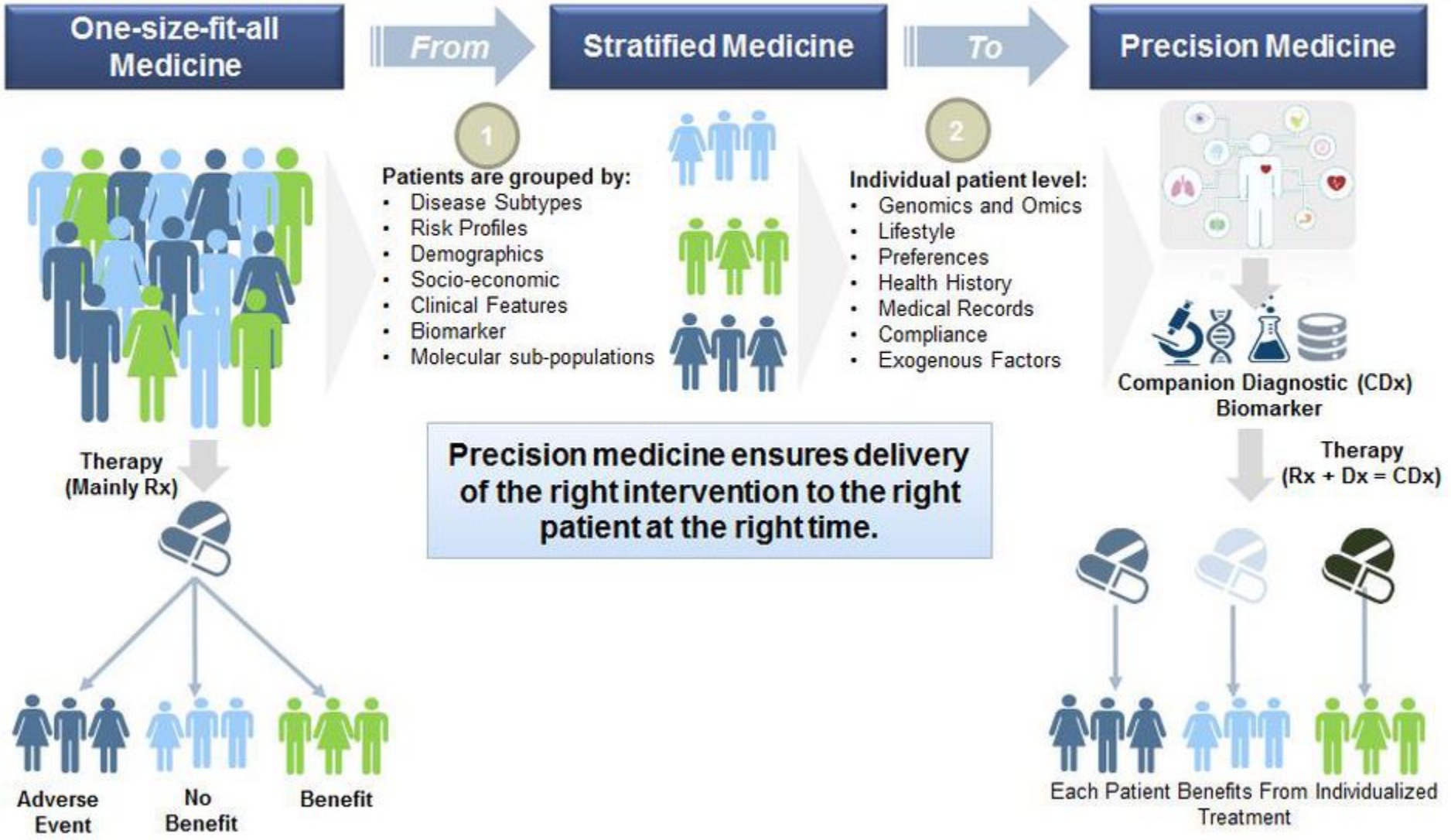
3. HUMIRA (adalimumab)
Arthritis



4. CRESTOR (rosuvastatin)
High cholesterol

https://www.nature.com/news/personalized-medicine-time-for-one-person-trials-1.17411

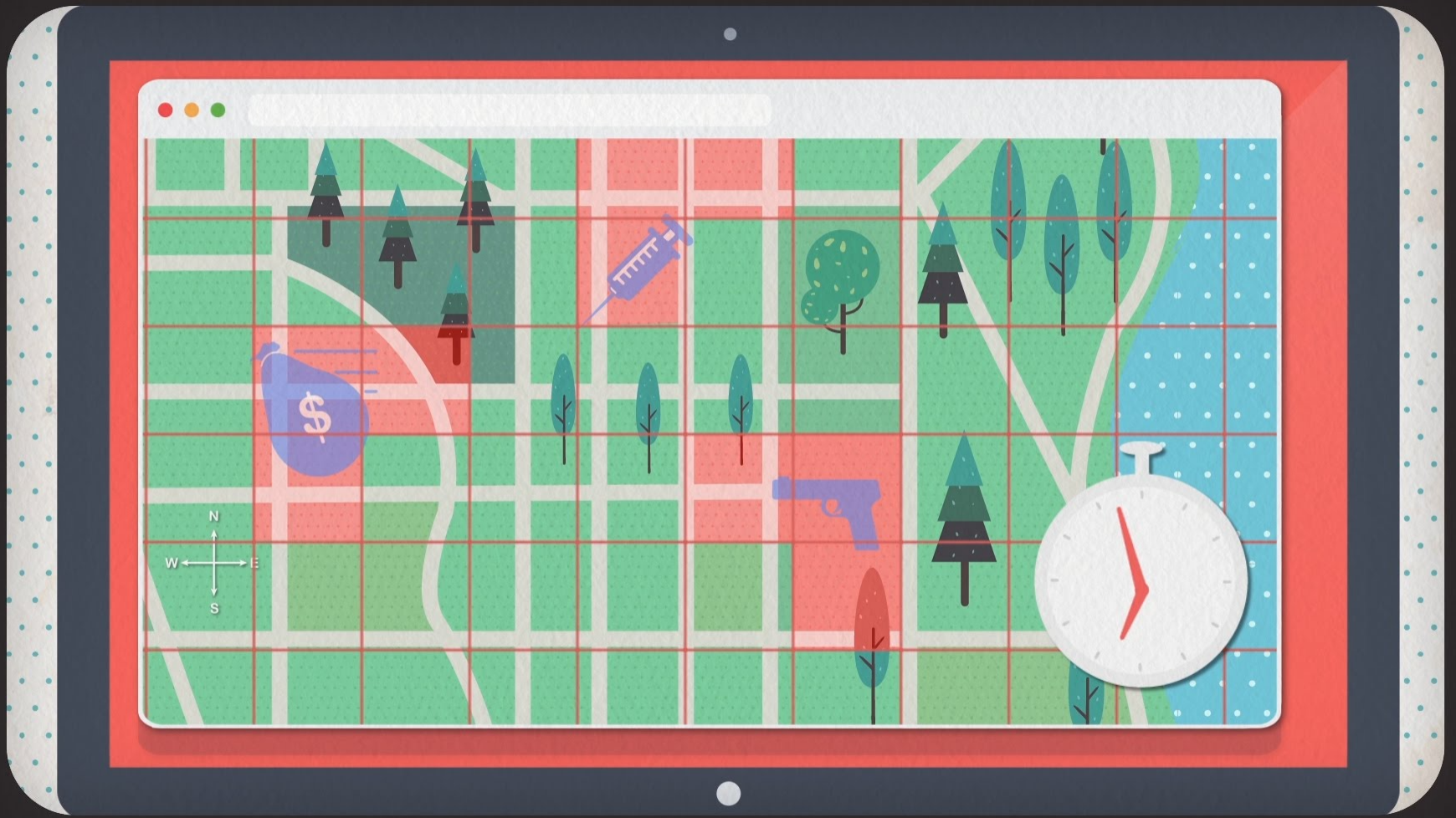
Precision Medicine



Source: forbes.com

Predictive Policing

Predictive Policing



The dark side of the force…

http://ragekg.deviantart.com/art/The-Dark-Side-of-the-Force-174559980

39% of the experts agree…

Thanks to many changes, including the building of "the Internet of Things," human and machine analysis of ***Big Data will cause more problems than it solves*** *by 2020. The existence of huge data sets for analysis will* ***engender false confidence in our predictive powers*** *and will lead many to make* ***significant and hurtful mistakes****. Moreover, analysis of Big Data will be* ***misused by powerful people and institutions with selfish agendas*** *who manipulate findings to make the case for what they want. And the advent of Big Data has a harmful impact because it* ***serves the majority (at times inaccurately) while diminishing the minority*** *and ignoring important outliers. Overall, the rise of Big Data is a big negative for society in nearly all respects.*

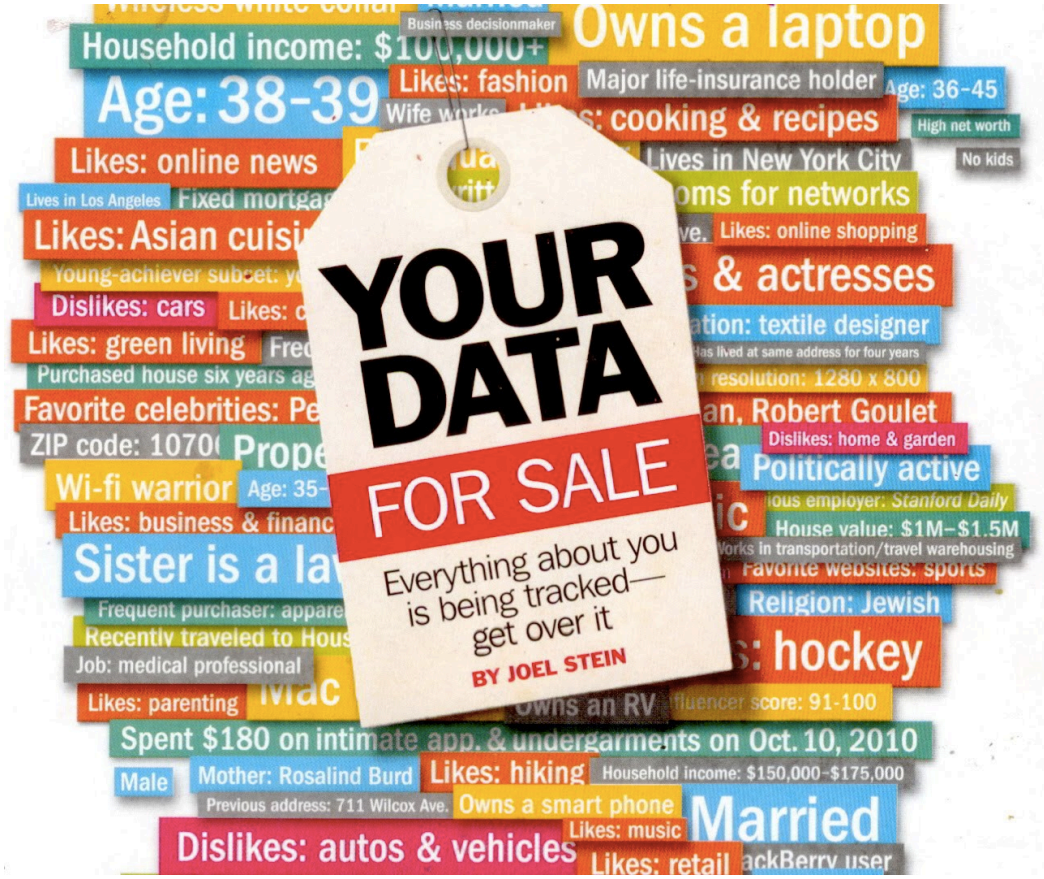
— 2012 Pew Research Center Report

http://pewinternet.org/Reports/2012/Future-of-Big-Data/Overview.aspx

Harm due to personalized data analytics …

- Privacy
- Fairness

Where is the data coming from?



Where is the data coming from?

- Census surveys
- IRS Records

- Medical records
- Insurance records

- Search logs
- Browse logs
- Shopping histories

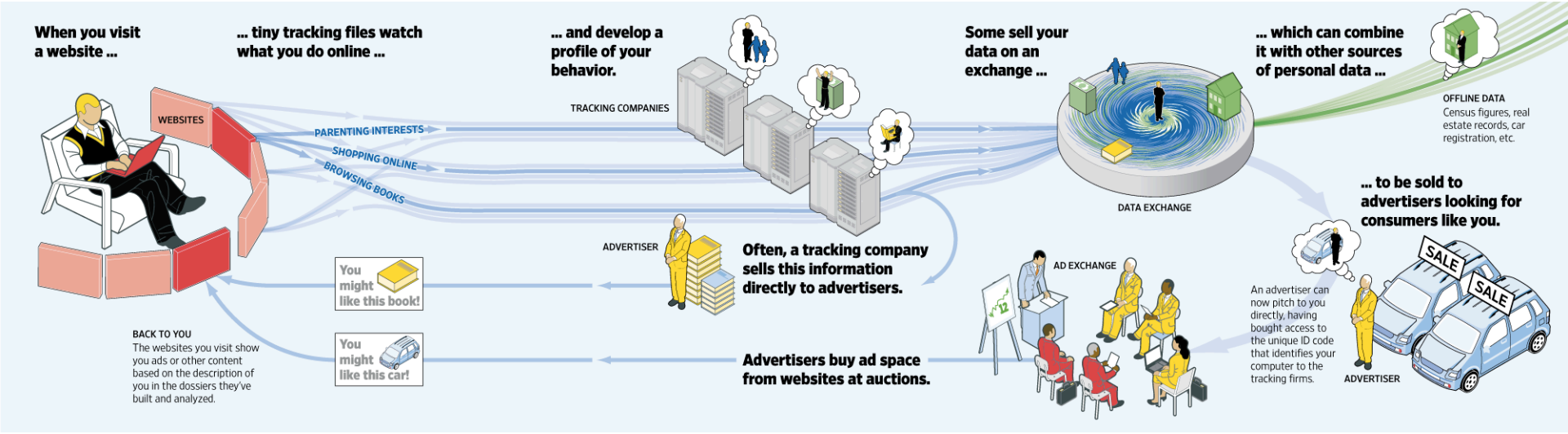
- Photos
- Videos

- Smart phone Sensors
- Mobility trajectories

- …

Very sensitive information …

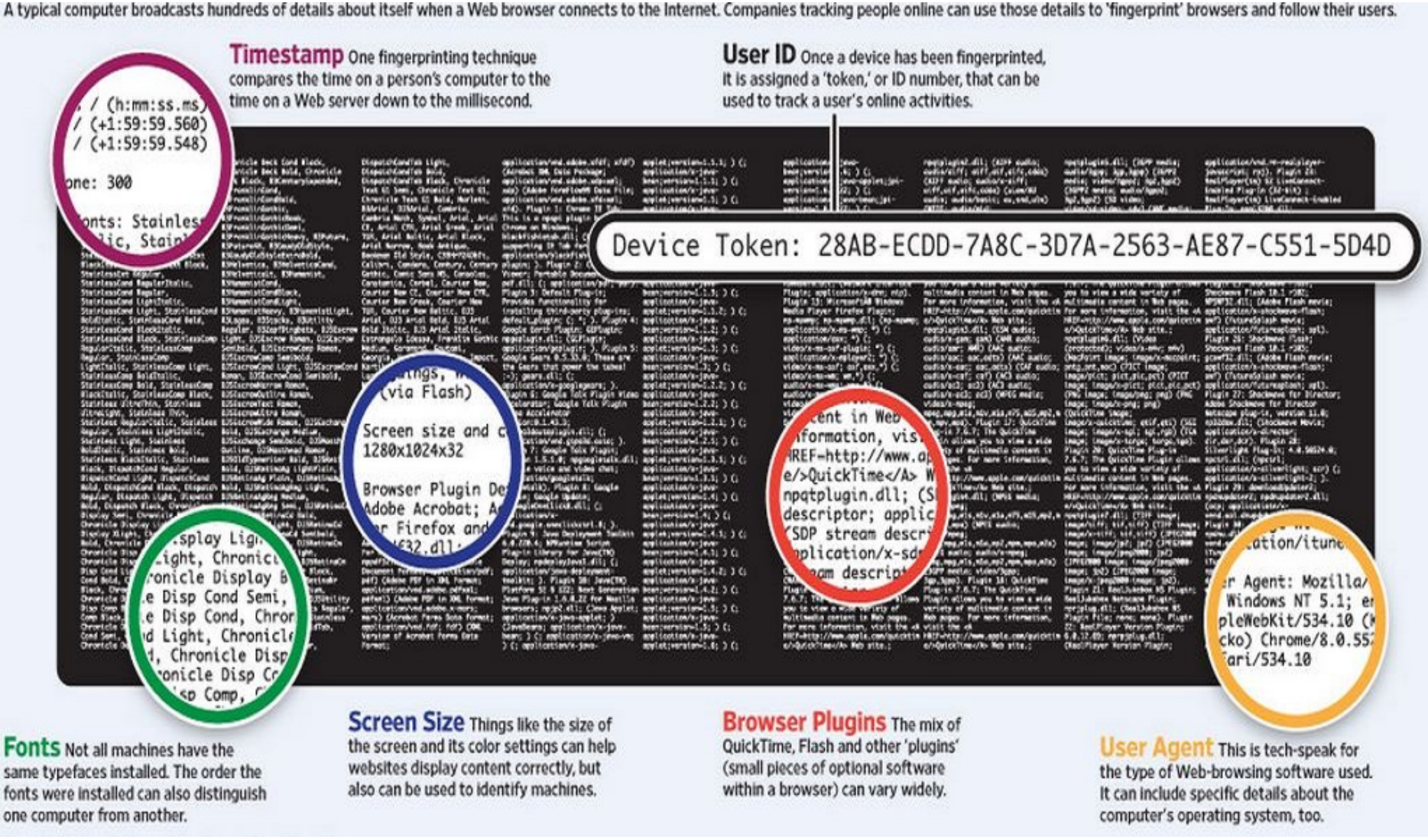
How is this data collected?



http://graphicsweb.**wsj.com**/documents/divSlider/media/ecosystem100730.png

Isn’t my data anonymous ?

Device Fingerprinting



PANOPTICLICK 3.0

Is your browser safe against tracking?

Your browser fingerprint **appears to be unique** among the 2,050,572 tested in the past 45 days.

Currently, we estimate that your browser has a fingerprint that conveys **at least 20.97 bits of identifying information.**

https://panopticlick.eff.org/

Let's get rid of unique identifiers …

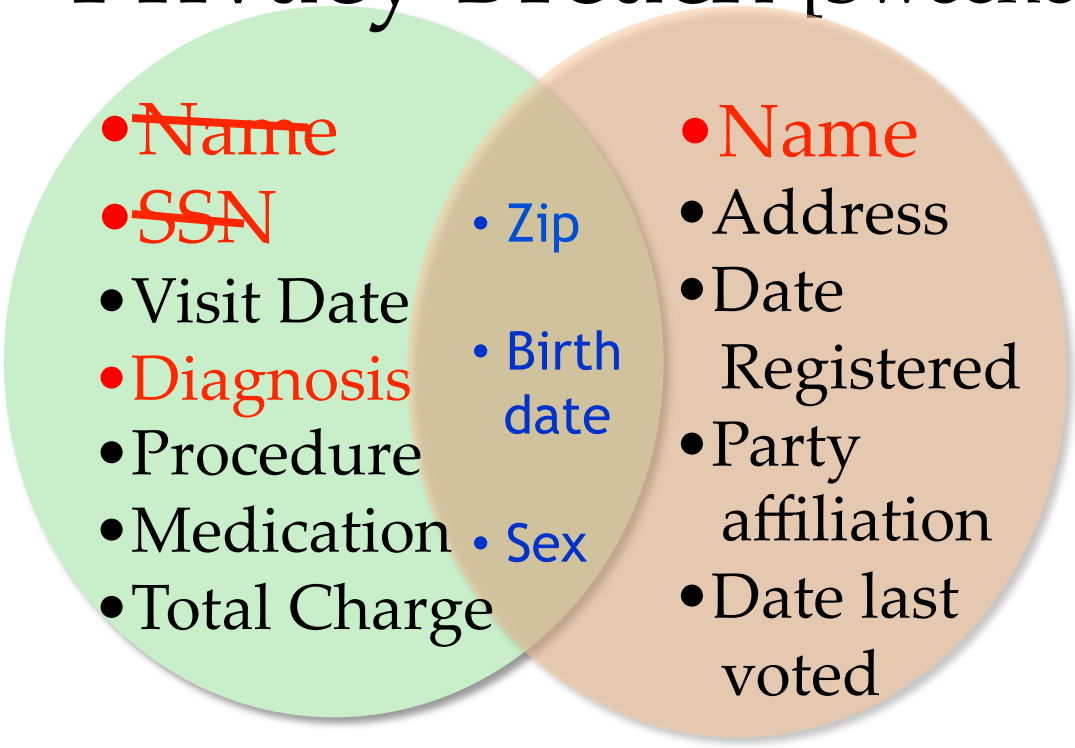
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]

- ~~Name~~
- ~~SSN~~
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge

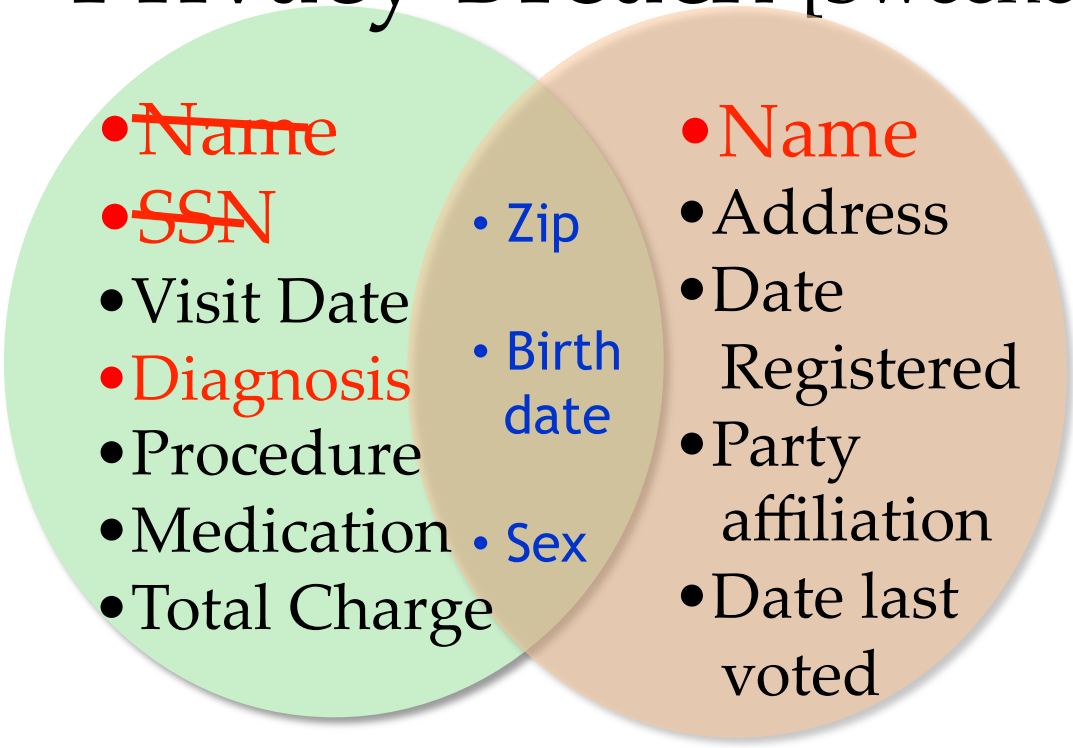
- Zip
- Birth date
- Sex

Medical Data

The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



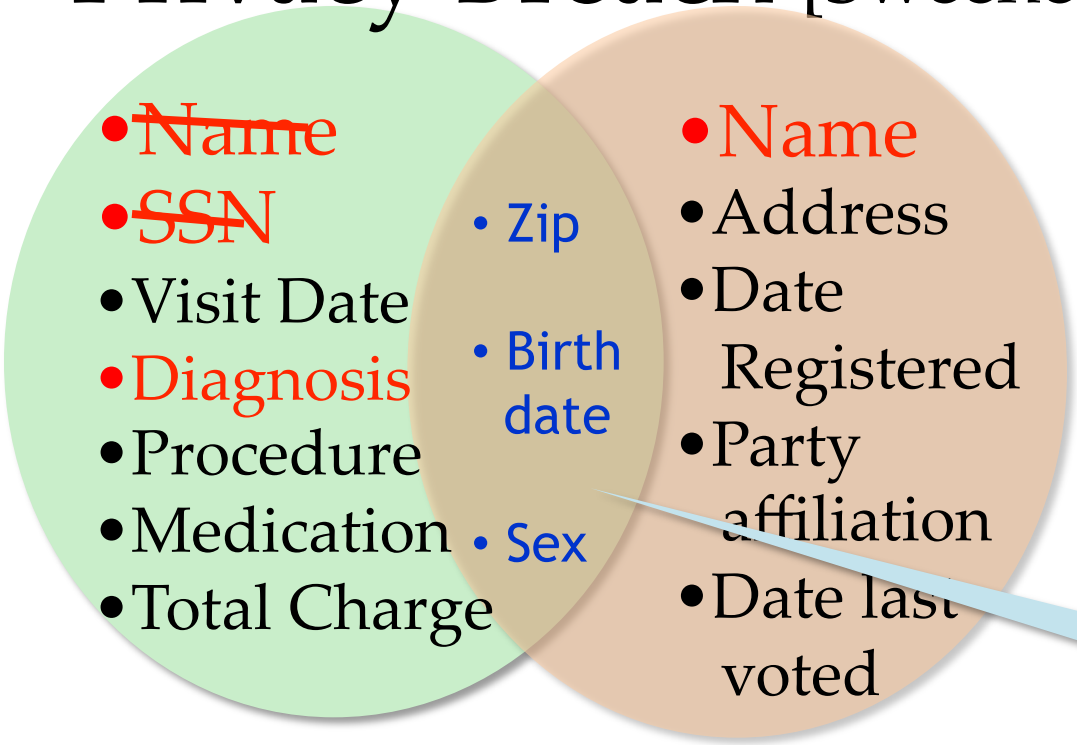
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



- Governor of MA **uniquely identified** using ZipCode, Birth Date, and Sex.

Name linked to Diagnosis

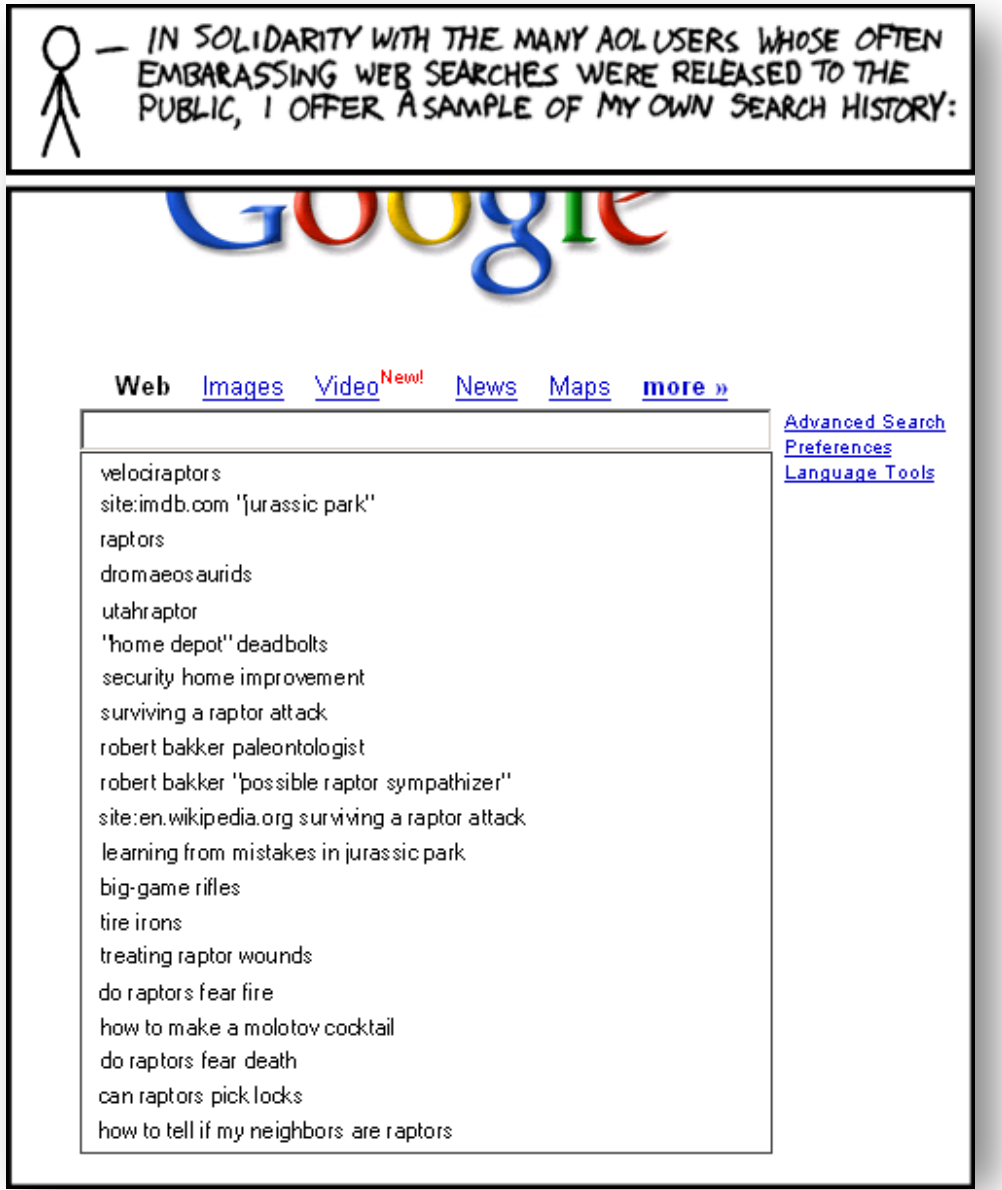
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



- 87 % of US population **uniquely identified** using ZipCode, Birth Date, and Sex.

Quasi Identifier

AOL data publishing fiasco



AOL data publishing fiasco …

Ashwin222	Uefa cup
Ashwin222	Uefa champions league
Ashwin222	Champions league final
Ashwin222	Champions league final 2013
Jun156	exchangeability
Jun156	Proof of deFinitti's theorem
Brett12345	Zombie games
Brett12345	Warcraft
Brett12345	Beatles anthology
Brett12345	Ubuntu breeze
Austin222	Python in thought
Austin222	Enthought Canopy

User IDs replaced with random numbers

865712345	Uefa cup
865712345	Uefa champions league
865712345	Champions league final
865712345	Champions league final 2013
236712909	exchangeability
236712909	Proof of deFinitti's theorem
112765410	Zombie games
112765410	Warcraft
112765410	Beatles anthology
112765410	Ubuntu breeze
865712345	Python in thought
865712345	Enthought Canopy

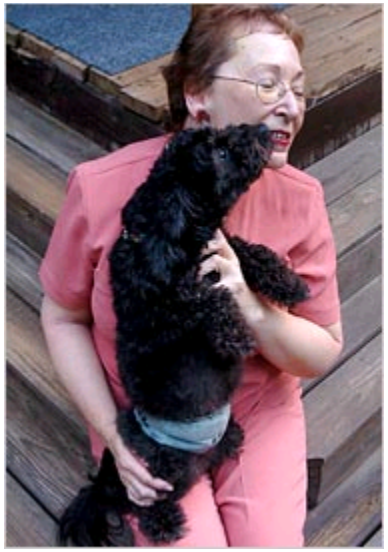
Privacy Breach

[NYTimes 2006]

A Face Is Exposed for AOL Searcher No. 4417749

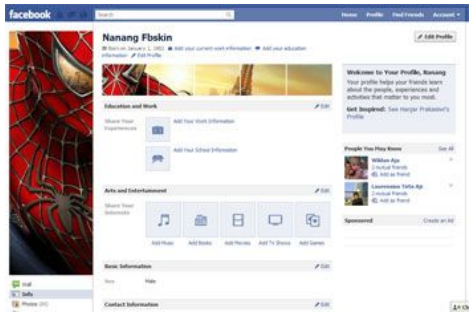
By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E-
THIS



Machine learning models can reveal sensitive information

Facebook Profile



+

Online Data

- who live in the **United States**
- who live within 50 miles of **Staten Island, NY**
- between the ages of **23** and **27** inclusive
- who are **female**
- who are connected to **DogAnd PonyShow**
- in one of the categories: **Pop Culture**, **Science Fiction/Fantasy**, **Alternative**, **Rock**, **Classic Rock** or **iPhone**

	Number of Impressions
+ Who are interested in **Men**	25
+ Who are interested in **Women**	0

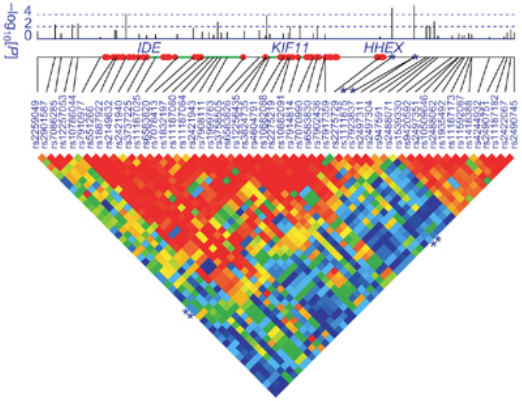
Facebook's learning algorithm uses private information to predict match to ad

[Korolova JPC 2011]

Genome wide association studies

[Homer et al PLOS Genetics 08]

Results of a GWAS study



High density SNP profile of Bob

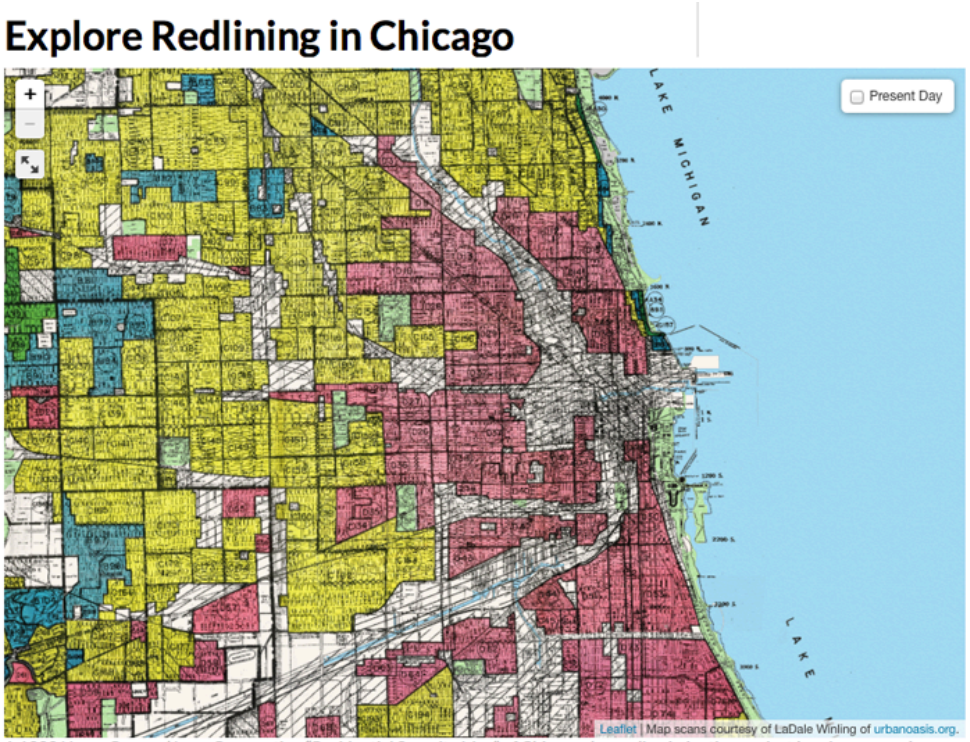
Did Bob participate in the study

Harm due to personalized data analytics …

- Privacy
- Fairness

The red side of learning

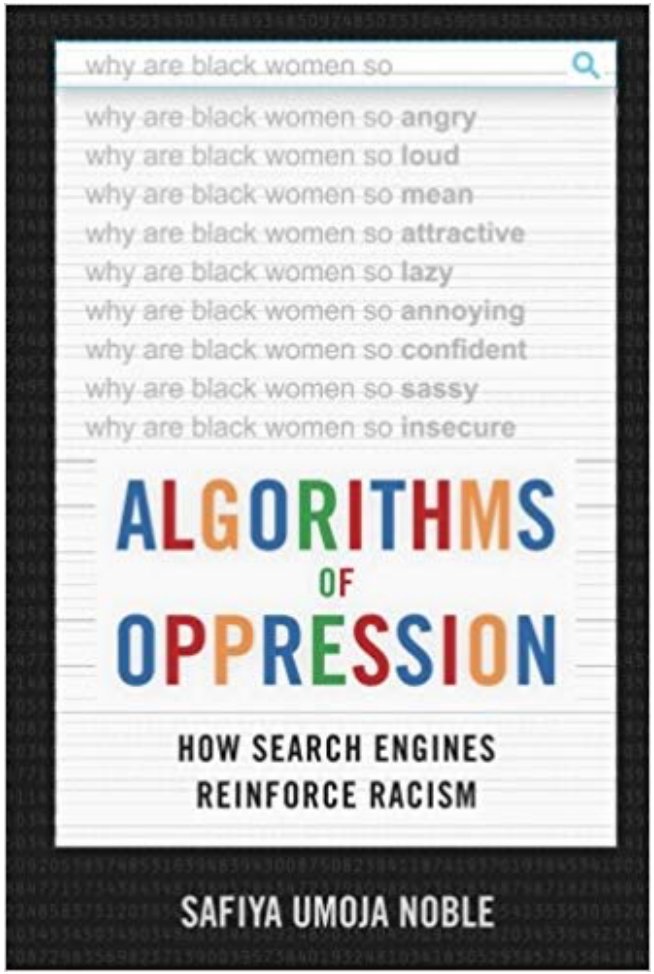
- **Redlining**: the practice of denying, or charging more for, services such as banking, insurance, access to health care, or even supermarkets, or denying jobs to residents in particular, often racially determined, areas.



A 1939 Home Owners' Loan Corporation "Residential Security Map" of Chicago shows discrimination against low-income and minority neighborhoods. The residents of the areas marked in red (representing "hazardous" real-estate markets) were denied FHA-backed mortgages. (Map development by Frankie Dintino)

Predictive Policing

- Predictive policing systems use machine learning algorithms to predict crime.
- But … the algorithms learn … patterns not about crime, per se, but about how police record crime.
- This can amplify existing biases



https://www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html

TheUpshot

HIDDEN BIAS

When Algorithms Discriminate

By Claire Cain Miller

July 9, 2015

147

The online world is shaped by forces beyond our control, determining the stories we read on Facebook, the people we meet on OkCupid and the search results we see on Google. Big data is used to make decisions about health care, employment, housing, education and policing.

But can computer programs be discriminatory?

There is a widespread belief that software and algorithms that rely on data are objective. But software is not free of human influence. Algorithms are written and maintained by people, and machine learning algorithms adjust what they do based on people's behavior. As a result, say researchers in computer science, ethics and law, algorithms can reinforce human prejudices.

Google's online advertising system, for instance, showed an ad for high-income jobs to men much more often than it showed the ad to women, a new study by Carnegie Mellon University researchers found.

Research from Harvard University found that ads for arrest records were significantly more likely to show up on searches for distinctively black names or a historically black fraternity. The Federal Trade Commission said advertisers are able to target people who live in low-income neighborhoods with high-interest loans.


BRACE YOURSELF
DEEP LEARNING IS COMING

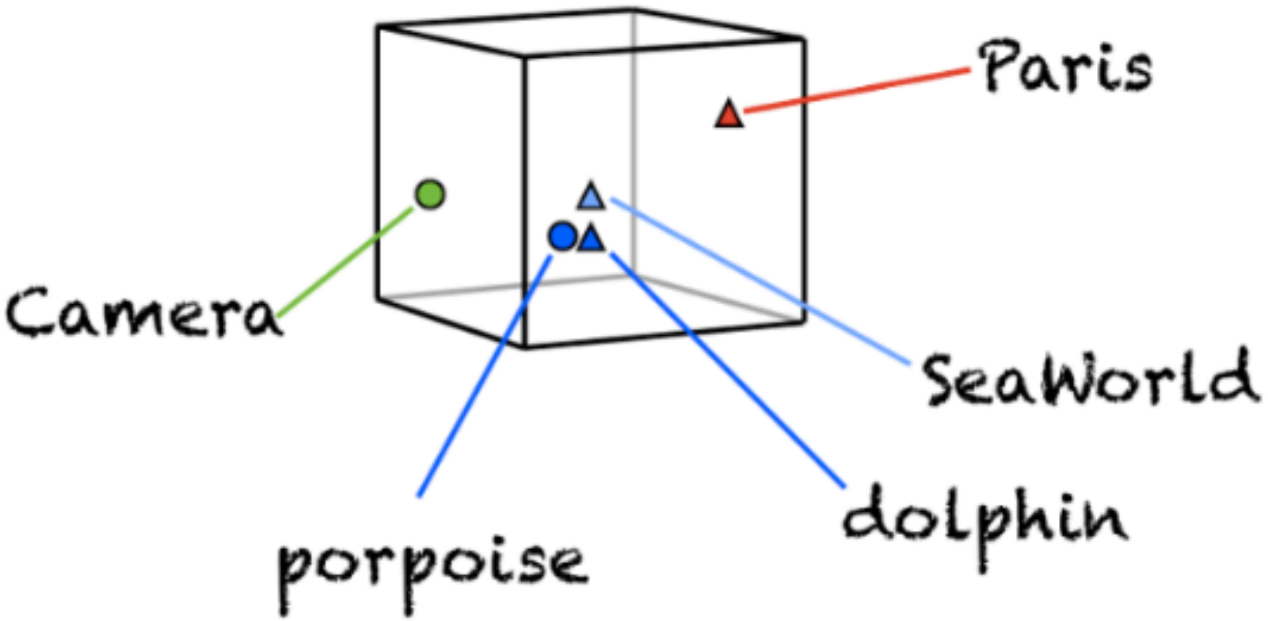
Deep Learning

Incredibly powerful tool for …

- Extracting regularities from data according to a given data

- Amplifying bias!

Word embeddings



Can convert words to vectors of numbers - at the hearth of most NLP applications with deep learning

http://slides.com/simonescardapane/the-dark-side-of-deep-learning

Embeddings are *highly* sexists!

Bolukbasi, T., Chang, K.W., Zou, J., Saligrama, V. and Kalai, A., 2016. **Quantifying and reducing stereotypes in word embeddings**. *arXiv preprint arXiv:1606.06121*.

http://slides.com/simonescardapane/the-dark-side-of-deep-learning

Deep Learning

Incredibly powerful tool for …

- Extracting regularities from data according to a given data

- Amplifying privacy concerns!

Given access to a black-box classifier, can we infer whether a specific example was part of the training dataset?

We can with **shadow training**:

Shokri, R., Stronati, M., Song, C. and Shmatikov, V., 2017, May. **Membership inference attacks against machine learning models.** In *2017 IEEE Symposium on Security and Privacy (SP),* (pp. 3-18). IEEE.

Dataset	*Training Accuracy*	*Testing Accuracy*	*Attack Precision*
Adult	0.848	0.842	0.503
MNIST	0.984	0.928	0.517
Location	1.000	0.673	0.678
Purchase (2)	0.999	0.984	0.505
Purchase (10)	0.999	0.866	0.550
Purchase (20)	1.000	0.781	0.590
Purchase (50)	1.000	0.693	0.860
Purchase (100)	0.999	0.659	0.935
TX hospital stays	0.668	0.517	0.657

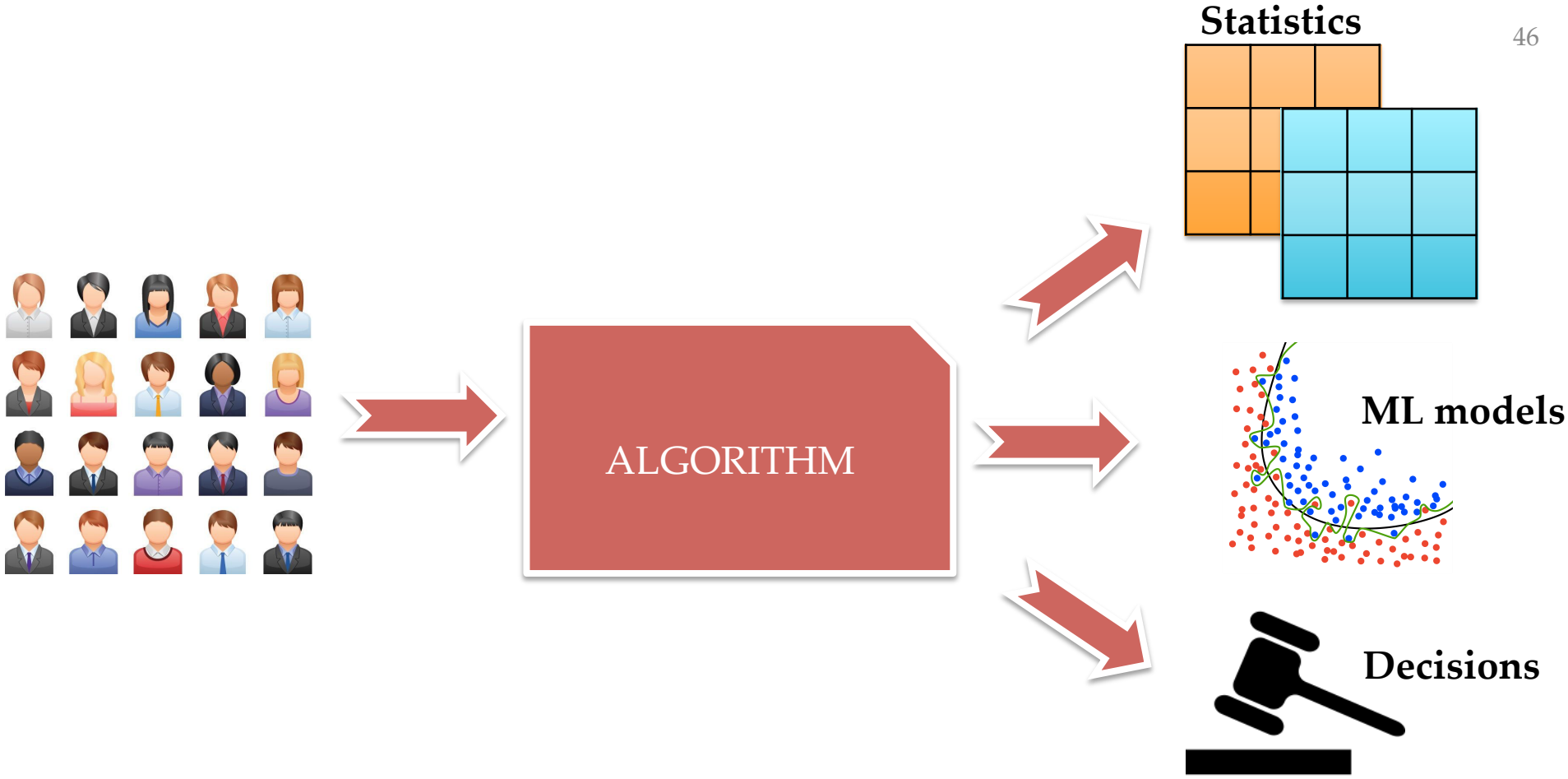
TABLE II: Accuracy of the Google-trained models and the corresponding attack precision.

This course:

http://www.webvisionsevent.com/userfiles/lightsabercrop_large_verge_medium_landscape.jpg

You will ...

- mathematically formulate privacy.
- mathematically formulate fairness.



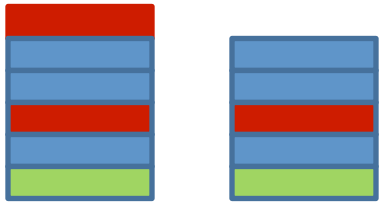
Privacy Question: Does releasing the output of the algorithm result increase the privacy risk (disclosure of identity, sensitive values, etc.) of an individual?

Fairness Question: Does the algorithm disadvantage or discriminate against an individual or a group of individuals?

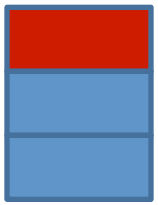
Differential Privacy

For every pair of inputs that differ in one row

For every output …



D_1 D_2



O

Adversary should not be able to distinguish between any D_1 and D_2 based on any O

$$\log\left(\frac{\Pr[A(D_1) = O]}{\Pr[A(D_2) = O]}\right) < \varepsilon \quad (\varepsilon > 0)$$

You will ...

- mathematically formulate privacy.
- mathematically formulate fairness.

- design algorithms to ensure privacy
- design algorithms to ensure fairness

Differential Privacy in practice

OnTheMap ***[ICDE 2008]***

[CCS 2014]

[Apple WWDC 2016]

You will ...

- mathematically formulate privacy.
- mathematically formulate fairness.

- design algorithms to ensure privacy
- design algorithms to ensure fairness

- do research into the interplay between privacy and fairness.

Course Format

- Module 1: Intro to Privacy
- Module 2: Intro to Fairness

In-class Exercise
Lectures
Mini-project

- Module 3: Algorithms for Privacy
- Module 4: Algorithms for Fairness

Read papers
Mini-critiques
Research Project

$$\forall i \in [n], d \in S, \left| \ln \frac{\Pr[T_i \in T | d_i = d]}{\Pr[T_i \in T | d_i = \text{NULL}]} \right|$$

$$\left| \frac{\mathcal{A}_{client}(d) = t]}{\mathcal{A}_{client}(\text{null}) = t]} \right| \leq \ln \left(\frac{e^{\epsilon}}{1 + e^{\epsilon}} \cdot \frac{1 + e^{\epsilon}}{1} \right) = \epsilon$$

$$\alpha = \frac{3k + 2c_{\epsilon}\sqrt{\ln(6mk/\beta)}}{\sqrt{n}} = O\left(\frac{\sqrt{\log(}}{\epsilon\sqrt{}} \right.$$

$$\alpha = \frac{3k + c_{\epsilon}\sqrt{\ln(4mk/\beta)}}{\sqrt{n}} = O\left(\frac{\sqrt{\log(p/\beta)}}{\epsilon\sqrt{n}} \right)$$

$$\left\{ \left(\frac{v[j] \cdot b[j] + 1}{2} \right), \forall j \in [m] \right\}$$

What we expect you to know …

- Strong background in
  - Probability
  - Proof techniques

- Some knowledge of
  - Programming with Python
  - Machine learning
  - Statistics
  - Algorithms

Misc. course info

- **Website**: http://sites.duke.edu/cs590f18privacyfairness/
  - Schedule (with links to lecture slides, readings, homeworks, etc.)
- **Grading**
  - Mini-projects: 20% (x2)
  - Mini-critiques: 10%
  - Class participation: 10%
    - Attending class!
  - Project: 40%
- **Sakai** for grades

Duke Community Standard

- See course website
- Mini-projects and paper critiques are individual work.
- Group discussion okay (and encouraged), but
  - Acknowledge help you receive from others
  - Make sure you "own" your solution
- All suspected cases of violation will be aggressively pursued

Privacy & Fairness in Data Science

CompSci 590.01 Fall 2018

DUKE
COMPUTER SCIENCE