

Privacy in an age of information abundance

Betsy Masiello

October 1, 2009

Princeton University

Center for Information Technology Policy

The transparent century

Malcom Gladwell: The Enron story is more important now than 10 years ago. It was the canary in the coal mine.*

- Fifty years ago, uncovering the unknown solved problems
- Solving today's problems requires making sense of large amounts of known information

When fixing the world involves analyzing information about *people*...how does privacy fit in?

21st century problems

- Medicine
 - Pandemic prevention
 - Recognizing local outbreaks quickly
 - Virus identification
 - Climate-sensitive disease prediction
- Environment
 - Managing electric power demand
 - Understanding climate change
- Predicting financial crises
- Uncovering corrupt governance
- Crop planning

Can Search information help solve some of these problems?



obama, clinton, edwards

Search Trends

[Sign in](#) to see and export additional Tre

Tip: Use commas to compare multiple search terms.

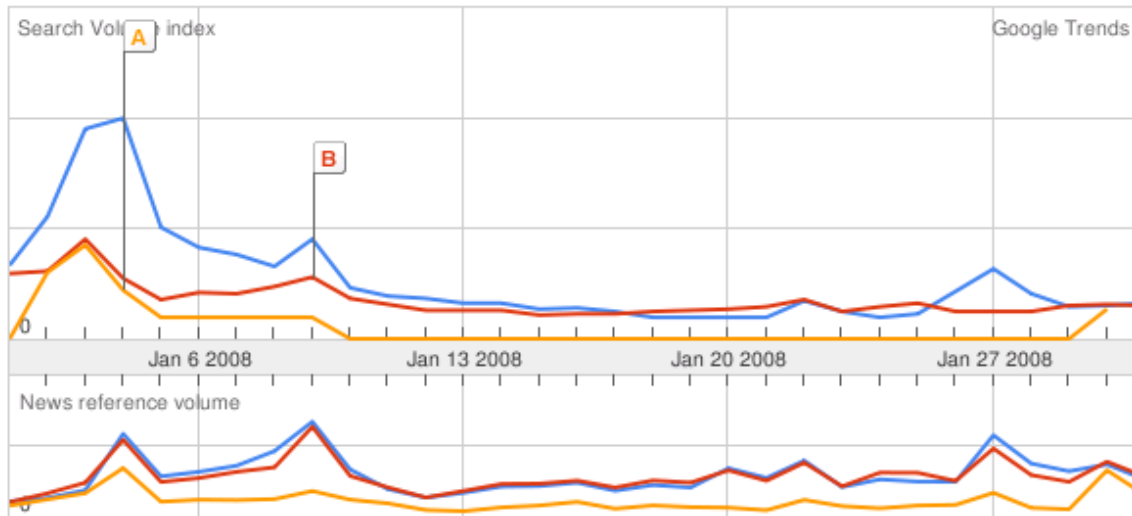
Searches [Websites](#)

United States

Iowa

Jan 2008

● obama ● clinton ● edwards



A [Edwards stumbles in Iowa](#)
Kentucky.com - Jan 4 2008

B [Clinton, McCain win in NH](#)
Maryville Daily Times - Jan 9 2008

[More news results »](#)

Rank by

Cities

1. Iowa City, IA, USA
2. Fairfield, IA, USA
3. Cedar Rapids, IA, USA
4. Des Moines, IA, USA

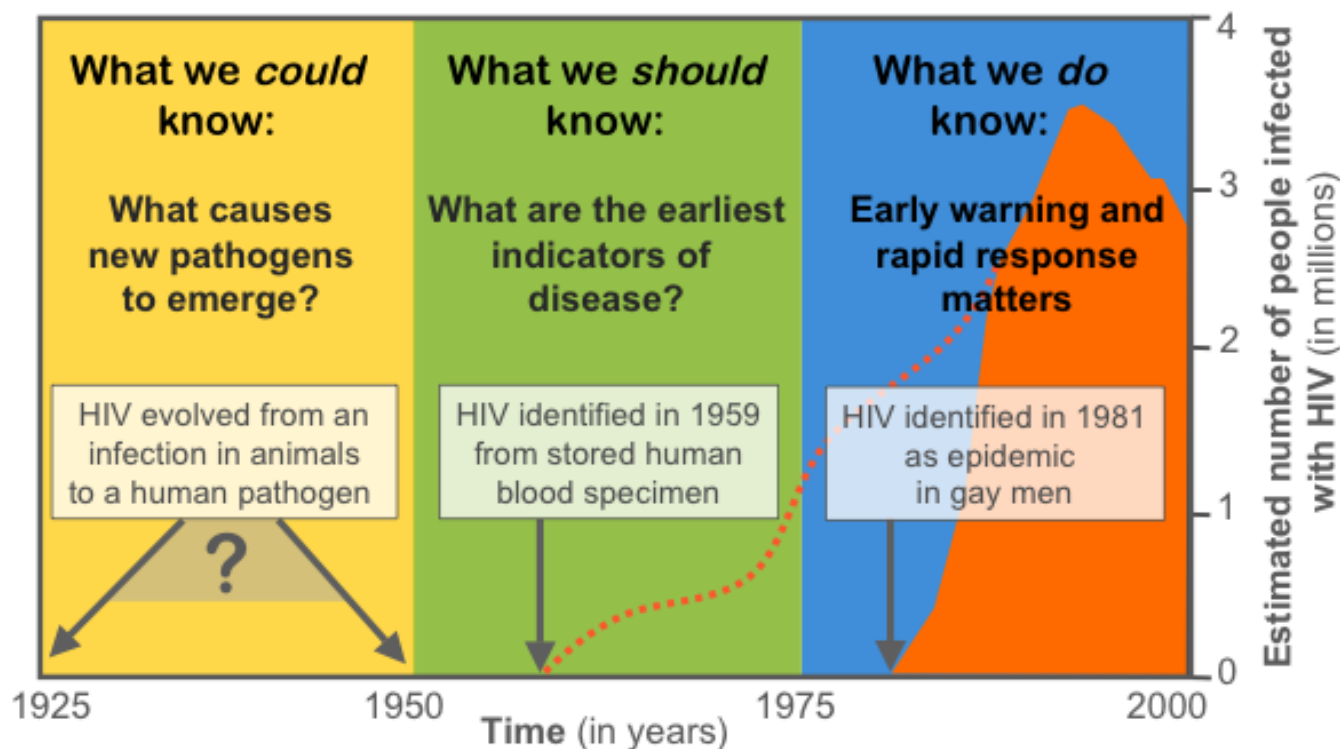
Languages

1. English

Predicting and preventing the next pandemic

1-2 new diseases have emerged every year for the last three decades, examples include **HIV**, **SARS**, and **bird flu (H5N1)**

75% of these new and re-emerging diseases are zoonosis, i.e., they spread from animals to humans



Google Flu Trends



[Google.org home](#)

Flu Trends

[Home](#)

[How does this work?](#)

[FAQ](#)

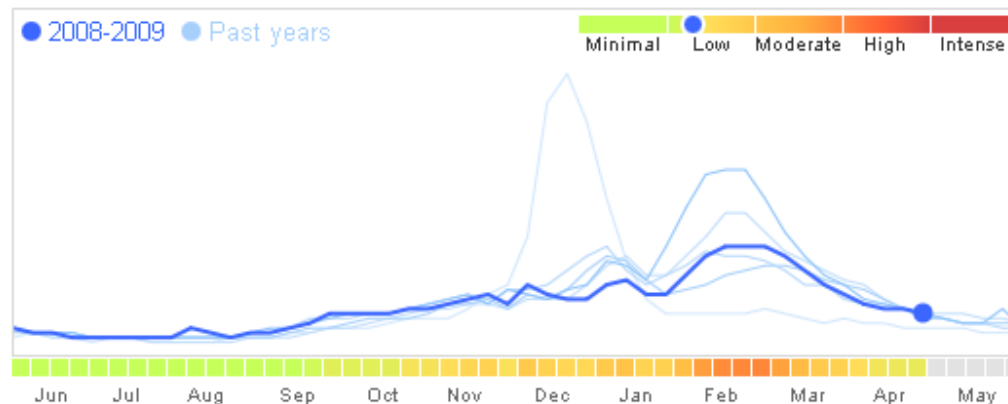
[Download raw data](#)

Explore flu trends across the U.S.

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity in your state up to two weeks faster than traditional systems. [Read more »](#)

United States flu activity: Low

Entire United States ▼



First step: flu vaccine

CDC recommends a yearly flu vaccine as the first and most important step in protecting against influenza. The flu vaccine protects against the three main flu strains that research indicates will cause the most illness during the flu season. And, the vaccine can protect against illness from similar flu viruses.



Flu shot locator

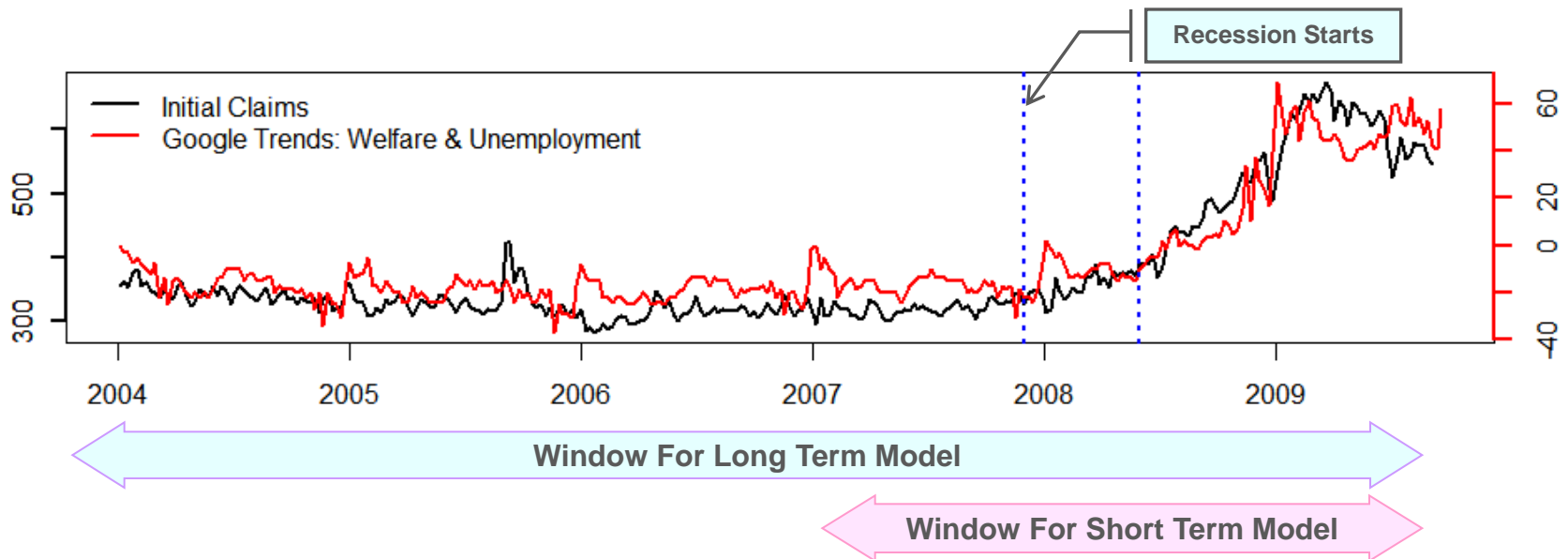
ZIP code:

Distance:

Source: American Lung Association

Flu in the news

Predicting unemployment



- According to the NBER, the current recession started December 2007.
- National unemployment rate passed 5% in mid 2008 and search queries on [Welfare and Unemployment] also increased at same time.

	Week	5/24/09 - 5/30/09	5/31/09 - 6/6/09	6/7/09 - 6/13/09	6/14/09 - 6/20/09	6/21/09 - 6/27/09	6/28/09 - 7/4/09
US Dept of Labor	Initial Claims(K)	625	605	612	630	617	Release at 7/9/09
	Continued Claims(MN)	8.84	6.71	6.76	6.72	6.90	
Google Trends	Jobs	-1%	-1%	0%	-1%	-2%	-3%
	Recruitment & Staffing	-33%	-30%	-27%	-28%	-29%	-37%
	Welfare & Unemployment	38%	41%	39%	43%	44%	44%

So what about privacy?

- Innovation historically has inspired new privacy frameworks
 - Novelty of the camera in the late 19th century
 - Telephony and wiretapping in the early 20th century
 - Databases in the mid-20th century
 - The Internet in the late 20th century
- Where is today's innovation leading us...?
 - Very cheap data processing
 - Crowd-sourcing
 - Social Web
 - Mobile everything
 - **...Information Abundance that can solve modern problems**



How should we think about privacy in this context?

Lenses on privacy

Reputation

Secrecy

Control

Security

Transparency

Lies

Anonymity

Impermanence

How will this affect the way others treat me?

One commonly discussed philosophy of privacy

- Data is personal if directly linkable to an individual
- People should be able to access and correct personal data
- Personal data should only be collected / retained / used for a clearly communicated purpose
- Personal data should only be collected with the individual's consent

How do we fix the world?

- Engineer solutions that enable users to protect their privacy without negating the possibility of useful aggregate analysis
- Rigorously discuss the more difficult questions
 - Should people have a right to obscure the linkage between themselves and data?
 - Should we enable lying?
 - ...under which circumstances?

Engineering challenge #1

How to support access and control to personal data that isn't authenticated?

Engineering challenge #2

How to inform users about both authenticated and unauthenticated personal data?

Engineering challenge #3

How to balance giving users control over data collection versus detecting and stopping abuse?

Engineering challenge #4

How to give users fine-grained control over their data without overwhelming them with options?

Engineering challenge #5

How to link sequential actions while preventing them from being linkable to a person?

Engineering challenge #6

How to make the benefits of aggregate data analysis apparent to the user?

Engineering challenge #7

How to avoid or detect inadvertent recording of data that can be linked to an individual?

Questions?
