



Admin stuff

PLEASE SORT OUT YOUR GROUPS!

Project 1 will be posted 12/9 – DUE 29/9 (Friday)

Project 2 will be posted 3/10 – DUE 3/11 (Friday)

Project 3 will be posted 7/11 – DUE 28/11 (Friday)

Exam due date – 18/12

plan for today

Start with the basics

What is ‘data’?

What is ‘big data’?

Critical questions for big data

Ethics, privacy, who cares?

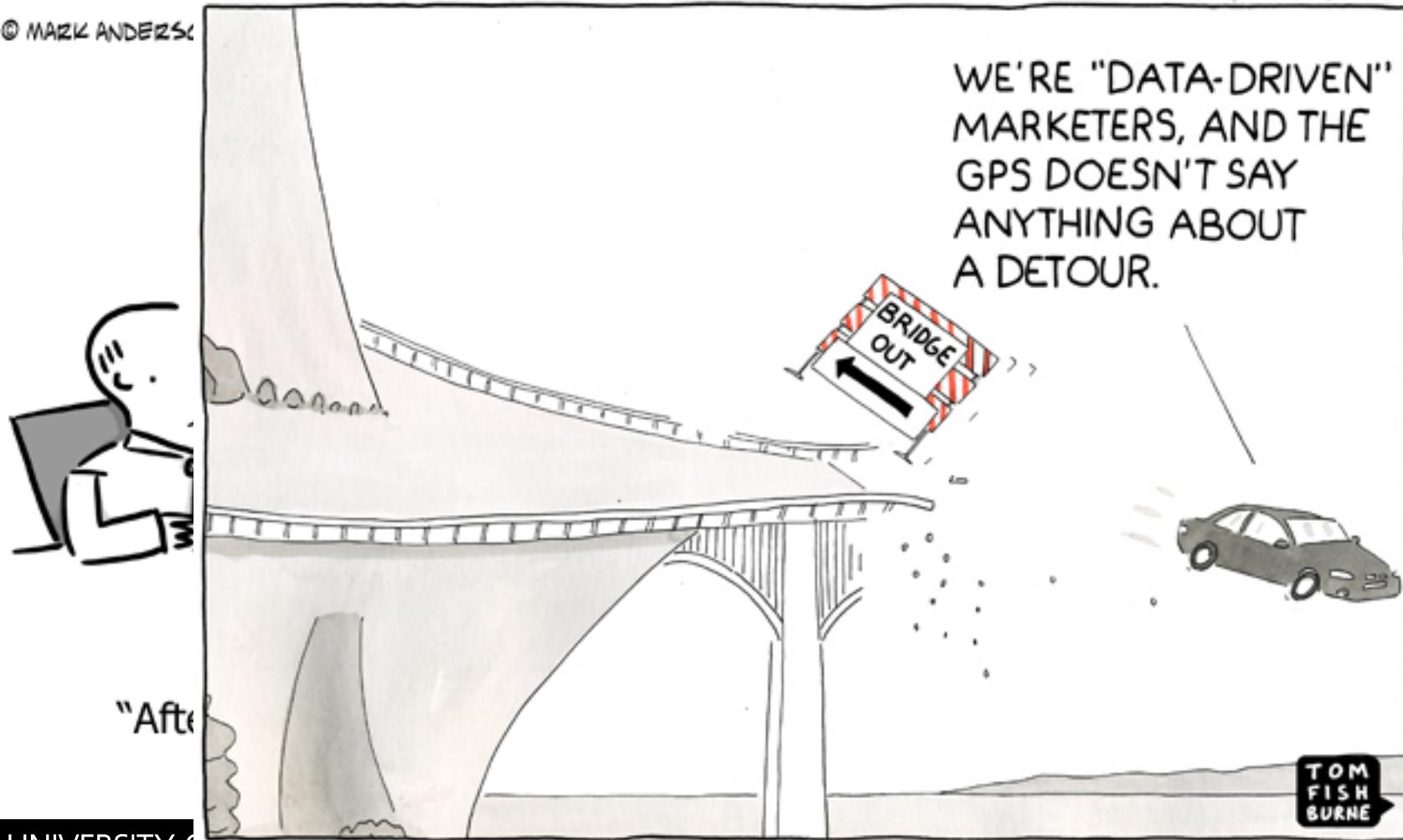
The data walk – looking for data “in the wild”

What is data?

What do YOU think?

What are data?

© MARK ANDERSON



IT UNIVERSITY OF COPENHAGEN

© marketoonist.com

What is data?

How we conceive of data, how we measure them,
what we do with them actively frames the nature of
data

Data arise out of particular contexts, framed and
shaped by the conditions of their production.

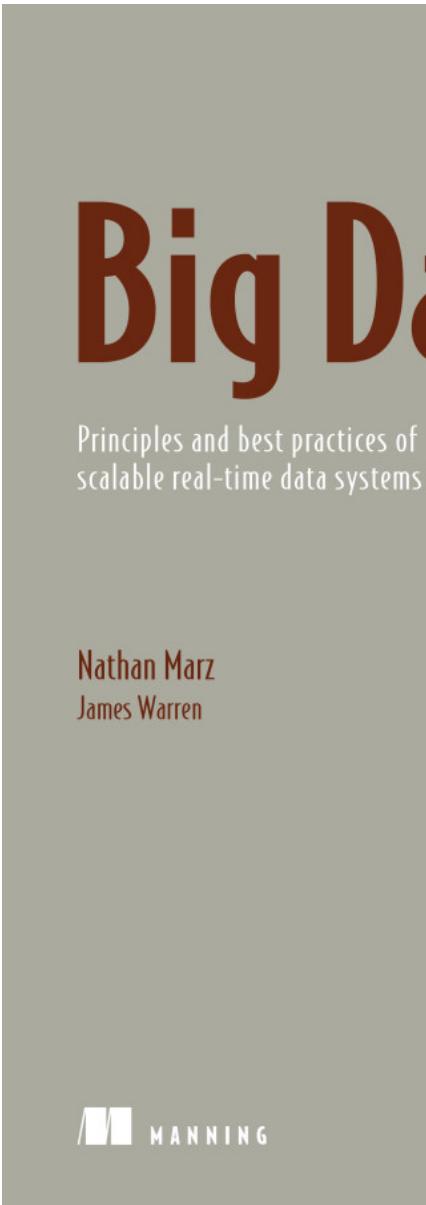
Properties of data

Data is raw

Data is immutable

Data is eternally true

IT UNIVERSITY OF COPENHAGEN

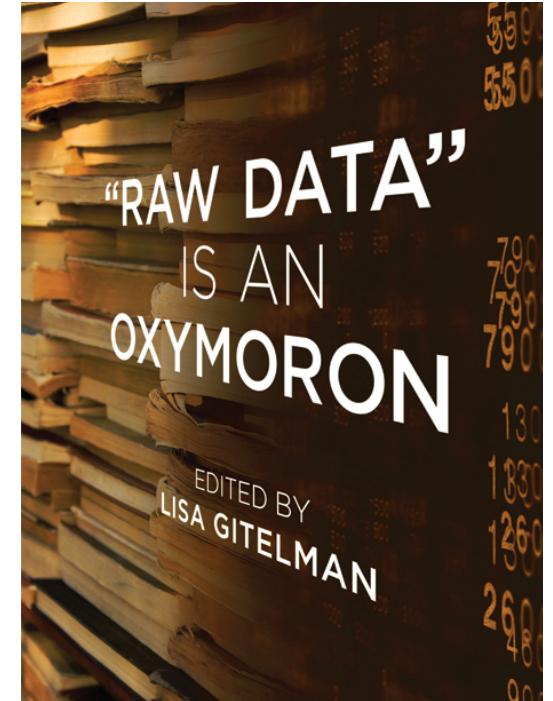


Can data be raw?

Data are constructions about the world

“Raw data is both an oxymoron and a bad idea; To the contrary, data should be cooked with care.”

- Goeff Bowker, 2005



To order is to classify

Categorization is a powerful semantic and political intervention:

What categories are, what belongs in a category, and who decides how to implement these categories in practice, are all powerful assertions about how things are and are supposed to be

Technology

Google search changes tackle fake news and hate speech

⌚ 25 April 2017 | Technology



Notification from Google Play



Google Play Support

To You and support@gab.ai

5:56 PM

...

Hi Developers at Gab AI Inc,

After review, Gab, ai.gab.android, has been suspended and removed from Google Play as a policy strike because it violates the **hate speech** policy.

Next Steps

1. Read through the [Hate Speech](#) policy for more details and examples of policy violations.
2. Make sure your app is compliant with all policies listed in the [Developer Program Policies](#). Remember additional enforcement could occur if there are further policy issues with your apps



Gab @getongab



Breaking news: Google has removed Gab's Android app from the Google Play Store for "hate speech."

12:45 AM - Aug 18, 2017

1,148 5,299 5,826

To store is to remember

We leave traces

Recorded traces may never be recalled again

In recording we frame the present through the act of consigning it to another medium for storage

Much memory practice is in the isolated act of recording

Recording is a reduction of complexity

The dead now have far richer afterlives than ever before

“The stuff of our experience is the aggregated experience of others” – G. Bowker (2005 p.5)

Databasing life – to remember is to classify

Classification is a political act – who gets to classify, why and how? With what outcomes?

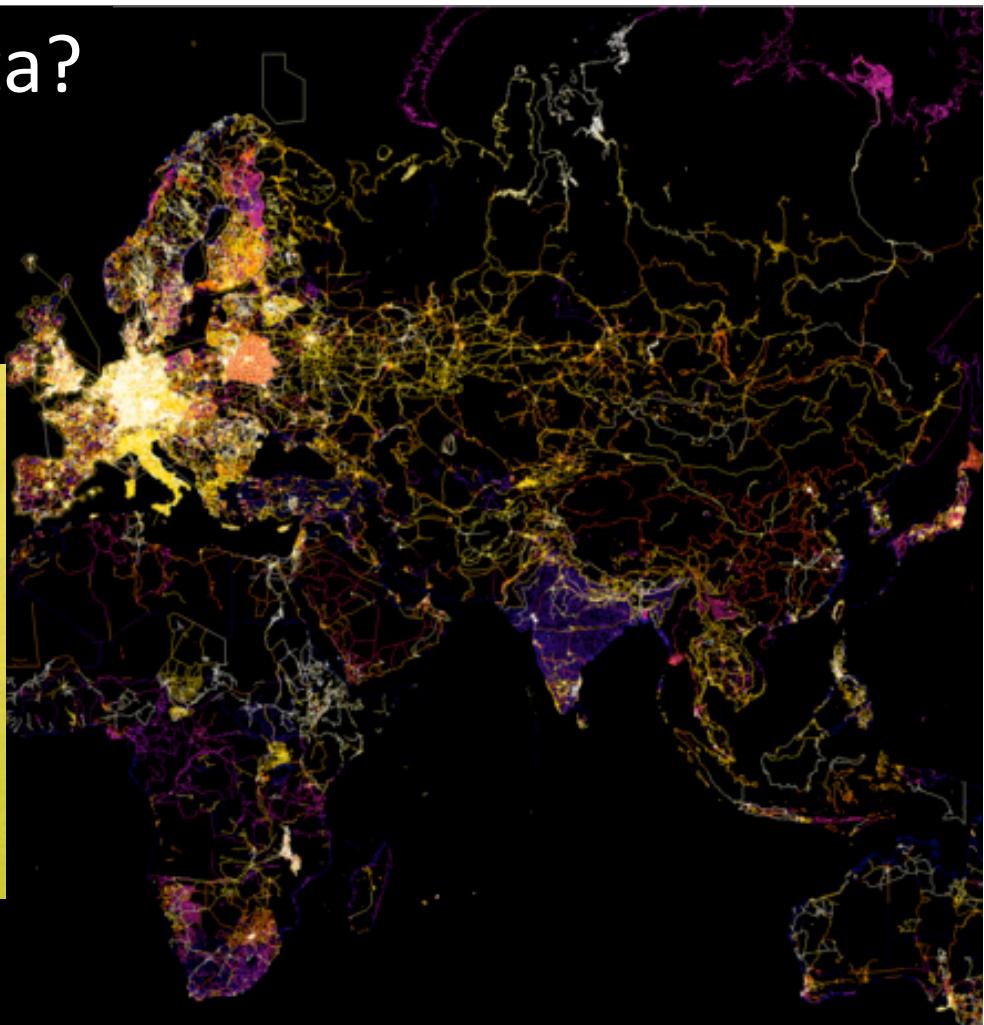
What is not classified is rendered invisible, insignificant, forgotten

“Memory itself gets configured differently in different information infrastructures” G. Bowker (2005, p.225)

Databasing life – to remember is to classify

“As a cultural form, the database represents the world as a list of items and it refuses to order this list. In contrast, a narrative creates a cause-and-effect trajectory of seemingly unordered items (events). Therefore, database and narrative are natural enemies. Competing for the same territory of human culture, each claims an exclusive right to make meaning out of the world.” (L. Manovich c.f. Bowker, p. 29)

What's so “big” about data?



Does size matter?

**“IT’S NOT THE SIZE OF THE DOG IN THE FIGHT,
IT’S THE SIZE OF THE FIGHT IN THE DOG.”**

MARK TWAIN

© Lifehack Quotes

LIFEHACK.QUOTE

What is “big data”?



What is “big data”?

U NEED BIG DATA?

BIG DATA IS JUST A FAD



**IT'S JUST
1960's BATCH
PROCESSING,
BUT IN COLOR**

WAIT, I GET IT 4 U

What is “big data”?

“We define Big Data as a cultural, technological, and scholarly phenomenon that rests on the interplay of:

- (1) **Technology**: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets.
- (2) **Analysis**: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.
- (3) **Mythology**: the widespread belief that **large data sets offer a higher form of intelligence and knowledge** that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.”

(boyd & Crawford 2012)

Six Provocations for Big Data

1. Big data changes the definition of knowledge
2. Claims to objectivity and accuracy are misleading
3. Bigger data are not always better data
4. Taken out of context, big data loses its meaning
5. Just because it is accessible does not make it ethical
6. Limited access to Big Data creates new digital divides

Big Data changes the definition of knowledge

Do numbers speak for themselves?

“Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.”

(Anderson 2008)

Correlation is not Causation
(and in science correlation is not enough)

Big Data changes the definition of knowledge



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."



© marketoonist.com

Claims to Objectivity & Accuracy are Misleading

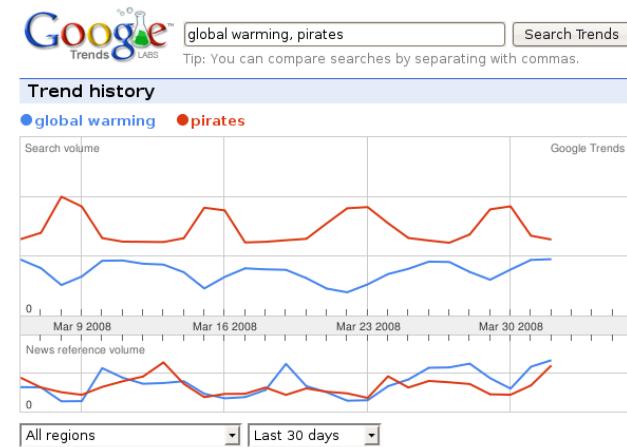
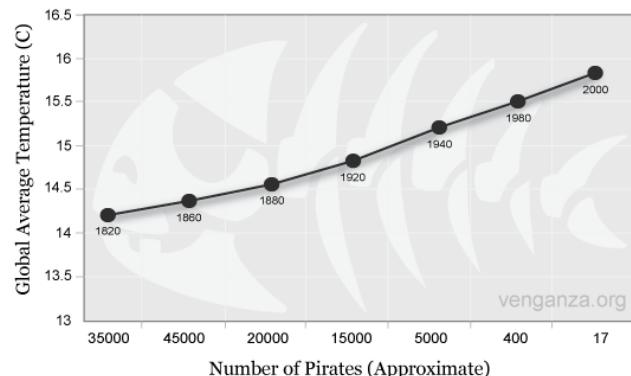
Large Datasets make social spaces quantifiable...

but information is not self explanatory.

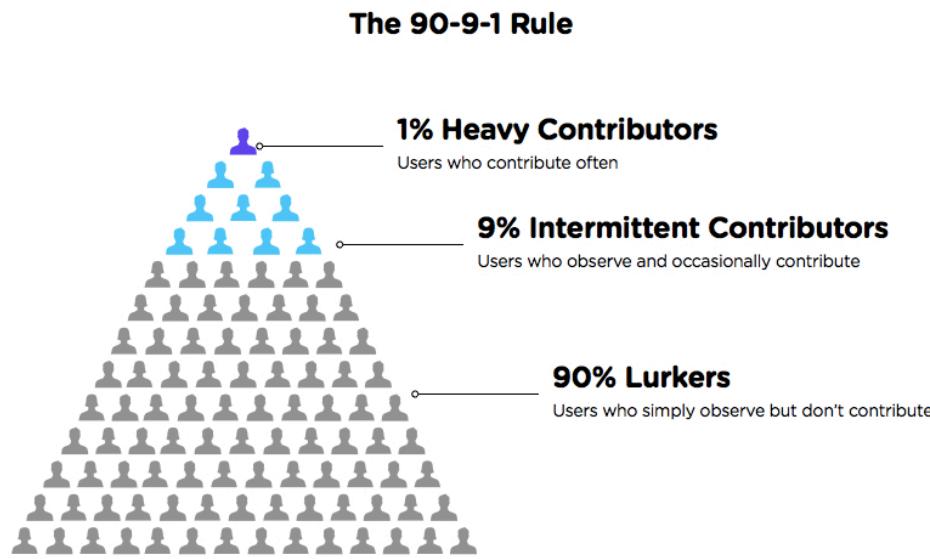
Theories, research design and interpretation are needed.

Otherwise Pirates will cause the global warming.

Global Average Temperature Vs. Number of Pirates



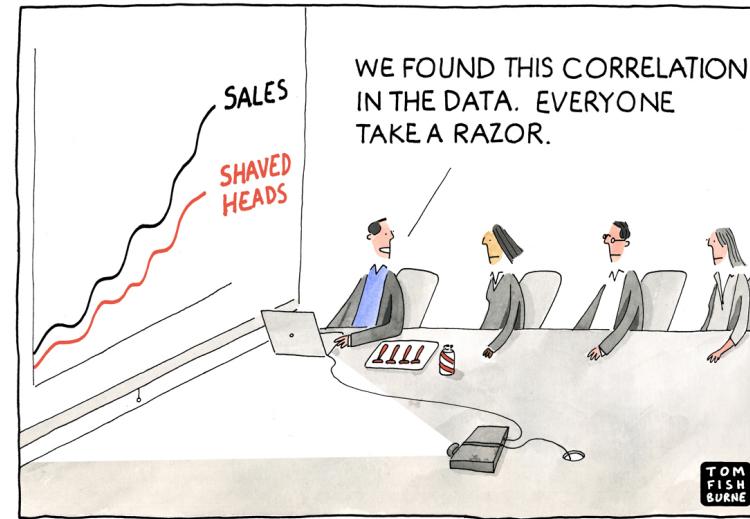
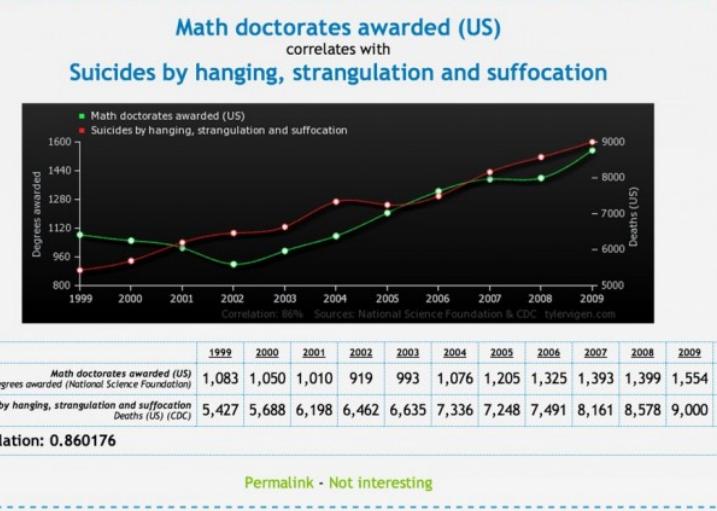
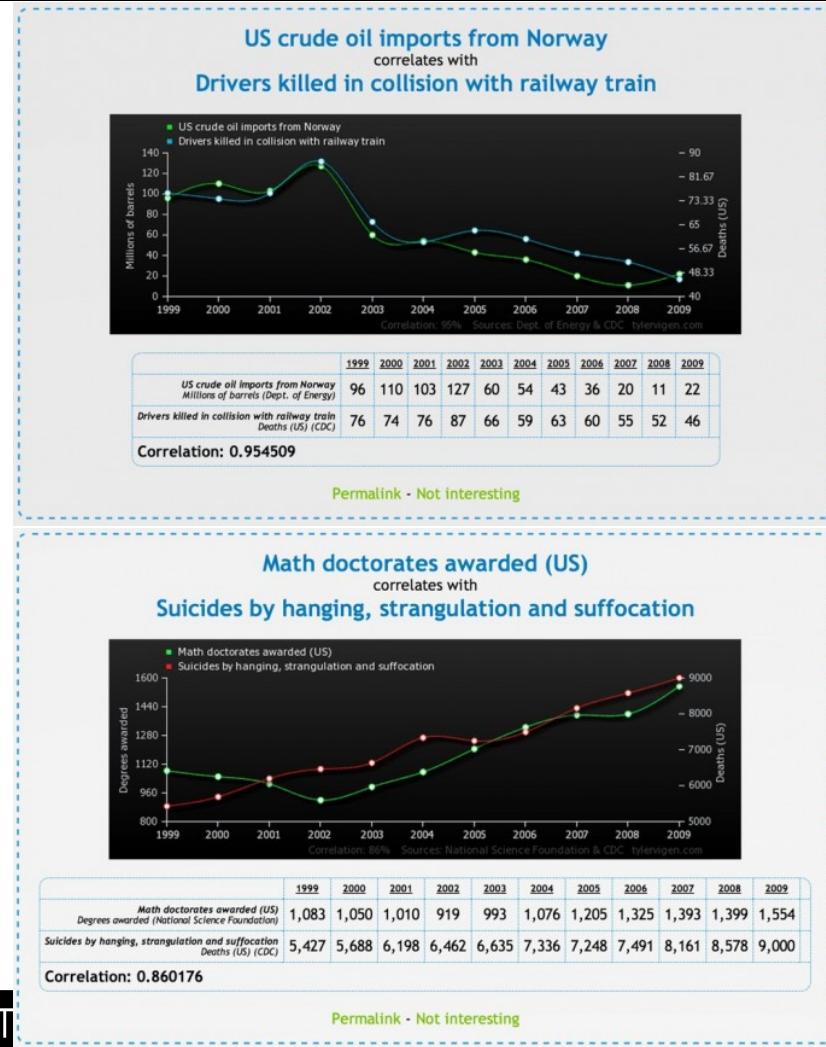
Bigger Data are not Always Better Data



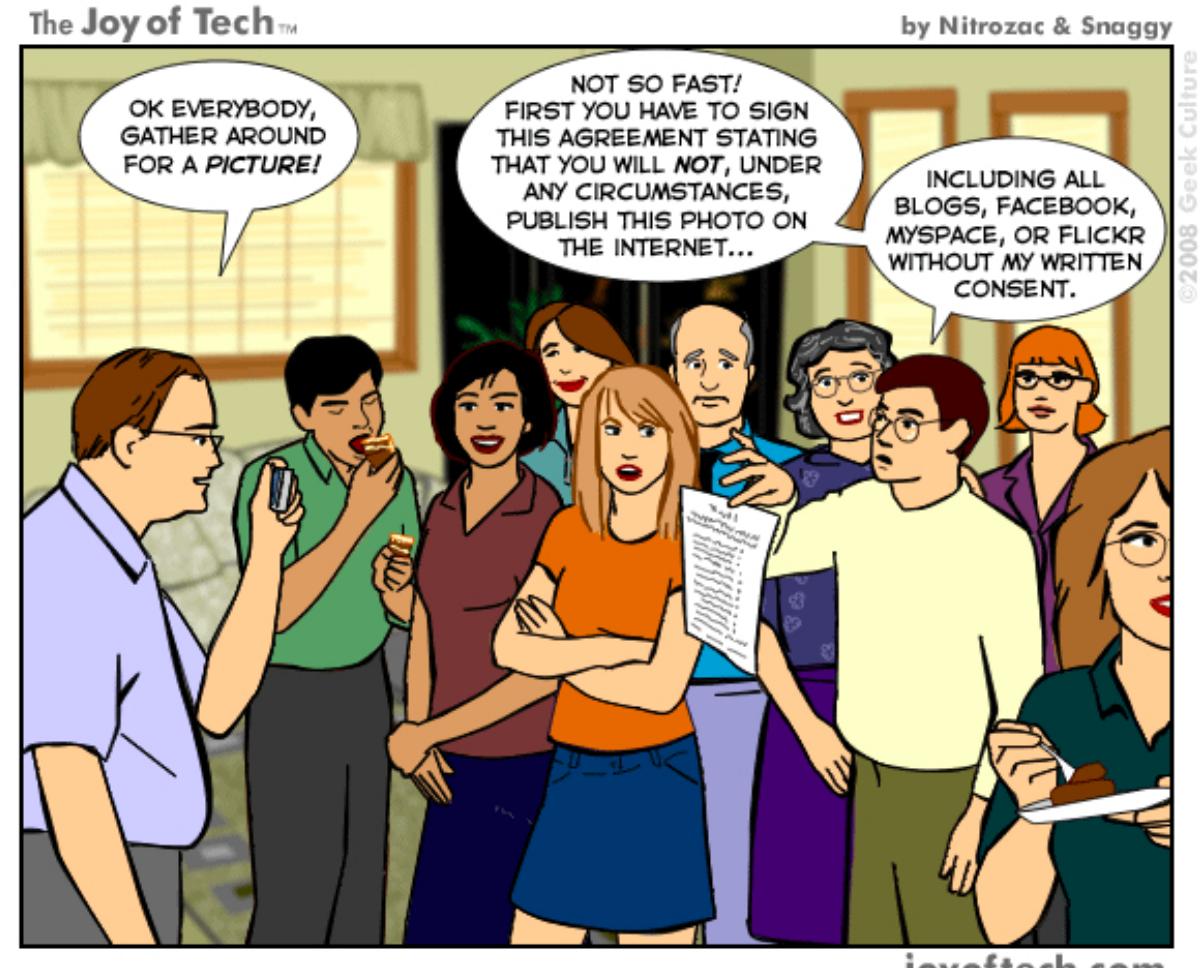
The
bigger
the better?

Large Social media Data have been used to examine a variety of patterns but “users” and “people” are not synonymous.

Taken out of Context Data Loses its Meaning



Just Because its Accessible Doesn't Make it Ethical



Just Because its Accessible Doesn't Make it Ethical

What is ETHICS?

Golden Rule

Hippocratic Oath



Norms of conduct that distinguish between acceptable and unacceptable behavior

Traditional ethical assumptions about individuality, free will and power

Just Because its Accessible Doesn't Make it Ethical

Moral Agency

1. **Causality:** an agent can be held responsible if the ethically relevant result is an outcome of its actions
2. **Knowledge:** An agent can be blamed for the result of its actions if it had (or should have had) knowledge of the consequences of its actions
3. **Choice:** An agent can be blamed for the result if it had the liberty to choose an alternative without greater harm for itself



Just Because its Accessible Doesn't Make it Ethical



Although the data is eternally true, you may prefer to “forget” the information either because you must or because it doesn’t provide enough value for the storage cost.

... or because it the right and ethical thing to do?

Categories of Big Data Stakeholders

Power imbalance between different stakeholders

- Collectors
- Utilizers
- Generators

Limited Access to Big Data Creates New Digital Divides

Who gets access to big data?

- People who work for social media/tech companies
- People who can pay
- Those with significant technical skills

Those without access can not reproduce or evaluate claims made by those with access

To Summarise

Data arise out of particular contexts, framed and shaped by the conditions of their production

Many definitions of ‘big data’ depending on goals and imagined futures

Ethics is an important human consideration when dealing with big data

BREAK!



IT UNIVERSITY OF COPENHAGEN