ICME Summer Workshop
Fundamentals of Data Science

Data Privacy and Ethics

Day 2 of 2

Prof. Johan Ugander
jugander@stanford.edu

Adapted from **MS&E 234** at Stanford,
see homepage for literature references and more:
http://msande234.stanford.edu/

ICME Stanford | Institute for Computational & Mathematical Engineering

Schedule

Wednesday

- **Part 4 - Differential Privacy**
  - 1:00p-2:05p Lecture
  - 2:05p-2:10p Break

- **Part 5 - Discussion: transparency & public records**
  - 2:10-3:00p Voting on examples, discussion
  - 3:00p-3:05p Break

- **Part 6 - GDPR, Regulation**
  - 3:05p-4:00p Lecture

Differential Privacy

Goal of differential privacy

- Data analysis can do a lot of good, privacy violations are a negative externality we want to manage.

- How can you allow meaningful usage of rich datasets while preserving individual privacy?

- General object of study: a “database”, a collection of records about individuals.

Example issues

- “I am releasing some useful statistic **f(D)** of a dataset **D**, and nothing more will be revealed.”

- What kind of statistic is “safe” to publish?

Example issues

- "I am releasing some useful statistic **f(D)** of a dataset **D**, and nothing more will be revealed."

- What kind of statistic is "safe" to publish?

- **<u>Obviously bad:</u> f(D)** = number of people named Johan who live in Palo Alto and work at Stanford and smoke.

- No obvious test for **f(D)**….

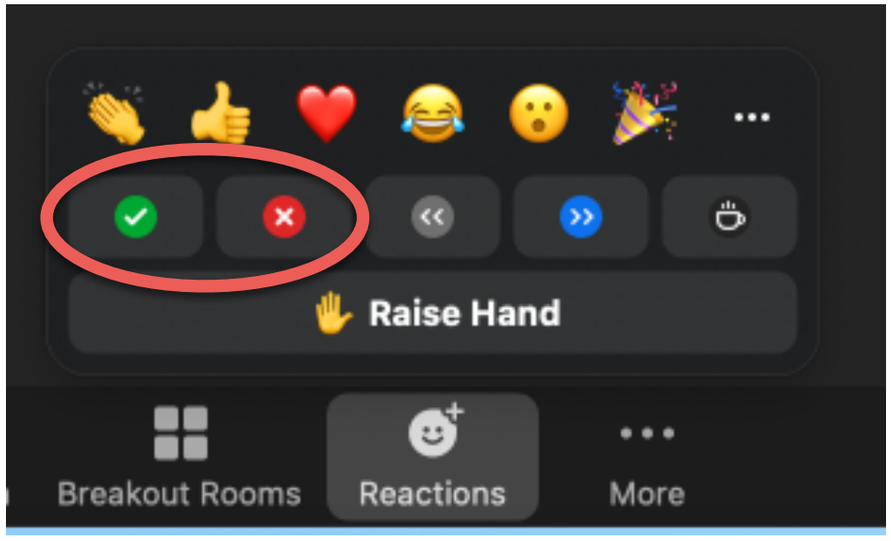
Example issues

- "I am releasing findings showing that people who smoke are very likely to get cancer."

- **<u>Possible claim:</u>** "You can't do that, it will violate my privacy. My insurance company knows that I am a smoker. My insurance will go up!"

- Is this a privacy violation?

Zoom reactions



Example issues

- "I am releasing findings showing that people who smoke are very likely to get cancer."

- **Possible claim:** "You can't do that, it will violate my privacy. My insurance company knows that I am a smoker. My insurance will go up!"

- Is this a privacy violation?

    - Differential privacy says **no**. Key point: my study didn't use your data.
    - The above is still an ethical issue about bringing information into the world, but it's not **privacy**.

Disclosure

- Definition due to (Dalenius,1977): “Towards a methodology for statistical disclosure control”

- “Consider an object O_k, and a characteristic D, which is a survey characteristic. For the object O_k this characteristic assumes the value D_k…”

If the release of the statistics S makes it possible to determine the value of D_k more accurately than it is possible without access to S, a **disclosure** has taken place.

Disclosure

- Definition due to (Dalenius,1977): “Towards a methodology for statistical disclosure control”

- “Consider an object O_k, and a characteristic D, which is a survey characteristic. For the object O_k this characteristic assumes the value D_k…”

  If the release of the statistics S makes it possible to determine the value of D_k more accurately than it is possible without access to S, a **disclosure** has taken place.

- Dalenius authored a typology with 6 binary traits of disclosures, $2^6 = 64$ types.

Dalenius to Dwork

- Dalenius 1977 goal: access to a statistical database should not enable one to learn anything about an individual that could not be learned without access.
- Dwork 2006: Auxiliary information makes this an impossible goal.

Dalenius to Dwork

- Dalenius 1977 goal: access to a statistical database should not enable one to learn anything about an individual that could not be learned without access.
- Dwork 2006: Auxiliary information makes this an impossible goal.

- **Database D:** Heights of individuals in world
- **Statistic f(D):** Average heights of women in the world
- **Auxiliary information:** "**Terry Gross** is two inches shorter than the average Lithuanian woman"

- A **disclosure** regardless of whether Terry Gross is in the database.

Dalenius to Dwork

- Dalenius 1977 goal: access to a statistical database should not enable one to learn anything about an individual that could not be learned without access.
- Dwork 2006: Auxiliary information makes this an impossible goal.

If the release of the statistics S makes it possible to determine the value of D_k more accurately than it is possible without access to S, a **disclosure** has taken place.

- A **disclosure** regardless of whether Terry Gross is in the database.

Differential Privacy

- Dwork 2006: “the risk to one’s privacy, or in general, any type of risk, such as the risk of being denied automobile insurance, should not substantially increase as a result of participating in a statistical database. This is captured by *differential* privacy.”

Differential Privacy

- Dwork 2006: "the risk to one's privacy, or in general, any type of risk, such as the risk of being denied automobile insurance, should not substantially increase as a result of participating in a statistical database. This is captured by *differential* privacy."

- Framed in terms of participation in a dataset, e.g. responding to a survey, and consequences.
  - My answer shouldn't impact results

    $f(D_{S\text{-}me}) = f(D_S)$
  - Published results don't let you learn about me

    $Pr[secret(me) \mid D] = Pr[secret(me)]$

Differential Privacy

- Dwork 2006: "the risk to one's privacy, or in general, any type of risk, such as the risk of being denied automobile insurance, should not substantially increase as a result of participating in a statistical database. This is captured by *differential* privacy."

- Framed in terms of participation in a dataset, e.g. responding to a survey, and consequences.
  - My answer shouldn't impact results
    $f(D_{S-me}) = f(D_S)$ **$\Rightarrow f(D_S) = f(D_\emptyset)$ by induction**
  - Published results don't let you learn about me
    $Pr[secret(me) \mid D] = Pr[secret(me)]$

Impossible

Differential Privacy

- Dwork 2006: "the risk to one's privacy, or in general, any type of risk, such as the risk of being denied automobile insurance, should not substantially increase as a result of participating in a statistical database. This is captured by *differential* privacy."

- Framed in terms of participation in a dataset, e.g. responding to a survey, and consequences.
  - My answer shouldn't impact results
    $f(D_{S-me}) = f(D_S)$ **$\Rightarrow f(D_S) = f(D_\emptyset)$ by induction**
  - Published results don't let you learn about me
    $Pr[secret(me) \mid D] = Pr[secret(me)]$
    but $Pr[secret(me) \mid f(D)] > Pr[secret(me)]$,
    possible even without me in D.

Impossible

Impossible

DP agenda

- Define **ε-differential privacy**.
- Define **sensitivity** of a query.

- How to use **Laplace mechanism** to provide ε-DP for a given query.

- The trouble with many queries (“significant” queries).

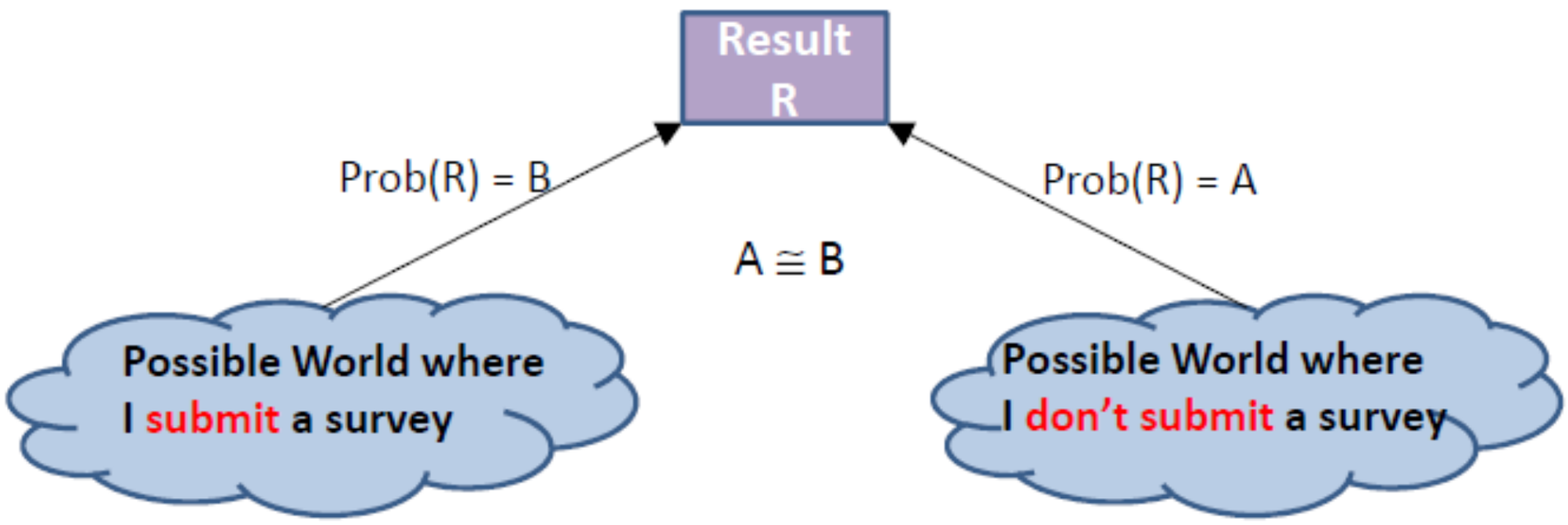
- DP in machine learning

Differential Privacy

- Definition
  Let D be a dataset, f(D) be a statistic. Then f() is **ε-differentially private** if

$$\Pr[f(D) = C] < e^{\varepsilon} \Pr[f(D_{\pm i}) = C]$$

for all $D_{\pm i}$ such that $|D_{\pm i} - D| \leq 1$ and any C in range(f).



Differential Privacy over-claims

- Overclaim: "DP protects individual against ALL harms regardless of prior knowledge."

- Example issue:
    - Study shows smoking causes cancer.
    - Mary's insurance premium rises.
    - Her insurance will rise regardless of whether she participated in the survey.

Sensitivity of a function

- **Sensitivity** Δf: $\Delta f = \max_{\text{adjacent } x, x'} |f(x)-f(x')|$

- Adjacent databases differ in at most one row.
- Captures how much one person's data can affect output.

- Counting queries have sensitivity 1.

Sensitivity of a function

- **Sensitivity** Δf: $\Delta f = \max_{\text{adjacent } x, x'} |f(x)-f(x')|$

- Adjacent databases differ in at most one row.
- Captures how much one person's data can affect output.

- Counting queries have sensitivity 1.

- How many survey takers are female? Sensitivity:
- In total, how many Taylor Swift studio albums are bought by survey takers? Sensitivity:

Sensitivity of a function

- **Sensitivity** Δf: $\Delta f = \max_{\text{adjacent } x, x'} |f(x)-f(x')|$

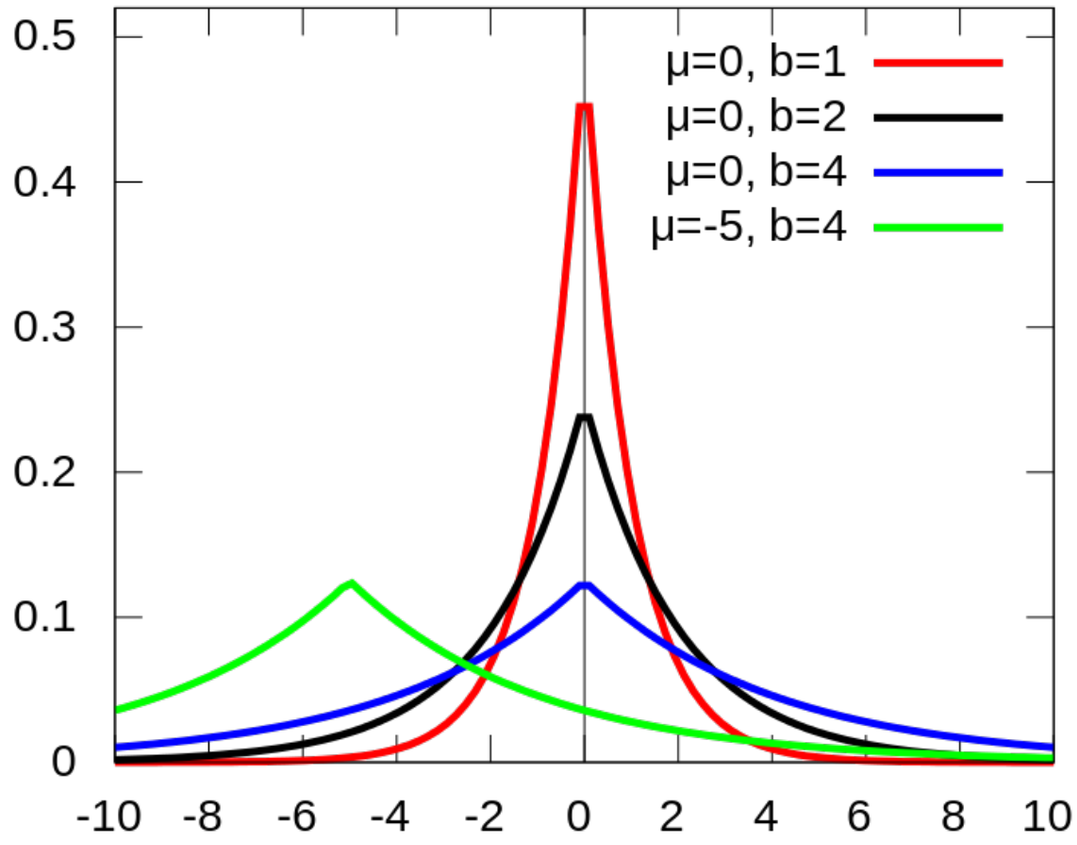
- Adjacent databases differ in at most one row.
- Captures how much one person's data can affect output.

- Counting queries have sensitivity 1.

- How many survey takers are female? Sensitivity: 1
- In total, how many Taylor Swift studio albums are bought by survey takers? Sensitivity: 9

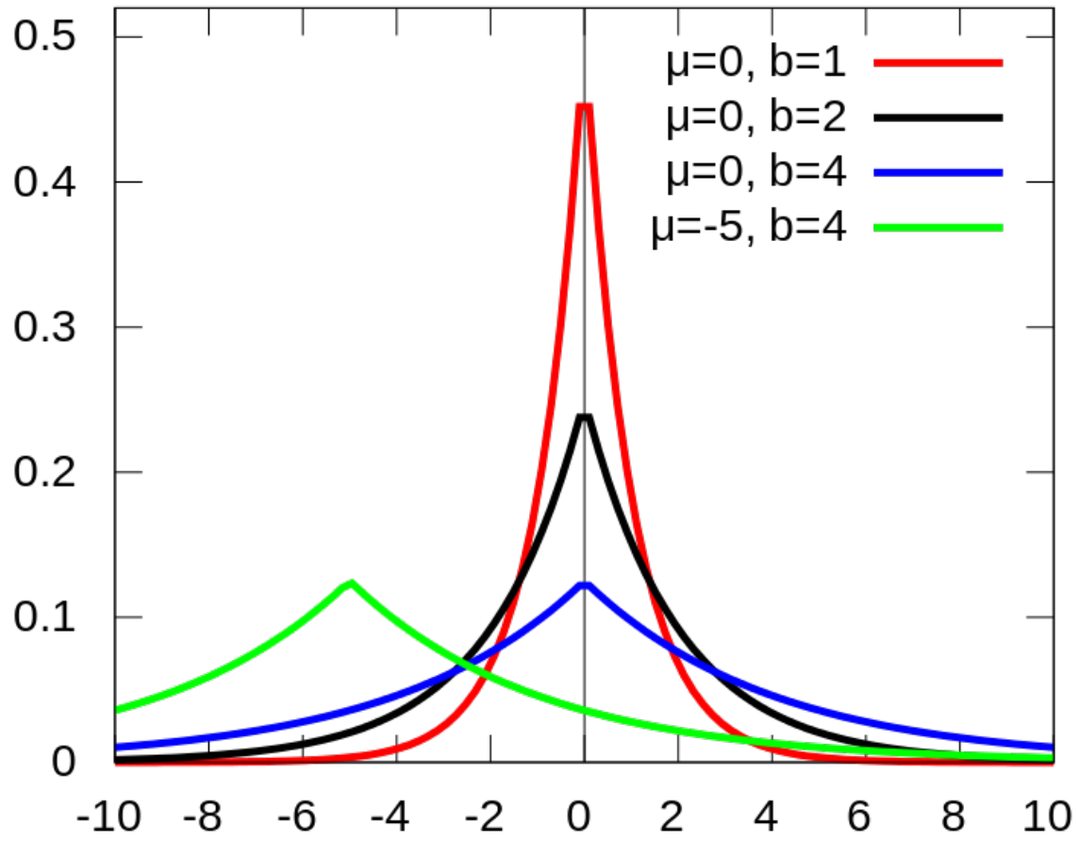
Laplace noise

- PDF: $p(z;b)=(1/2b)*e^{-|z|/b}$
- Variance: $2b^2$
- S.D. = $\sqrt{2}\ b$



Laplace noise

- PDF: $p(z;b)=(1/2b)*e^{-|z|/b}$
- Variance: $2b^2$
- S.D. = $\sqrt{2}$ b



- **Theorem** [DMNS06]: On query f, if we add scaled symmetric noise Lap(b) with b= Δf/ε, we achieve ε-DP.

- Noise depends on f and ε, but not on database!

Laplace noise example

- How many people in the database are female?
- Sensitivity: 1
- Instead of responding **f(D)**, respond **f(D)+Lap(1/ε)**

Laplace noise example

- How many people in the database are female?
- Sensitivity: 1
- Instead of responding **f(D)**, respond **f(D)+Lap(1/ε)**

- **<u>Local DP vs. Centralized DP</u>**
  - In the above example, DP is only achieved at the level of the query. Called **centralized model.**
  - Could have also added noise (once) to the entries in original data. Called **local model.**
    - Would flip female/male bits with probability **p.**
    - Guarantees **ε-DP** where **ε** is function of **p.**
  - Local much stronger guarantee.

Laplace noise example

- How many people in the database are female?
- Sensitivity: 1
- Instead of responding **f(D)**, respond **f(D)+Lap(1/ε)**

- **<u>Local DP vs. Centralized DP</u>**
  - In the above example, DP is only achieved at the level of the query. Called **centralized model.**
  - Could have also added noise (once) to the entries in original data. Called **local model.**
    - Would flip female/male bits with probability **p.**
    - Guarantees **ε-DP** where **ε** is function of **p.**
  - Local much stronger guarantee.

Laplace noise example

- How many people in the database are female?
- Sensitivity: 1
- Instead of responding **f(D)**, respond **f(D)+Lap(1/ε)**

- **<u>Local DP vs. Centralized DP</u>**

- In the above example, DP is only achieved at the level of the query. Called **centralized model.**
- Could have also added noise (once) to the entries in original data. Called **local model.**
  - Would flip female/male bits with probability **p.**
  - Guarantees **ε-DP** where **ε** is function of **p.**
- Local much stronger guarantee.
- Connection to **randomized response** survey design.

Laplace and multiple queries

- **<u>What about multiple counting queries?</u>**

    - To be ε-DP with q queries, add noise ~ Lap(qΔf/ε).
    - Alternatively: tolerated reduced ε (by factor q).

- **Problem: That's a lot of noise!**

Laplace and multiple queries

- **<u>What about multiple counting queries?</u>**

    - To be ε-DP with q queries, add noise ~ Lap(qΔf/ε).
    - Alternatively: tolerated reduced ε (by factor q).

- **Problem: That's a lot of noise!**
- **Further problem: consider sequential setting.** If you already answered first query, you can't answer any more queries…

Laplace and multiple queries

- **<u>What about multiple counting queries?</u>**

    - To be ε-DP with q queries, add noise ~ Lap(qΔf/ε).
    - Alternatively: tolerated reduced ε (by factor q).

- **Problem: That's a lot of noise!**
- **Further problem: consider sequential setting.** If you already answered first query, you can't answer any more queries…
    - Observe: "ε is additive"
    - Could start at Δf/2, then Δf/4, then Δf/8, but…

A solution: significant queries

- Solution(?): Only answer “significant” queries

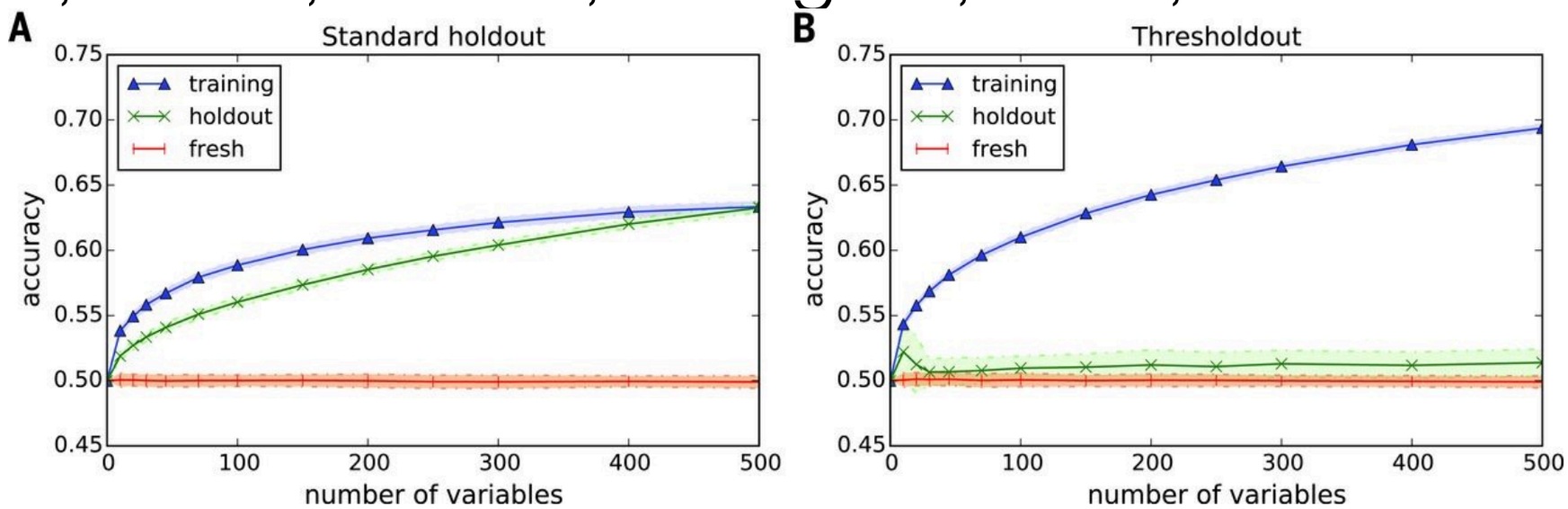
- Database of size n.
- m counting queries, m >> n.
- k significant queries, k = O(n), where count exceeds publicly known threshold T.

- **Goal:** find, and optionally release, counts for significant queries, paying only for significant queries.
- **Challenge:** Conditional branch leaks private information!
- **Solution:** Need noisy threshold T + Lap(x)

DP applications

- DP in auctions.
- Uber's DP SQL database.
- Private machine learning
    - DP Logistic regression (Chaudhuri et al. NeurIPS09)
    - DP PCA (Chaudhuri et al., NeurIPS12)
    - DP Low-rank approximation (Hardt et al., STOC12)

- "Reusable holdout": DP to prevent over-fitting. [Dwork, Feldman, Hardt, Pitassi, Reingold, Roth, Science 2015.]



DP Summary

- DP is called *differential* because it concerns itself with whether an analysis changes when a subject participates (vs. not) in a dataset.

- Building DP into systems does not protect all persons from all privacy harms. Only per above.

- **Centralized model:** corrupt query output.
- **Local model:** corrupt data.

- DP can **prevent overfitting** in ML. Said differently: overfitting is bad for privacy! Related to DPs strong ties to robust statistics.

DP Summary

- DP is called *differential* because it concerns itself with whether an analysis changes when a subject participates (vs. not) in a dataset.

- Building DP into systems does not protect all persons from all privacy harms. Only per above.

- **Centralized model:** corrupt query output.
- **Local model:** corrupt data.

- DP can **prevent overfitting** in ML. Said differently: overfitting is bad for privacy! Related to DPs strong ties to robust statistics.

Break!

Discussion:

Data transparency

Agenda

- A tour of complicated case studies in data sharing.

- Examples:
    - HuffPo Fundrace
    - California Prop 8 “eightmaps”
    - Sandy hook gun owners
    - Mug shots
    - Wikileaks & DNC/Clinton emails

FOIA and public records

- Johnson et al. (2011): **“Public records”** and changing nature of **“practical obscurity”**.

- **US Freedom of Information Act (FOIA)** enacted 1967.
- **E-FOIA** amendments of 1996 added electronic requirements.

- Many complexities, e.g. Megan’s law, campaign finance.

FOIA and public records

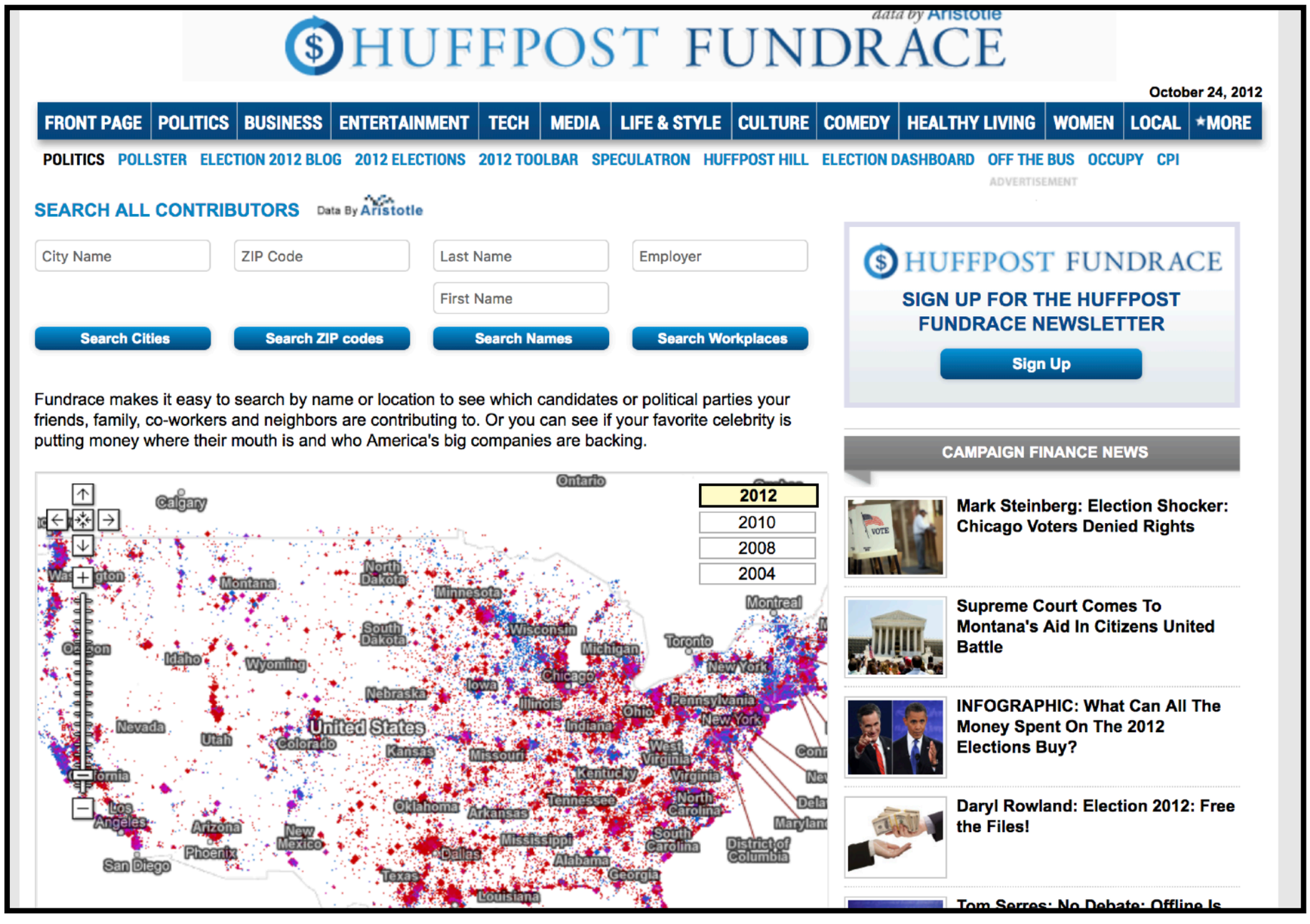
- Johnson et al. (2011): **“Public records”** and changing nature of **“practical obscurity”**.
- **US Freedom of Information Act (FOIA)** enacted 1967.
- **E-FOIA** amendments of 1996 added electronic requirements.
- Many complexities, e.g. Megan’s law, campaign finance.
- 1989 SCOTUS decision, general comment:
    - “there is a vast difference between **(a)** the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and **(b)** a computerized summary located in a single clearinghouse of information.”

FOIA and public records

- *DOJ vs. Reporters Comm. for Freedom of the Press,* 1989 SCOTUS ruling about “rap sheets”:
    - “as a categorical matter, the granting of a third party's FOIA request for the disclosure of law enforcement records or information about a private citizen can reasonably be expected to invade that citizen's privacy for the purpose of Exemption 7(C), and **when such a request** seeks no official information about a Federal Government agency, but **merely seeks records that the Federal Government happens to be storing**, **the invasion of privacy is "unwarranted"** for the purpose of Exemption 7(C);”

Huffington Post FundRace

- Johnson et al. (2011) specifically about campaign finance disclosure.



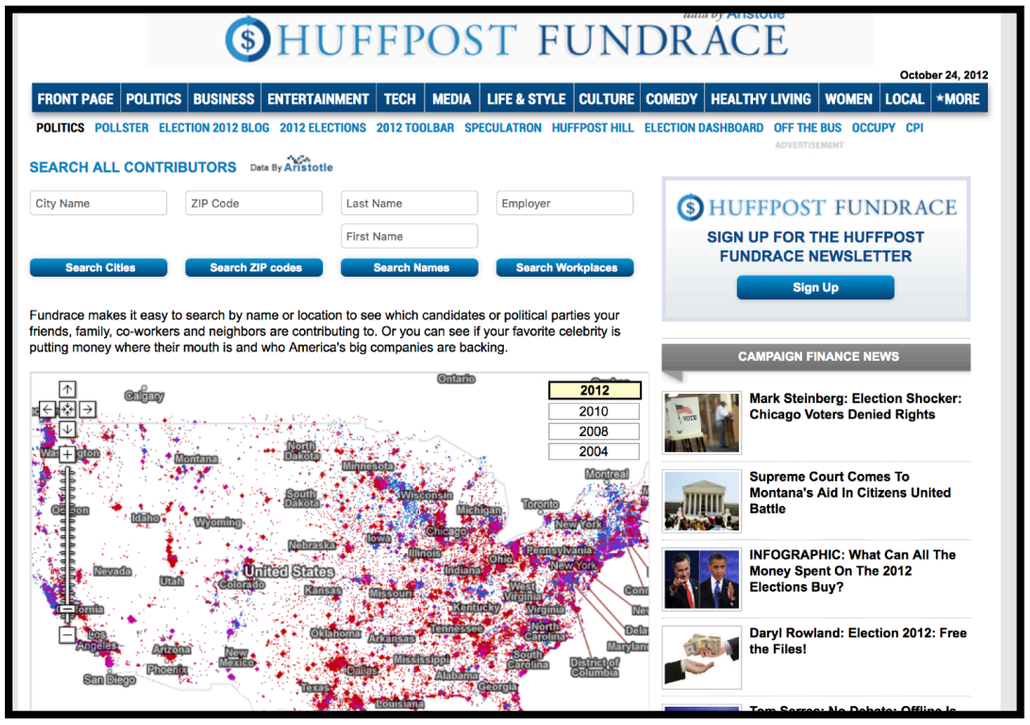
Huffington Post FundRace

- Johnson et al. (2011) specifically about campaign finance disclosure.

The highlighting and shading can be illustrated with a simple example. If one searches on Google for "Kent Wayland" (one of the authors of this paper), one of the top results links to a database available at the *Huffington Post*, an online newspaper, where users may browse campaign donations with specially designed Web tools, by name, zip code, date of donation, campaign season, etc.[108] The database further displays recent political donations, downloaded from the FEC, repackaged and subject to indexing by Google's Web crawler.[109] Although Wayland's political activity and campaign donations are relatively minor, they make up a significant component of his online identity due to the high ranking Google gives these search results. Information on Wayland's contributions is not just bounced from site to site; his contributions become a highlighted aspect of his Web presence because of the combination of the way Google works, the *Huffington Post*'s popularity, and Wayland's other (in)activity: Wayland's name is not especially common, his Web presence is not especially extensive, and the *Huffington Post* is recognized by Google's search engine as an especially popular site. This illustrates how the architecture of the Internet may shape a person's identity in a unpredictable way.

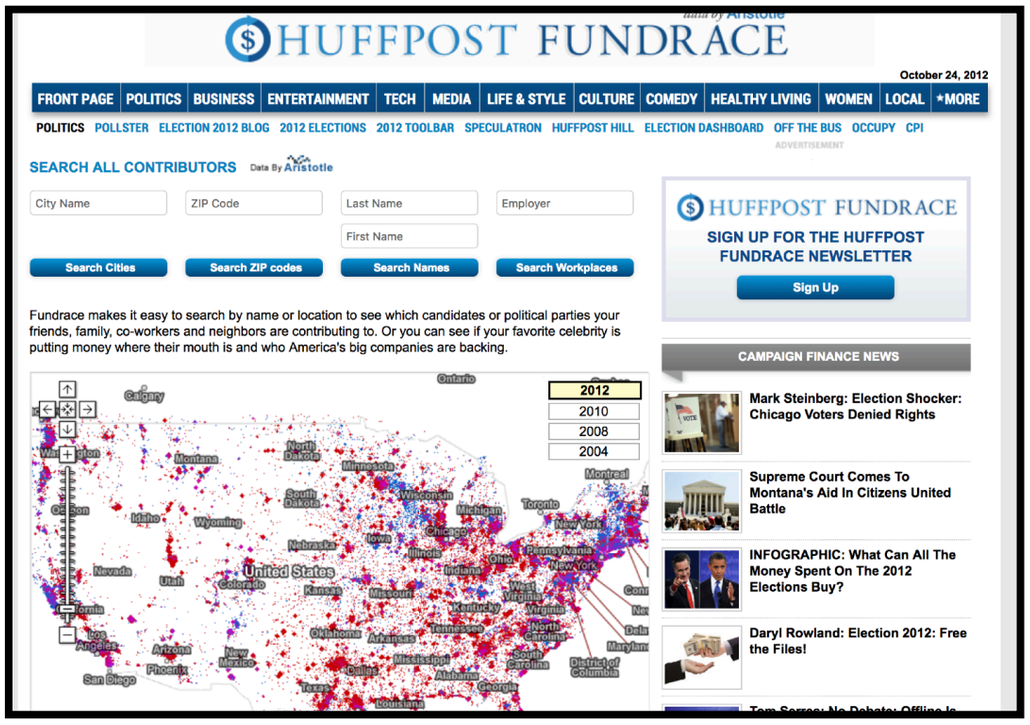
Huffington Post FundRace

- The internet changed the nature of “practical obscurity”.
- Search engines changed “practical obscurity” again (Cf. “right to be forgotten”).

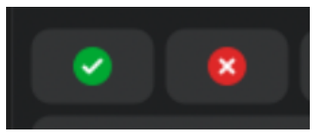


Huffington Post FundRace

- The internet changed the nature of “practical obscurity”.
- Search engines changed “practical obscurity” again (Cf. “right to be forgotten”).



- Appropriate?



Right to be forgotten

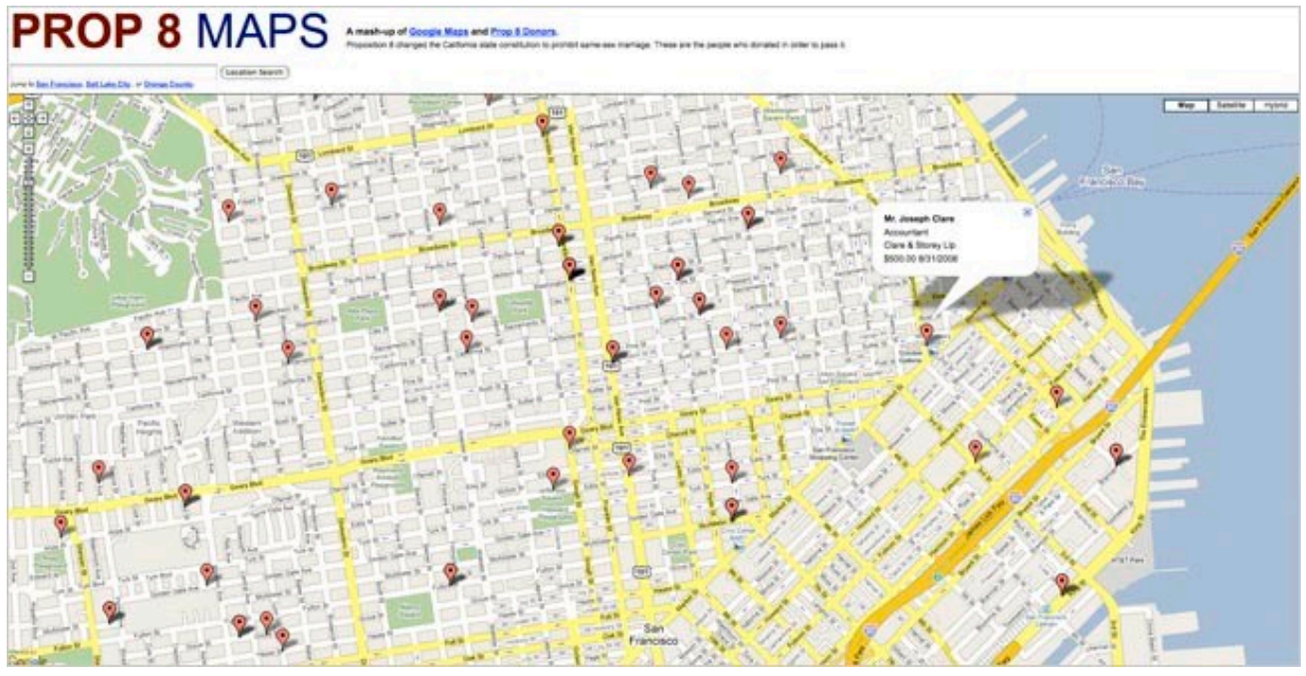
- Desire of people to "determine the development of their life in an autonomous way, without being perpetually or periodically stigmatized as a consequence of a specific action performed in the past.”

- Many “right to privacy”-aligned arguments for; arguments against tend to be more subtle, e.g. speech/censorship and “social value” of celebrity.

- Recommended: Google’s 2018 paper: “5 years of R2BF”

Eightmaps

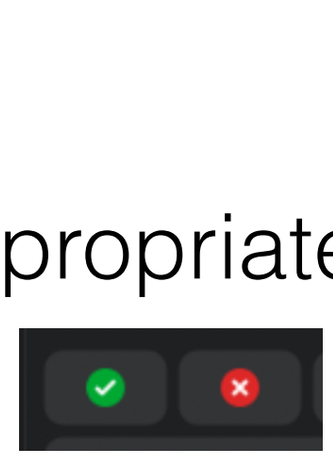
- CA Political Reform Act of 1974 (also other states): sunlight laws about campaign finance.
- Parallels thorny issues about political donations being “free speech” (*Citizens United vs. FEC*).
- “free speech” != “anonymous speech”

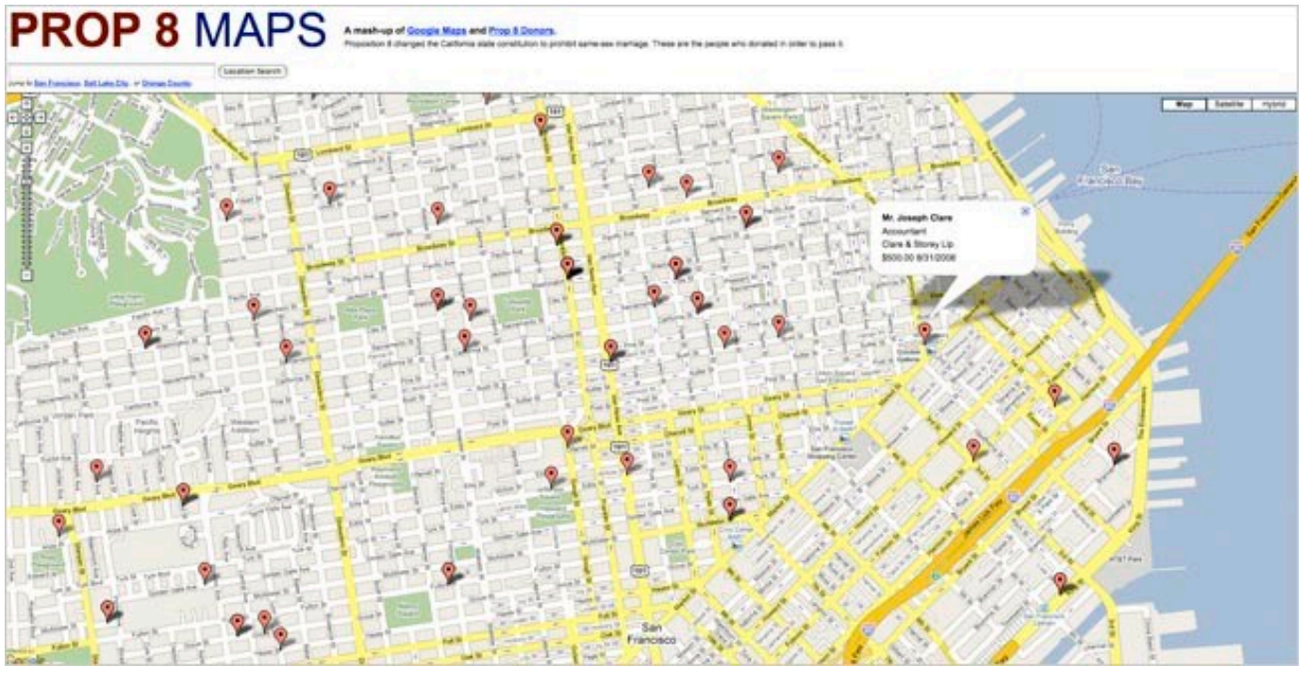


Eightmaps

- CA Political Reform Act of 1974 (also other states): sunlight laws about campaign finance.
- Parallels thorny issues about political donations being “free speech” (*Citizens United vs. FEC*).
- “free speech” != “anonymous speech”

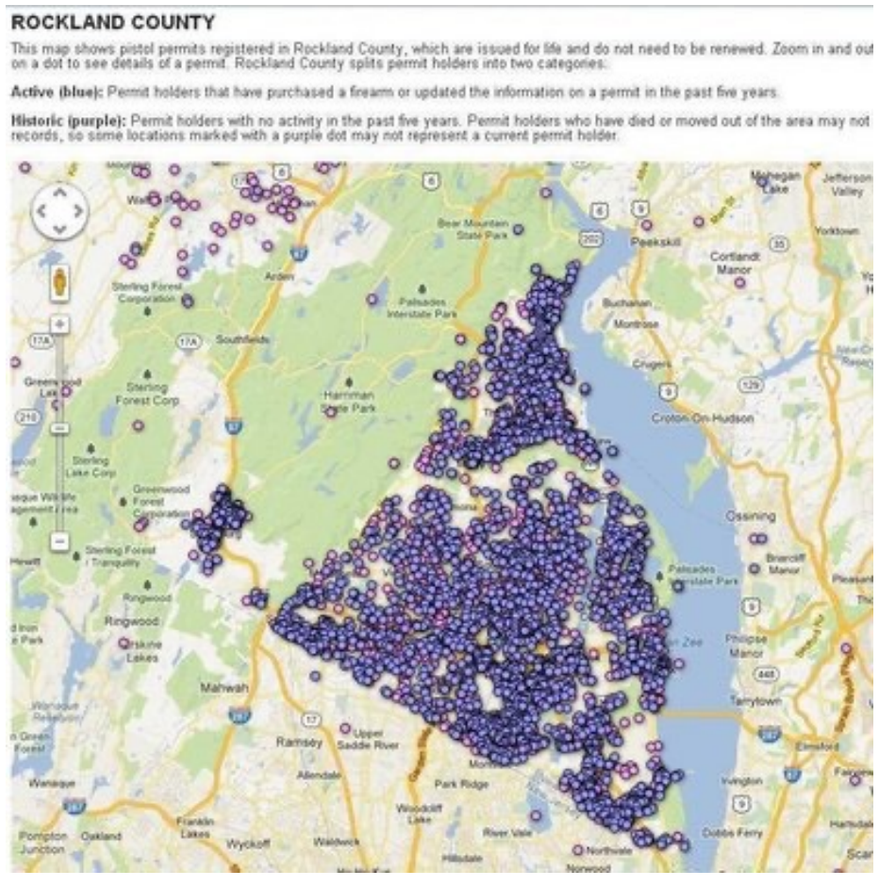
- Appropriate?





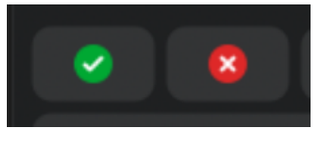
Sandy hook gun map

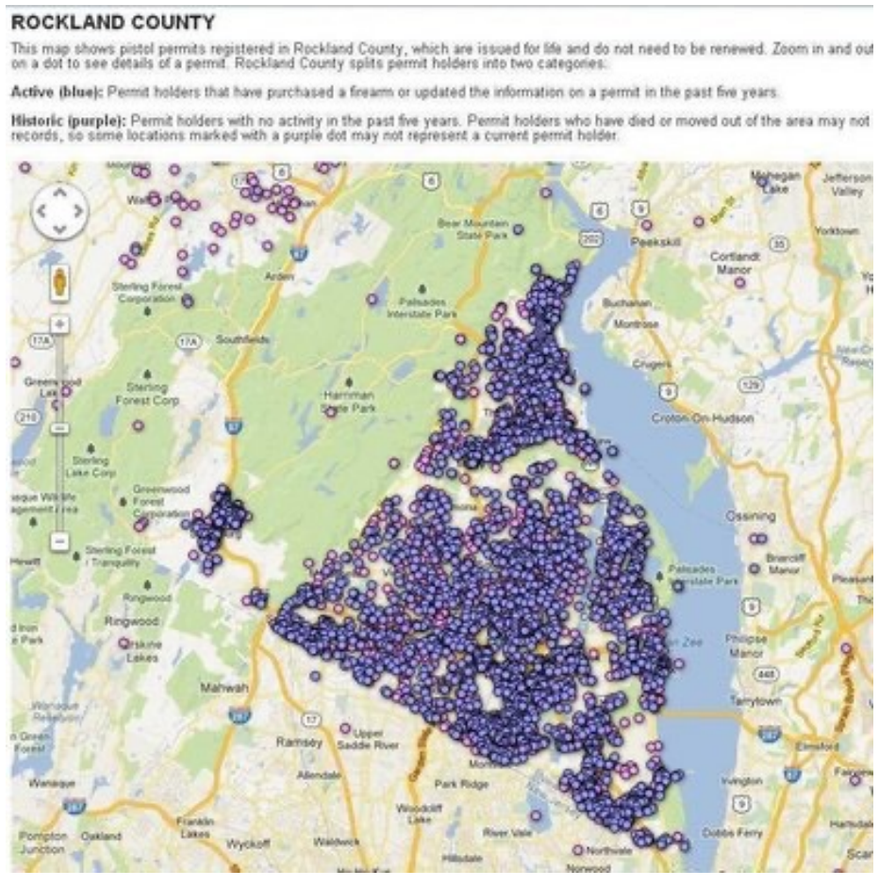
- Journal News (local NY newspaper) story: “Map: Where are the gun permits in your neighborhood?”
- Based on FOIA requests for gun permits.
- Lead to several break-ins and stolen guns.
- Response “Map: Where are the Journal News employees in your neighborhood?”



Sandy hook gun map

- Journal News (local NY newspaper) story: “Map: Where are the gun permits in your neighborhood?”
- Based on FOIA requests for gun permits.
- Lead to several break-ins and stolen guns.
- Response “Map: Where are the Journal News employees in your neighborhood?”
- Appropriate?





Mug shots

- Mugshots.com, JustMugshots, etc.
- Lee 2018: “majority of states deem mugshots **open records** under public records laws … mugshot companies & press have constitutional right to publish”
- Exceptions, e.g. Ohio has a right-to-publicity statute.
- “Removal fees” business model, but also “vigilante”.

JOHN LOVETT: ACCORDING TO CHRON.COM, SAN JACINTO COUNTY JUDGE ARRESTED ON THE COUNTS OF BURGLARY, TAMPERING WITH AN OFFICIAL GOVERNMENT INSTRUMENT AND FORGERY — 4/24/2018

Coldspring, TX - Coldspring residents near the courthouse square Monday afternoon were surprised to see San Jacinto County Sheriff Greg Capers leading County Judge John Lovett out of the courthouse in handcuffs.

Lovett was arrested at 12:23 p.m. after charges stemming from a July 2017 incident went before the grand jury and Lovett was indicted.

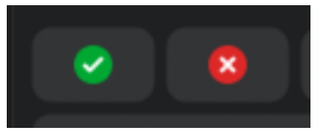
"The grand jury returned a true bill for an indictment against John Lovett on ...

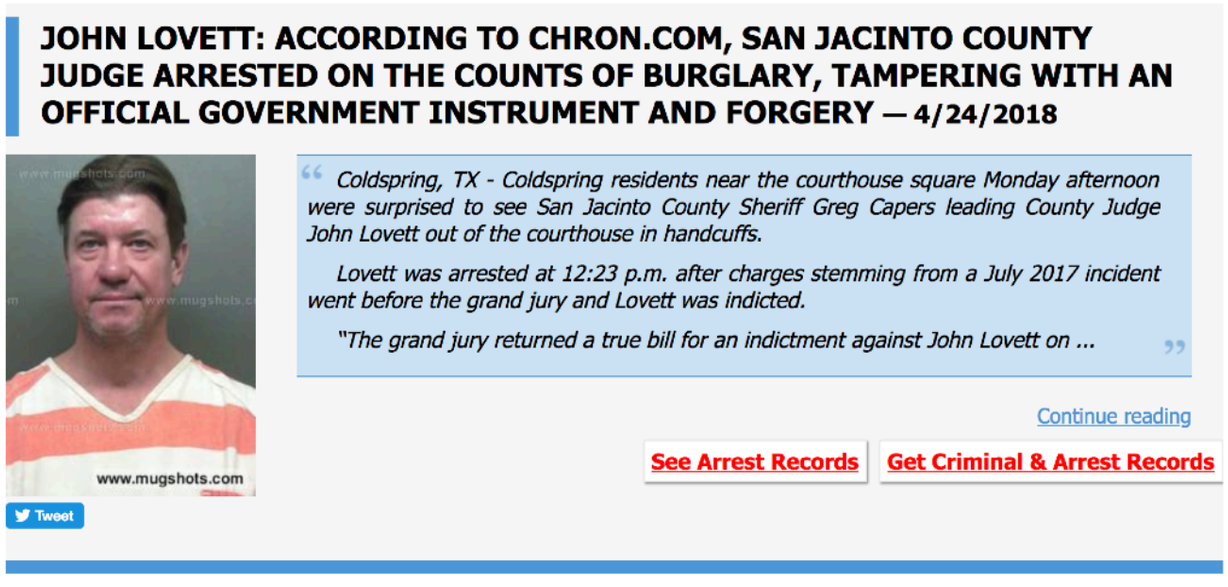
Continue reading

See Arrest Records | Get Criminal & Arrest Records

Tweet

Mug shots

- Mugshots.com, JustMugshots, etc.
- Lee 2018: “majority of states deem mugshots **open records** under public records laws … mugshot companies & press have constitutional right to publish”
- Exceptions, e.g. Ohio has a right-to-publicity statute.
- “Removal fees” business model, but also “vigilante”.
- Appropriate? 



Latanya Sweeney & Ads

- Google ads for “Latanya Sweeney” vs. “Jill Schneider”

Ads by Google

Latanya Sweeney, Arrested?
1) Enter Name and State. 2) Access Full Background Checks Instantly.
www.instantcheckmate.com/

Latanya Sweeney
Public Records Found For: **Latanya Sweeney**. View Now.
www.publicrecords.com/

La Tanya
Search for La Tanya Look Up Fast Results now!
www.ask.com/La+Tanya

Ads related to **Jill Schneider** ⓘ

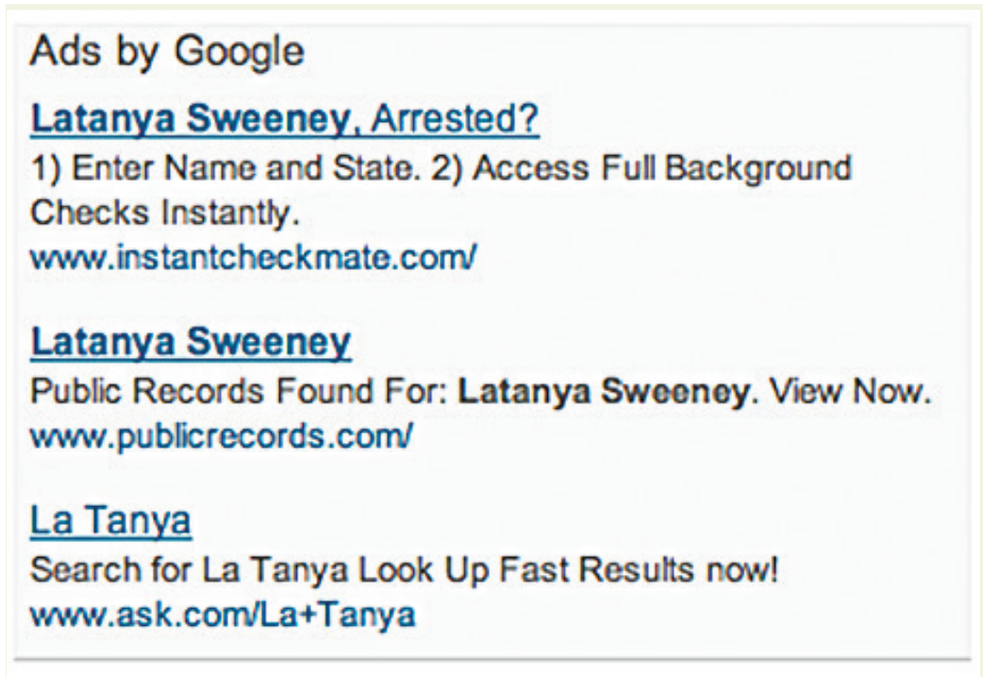
Jill Schneider Art
www.posters2prints.com/
Custom Frame Prints and Canvas. Shop Now, SAVE Big + Free Shipping!

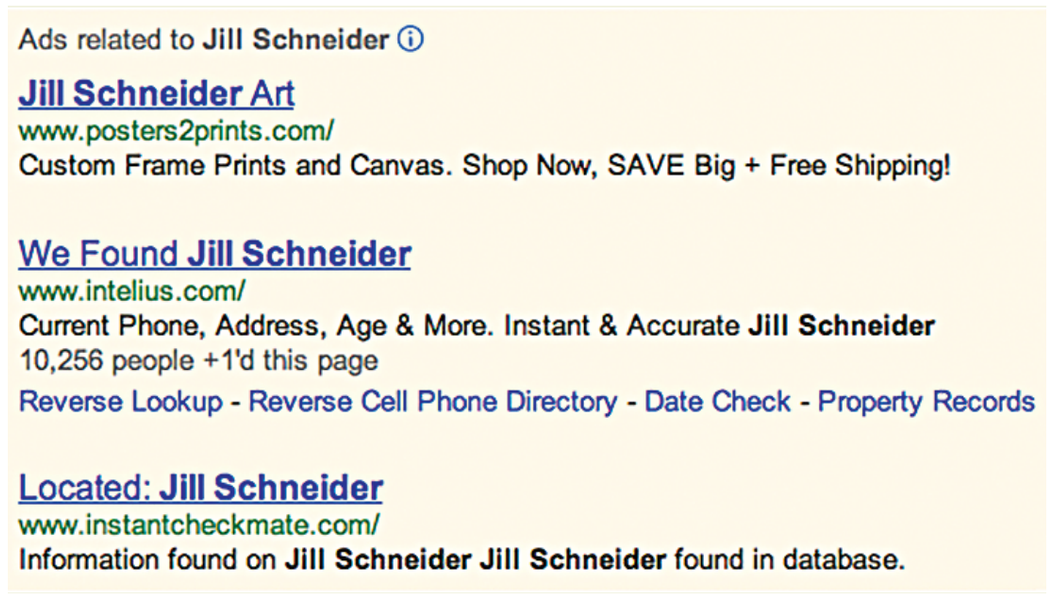
We Found **Jill Schneider**
www.intelius.com/
Current Phone, Address, Age & More. Instant & Accurate **Jill Schneider**
10,256 people +1'd this page
Reverse Lookup - Reverse Cell Phone Directory - Date Check - Property Records

Located: **Jill Schneider**
www.instantcheckmate.com/
Information found on **Jill Schneider Jill Schneider** found in database.

Latanya Sweeney & Ads

- Google ads for “Latanya Sweeney” vs. “Jill Schneider”





- Greater percentage of InstantCheckmate ads used **“arrest”** for black-identifying first names than white.
- Why?

- L. Sweeney (2013) “Discrimination in Online Ad Delivery”. ACM Queue.

Latanya Sweeney & Ads

- "**Is Instant Checkmate, Google, or society to blame?** We do not yet know. Google understands that an advertiser may not know which ad copy will work best, so the advertiser may provide multiple templates for the same search string, and the 'Google algorithm' learns over time which ad text gets the most clicks from viewers."

Ads by Google

Latanya Sweeney, Arrested?
1) Enter Name and State. 2) Access Full Background Checks Instantly.
www.instantcheckmate.com/

Latanya Sweeney
Public Records Found For: **Latanya Sweeney**. View Now.
www.publicrecords.com/

La Tanya
Search for La Tanya Look Up Fast Results now!
www.ask.com/La+Tanya

Ads related to **Jill Schneider** ⓘ

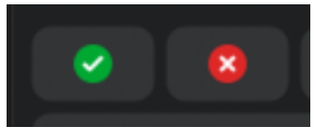
Jill Schneider Art
www.posters2prints.com/
Custom Frame Prints and Canvas. Shop Now, SAVE Big + Free Shipping!

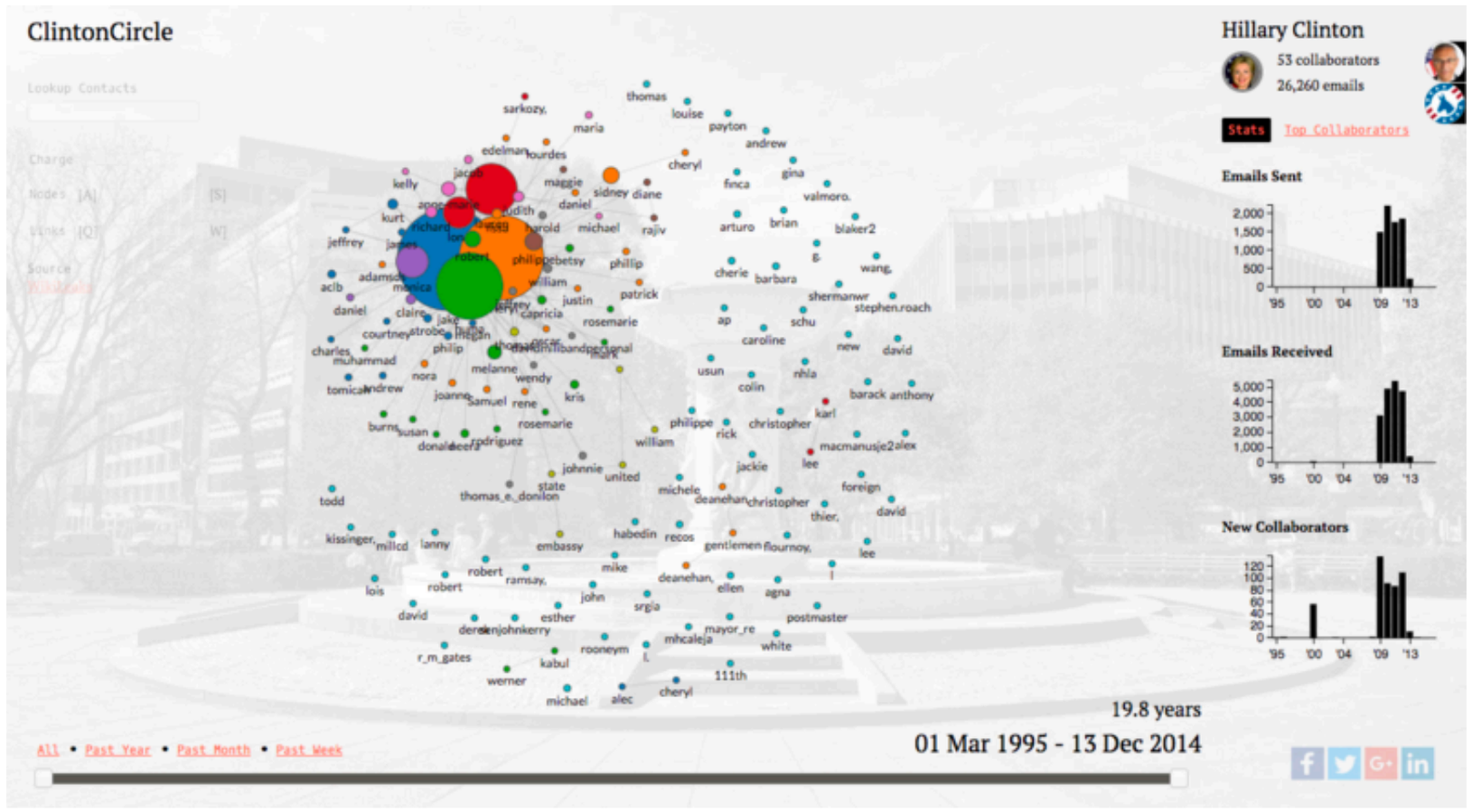
We Found **Jill Schneider**
www.intelius.com/
Current Phone, Address, Age & More. Instant & Accurate **Jill Schneider**
10,256 people +1'd this page
Reverse Lookup - Reverse Cell Phone Directory - Date Check - Property Records

Located: **Jill Schneider**
www.instantcheckmate.com/
Information found on **Jill Schneider Jill Schneider** found in database.

Wikileaks

- MIT Media Lab researchers built a visualization tool to explore the Clinton/DNC/Podesta emails.
- Based on Wikileaks dump (not FOIA dump).

- Appropriate? 



Other public records

- Real estate listings: good/bad?
- What else have you found about yourself?

INTERNET | LINK BY LINK

Law Students Teach Scalia About Privacy and the Web

By **NOAM COHEN** MAY 17, 2009

This spring, the students of an elective course on Internet privacy at Fordham Law School experienced a number of fascinating "teaching moments" during an assignment meant to demonstrate how much personal information is floating around online.

The assignment from the class's professor, Joel R. Reidenberg, was, admittedly, a bit provocative: create a dossier about Supreme Court Justice Antonin Scalia from what can be found on the Internet.

Why Justice Scalia? Well, the class had been discussing his recent dismissive comments about Internet privacy concerns at a conference. His summation, as reported by The Associated Press: "Every single datum about my life is private? That's silly."

Other public records

- Real estate listings: good/bad?
- What else have you found about yourself?

INTERNET | LINK BY LINK

Law Students Teach Scalia About Privacy and the Web

By **NOAM COHEN** MAY 17, 2009

This spring, the students of an elective course on Internet privacy at Fordham Law School experienced a number of fascinating "teaching moments" during an assignment meant to demonstrate how much personal information is floating around online.

The assignment from the class's professor, Joel R. Reidenberg, was, admittedly, a bit provocative: create a dossier about Supreme Court Justice Antonin Scalia from what can be found on the Internet.

Why Justice Scalia? Well, the class had been discussing his recent dismissive comments about Internet privacy concerns at a conference. His summation, as reported by The Associated Press: "Every single datum about my life is private? That's silly."

Break!

General Data Protection Regulation

GDPR Timeline

- **Jan 2012**: Proposal released
- **April 2016**: Adopted by European Parliament
- **May 2018**: GDPR became enforceable.

- Enforced by each national data protection authority.

- See: www.gdpr-info.eu

Other regulations

- US: Health Ins. Portability and Accountability Act (HIPAA)
    - 1996
- European Union: GDPR <— **Focus today**
    - 2016
- Vermont: Vermont Data Broker regulation (“Act 171”)
    - 2018
- California: CCPA (California Consumer Privacy Act)
    - Jan 2020
    - Nov 2020: Modified by ballot initiative (Prop 24)

GDPR Outline

- **Articles 1-11:** General provisions, principles
- **Articles 12-23:** Rights of the data subject
- **Articles 24-42:** Controller and processor
- **Articles 44-76:** Third countries, supervision, cooperation
- **Articles 77-84:** Penalties
- **Articles 85-91:** Exceptions (e.g. employment)
- **Articles 92-99:** Eurocrat details

- Today will focus on 1-23, background and rights.

Articles 1-11:

General provisions and principles

Articles 2-3: Scope

- “This Regulation applies to the processing of personal data wholly or partly by automated means…”
- Except:
    - By EU member states
    - “by a natural person in the course of a purely personal or household activity”
    - Law enforcement exceptions

- Applies to any processor “offering goods or services” (regardless of payment or not) to EU data subjects (regardless of location of processor or processing).

Article 4: Definitions

- **Personal data:** any information relating to an identified or identifiable natural person.

- **Identifiable natural person:** one who can be identified, directly or indirectly.

- **Profiling:** “any form of automated processing of personal data […] to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements”

Article 5: Principles

- Specific purpose *and* in service of that purpose.

- **Purpose limitation:** collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes

- **Data minimisation:** Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

Article 6-8: Lawfulness, Consent

- Processing is lawful only **if data subject has given consent** or processing is necessary for:
    - performance of a contract (data subject is party)
    - legal obligations of controller
    - to protect the vital interests of the data subject
    - for the performance of a task in the public interest
    - Legitimate interests of the controller except where in conflict with rest of GDPR.
- **Consent**:
    - the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language.
    - Can be withdrawn at any time. It shall be as easy to withdraw as to give consent.
- Without consent, utmost account shall be taken of whether processing of personal data that is necessary for the contract/service.

Article 6-8: Lawfulness, Consent

- Processing is lawful only **if data subject has given consent** or processing is necessary for:
    - performance of a contract (data subject is party)
    - legal obligations of controller
    - to protect the vital interests of the data subject
    - for the performance of a task in the public interest
    - Legitimate interests of the controller except where in conflict with rest of GDPR.
- **Consent**:
    - the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language.
    - Can be withdrawn at any time. It shall be as easy to withdraw as to give consent.
- Without consent, utmost account shall be taken of whether processing of personal data that is necessary for the contract/service.

Article 9-10: Special data

- Processing of personal data revealing **racial or ethnic origin**, **political opinions**, **religious** or **philosophical beliefs**, or **trade union membership**, and the processing of **genetic data**, **biometric data** for the purpose of uniquely identifying a natural person, data concerning **health** or data concerning a natural person's **sex life or sexual orientation** shall be **prohibited**.
- Exceptions:
  - explicit consent for specified purposes
  - processing relates to personal data which are manifestly made public by the data subject
  - Medical, criminal, public health, historical purposes.

Article 11: Non-identifying

- **The article, In full:**
  "If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject **for the sole purpose of complying** with this Regulation.

  Where, in cases referred to in paragraph 1 of this Article, the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 (rights to access) to 20 (rights to portability) shall not apply except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification."

Article 12-23: Rights!

Rights of the data subject

- **Articles 12-14:**
  - Details of how to exercise rights
  - Providing contact details for data protection officer.
  - Transparency about appeal process, “meta-privacy”.

- **15**: Right of access
- **16**: Right to rectification
- **17**: Right to be forgotten
- **18-19**: Right to restriction of processing
- **20**: Right to data portability
- **21**: Right to object
- **22**: Profiling, right to an explanation
- **23**: Restrictions (super-exceptions)

Article 15: Rights of access

- The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, **access to the personal data** and the purpose of the processing.

Article 16: Right to rectification

- Right to correct errors in private data in a timely manner.

Article 17: Right to be forgotten

- Discussed Google R2BF paper.
- Erasure when one of the following applies:
    - personal data are no longer necessary
    - withdrawing consent
    - Data has been unlawfully processed
    - Legal obligation in EU or member state law
- Exceptions:
    - Right of freedom of expression and information
    - Processing is part of obligations of controller: legal, health, public interest, history.

Art. 18-19: restriction, notification

- Right to restrict processing if:
  - Objection (Article 21) has been filed
  - Also edge cases like: the accuracy of the data is contested, for a period of time needed to determine accuracy.

- Controller required to keep data subject notified in any restriction, erasure, etc. process.

Art. 20: Right to data portability

- Extension of Article 15, Right to access. Like E-FOIA.

- The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, **in a structured, commonly used and machine-readable format** and have the right to transmit those data to another controller without hindrance from the controller.

- The data subject shall have the right to have the personal data **transmitted directly from one controller to another**, where technically feasible.

Article 21: Right to object

- The data subject shall have the right to object at any time to processing of personal data concerning him or her.

- Where personal data are processed for **direct marketing purposes**, the data subject shall have the right to object at any time to processing of personal data concerning him or her for such marketing, which includes profiling to the extent that it is related to such direct marketing.

Article 22: Right to explanation

- The data subject shall have the right not to be subject to a decision based solely on automated processing, including **profiling**, which produces legal effects concerning him or her or similarly significantly affects him or her.

- Exceptions:
  - Explicit consent.
  - Profiling is necessary for the contract/service
  - Coverage of other EU/member state law

- Safeguard data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

GDPR Outline

- **<u>Articles 1-11:</u>** General provisions, principles
- **<u>Articles 12-23:</u>** Rights of the data subject
- **Articles 24-42:** Controller and processor
- **Articles 44-76:** Third countries, supervision, cooperation
- **Articles 77-84:** Penalties
- **Articles 85-91:** Exceptions (e.g. employment)
- **Articles 92-99:** Eurocrat details

Article 83: Fines

- Articles 8 (child extras), 11 (identifying), 25-39 (breach), 42-43 (certification):
    - Fines up to max(10M EUR, **2%** worldwide turnover)

- Articles 5-7 (basics), 9 (special data), 12-22 (rights), 44-49 (third countries):
    - Fines up top max(20M EUR, **4%** worldwide turnover)

Article 83: Fines

- Articles 8 (child extras), 11 (identifying), 25-39 (breach), 42-43 (certification):
    - Fines up to max(10M EUR, **2%** worldwide turnover)

- Articles 5-7 (basics), 9 (special data), 12-22 (rights), 44-49 (third countries):
    - Fines up top max(20M EUR, **4%** worldwide turnover)

- **First lawsuits** filed over informed consent against:
    - FB, Instragram, WhatsApp: seeking 3.9B EUR each.
    - Google Android: seeking 3.7B EUR

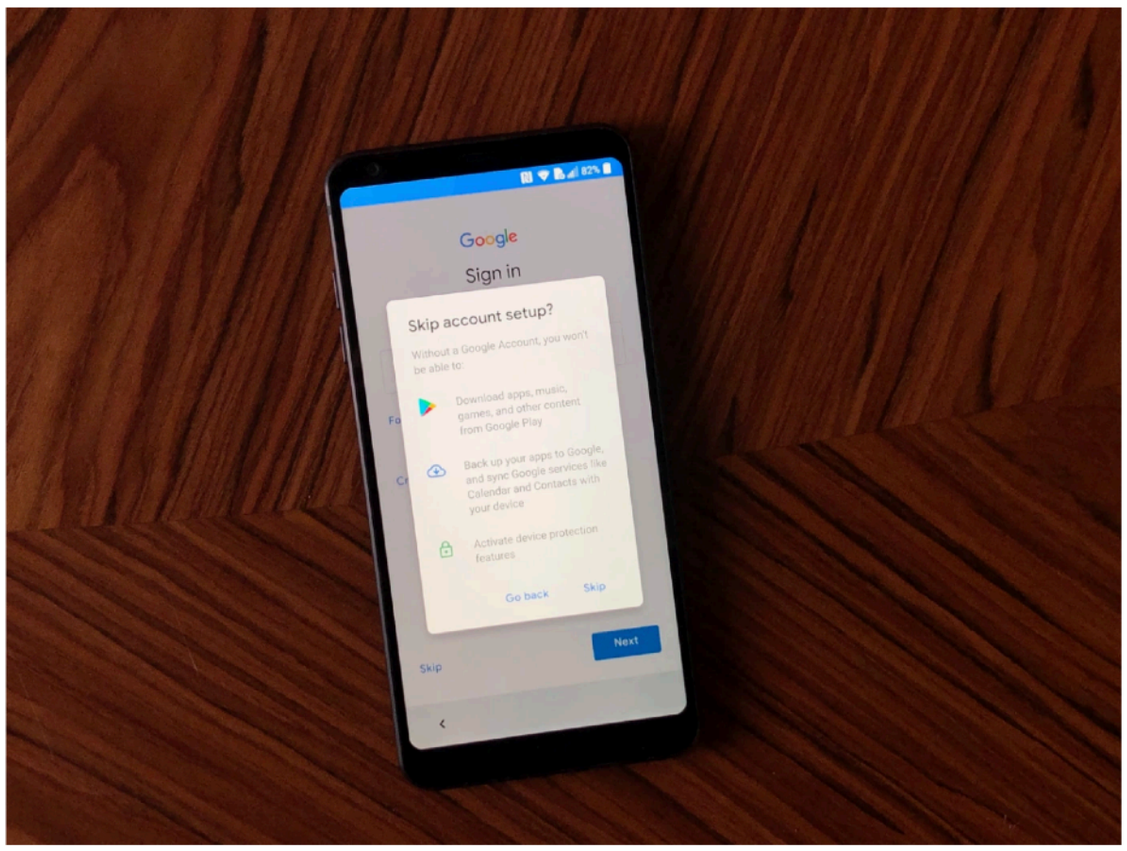
GDPR Fines, 2019

Consent:

French data protection watchdog fines Google $57 million under the GDPR

Romain Dillet @romaindillet / 7:46 AM PST • January 21, 2019

Comment



Breaches:

British Airways faces record £183m fine for data breach

8 July 2019

British Airwa
security syste

Marriott to face $123 million fine by UK authorities over data breach

Zack Whittaker @zackwhittaker / 7:36 AM PDT • July 9, 2019

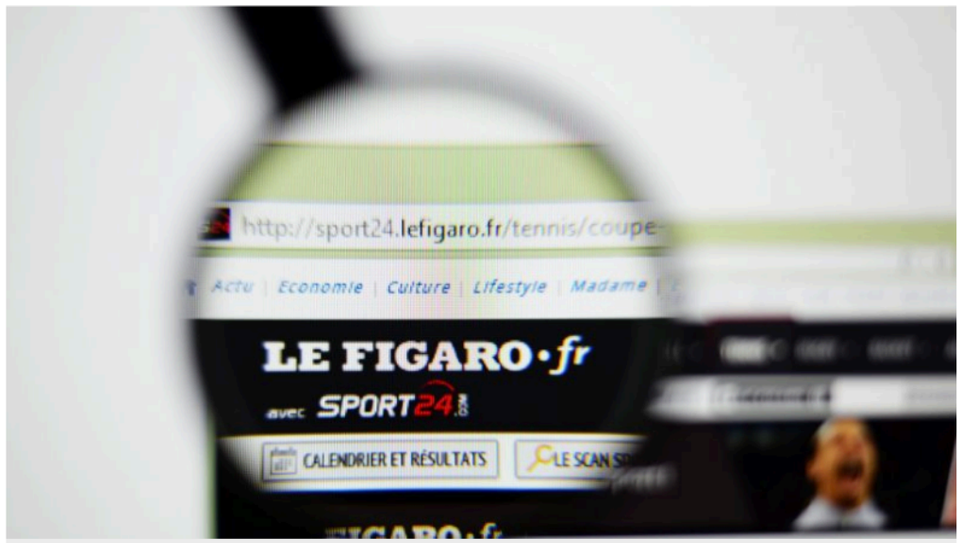
BA fine reduced to £20m and Marriot to $23m. Both haircuts cited the pandemic.

GDPR Fines, 2021

Cookies and consent:

Le Figaro publisher fined €50,000 for GDPR violation

By Molly Killeen | EURACTIV.com — Jul 29, 2021 (updated: Jul 29, 2021)



The CNIL found that advertising cookies had been placed on the computers of visitors to the lefigaro.fr website without their consent. [Gil C / Shutterstock]

Languages: Français — Comments | Print

French publisher *Le Figaro* has been fined €50,000 by the country's data protection authority after its website was discovered to be installing third-party advertising cookies without the users' consent.

EURACTIV

CCIA – Compute
CECIMO - Europ
CEN & CENELEC
EDAA - Europea
Eupportunity
GIGAEurope
Photonics21
PostEurop – You
Red Flag

Details emerging:

EU hits Amazon with record-breaking $887M GDPR fine over data misuse

Carly Page @carlypage_ / 8:06 AM PDT • July 30, 2021 — Comment

Revealed in Amazon's quarterly earnings report, Summer 2021. Silence during last year.

Summary, Day 2

- Data privacy is hard! **Differential privacy** can address concerns that your participation (vs non-participation) in a data product will bring you harm.

- **Transparency** is hard. Reasonably people can disagree about how “public” public records should be.

- The EU’s **GDPR** has been laying the groundwork for data privacy regulation worldwide. Three years on, laws are still largely untested.

Thank you!

Homepage of full Stanford course: http://msande234.stanford.edu/