Deloitte.

UCSB

Actuarial Science as Data Science

A vision for actuarial science in the 21st century

California Actuarial Student Congress
UC – Santa Barbara
April 4, 2015

James Guszcza, PhD, FCAS, FSA
Deloitte Consulting | US
jguszcza@deloitte.com

James Guszcza – US Chief Data Scientist, Deloitte Consulting

Deloitte.

Deloitte Consulting LLP
350 South Grand Avenue
Los Angeles, CA 90071

James Guszcza, PhD, FCAS, FSA
Chief Data Scientist
Deloitte Consulting | US

Tel: +1 310 883 4042

jguszcza@deloitte.com

Member of
Deloitte Touche Tohmatsu

James Guszcza is the Chief Data Scientist of Deloitte Consulting in the United States, as well as a member of Deloitte's Advanced Analytics and Modeling practice.

Jim has applied statistical and machine learning methods to such diverse business problems as healthcare utilization, customer and employee retention, talent management, insurance agent recruiting, customer segmentation, insurance pricing and underwriting, credit scoring, child support enforcement, medical malpractice and patient safety, claims management, and fraud detection. He has also explored the use of behavioral nudge tactics to more effectively act on model indications.

A frequent author and conference speaker, Jim designed and teaches hands-on business analytics training seminars for both the Casualty Actuarial Society and the Society of Actuaries.

Jim is a former professor at the University of Wisconsin-Madison business school, and he holds a PhD in the Philosophy of Science from The University of Chicago. Jim is a Fellow of both the Casualty Actuarial Society and Society of Actuaries.

Agenda

Actuarial science and data science

Why analytics is everywhere

A small note about big data

A new mindset for data science

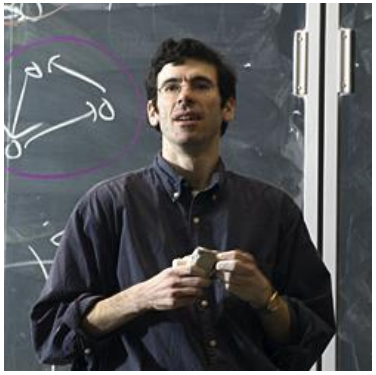
Data science and actuarial science

“The potential to transform everything”

“The term itself is vague, but it is getting at something that is real…

Big Data is a tagline for ***a process that has the potential to transform everything.”***

— Jon Kleinberg, Cornell University



nature International weekly journal of science

Computational social science: Making the links

From e-mails to social networks, the digital traces left by life in the modern world are transforming social science.

Glamorous models

(No, I'm not making this up)

At the center of it all: data science

Or: "The Collision between Statistics and Computation"

An intuitive definition of "data science"...

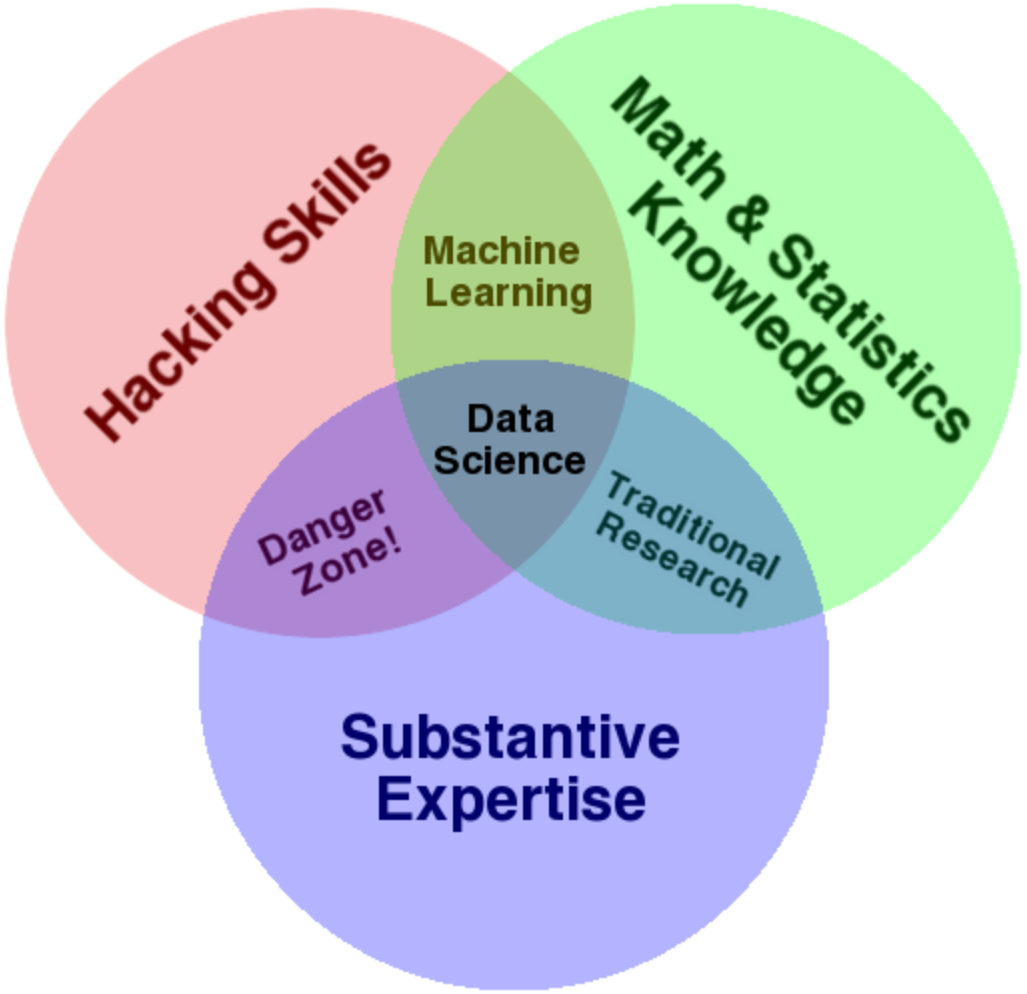


Image borrowed from Drew Conway's blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

At the center of it all: data science

Or: "The Collision between Statistics and Computation"

Is the actuarial profession here?

(Should it be?)

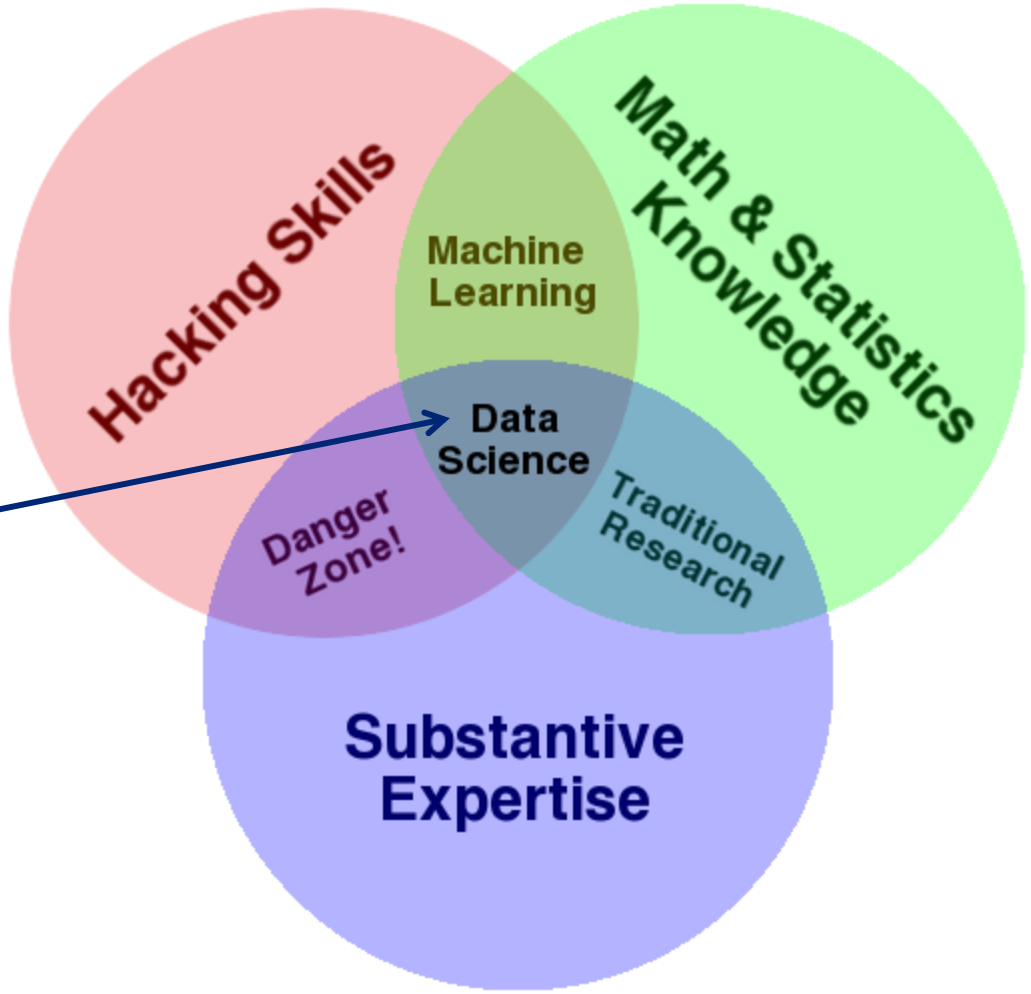


Image borrowed from Drew Conway's blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

At the center of it all: data science

Or: "The Collision between Statistics and Computation"

Or are we here?

(Is that OK?)

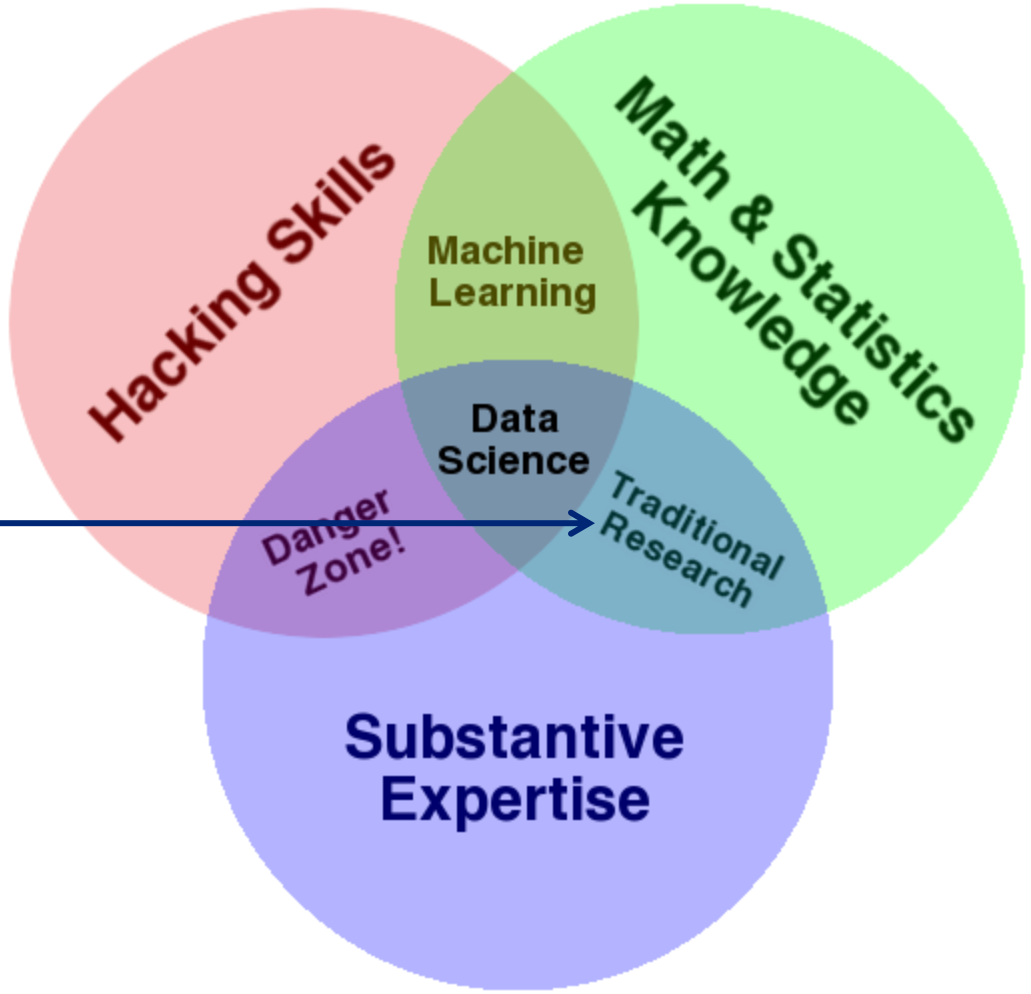


Image borrowed from Drew Conway's blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

From *Fast Company*

(December, 2013)

THE HEALTH CARE INDUSTRY: BIG DATA IS GREAT, BUT WE CAN'T FIND ENOUGH TALENT

EIGHTY-FOUR PERCENT SAY THEY HAVE TROUBLE FINDING STAFF WHO CAN CRUNCH BIG NUMBERS AND GET ACTIONABLE RESULTS.

BY ANYA KAMENETZ

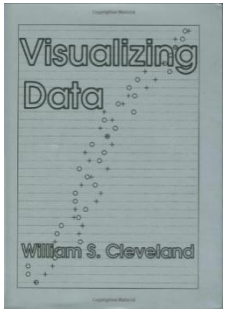
A new survey out from the Society of Actuaries confirms that leaders in the health care industry anticipate big benefits from big data. Eighty-seven percent agreed that big data will impact the business in the future, and 66% say they are "excited" about the future potential.

But while the industry may be excited about big data's prospects, many say the era of big data isn't quite here yet. The reason: a lack of talent.

Eighty-four percent of those surveyed said they'd had at least some difficulty finding staff with the technical qualifications to handle large datasets. Forty-five percent said they'd be looking to hire folks with those skills in 2014. Health care is just one crucial example of the need for more data scientists--even the White House recently announced a $37 million university partnership to steer more young people into these fields.

Examples of big data's potential in health care include companies like Dell, which is partnering with children's hospitals to help determine the optimal personalized cancer drug cocktail for each patient--a feat that requires processing 30 terabytes of data, for which the time has now been driven down from months to days. And the Durkheim Project, an attempt to apply predictive analytics to social media to lower the tragically high suicide rate among veterans.

The origin of “data science”



Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics

William S. Cleveland
Statistics Research, Bell Labs
wsc@bell-labs.com

Abstract

An action plan to enlarge the technical areas of statistics focuses on the data analyst. The plan sets out six technical areas of work for a university department and advocates a specific allocation of resources devoted to research in each area and to courses in each area. The value of technical work is judged by the extent to which it benefits the data analyst, either directly or indirectly. The plan is also applicable to government research labs and corporate research organizations.

Just one letter, young man

R is an open-source, object-oriented statistical programming language. In the past decade, it has become the global lingua franca of statistics.

History:

- Original language (S) developed by **John Chambers** at Bell labs
- R is an open-source implementation of the S language
- Developed by **Robert Gentlemen** and **Ross Ihaka** at U Auckland

Just one letter, young man

R is an open-source, object-oriented statistical programming language. In the past decade, it has become the global lingua franca of statistics.

History:

- Original language (S) developed by **John Chambers** at Bell labs
- R is an open-source implementation of the S language
- Developed by **Robert Gentlemen** and **Ross Ihaka** at U Auckland

"The great beauty of R is that you can modify it to do all sorts of things," said Hal Varian, chief economist at Google. "And you have a lot of prepackaged stuff that's already available, so you're standing on the shoulders of giants."

The culture of data science

"The best thing about being a statistician is that you get to play in everyone's back yard."

-- John Tukey
Princeton/Bell Labs

The culture of data science

"The best thing about being a statistician is that you get to play in everyone's back yard."

-- John Tukey
Princeton/Bell Labs

"The dominant trait among data scientists is an intense curiosity... This often entails the associative thinking that characterizes the most creative scientists in any field."

-- D.J. Patil

Precision medicine

Precision medicine

A Memo to the American People from U.S. Chief Data Scientist Dr. DJ Patil:

Ed. note: This is cross-posted on Medium.

DJ Patil
February 20, 2015
12:01 PM EDT

Memorandum: Unleashing the Power of Data to Serve the American People
To: The American People
From: Dr. DJ Patil, Deputy U.S. CTO for Data Policy and Chief Data Scientist
Date: February 20, 2015

Precision medicine. Medical and genomic data provides an incredible opportunity to transition from a "one-size-fits-all" approach to health care towards a truly personalized system, one that takes into account individual differences in people's genes, environments, and lifestyles in order to optimally prevent and treat disease. We will work through collaborative public and private efforts carried out under the President's new Precision Medicine Initiative to catalyze a new era of responsible and secure data-based health care.

Why analytics is everywhere

The technological answer ("Moore, Moore, Moore")

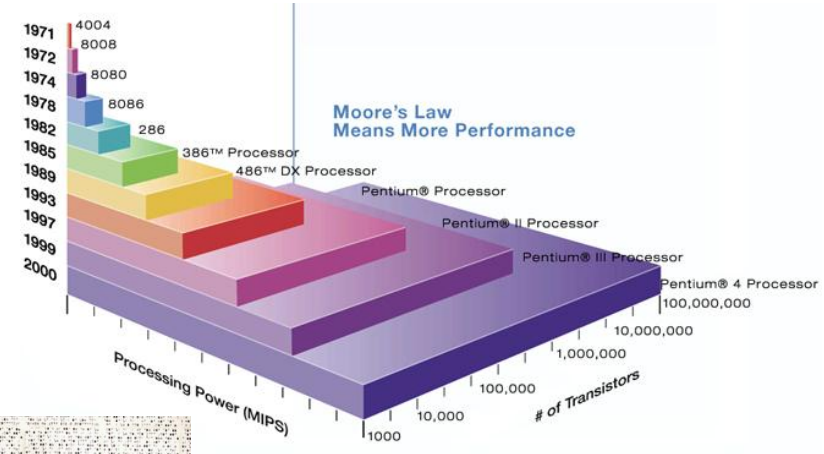
Technology (Moore's Law)

- Cost of storage and computing power has decreased exponentially

The technological answer ("Moore, Moore, Moore")

Technology (Moore's Law)

- Cost of storage and computing power has decreased exponentially



Data

- It's everywhere
- Mobile devices, the internet of things, cloud computing, …

The technological answer ("Moore, Moore, Moore")

Technology (Moore's Law)

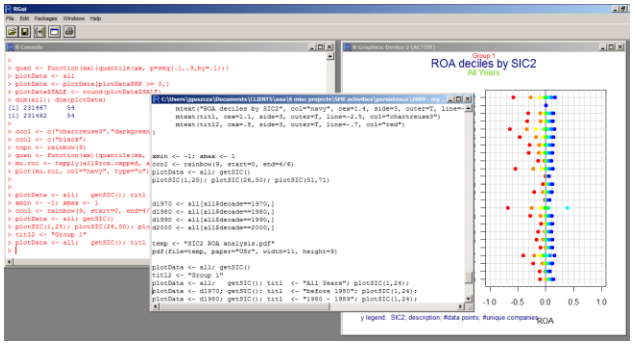
- Cost of storage and computing power has decreased exponentially

Data

- It's everywhere
- Mobile devices, the internet of things, cloud computing, …

Software and algorithms

- Great analytic ideas keep coming from statistics, economics, machine learning, marketing, …
- Free tools like R, Python



The business answer

Strength in Numbers:

How Does Data-Driven Decisionmaking Affect Firm Performance?

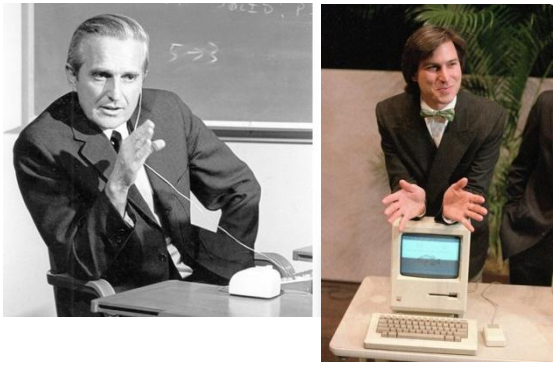
Erik Brynjolfsson, MIT and NBER
Lorin Hitt, University of Pennsylvania
Heekyung Kim, MIT

Abstract

We examine whether performance is higher in firms that emphasize decisionmaking based on data and business analytics (which we term a data-driven decisionmaking approach or DDD). Using detailed survey data on the business practices and information technology investments of 179 large publicly traded firms, we find that firms that adopt DDD have output and productivity that is 5-6% higher than what would be expected given their other investments and information technology usage. Using instrumental variables methods, we find evidence that these effects do not appear to be due to reverse causality. Furthermore, the relationship between DDD and performance also appears in other performance measures such as asset utilization, return on equity and market value. Our results provide some of the first large scale data on the direct connection between data-driven decisionmaking and firm performance.

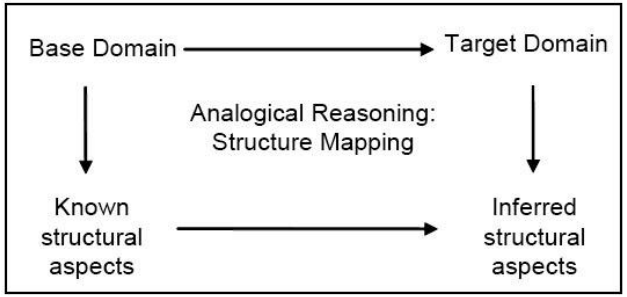
And also… a conduit for innovation

- It's often about **incremental improvements**, not rare, transformative, breakthrough ideas.



- **Strategic and Design thinking** is important.

- It often happens when you **think by analogy**.
  - Borrow and recombine ideas from other domains
  - Analytics is the ultimate transferrable skill
  - Contemplate analytics applications from outside your domain



“Clinical versus Actuarial Judgment: the Motion Picture”

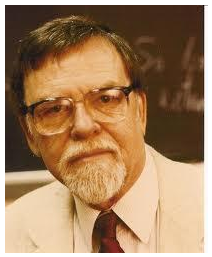
Science 31 March 1989:
Vol. 243 no. 4899 pp. 1668-1674
DOI: 10.1126/science.2648573

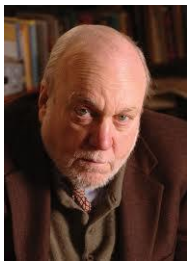
Clinical versus actuarial judgment

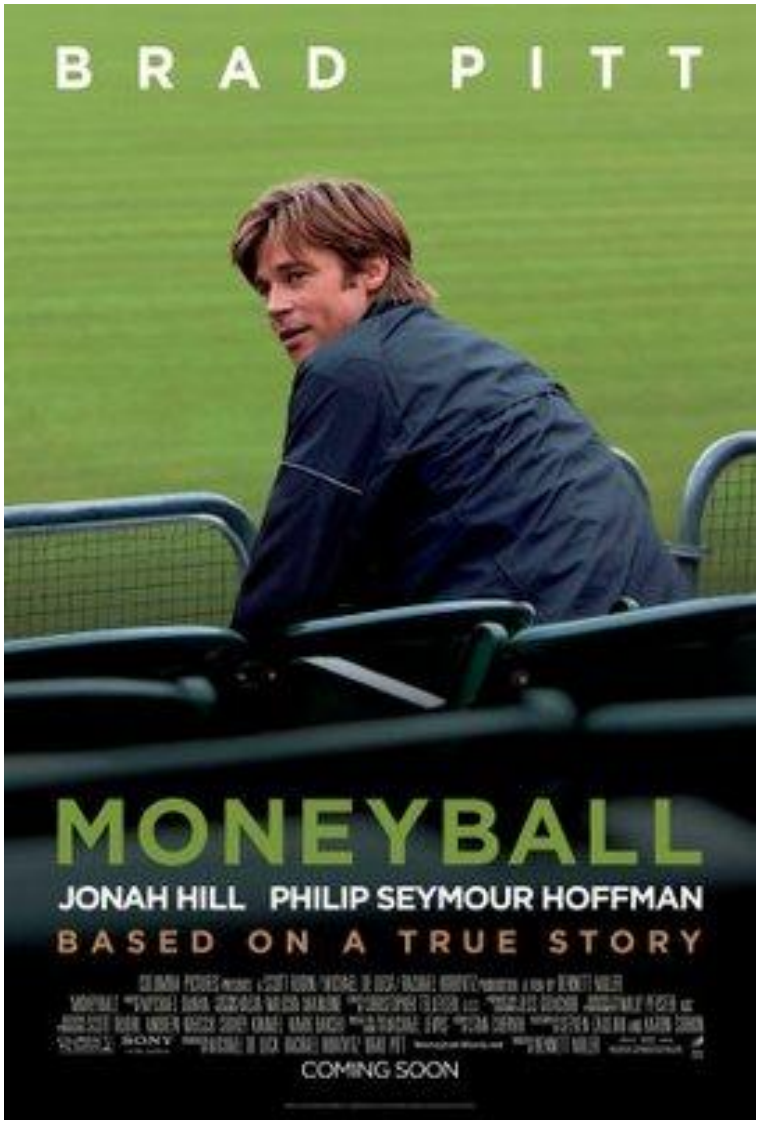
RM Dawes, D Faust and PE Meehl

ABSTRACT

Professionals are frequently consulted to diagnose and predict human behavior; optimal treatment and planning often hinge on the consultant's judgmental accuracy. The consultant may rely on one of two contrasting approaches to decision-making--the clinical and actuarial methods. Research comparing these two approaches shows the actuarial method to be superior. Factors underlying the greater accuracy of actuarial methods, sources of resistance to the scientific findings, and the benefits of increased reliance on actuarial approaches are discussed.







The city of New York does actuarial science

Big Data in the Big Apple

How New York's first "director of analytics" revolutionized the city's building inspections.

407

By Viktor Schönberger and Kenneth Cukier

A new way to figure out which old buildings are most at risk

Dry Idea

In the news

California Imposes First Mandatory Water Restrictions to Deal With Drought

New York Times - 2 days ago

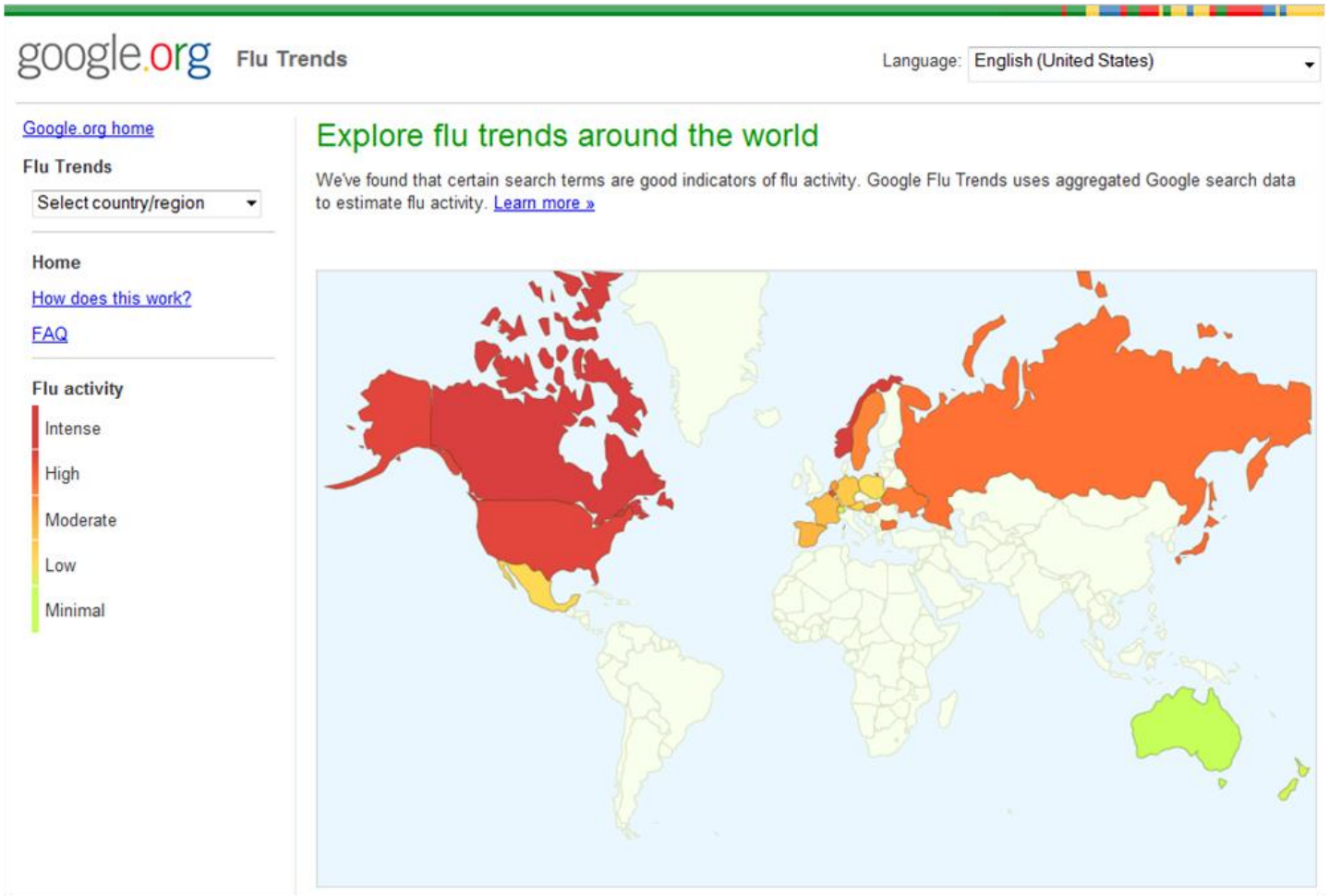
Gov. **Jerry Brown** on Wednesday ordered mandatory **water** use reductions for the first time in ...

Low California snowpack ushers mandatory water restrictions

CNN - 3 days ago

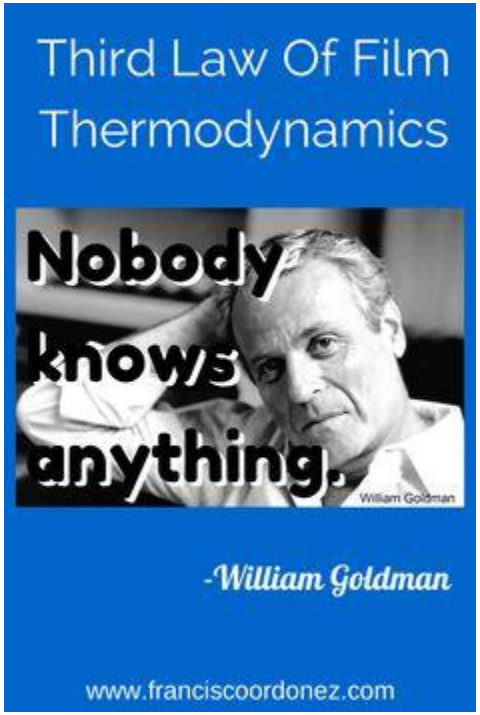
Will California do something similar?

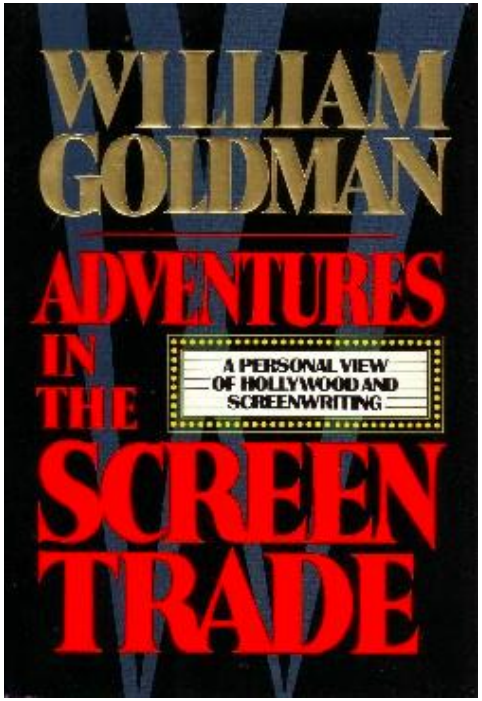
The pulse of the nation



Google data repurposed to track flu hot-spots

In Hollywood, “Nobody Knows Anything”





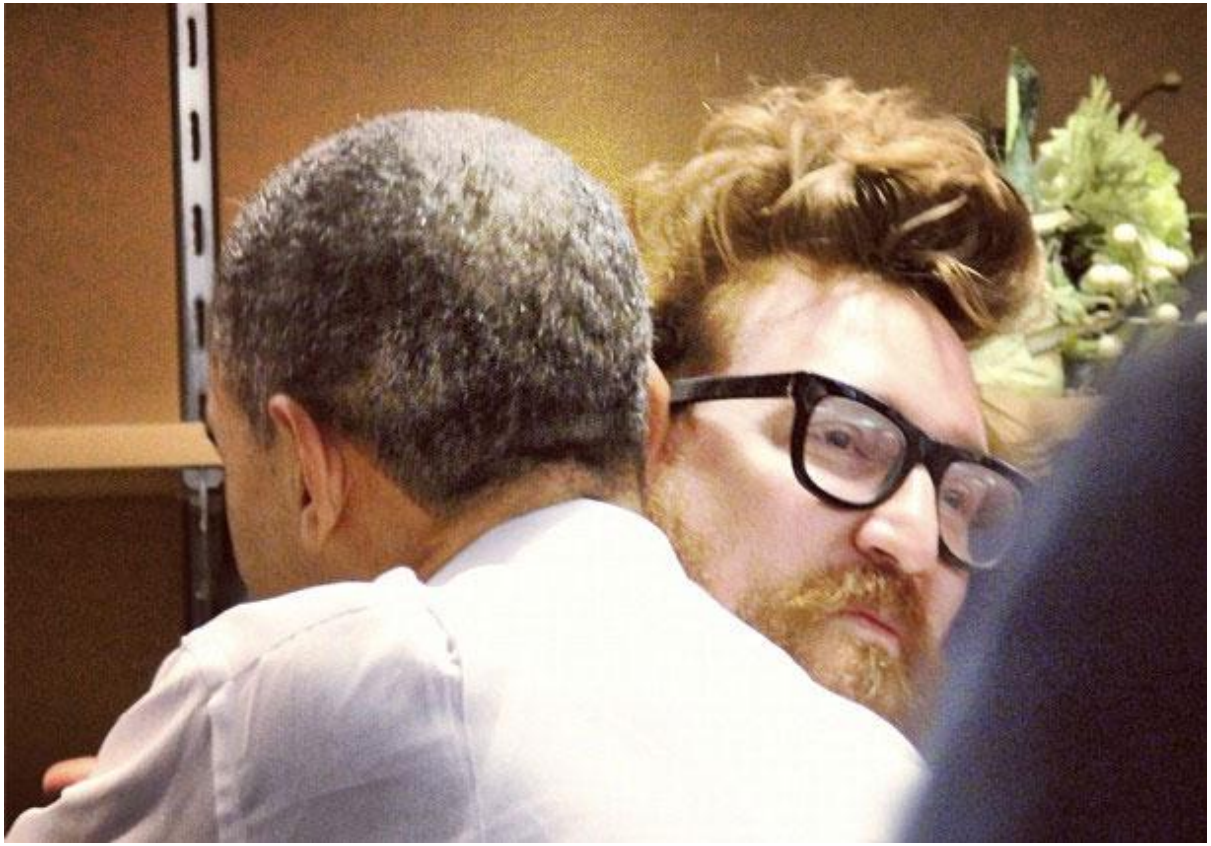
"Nobody Knows Anything"

Perhaps the most famous quotation from the book. It is one of his two "Roman numeral I's" and is repeated throughout the book. Now widely quoted, it is often inaccurately used to suggest that Hollywood executives are stupid, but in fact refers to Goldman's belief that, prior to a movie's release, Hollywood has no real idea how well a film will do.

(except Netflix)

Better viewing through "datafication"

Some celebrated examples of “big data”

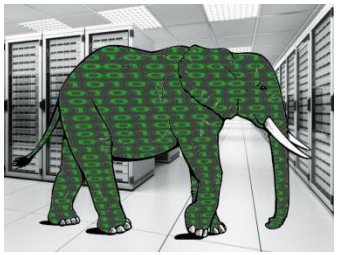


Harper Reed - The CTO of the 2012 Obama re-election campaign
(pictured with unidentified staffer)

A small note about big data

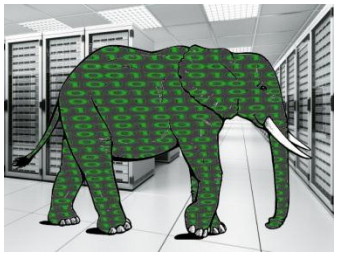
Three definitions of big data

1. Data sets with sizes beyond the capability of standard IT tools to capture, process, and analyze in reasonable time frames.



Three definitions of big data

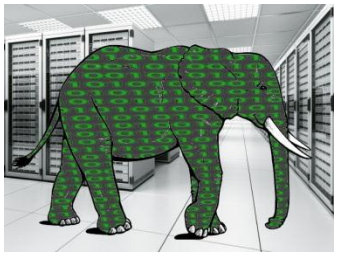
1. Data sets with sizes beyond the capability of standard IT tools to capture, process, and analyze in reasonable time frames.



2. Data with high Volume, Velocity, Variety
   - Huge datasets
   - … emanating continuously from smart phones, sensors, cameras, GPS devices, computers, TVs, …
   - … involving all manner of numeric, text, photographic data

Three definitions of big data

1. Data sets with sizes beyond the capability of standard IT tools to capture, process, and analyze in reasonable time frames.

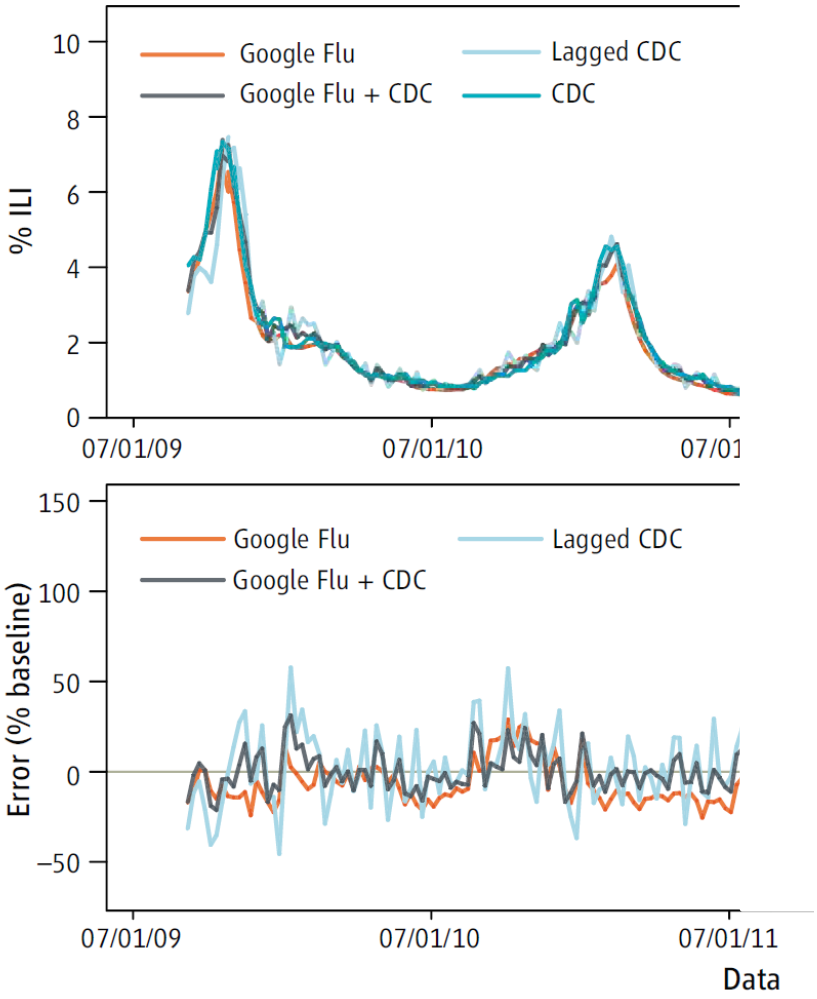


2. Data with high Volume, Velocity, Variety
   - Huge datasets
   - … emanating continuously from smart phones, sensors, cameras, GPS devices, computers, TVs, …
   - … involving all manner of numeric, text, photographic data

3. "Anything that doesn't fit in Excel"

Back to Google Flu Trends:

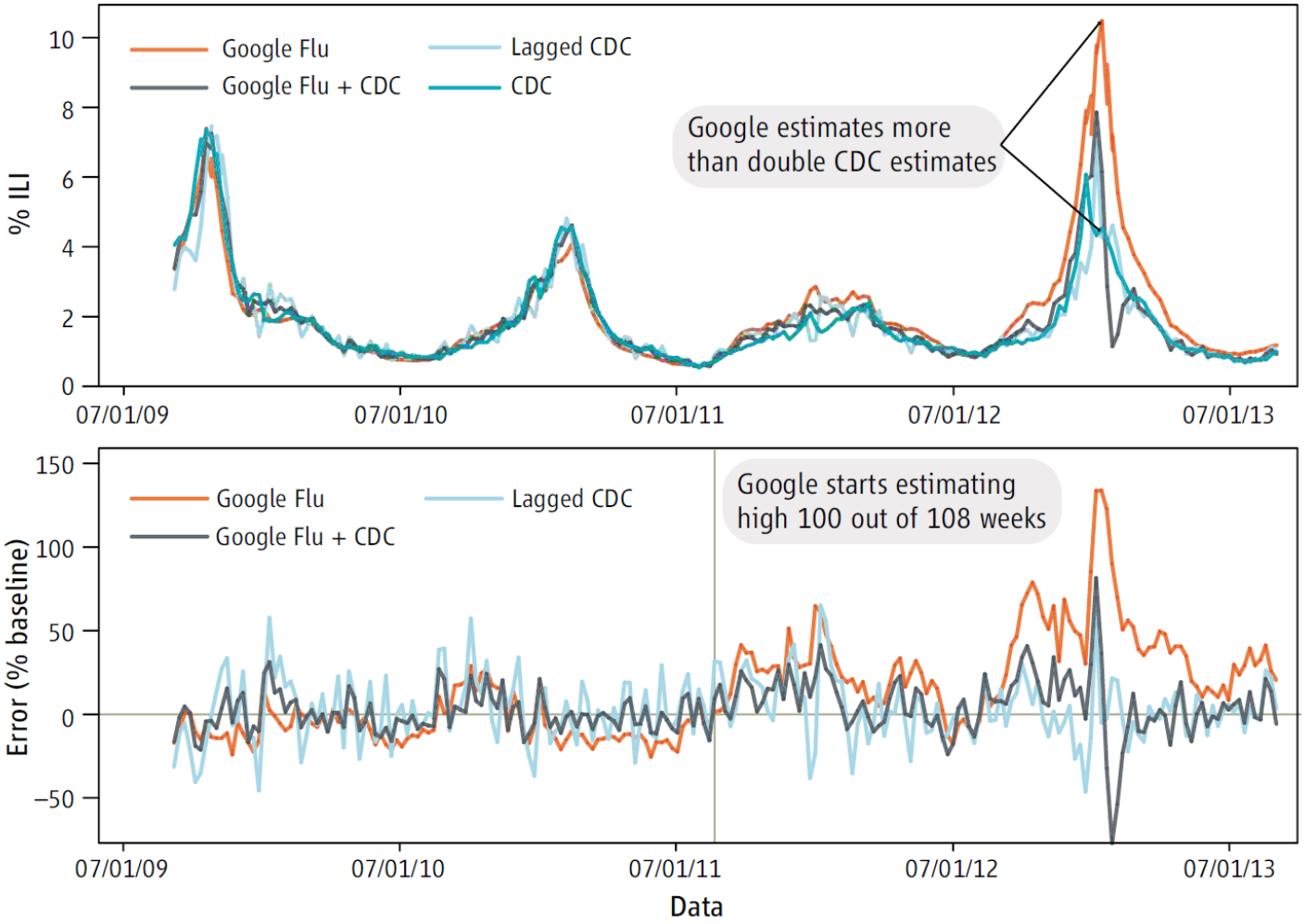
From poster child…



From poster child…to parable

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]



Google Flu Trends and “big data hubris”

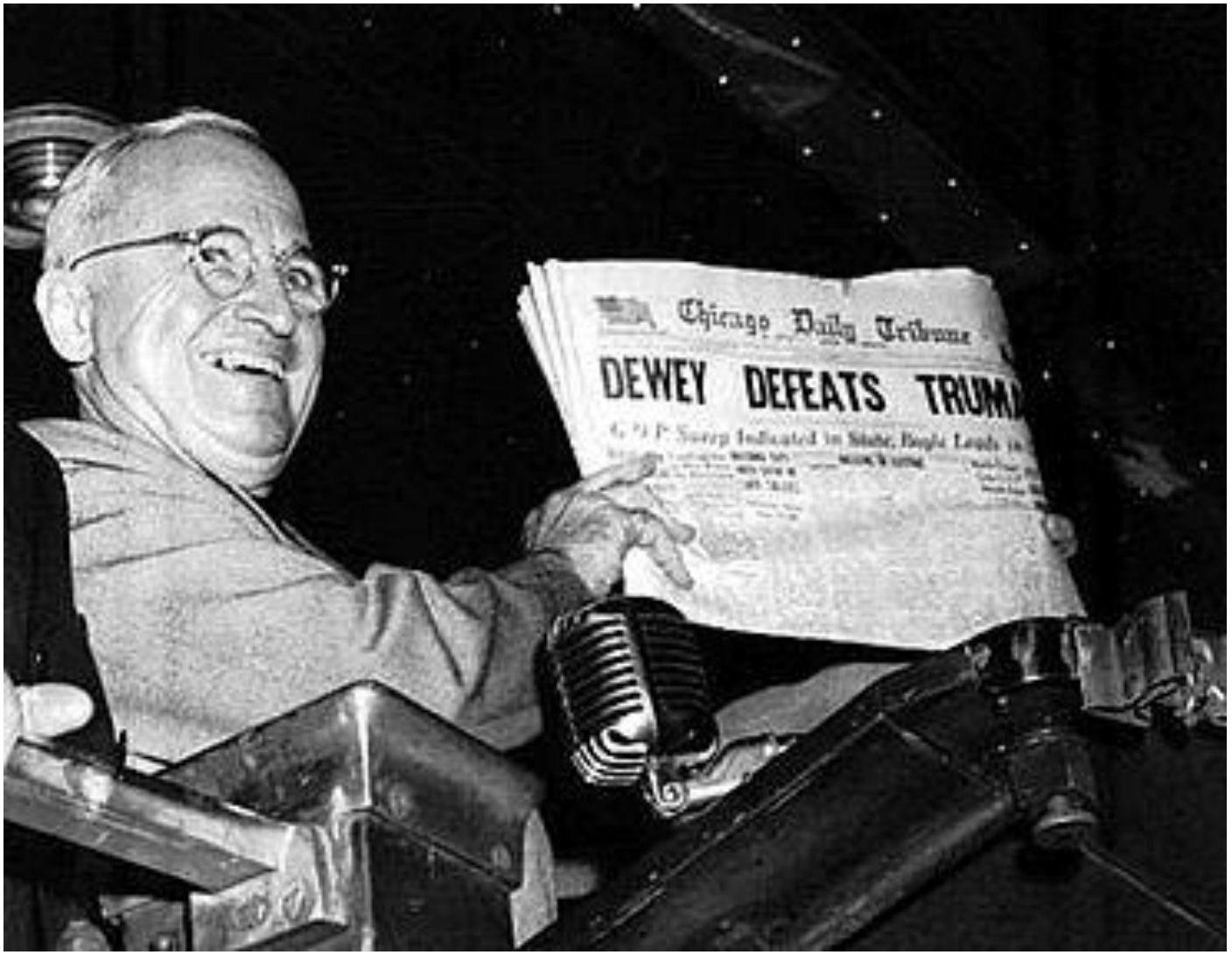
The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]

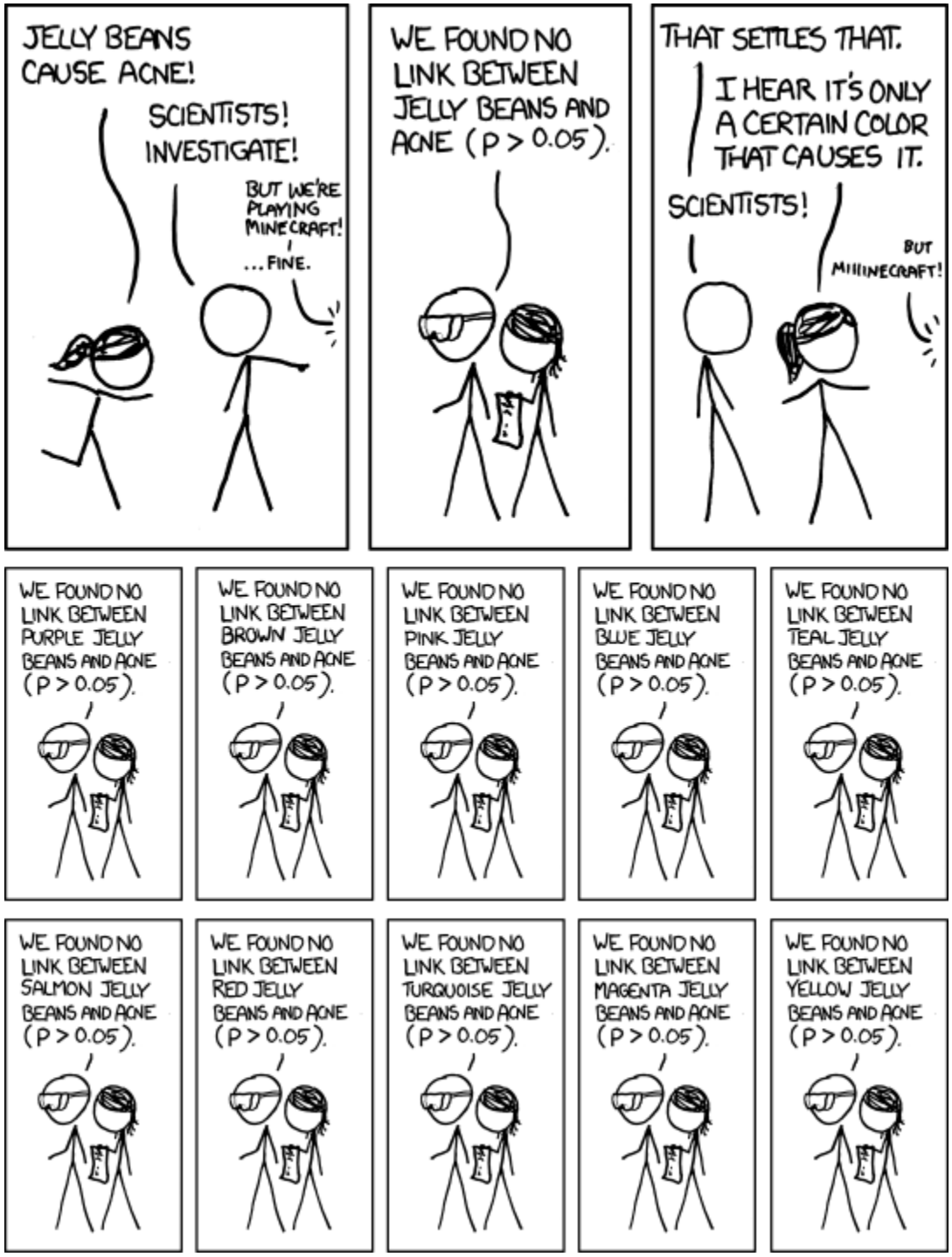
Big Data Hubris

“Big data hubris” is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.

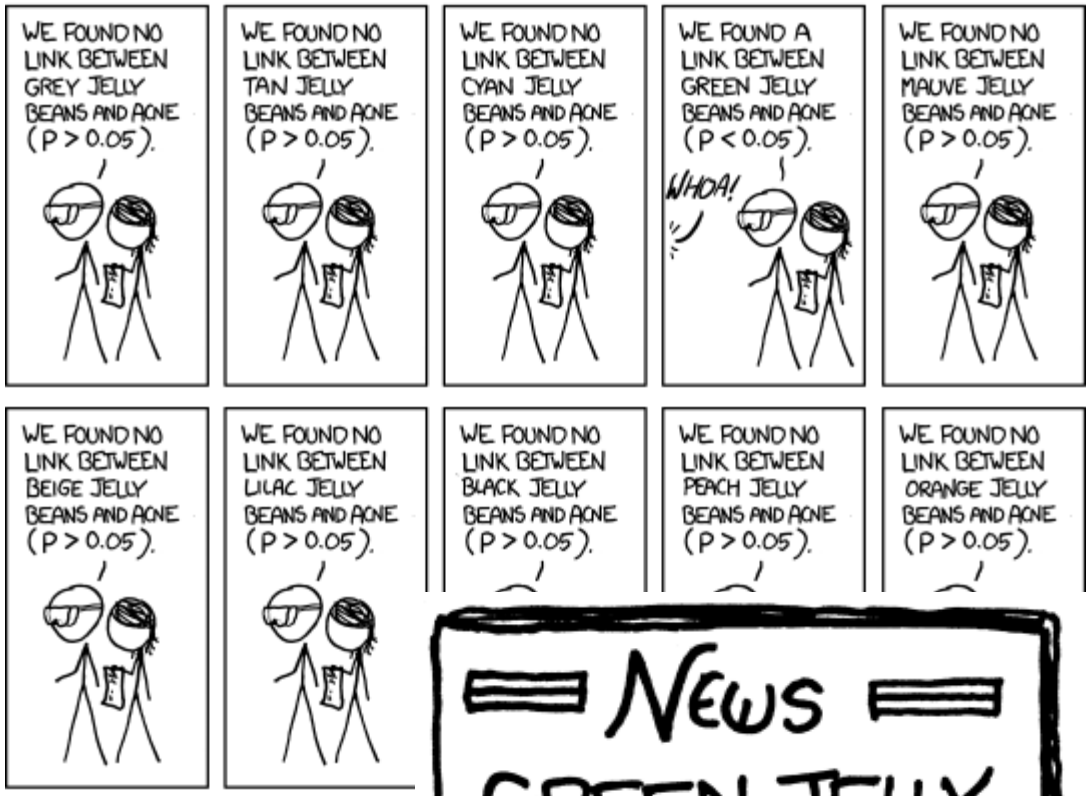
History doesn’t repeat itself but it does rhyme

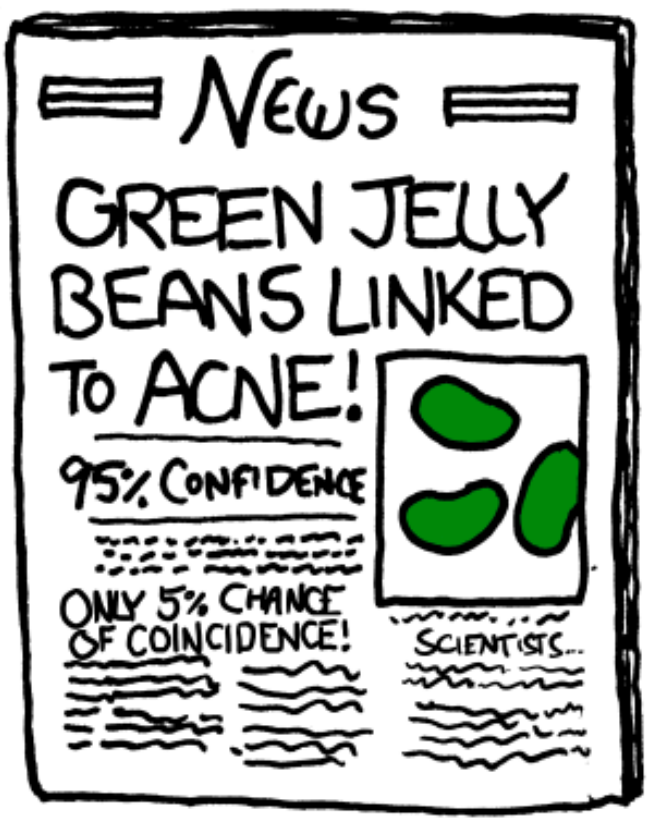


The truth…


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).

… wears off





THE NEW YORKER

ANNALS OF SCIENCE | DECEMBER 13, 2010 ISSUE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

BY JONAH LEHRER

Not just green jelly beans… red shirts too

SCIENCE THE STATE OF THE UNIVERSE

JULY 24 2013 12:37 PM

Too Good to Be True

Statistics may say that women wear red when they're fertile … but you can't always trust statistics.

By Andrew Gelman

Does the red shirt mean she's ovulating? Not so fast …

Photo by Gheradie Ruzu/Thinkstock

Are women three times more likely to wear red or pink when they are most fertile? No, probably not. But here's how hardworking researchers, prestigious scientific journals, and gullible journalists have been fooled into believing so.

MIT Technology Review

Big Data Mining

A Q&A with one of the leading inventors of tools for medical data analytics.

"I am concerned that it's all too easy to see the data and say, 'I've been doing big-data analysis for Target and now I can do it for medicine.' That turns out not to be true. You really need to know something about medicine. If statistics lie, then big data can lie in a very, very big way."

Isaac Kohane, co-director of the Center for Biomedical Informatics at Harvard Medical Schoo

Necessary, but not sufficient

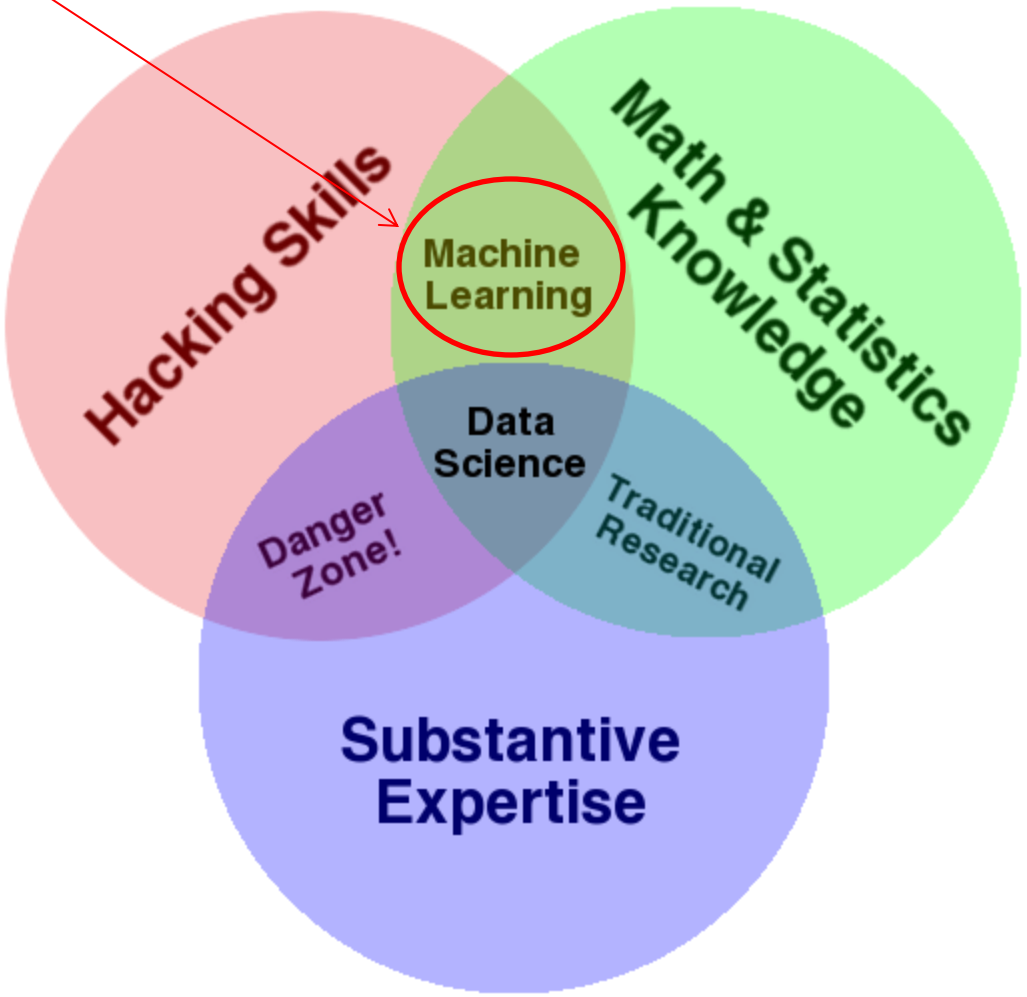
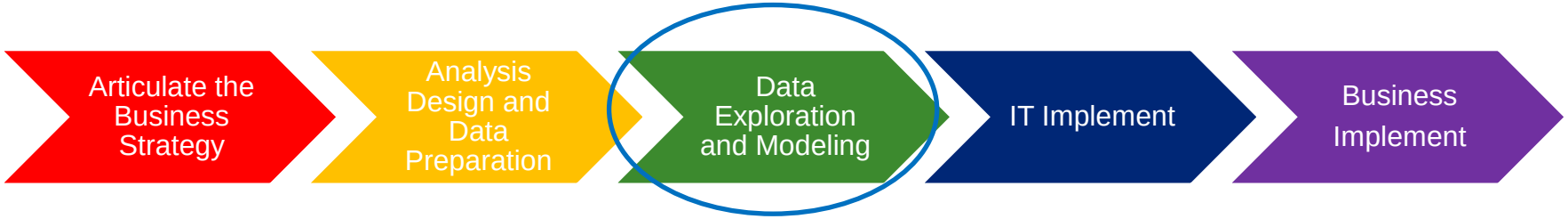


Image borrowed from Drew Conway's blog
http://www.dataists.com/2010/09/the-data-science-venn-diagram

Big data in context



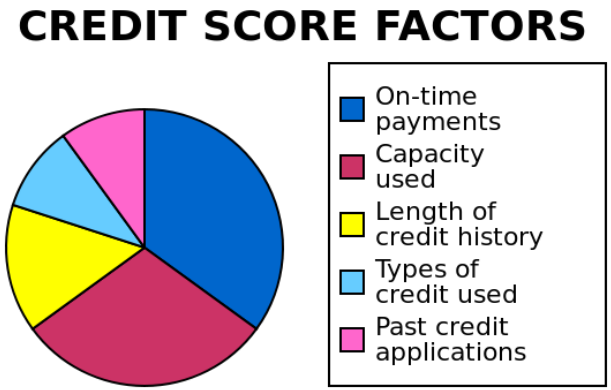
- In real-world analytics projects, the bulk of the time is *not* spent on analytical techniques / methodology.

- Most of the effort is spent on activities like:
  - Researching, ordering, loading and auditing raw data
  - Creating data features
  - Documentation
  - Meetings / communications
  - Project management
  - Quality control
  - Technical implementation
  - Business implementation

Modeling accounts for ≈10-20% of a strategic "modeling" project

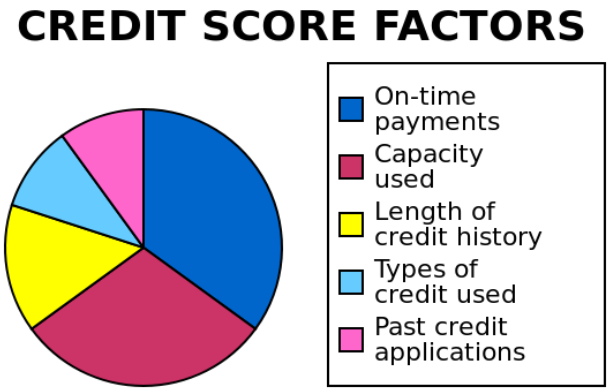
Big data and behavioral data

An early example of business analytics

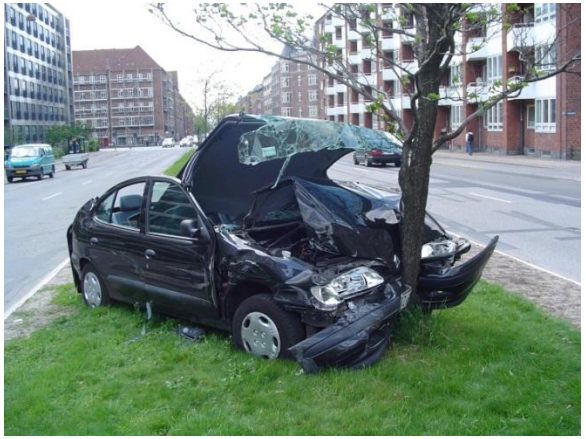


(This we know)

A more striking correlation

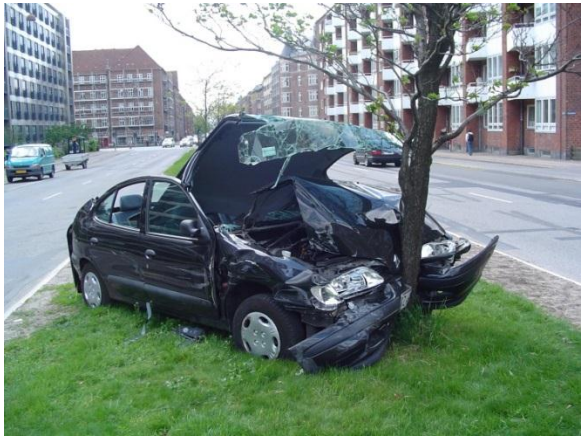


(!)



More food for thought

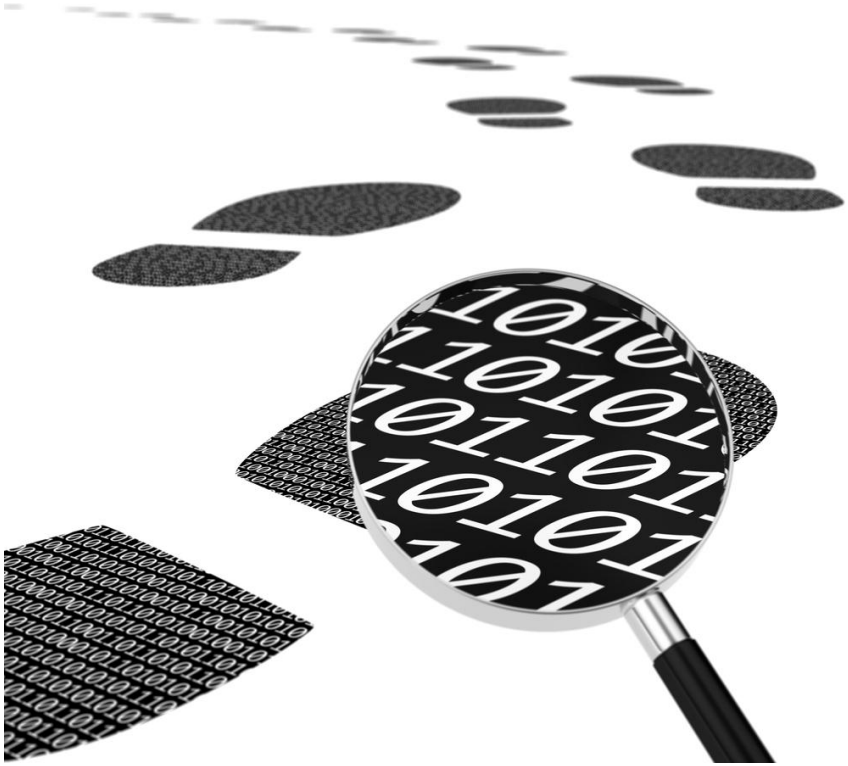
(!!)



Digital footprints

➔ In our daily lives, we increasingly leave behind digital traces of:

- How we drive
- What we buy
- What we eat
- What we watch, read
- What / how we opine
- Where we travel
- Whom we know / networks
- How we socialize
- How we surf the web



The resulting data can be a major source of operational improvements and business innovation…

… and societal change…

Like, you know

Researchers at the Cambridge University Psychometrics Centre built predictive models of personal details based purely on social network “Likes” of a sample of 58,000 people.

• Relationship status, substance abuse	65-73% accurate
• Political leanings (democrat vs Republican)	85% accurate
• Religion (Christian vs Muslim)	82% accurate
• Male sexual orientation	88% accurate
• Ethnicity (African-American vs Caucasian)	95% accurate

“Observation of Likes alone was nearly is roughly as informative as using an individual’s actual personality test score.”

“Similar predictions could be made from all manner of digital data, with this kind of secondary ‘inference’ made with remarkable accuracy”

-- “Digital Records Could Expose Intimate Details and Personality Traits of Millions”
University of Cambridge Research News

http://www.cam.ac.uk/research/news/digital-records-could-expose-intimate-details-and-personality-traits-of-millions

Why big data is a big deal

“I believe that the power of Big Data is that it is information about people's behavior instead of information about their beliefs. It's about the behavior of customers, employees, and prospects for your new business. It's not about the things you post [online] … which is what most people think about, and it's not data from internal company processes and RFIDs. This sort of Big Data comes from things like location data off of your cell phone or credit card, it's the little **data breadcrumbs** that you leave behind you as you move around in the world.”

—Sandy Pentland, MIT Media Lab

“Reinventing Society in the Wake of Big Data”

edge.org conversation

And why this matters to society

"Since this data is mostly about people, there are enormous issues about privacy, data ownership, and data control. You can imagine using Big Data to make a world that is incredibly invasive, incredibly 'Big Brother'… **George Orwell was not nearly creative enough when he wrote *1984*.**"

—Sandy Pentland, MIT Media Lab
"Reinventing Society in the Wake of Big Data"
edge.org conversation

A new mindset for data science

Can we do better by doing good?

"The best minds of my generation are thinking about how to make people click ads... that sucks."

-- Jeff Hammerbacher
Cloudera Founder

"For all the damage that misapplied data can do, data used correctly is a powerful ***positive force****."*

-- Cathy O'Neil, mathbabe.org
"On Being a Data Skeptic"

What are companies for?

"There is one and only one social responsibility of business – to use its resources and engage in activities designed to increase its profits… so long as it engages on open and free competition without deception or fraud."

-- **Milton Friedman**

"The Social Responsibility of Business is to Increase its Profits"

"The only valid purpose of a firm is to create a customer."

-- **Peter Drucker**

Management: Tasks, Responsibilities, Practices

What are companies for?

“There needs to be a completely new approach to how we operate as business leaders, one that clearly puts people at the centre of all we do.”

-- **B Team Leadership Statement**

Davos, January 22, 2014

Copies available in the lobby

For the full story, go to:
http://dupress.com/articles/dr14-personalized-and-personal/