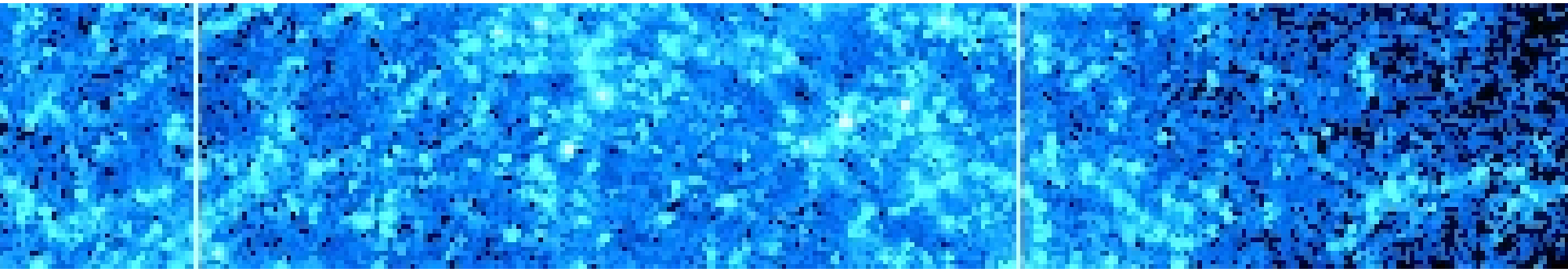
Big Data: the Big Picture

Massive new data resources are changing the way we approach data analysis, inference, and learning.

Massive Computational Challenges



- Computation, algorithms, and software
- Hardware (distributed/parallel processing, GPUs, heterogeneous computing, programmable hardware, etc.)
- Methods that can take advantage of state-of-the-art hardware
- Data management, retrieval, collating, and archiving.

We have focused largely on computational challenges.

This is just the tip of the iceberg.

Big Data and Privacy

POLITICO

Who watches the watchers? Big Data goes unchecked

By **JOSH GERSTEIN** and **STEPHANIE SIMON** | 5/14/14 5:01 AM EDT
Updated: 5/17/14 1:29 PM EDT

The National Security Agency might be tracking your phone calls. But private industry is prying far more deeply into your life.

Commercial data brokers know if you have diabetes. Your electric company can see what time you come home at night. And tracking companies can tell where you go on weekends by snapping photos of your car's license plate and cataloging your movements.

Statistical methods have long been available to "anonymizing" individual data while preserving patterns.

But now goal is often to learn about individuals.

White House May 2014 Report

BIG DATA:
SEIZING OPPORTUNITIES, PRESERVING VALUES

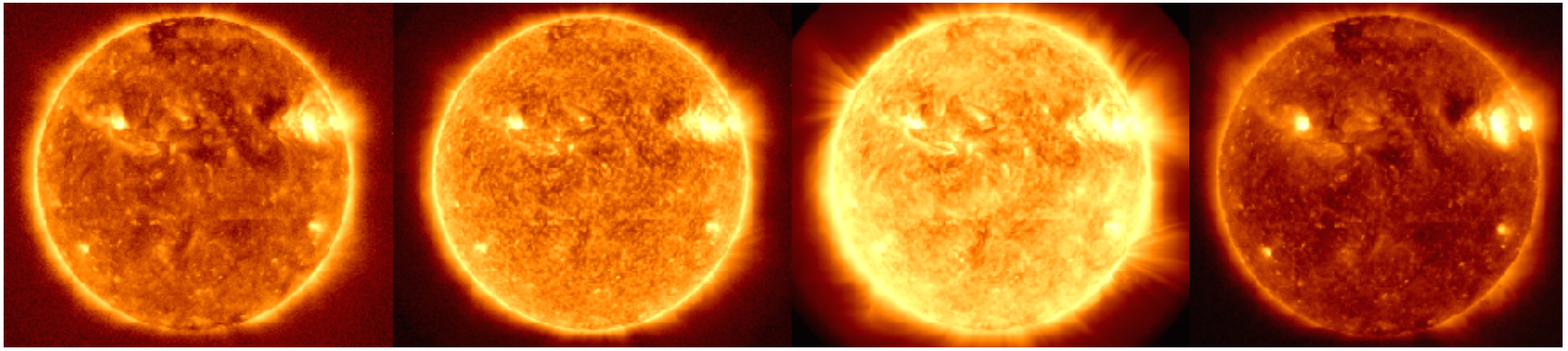
Executive Office of the President

Big Data = Big Brother

Revelations of NSA/GCHQ data gathering = bad publicity!

Will public opinion affect data availability?

Big Data and Scientific Learning



- Data is not just big: it is deep and rich.
  - Enables more interesting / sophisticated statistical models
- Predictive models versus Descriptive models
  - Data-driven versus Science-driven methods
  - Many “learning techniques” focus on prediction.
  - Modeling mechanisms giving rise to data is challenging.
  - Tradeoff: computational speed and statistical principles
  - How best to integrate predictive and descriptive models?
  - How best to scale up methods for descriptive models?

Misconceptions about Big Data

"With Big Data there is no longer a difference between Correlation and Causation."

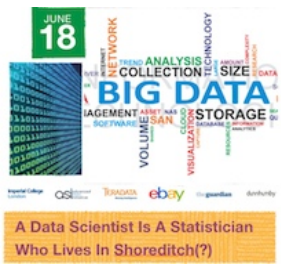
- What about mechanisms? E.g., for scientific arguments.
- What if we want to predict how an outcome will be affected by an intervention?

"With Big Data, there is no need for a random or representative sample."

- What about sampling bias, non-response bias, response bias, voluntary-response bias.
- Big Data is sometimes (often??) cheaper than quality data.
- Comprehensive review by Jon A. Krosnick (Stanford).

Big Data and Inferential Challenges

- Combining multiple siloed data streams.
  - Different data types and quality; unstructured data.
  - Do data sources agree? give conflicting predictions?
- Need methods with predictable mathematical properties.
  - Quantify precision & accuracy of predictions & estimators.
  - Can we quantify the effects of non-representative data?
  - New frameworks: hypothesis testing is silly with big data.
- With a plethora of new methods, how to pick "right one"?
  - Do we believe that there a silver bullet? e-science.



Why Do We Need Data Science when We've Had Statistics for Centuries?

By Irving Wladawsky-Berger *Wednesday, April 30, 2014*

— USC Annenberg Innovation Lab

Big Data: A PR Boon for Data Scientists

Ben Kepes (http://www.forbes.com/sites/benkepes/) Contributor
I cover how technology helps business compete.
Opinions expressed by Forbes Contributors are their own.
FOLLOW

TECH (/TECHNOLOGY) 6/18/2014 @ 7:30AM | 1,453 views

Because Big Data Is Hot -- ThoughtSpot Raises $30M From Khosla

Comment Now Follow Comments

There is one sure fire way to raise a truckload of money from Venture Capitalists – deliver a compelling story that promises something related to big data. And this isn't just hype fuelled funding – the promise that big data solutions hold to deliver actionable insights from the ever increasing amount of data organizations are awash in is an attractive funding proposition. The latest company to leverage

Safari Po
Click to S
Pac

Data Scientist:
The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

70 Harvard Business Review October 2012

Not long ago statisticians were just "bean counters" and there were no machine learners, data miners, or data scientists.

Data Scientists: Top Career Choice

SmartData Collective
The World's Best Thinkers on Data

Home Post Here Analytics Big Data Business Intel Data Management IT Software

Tell Your Kids to be Data Scientists, Not Doctors

Like it? 15

Posted June 11, 2014 0 comments

Keywords: Analytics, Data Mining, Big Data

48 188 104 g+1 8 340

Posted by:

Linda Burtch

Posts Profile
Comments Follow

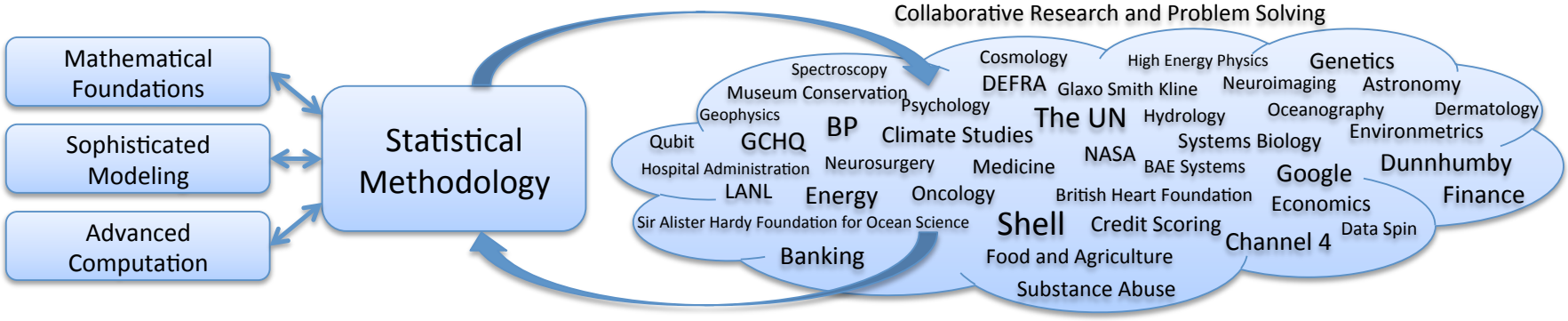
Recently I had the pleasure of being interviewed by John Phillips at CNBC about our data scientist salary study. His article, Why Your Kids Will Want to be Data Scientists, was published at the end of May, and in it he raised a very interesting point:

*"According to Burtch Works' 2014 study of salaries for data scientists… those responsible for a team of 1-3 earn [a median salary of] $140,000 and those responsible for a team of 10 or more earn **$232,500**.*

Can we attract top young talent to data science?

Big Data: A Quickly Evolving Research Envronment

Massive new data streams are opening up a world of opportunities for data scientists!



Real opportunities to contribute to science, buisness, industry, government, R&D, etc. etc.