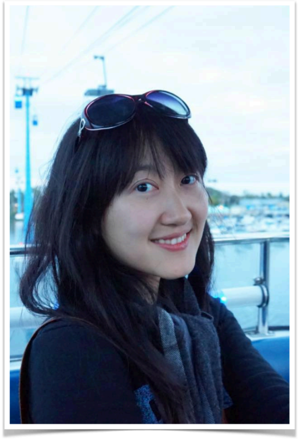
Privacy & Fairness in Data Science

CS848 Fall 2019

UNIVERSITY OF WATERLOO | DSg Data Systems Group

Instructor



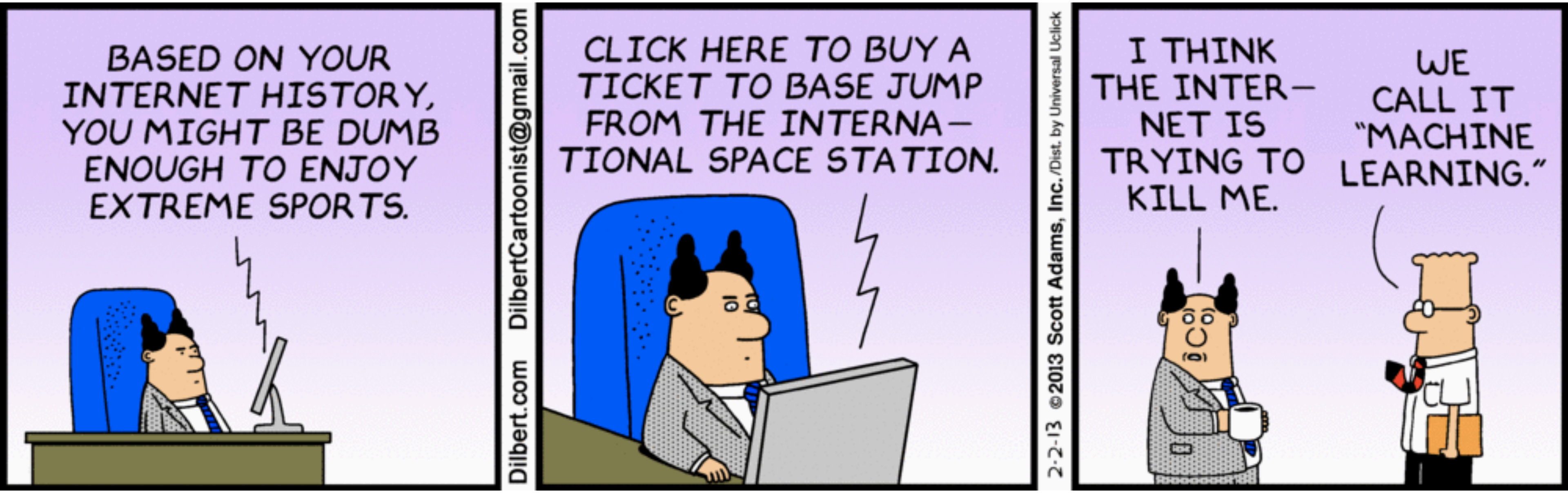
Xi He:

- Research interest: privacy and fairness for big-data management and analysis
- CS848, Fall 2019:
  - Tue: 3:00pm - 5:50pm (DC2568)

Tell me …

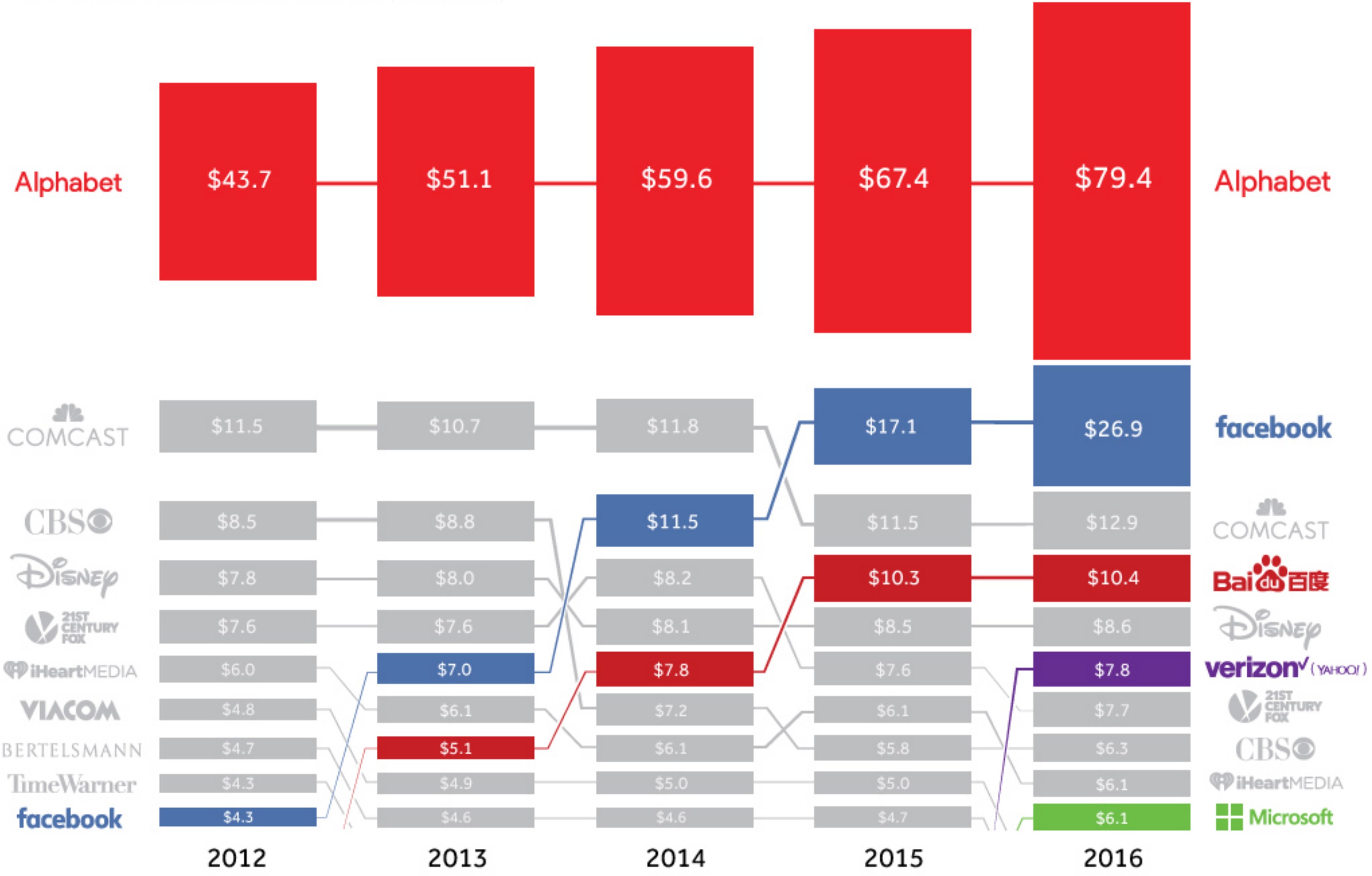
… why do you want to do this course?

Personalization …


BASED ON YOUR INTERNET HISTORY, YOU MIGHT BE DUMB ENOUGH TO ENJOY EXTREME SPORTS.
Dilbert.com DilbertCartoonist@gmail.com
CLICK HERE TO BUY A TICKET TO BASE JUMP FROM THE INTERNATIONAL SPACE STATION.
2-2-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick
I THINK THE INTERNET IS TRYING TO KILL ME.
WE CALL IT "MACHINE LEARNING."

Online Advertising

TOP 10: GLOBAL ADVERTISING REVENUE (IN BILLIONS)

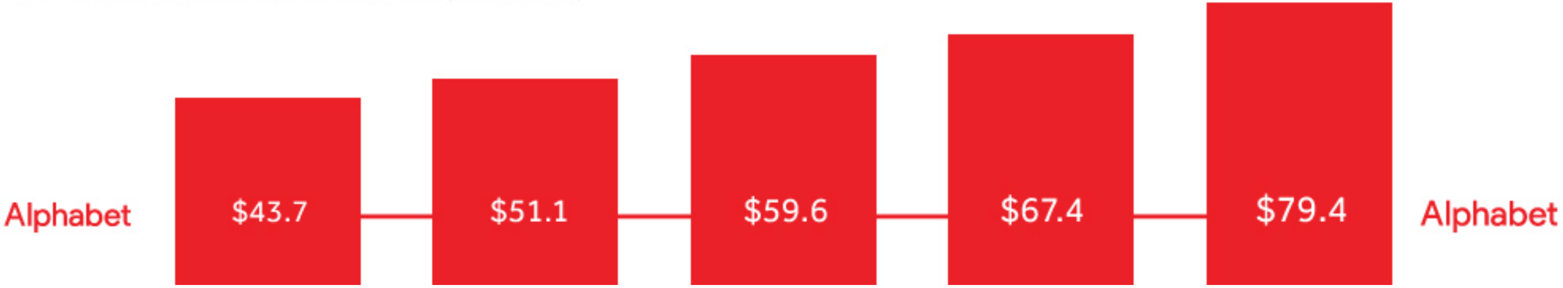


SOURCE: Bloomberg, Zenith Media

visualcapitalist.com

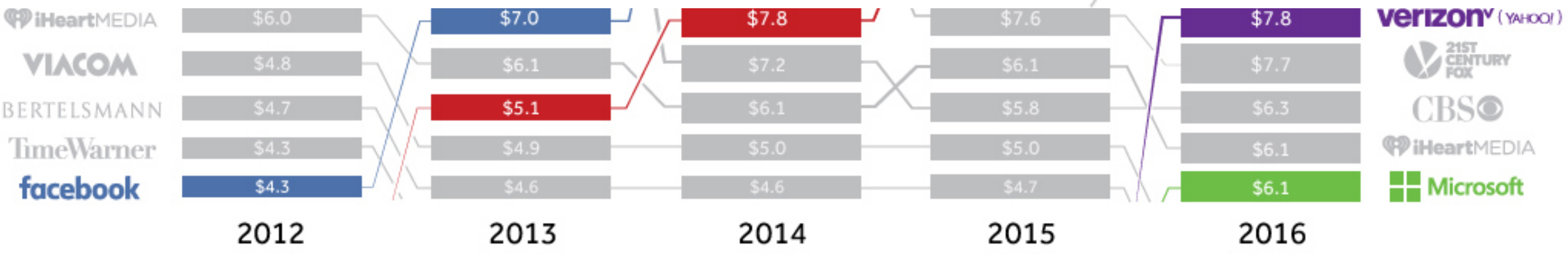
Online Advertising

TOP 10: GLOBAL ADVERTISING REVENUE (IN BILLIONS)



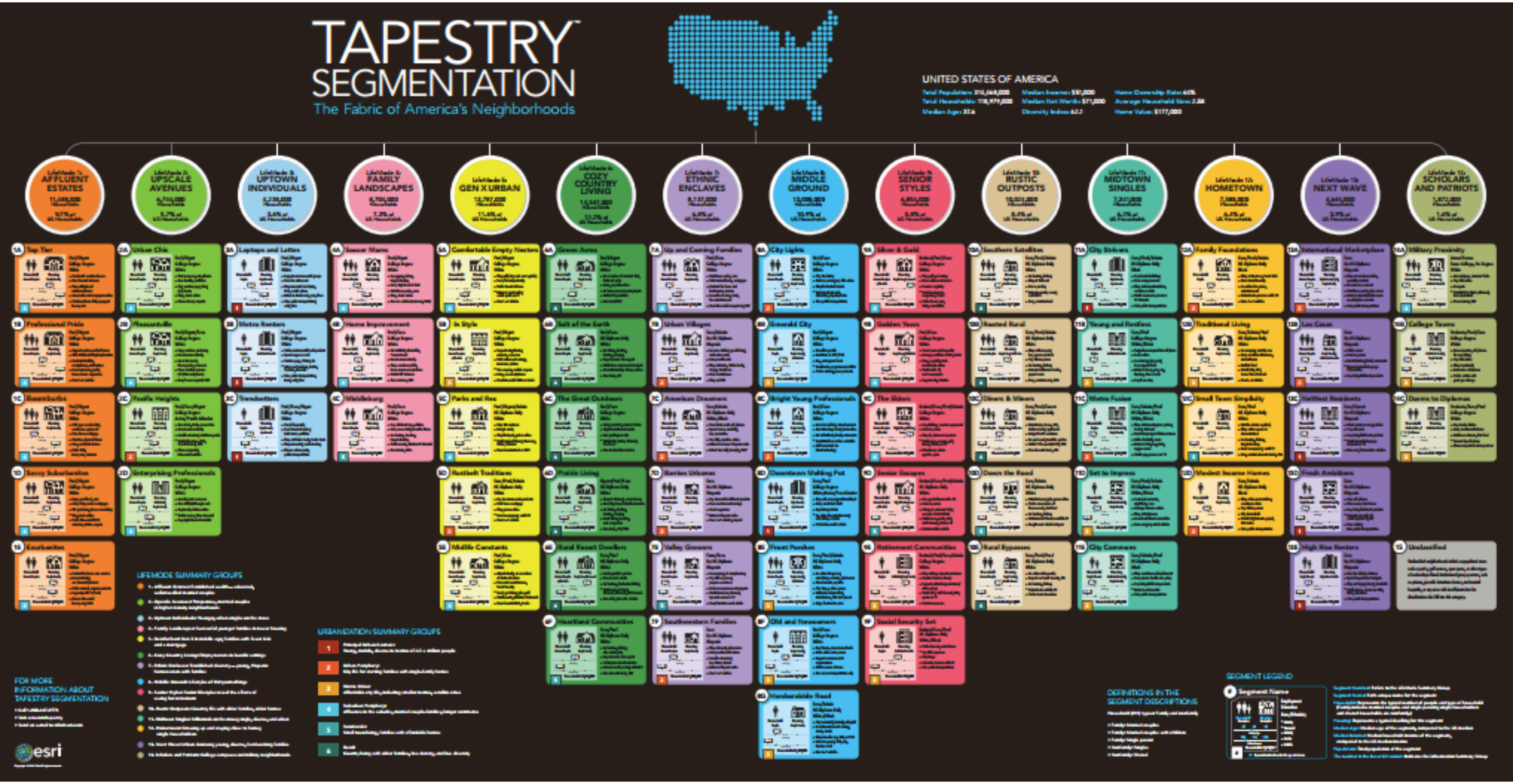
Ad-Supported Internet Brings Over $1 Trillion To The U.S. Economy, Representing 6 Percent Of Country's Total GDP, According To IAB Study Led By Harvard Business School Professor

03.15.17

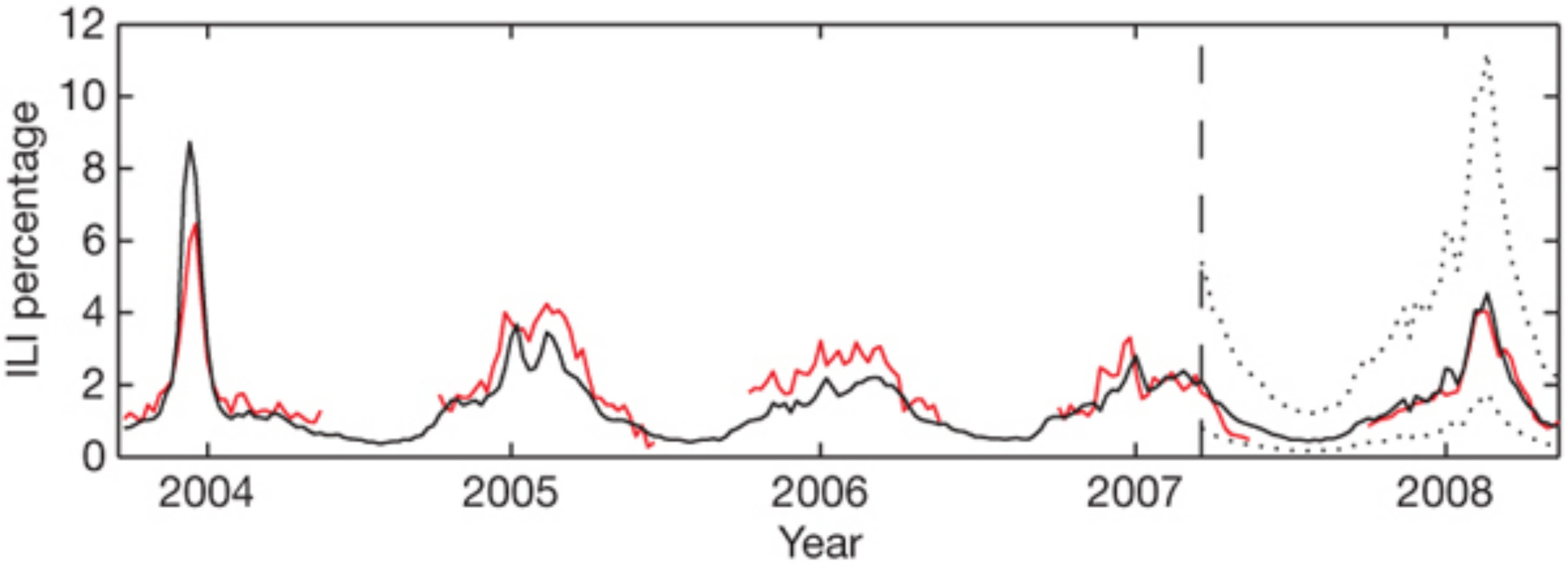


SOURCE: Bloomberg, Zenith Media

visualcapitalist.com


TAPESTRY™
SEGMENTATION
The Fabric of America's Neighborhoods
UNITED STATES OF AMERICA
Total Population: 316,468,000
Total Households: 118,979,000
Median Age: 37.6
Median Income: $51,000
Median Net Worth: $71,000
Diversity Index: 62.2
Home Ownership Rate: 64%
Average Household Size: 2.58
Home Value: $177,000
LifeMode 1: AFFLUENT ESTATES 11,088,000 Households 9.7% of US Households
LifeMode 2: UPSCALE AVENUES 6,716,000 Households 5.7% of US Households
LifeMode 3: UPTOWN INDIVIDUALS 4,238,000 Households 3.6% of US Households
LifeMode 4: FAMILY LANDSCAPES 8,704,000 Households 7.3% of US Households
LifeMode 5: GEN X URBAN 13,787,000 Households 11.6% of US Households
LifeMode 6: COZY COUNTRY LIVING 14,547,000 Households 12.2% of US Households
LifeMode 7: ETHNIC ENCLAVES 8,137,000 Households 6.8% of US Households
LifeMode 8: MIDDLE GROUND 13,008,000 Households 10.9% of US Households
LifeMode 9: SENIOR STYLES 6,854,000 Households 5.8% of US Households
LifeMode 10: RUSTIC OUTPOSTS 10,024,000 Households 8.4% of US Households
LifeMode 11: MIDTOWN SINGLES 7,341,000 Households 6.2% of US Households
LifeMode 12: HOMETOWN 7,588,000 Households 6.4% of US Households
LifeMode 13: NEXT WAVE 4,646,000 Households 3.9% of US Households
LifeMode 14: SCHOLARS AND PATRIOTS 1,873,000 Households 1.6% of US Households
1A Top Tier
1B Professional Pride
1C Boomburbs
1D Savvy Suburbanites
1E Exurbanites
2A Urban Chic
2B Pleasantville
2C Pacific Heights
2D Enterprising Professionals
3A Laptops and Lattes
3B Metro Renters
3C Trendsetters
4A Soccer Moms
4B Home Improvement
4C Middleburg
5A Comfortable Empty Nesters
5B In Style
5C Parks and Rec
5D Rustbelt Traditions
5E Midlife Constants
6A Green Acres
6B Salt of the Earth
6C The Great Outdoors
6D Prairie Living
6E Rural Resort Dwellers
6F Heartland Communities
7A Up and Coming Families
7B Urban Villages
7C American Dreamers
7D Barrios Urbanos
7E Valley Growers
7F Southwestern Families
8A City Lights
8B Emerald City
8C Bright Young Professionals
8D Downtown Melting Pot
8E Front Porches
8F Old and Newcomers
8G Hardscrabble Road
9A Silver &amp; Gold
9B Golden Years
9C The Elders
9D Senior Escapes
9E Retirement Communities
9F Social Security Set
10A Southern Satellites
10B Rooted Rural
10C Diners &amp; Miners
10D Down the Road
10E Rural Bypasses
11A City Strivers
11B Young and Restless
11C Metro Fusion
11D Set to Impress
11E City Commons
12A Family Foundations
12B Traditional Living
12C Small Town Simplicity
12D Modest Income Homes
13A International Marketplace
13B Las Casas
13C NeWest Residents
13D Fresh Ambitions
13E High Rise Renters
14A Military Proximity
14B College Towns
14C Dorms to Diplomas
15 Unclassified
LIFEMODE SUMMARY GROUPS
URBANIZATION SUMMARY GROUPS
FOR MORE INFORMATION ABOUT TAPESTRY SEGMENTATION
DEFINITIONS IN THE SEGMENT DESCRIPTIONS
SEGMENT LEGEND
Segment Name
esri

Health



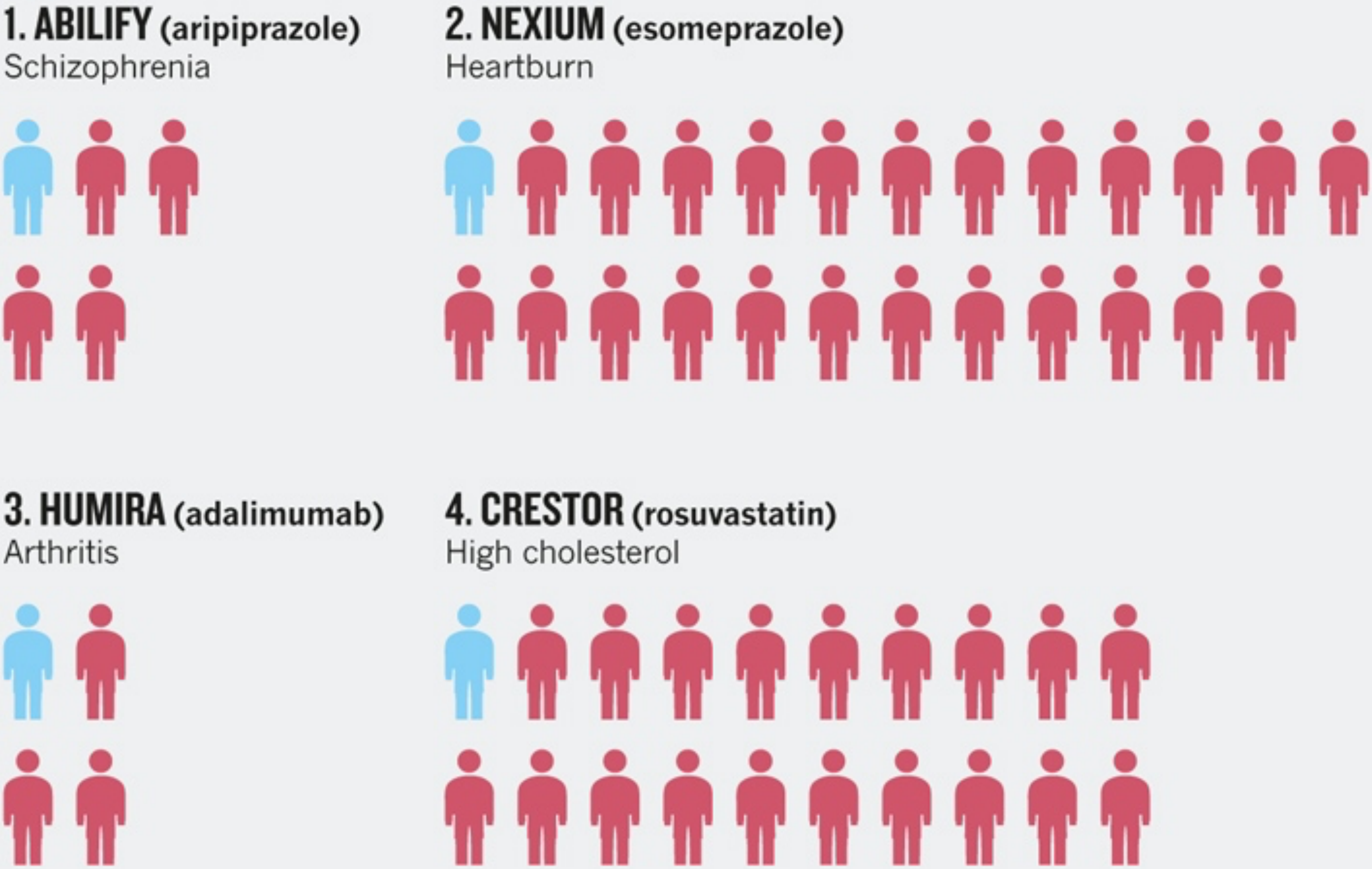
Red: official numbers from Center for Disease Control and Prevention; weekly
Black: based on Google search logs; daily (potentially instantaneously)

Detecting influenza epidemics using search engine query data

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

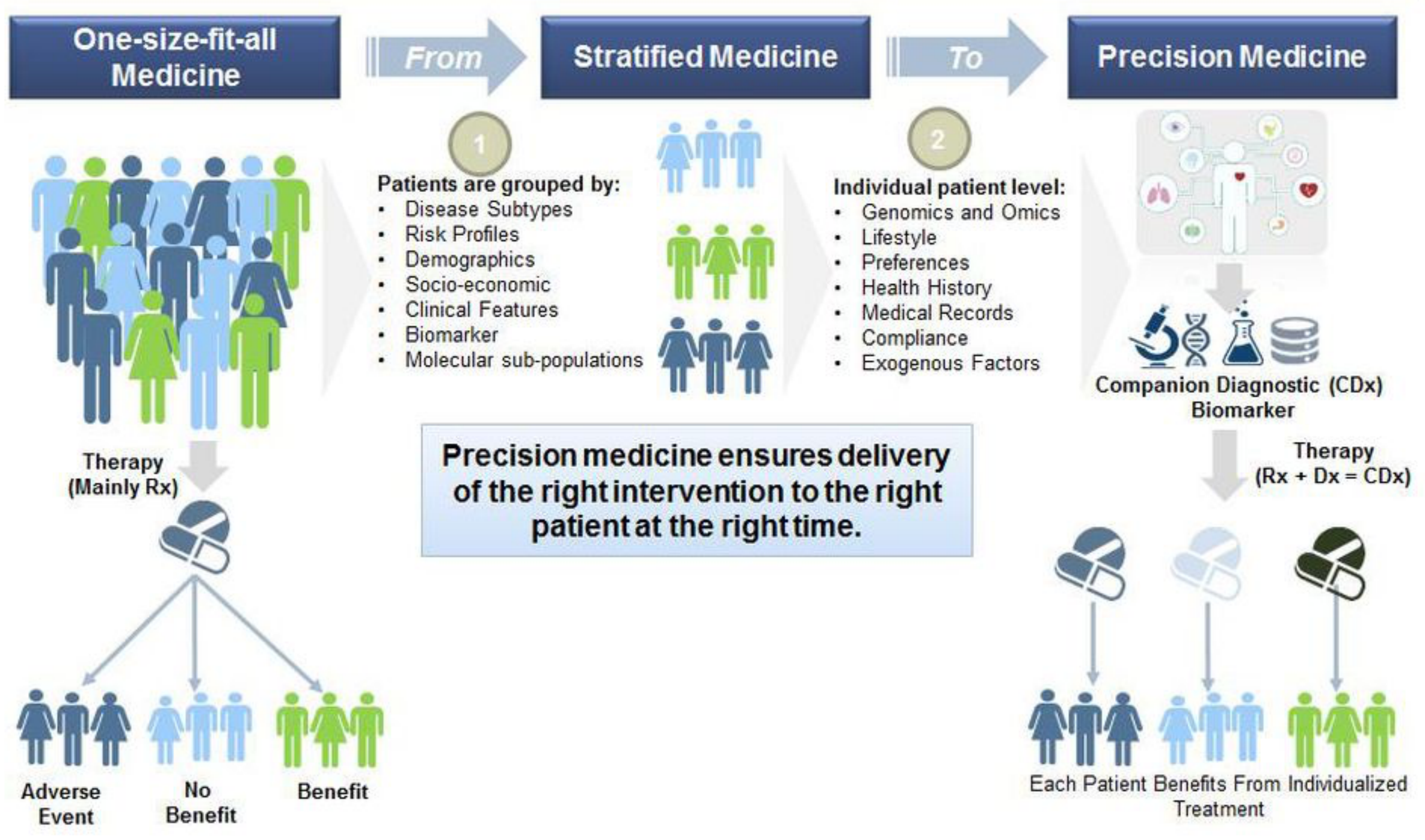
IMPRECISION MEDICINE

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).



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

Precision Medicine



One-size-fit-all Medicine

From

Stratified Medicine

To

Precision Medicine

1

Patients are grouped by:

- Disease Subtypes
- Risk Profiles
- Demographics
- Socio-economic
- Clinical Features
- Biomarker
- Molecular sub-populations

2

Individual patient level:

- Genomics and Omics
- Lifestyle
- Preferences
- Health History
- Medical Records
- Compliance
- Exogenous Factors

Companion Diagnostic (CDx) Biomarker

Therapy (Rx + Dx = CDx)

Therapy (Mainly Rx)

Precision medicine ensures delivery of the right intervention to the right patient at the right time.

Adverse Event

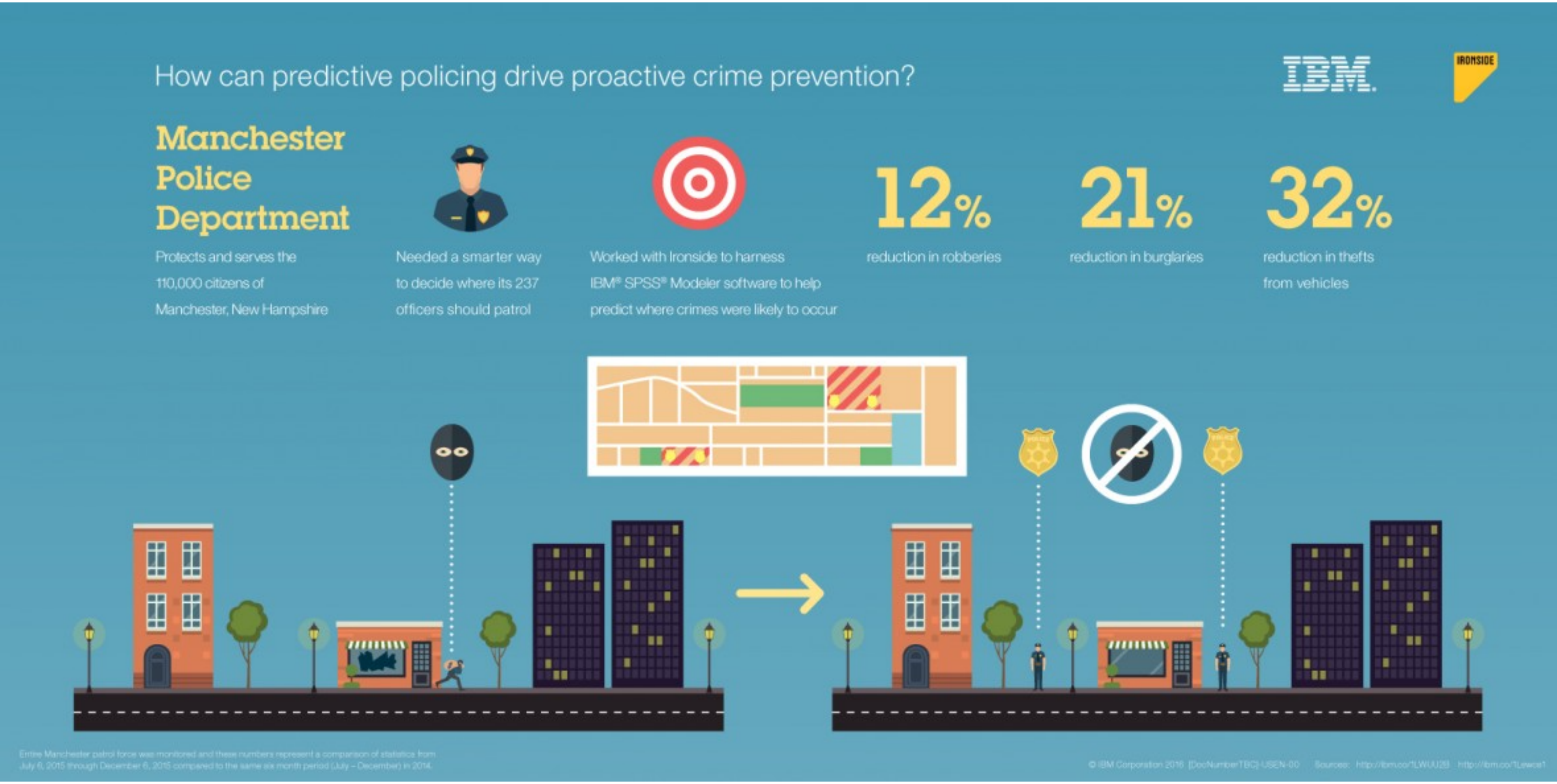
No Benefit

Benefit

Each Patient Benefits From Individualized Treatment

Source: forbes.com

Predictive Policing



How can predictive policing drive proactive crime prevention?

Manchester Police Department

Protects and serves the 110,000 citizens of Manchester, New Hampshire

Needed a smarter way to decide where its 237 officers should patrol

Worked with Ironside to harness IBM® SPSS® Modeler software to help predict where crimes were likely to occur

12% reduction in robberies

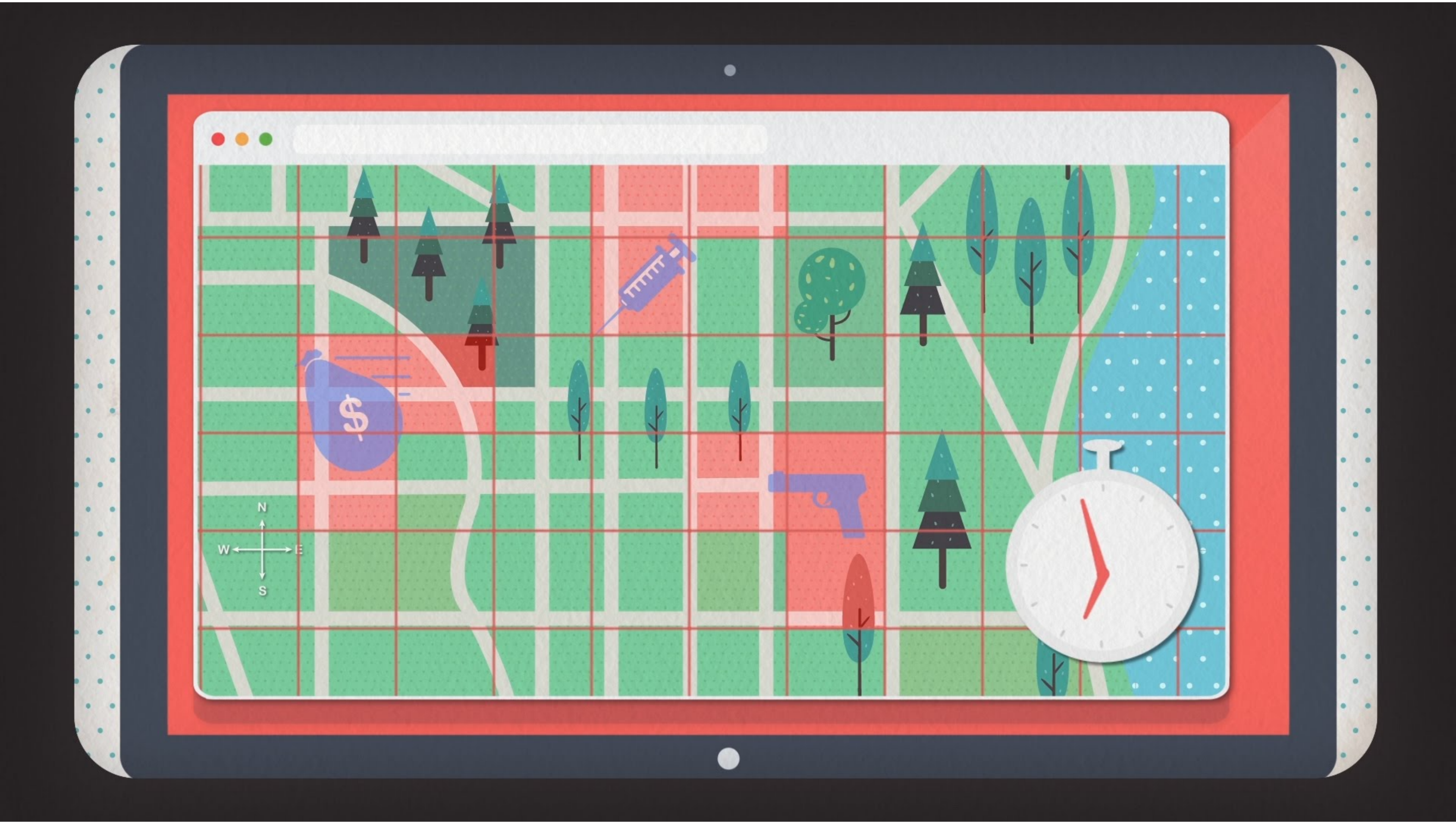
21% reduction in burglaries

32% reduction in thefts from vehicles

Entire Manchester patrol force was monitored and these numbers represent a comparison of statistics from July 6, 2015 through December 6, 2015 compared to the same six month period (July – December) in 2014.

[DocNumberTBC]-USEN-00 Sources: http://ibm.co/1LWUU28 http://ibm.co/1Lewce1

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

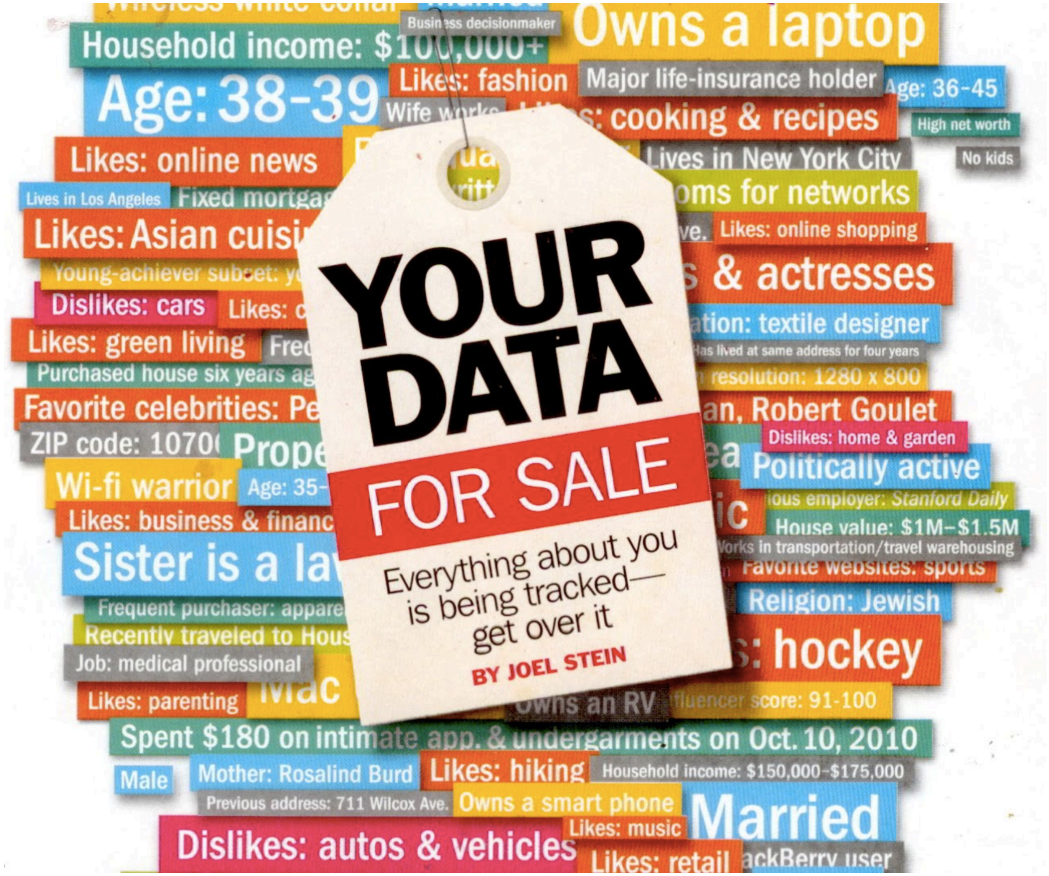
http://pewinternet.org/Reports/2012/Future-of-Big-Data/Overview.aspx

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

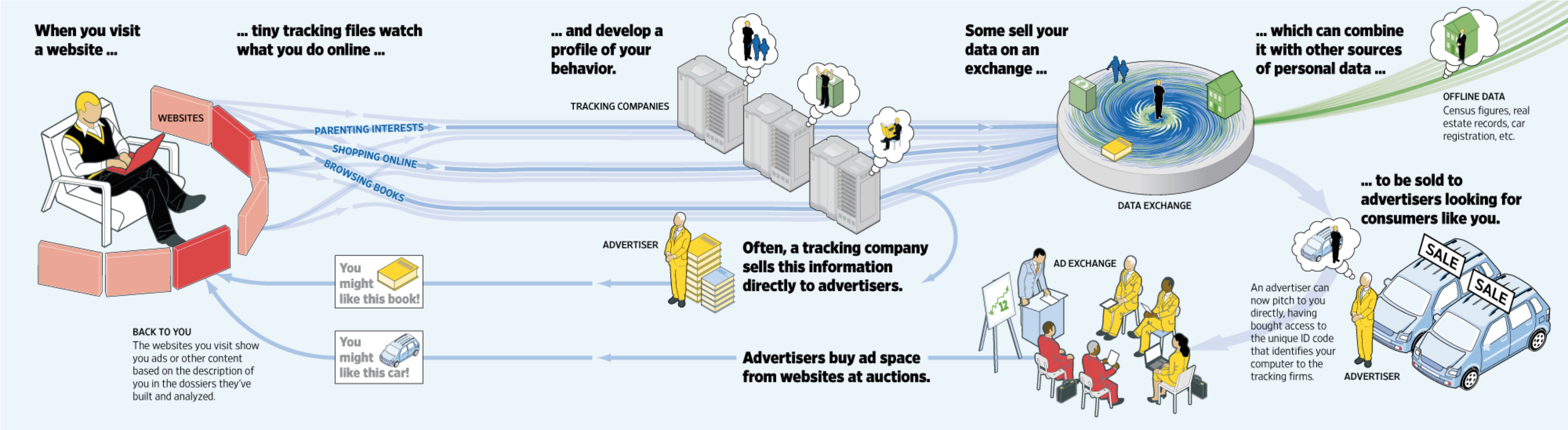
- Photos
- Videos

- Smart phone Sensors
- Mobility trajectories

- …

Very sensitive information …

How is this data collected?



When you visit a website …

… tiny tracking files watch what you do online …

… and develop a profile of your behavior.

Some sell your data on an exchange …

… which can combine it with other sources of personal data …

OFFLINE DATA
Census figures, real estate records, car registration, etc.

… to be sold to advertisers looking for consumers like you.

Often, a tracking company sells this information directly to advertisers.

Advertisers buy ad space from websites at auctions.

An advertiser can now pitch to you directly, having bought access to the unique ID code that identifies your computer to the tracking firms.

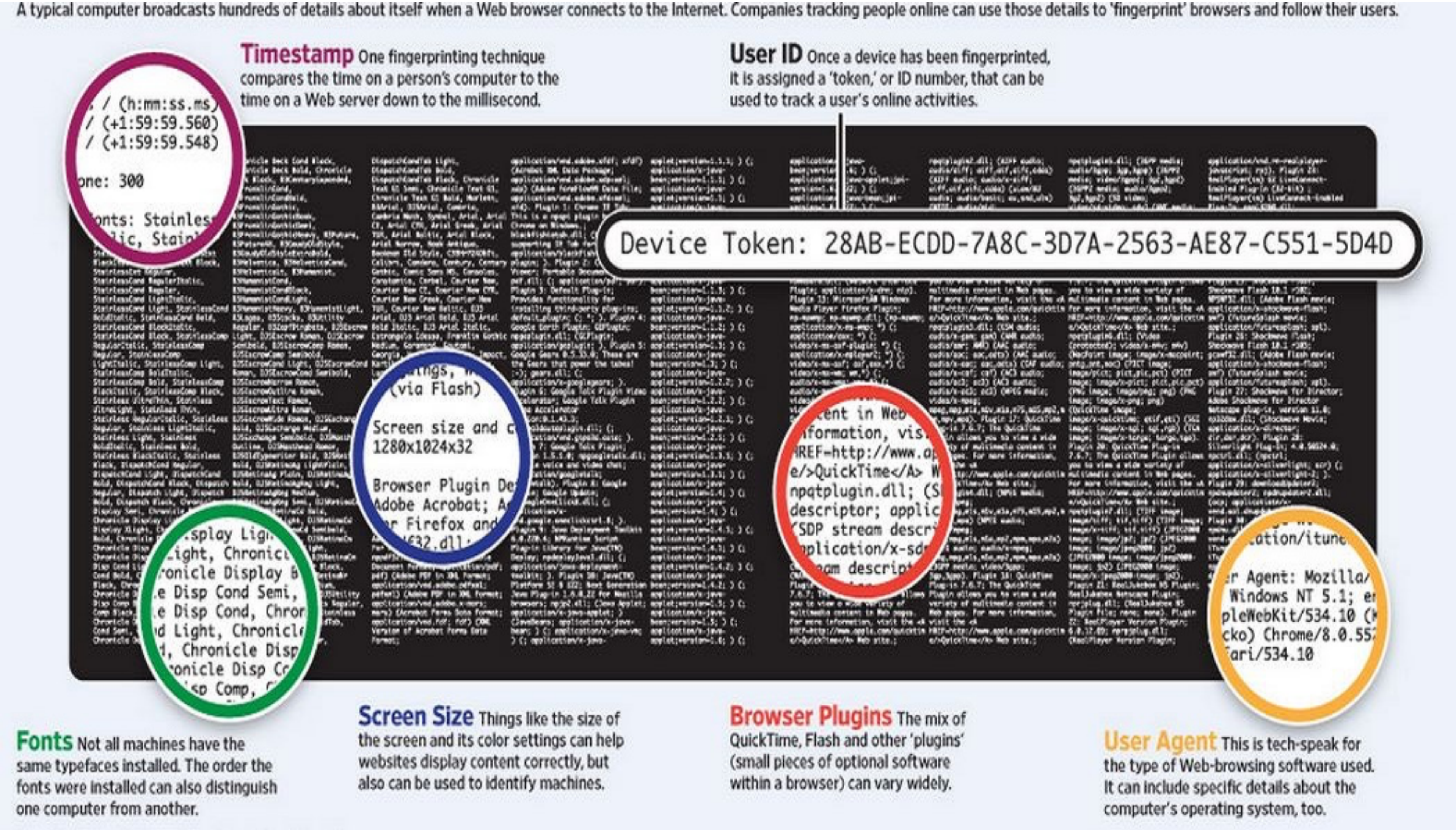
BACK TO YOU
The websites you visit show you ads or other content based on the description of you in the dossiers they've built and analyzed.

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

Isn't my data anonymous ?

WE ARE ANONYMOUS. WE ARE LEGION. WE ARE NO ONE AND EVERYONE.
AND WE ARE HERE TO FIGHT FOR WIKILEAKS.
http://184.50.42
WIKILEAKS
NEW LEAK:
NAMES, ADDRESSES, IPs, AND PHONE NUMBERS OF EVERYONE IN ANONYMOUS.
DOWNLOAD NOW
WIKILEAKS
... DAMMIT, JULIAN.

Device Fingerprinting



A typical computer broadcasts hundreds of details about itself when a Web browser connects to the Internet. Companies tracking people online can use those details to 'fingerprint' browsers and follow their users.

Timestamp One fingerprinting technique compares the time on a person's computer to the time on a Web server down to the millisecond.

User ID Once a device has been fingerprinted, it is assigned a 'token,' or ID number, that can be used to track a user's online activities.

Fonts Not all machines have the same typefaces installed. The order the fonts were installed can also distinguish one computer from another.

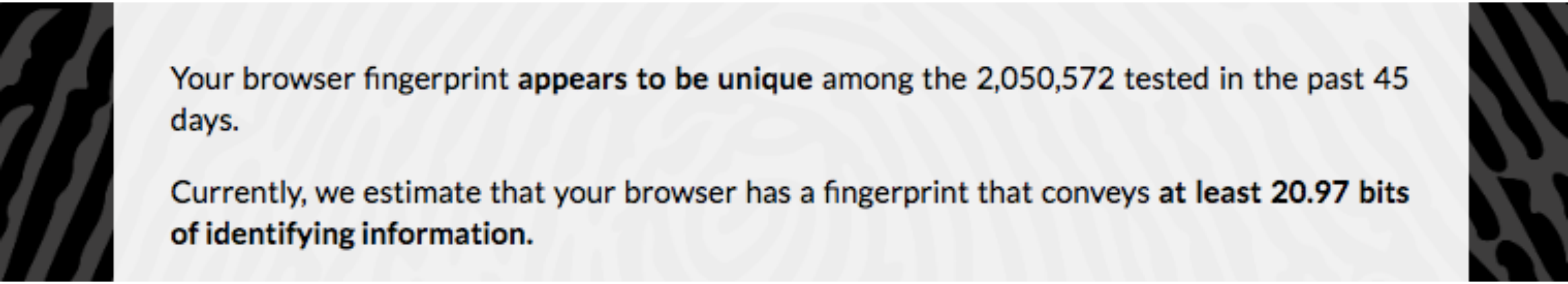
Screen Size Things like the size of the screen and its color settings can help websites display content correctly, but also can be used to identify machines.

Browser Plugins The mix of QuickTime, Flash and other 'plugins' (small pieces of optional software within a browser) can vary widely.

User Agent This is tech-speak for the type of Web-browsing software used. It can include specific details about the computer's operating system, too.

PANOPTICLICK 3.0

Is your browser safe against tracking?



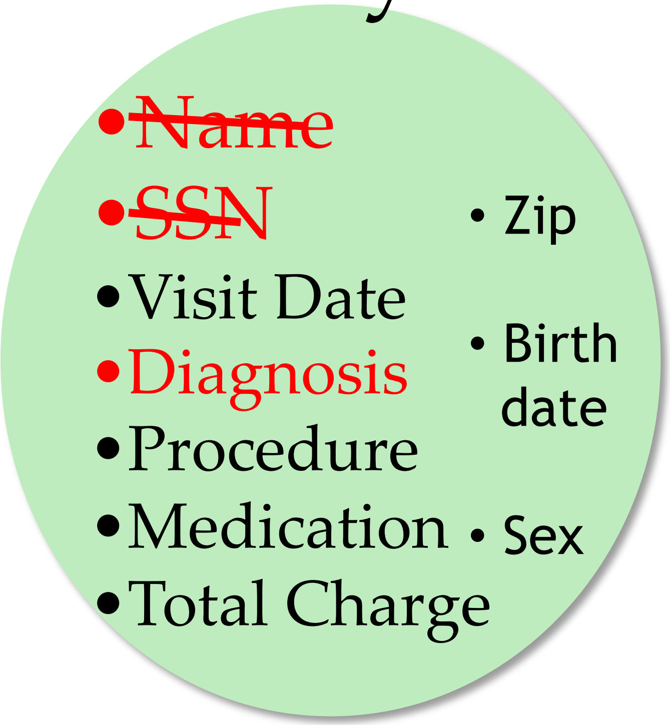
Your browser fingerprint **appears to be unique** among the 2,050,572 tested in the past 45 days.

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]

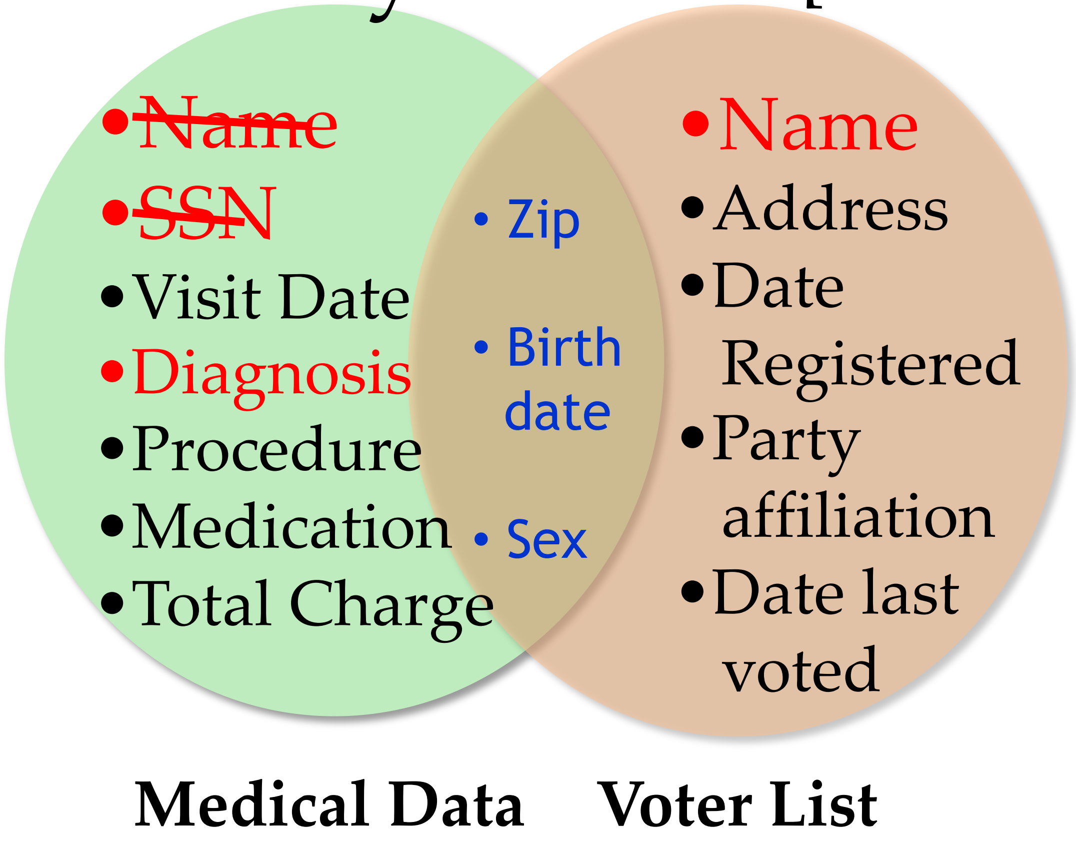


- ~~Name~~
- ~~SSN~~
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge

- Zip
- Birth date
- Sex

Medical Data

The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



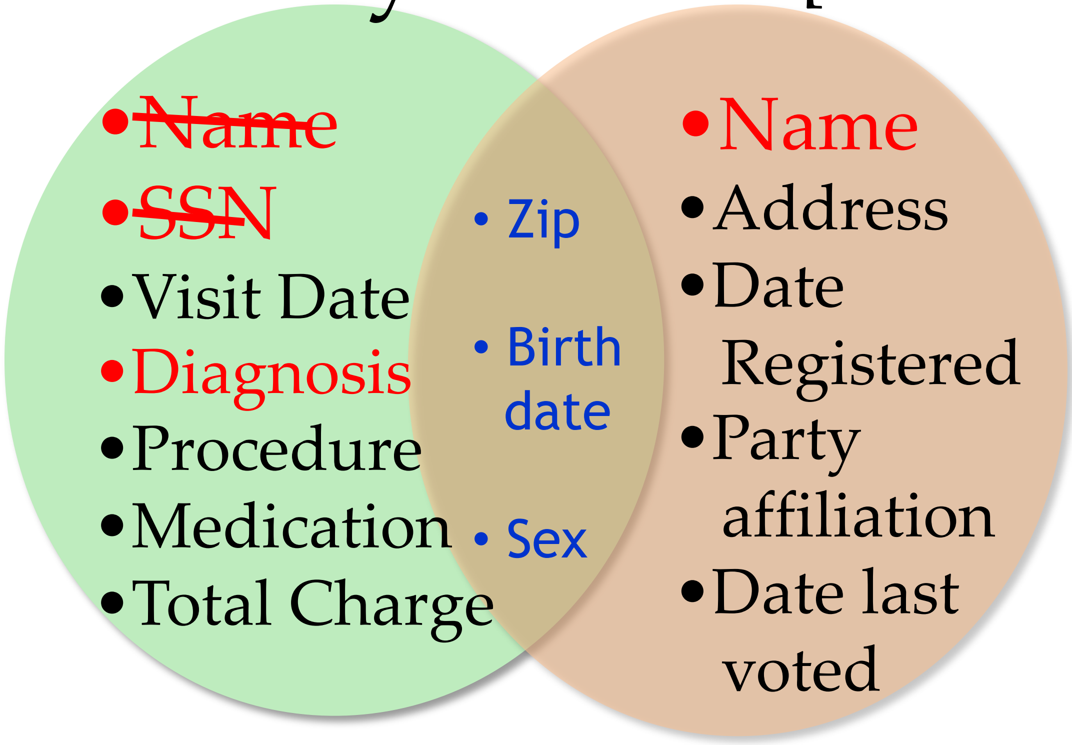
- ~~Name~~
- ~~SSN~~
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge

- Zip
- Birth date
- Sex

- Name
- Address
- Date Registered
- Party affiliation
- Date last voted

Medical Data **Voter List**

The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



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

- Zip
- Birth date
- Sex

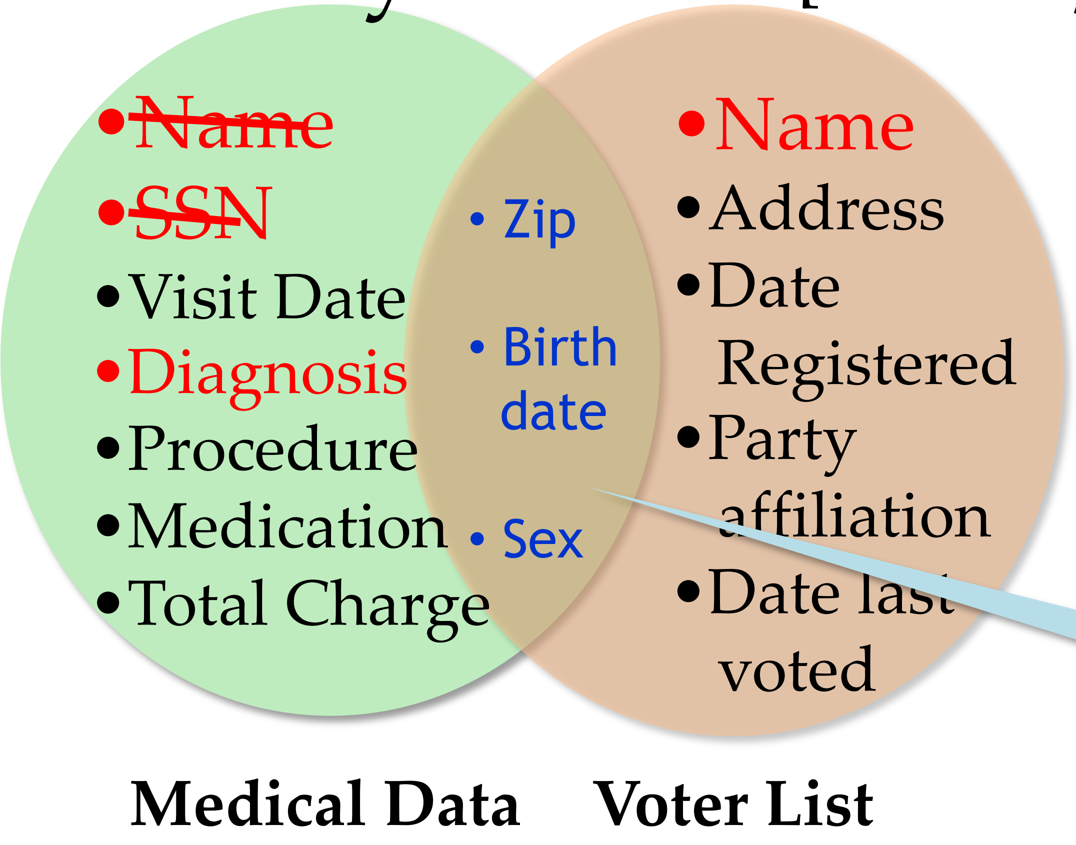
- Name
- Address
- Date Registered
- Party affiliation
- Date last voted

Medical Data **Voter List**

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

Name linked to Diagnosis

The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]



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

- Zip
- Birth date
- Sex

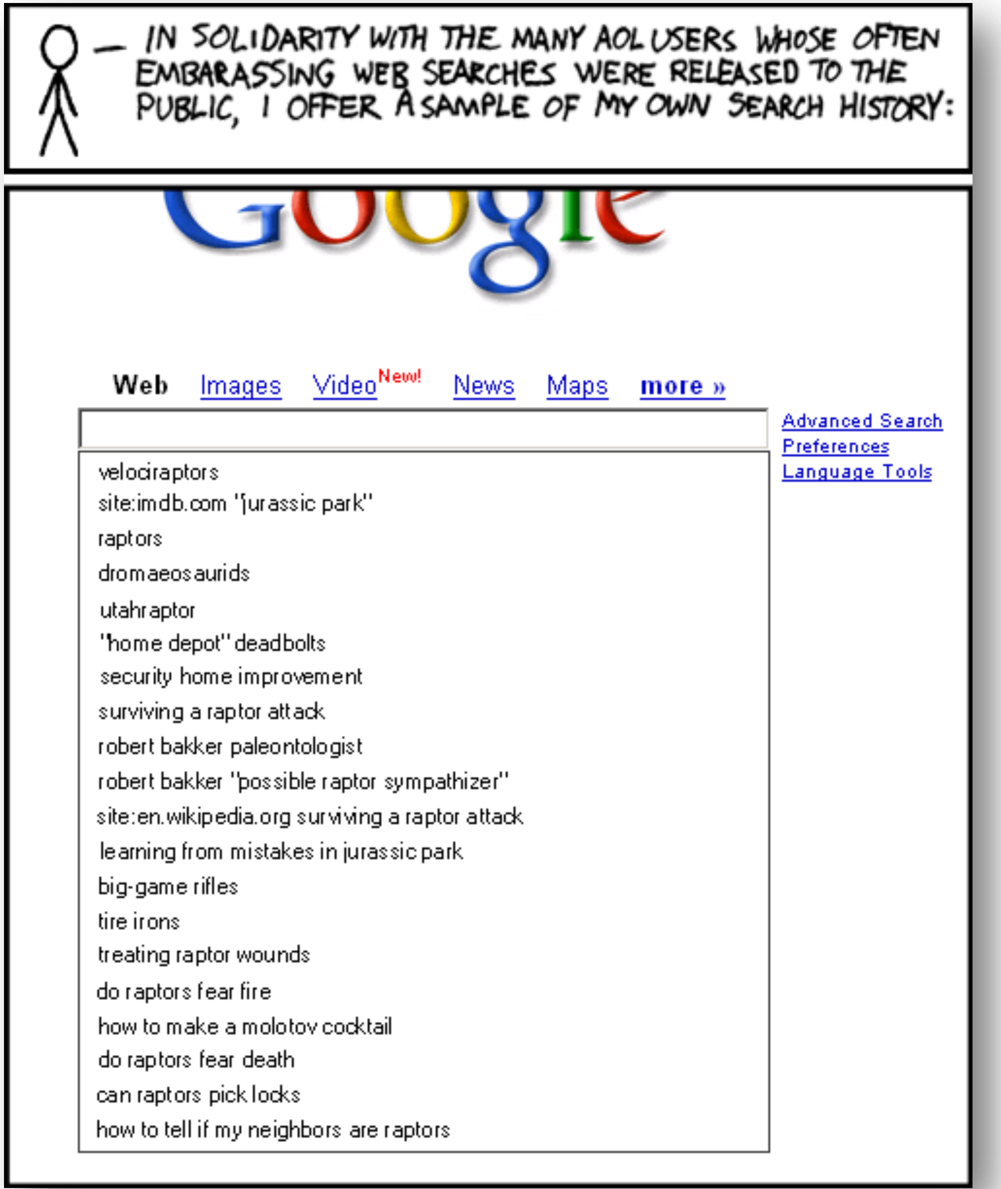
- Name
- Address
- Date Registered
- Party affiliation
- Date last voted

Medical Data **Voter List**

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

Quasi Identifier

AOL data publishing fiasco


— IN SOLIDARITY WITH THE MANY AOL USERS WHOSE OFTEN EMBARASSING WEB SEARCHES WERE RELEASED TO THE PUBLIC, I OFFER A SAMPLE OF MY OWN SEARCH HISTORY:
Web
Images
Video New!
News
Maps
more »
Advanced Search
Preferences
Language Tools

velociraptors
site:imdb.com "jurassic park"
raptors
dromaeosaurids
utahraptor
"home depot" deadbolts
security home improvement
surviving a raptor attack
robert bakker paleontologist
robert bakker "possible raptor sympathizer"
site:en.wikipedia.org surviving a raptor attack
learning from mistakes in jurassic park
big-game rifles
tire irons
treating raptor wounds
do raptors fear fire
how to make a molotov cocktail
do raptors fear death
can raptors pick locks
how to tell if my neighbors are raptors

AOL data publishing fiasco …

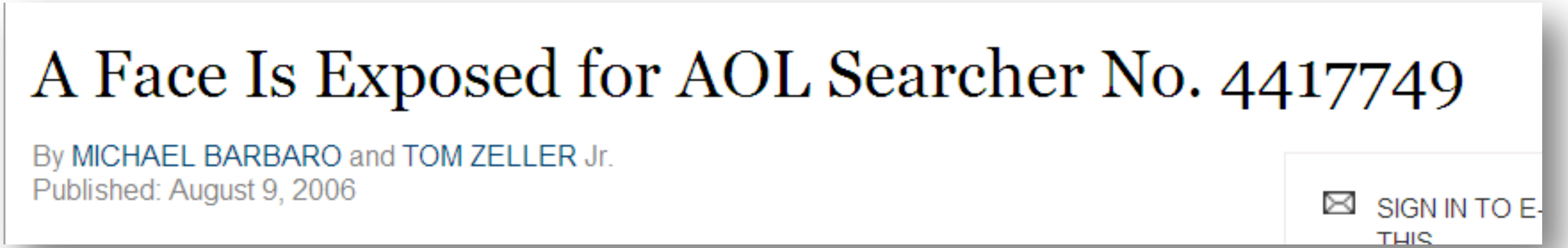
Xi222	Uefa cup
Xi222	Uefa champions league
Xi222	Champions league final
Xi222	Champions league final 2013
Abel156	exchangeability
Abel156	Proof of deFinitti's theorem
Jane12345	Zombie games
Jane12345	Warcraft
Jane12345	Beatles anthology
Jane12345	Ubuntu breeze
Bob222	Python in thought
Bob222	Enthought 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

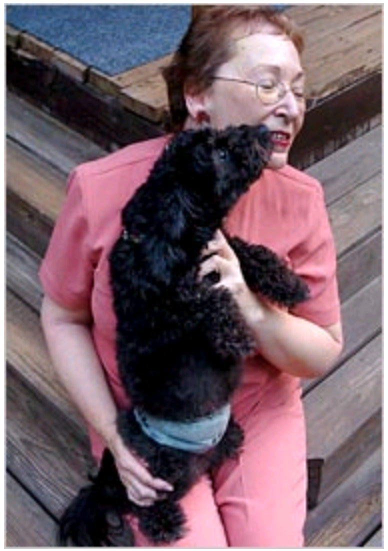
Privacy Breach

[NYTimes 2006]



A Face Is Exposed for AOL Searcher No. 4417749

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

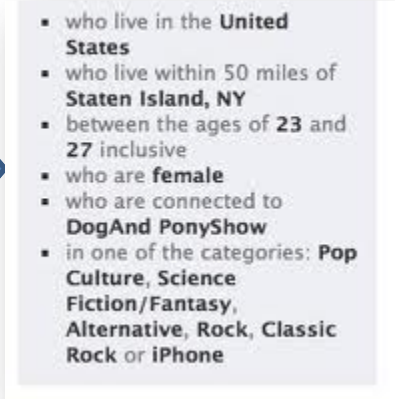


Machine learning models can reveal sensitive information

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

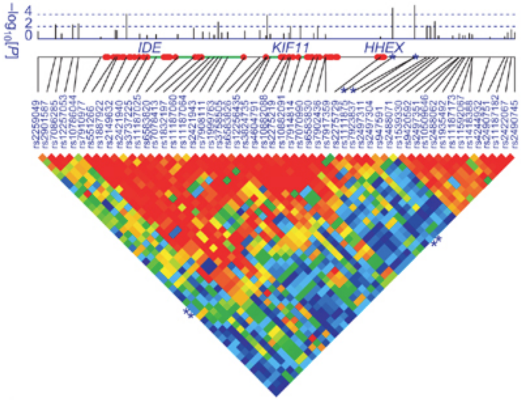
Facebook's learning algorithm uses private information to predict match to ad

[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 …

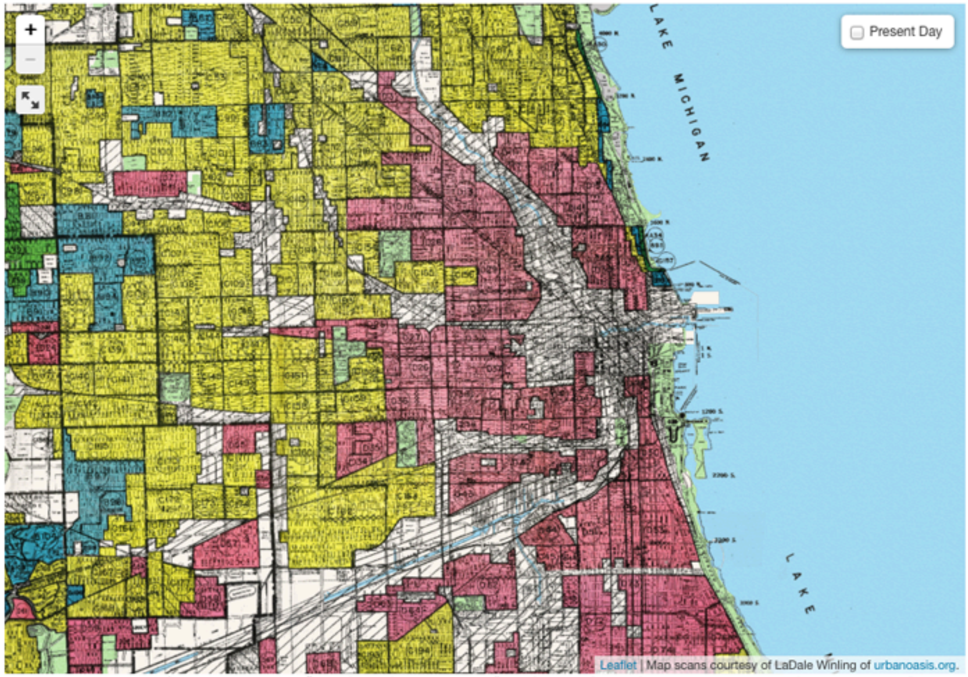
- Privacy

- Fairness

The red side of learning

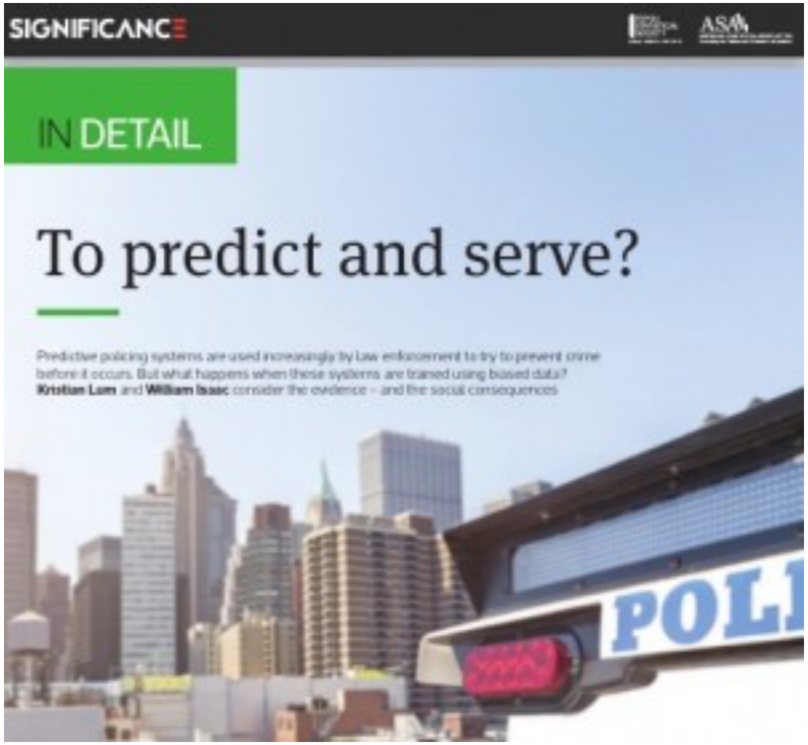
- **Redlining**: the practice of denying, or charging more for, services such as banking, insurance, access to health care, or even supermarkets, or denying jobs to residents in particular, often racially determined, areas.

Explore Redlining in Chicago


+
−
LAKE MICHIGAN
□ Present Day
LAKE
Leaflet | Map scans courtesy of LaDale Winling of urbanoasis.org.

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



SIGNIFICANCE

IN DETAIL

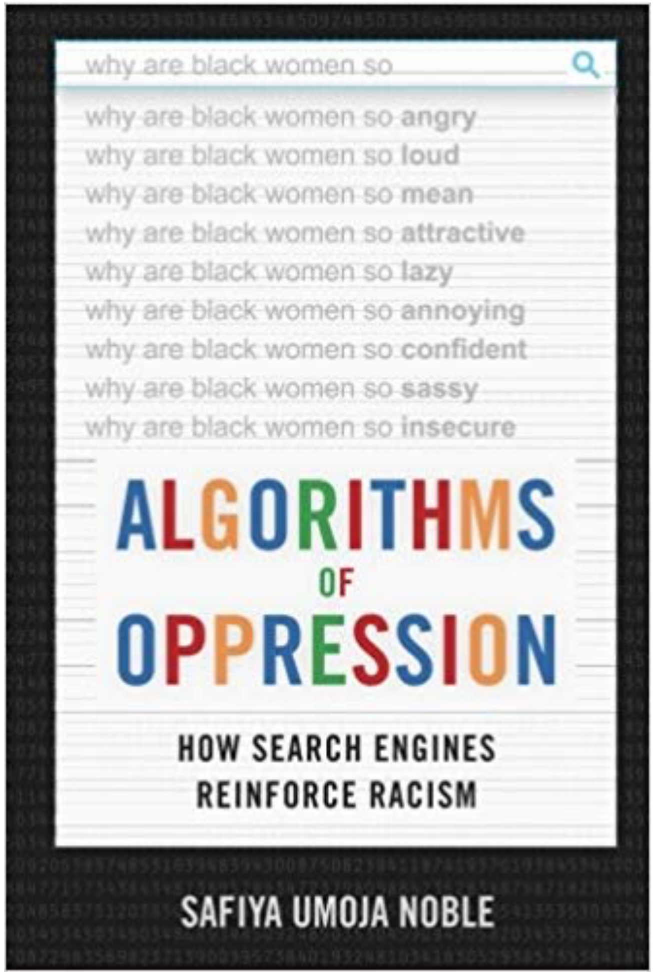
To predict and serve?

Predictive policing systems are used increasingly by law enforcement to try to prevent crime before it occurs. But what happens when these systems are trained using biased data? **Kristian Lum** and **William Isaac** consider the evidence – and the social consequences

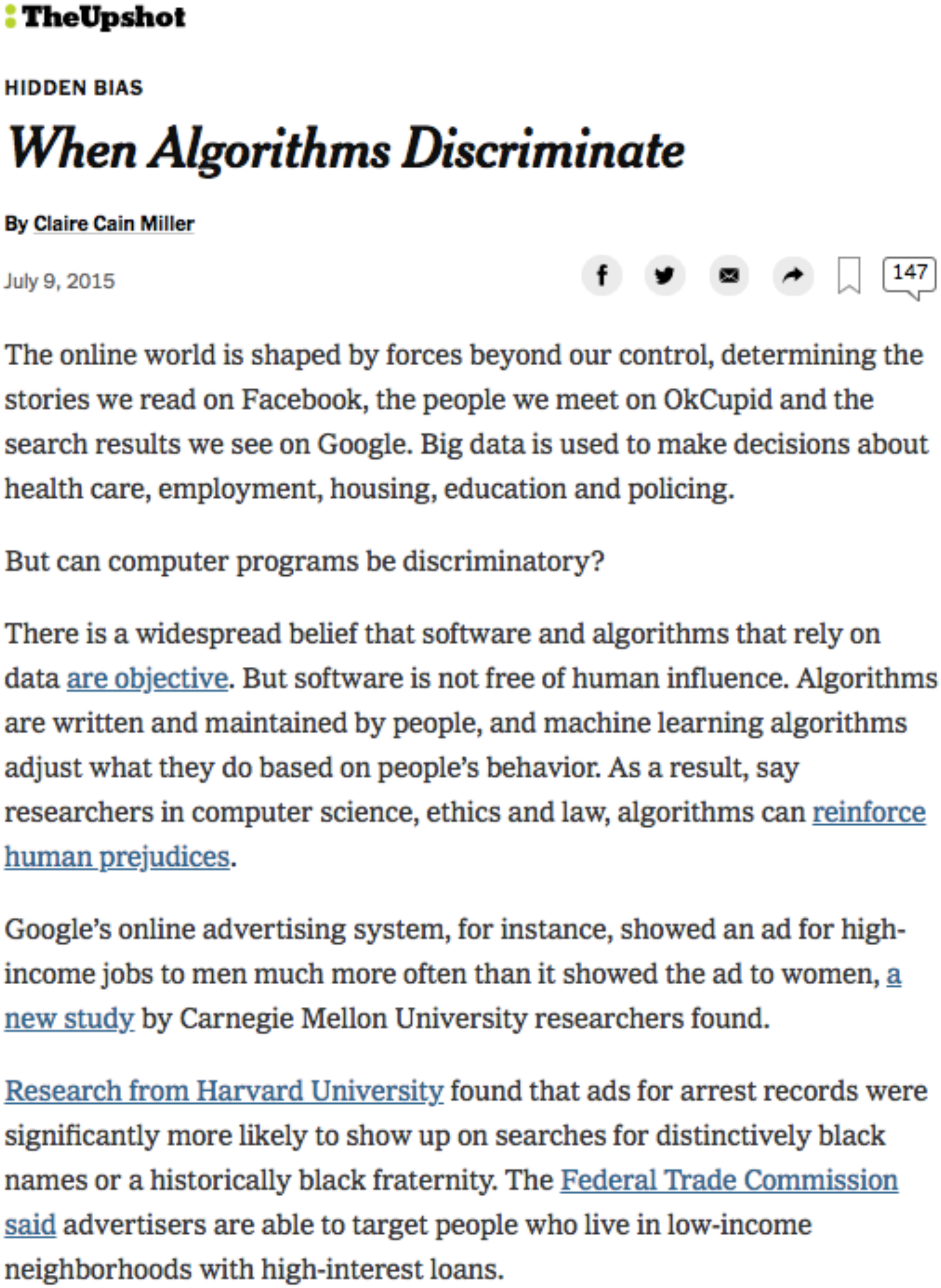
- Predictive policing systems use machine learning algorithms to predict crime.

- But … the algorithms learn … patterns not about crime, per se, but about how police record crime.

- This can amplify existing biases


why are black women so
why are black women so angry
why are black women so loud
why are black women so mean
why are black women so attractive
why are black women so lazy
why are black women so annoying
why are black women so confident
why are black women so sassy
why are black women so insecure
ALGORITHMS
OF
OPPRESSION
HOW SEARCH ENGINES
REINFORCE RACISM
SAFIYA UMOJA NOBLE

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



HIDDEN BIAS

When Algorithms Discriminate

By Claire Cain Miller

July 9, 2015

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
memegenerator.net

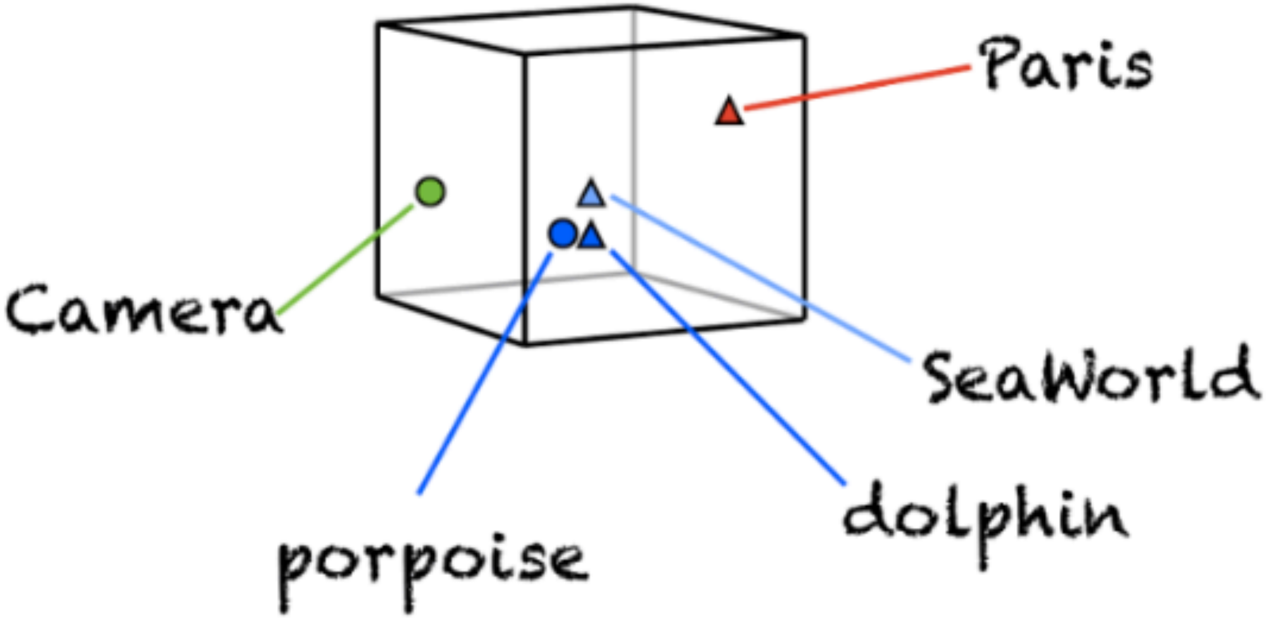
Deep Learning

Incredibly powerful tool for …

- Extracting regularities from data according to a given data

- Amplifying bias!

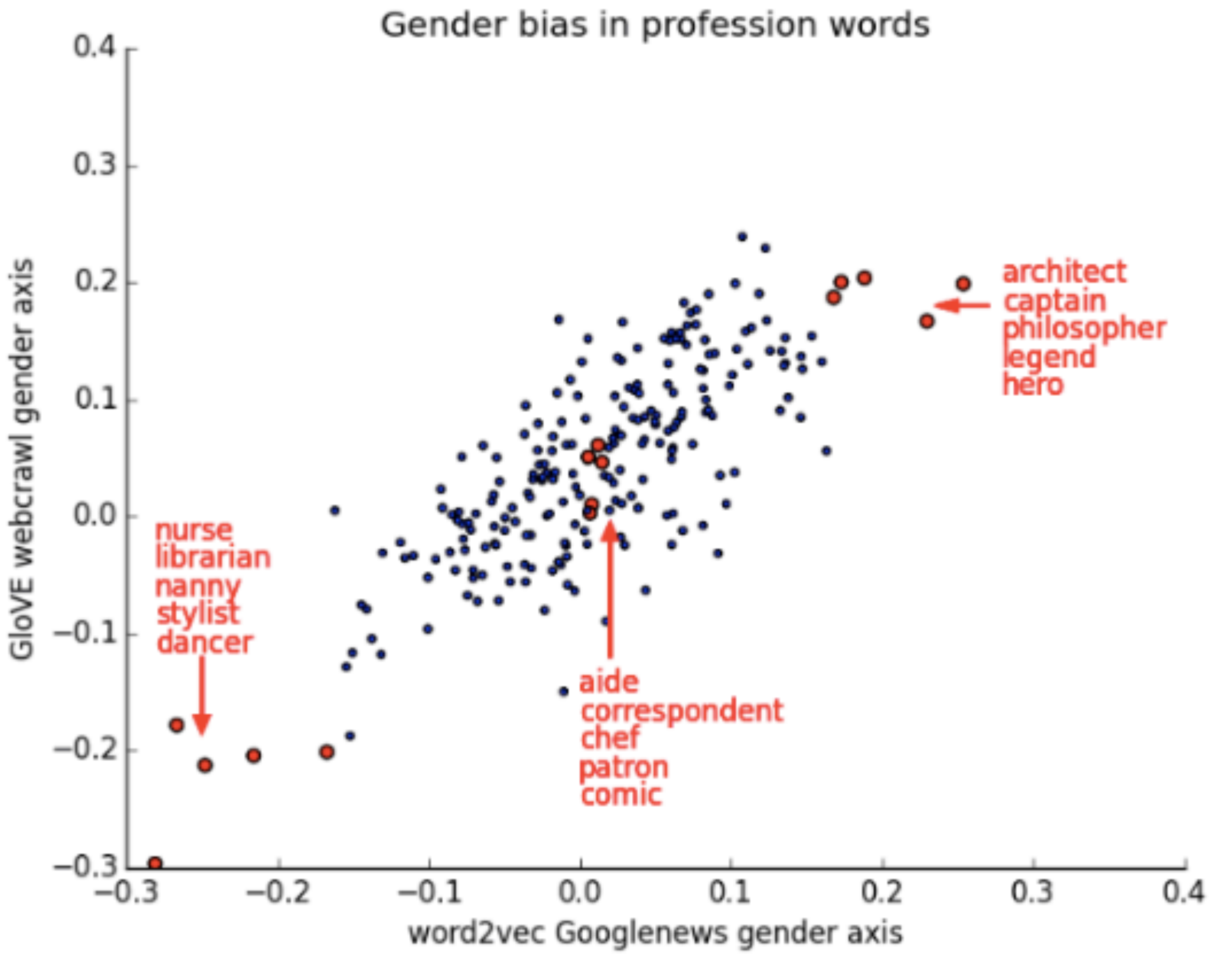
Word embeddings



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 …

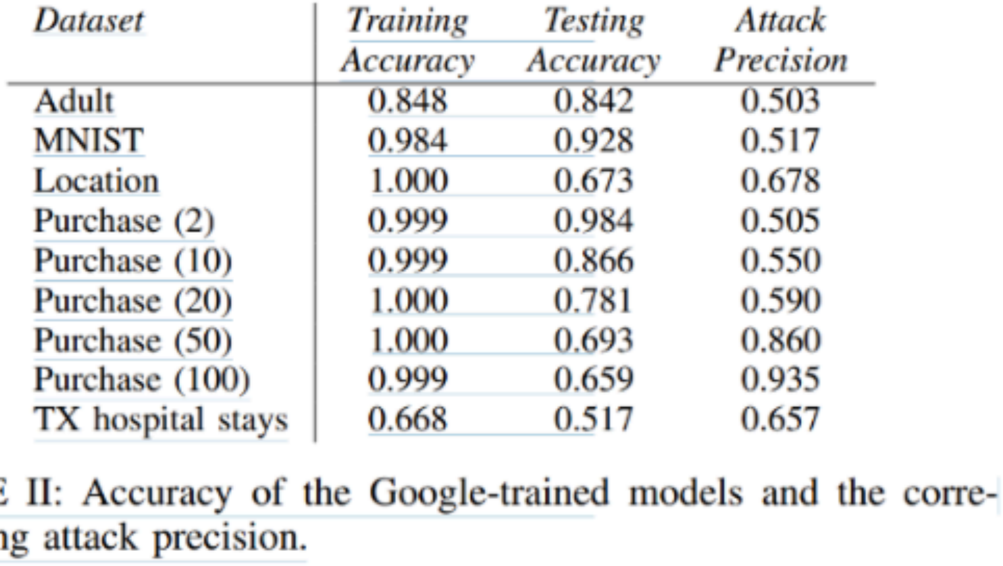
- Extracting regularities from data according to a given data

- 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**:

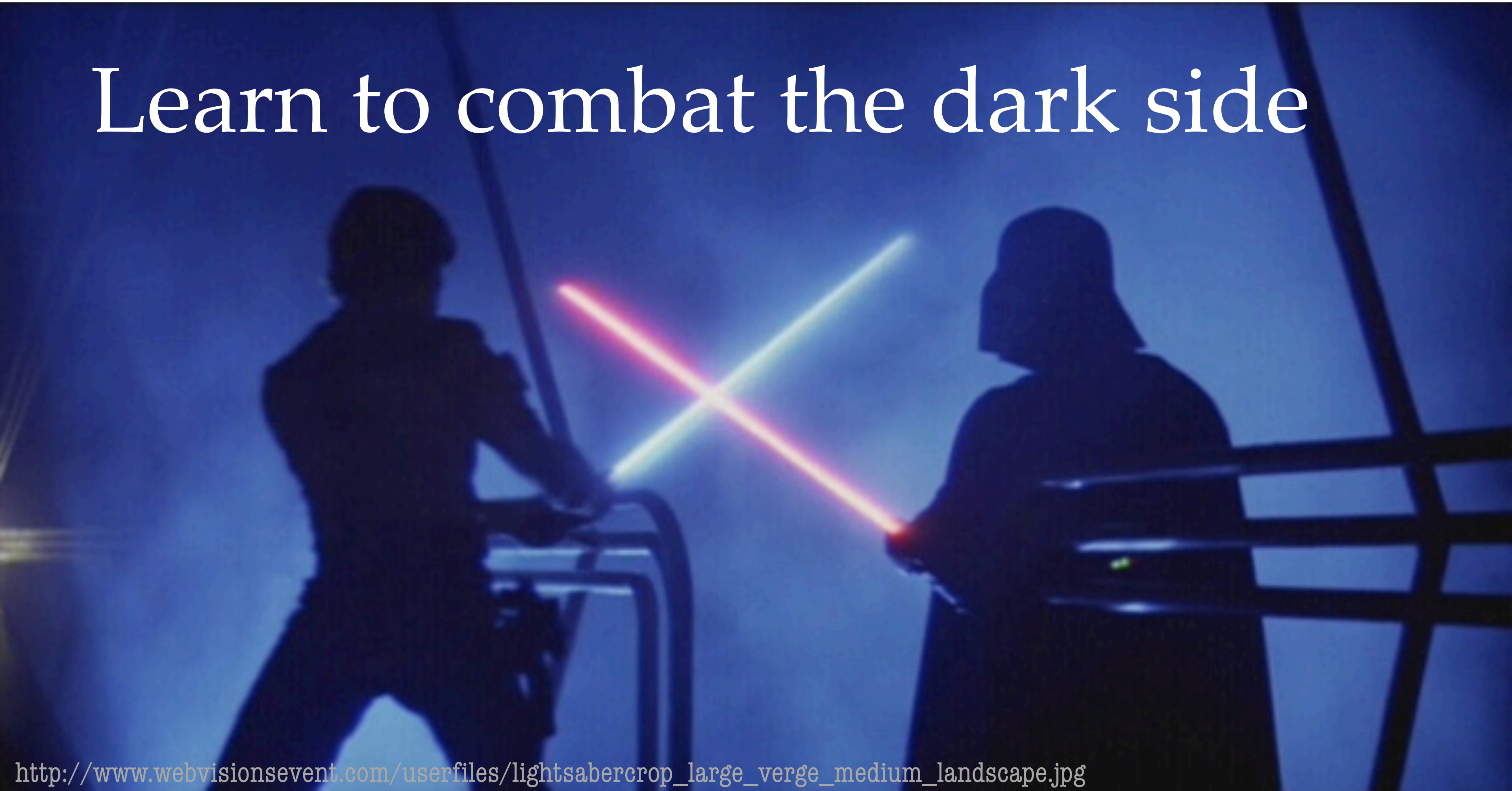
Shokri, R., Stronati, M., Song, C. and Shmatikov, V., 2017, May. **Membership inference attacks against machine learning models**. In *2017 IEEE Symposium on Security and Privacy (SP),* (pp. 3-18). IEEE.



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:



Learn to combat the dark side

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

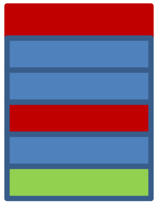
You will …

- mathematically formulate privacy.
- mathematically formulate fairness.

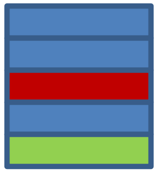
Differential Privacy

For every pair of inputs that differ in one row

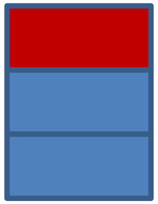
For every output …



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

United States Census Bureau

OnTheMap ***[ICDE 2008]***

chrome

[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
In-class Mini-project
Lectures

- Module 3:
  Paper Reading by Topics
  - privacy v.s. fairness
  - private machine learning
  - deployments of DP
  - sources of bias
  - fairness mechanisms

Read papers
Mini-critiques
Research Project

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

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

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

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

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

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**: https://cs.uwaterloo.ca/~xihe/cs848
  - Schedule (with links to lecture slides, readings, projects, etc.)
- **Grading**
  - In class mini-projects: 10% x 2
  - Mini-critiques: 10%
  - Class participation and presentation: 20%
    - Attending class!
  - Project: 50%
- **LEARN** for submission and grades:
  - https://learn.uwaterloo.ca/d2l/home/492027

Academic Integrity

- See course website
- Mini-project reports and paper critiques are individual work and submission.
- 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

Reference

- Course materials are adapted from: https://sites.duke.edu/cs590f18privacyfairness/