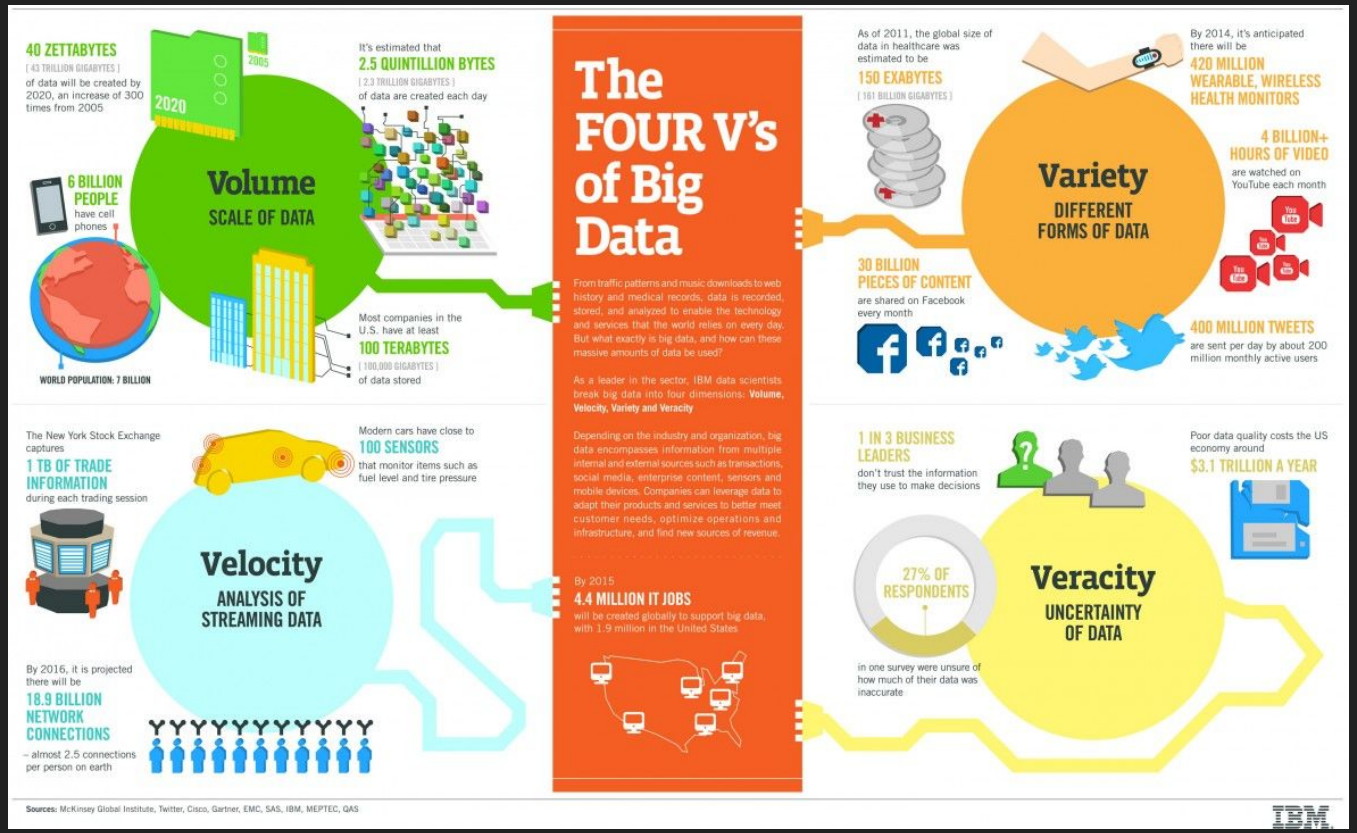
Towards Participatory Democracy in the Big Data Age

Raazesh Sainudiin, Dept. of Mathematics, Uppsala Univ.

Feb 26, 2018

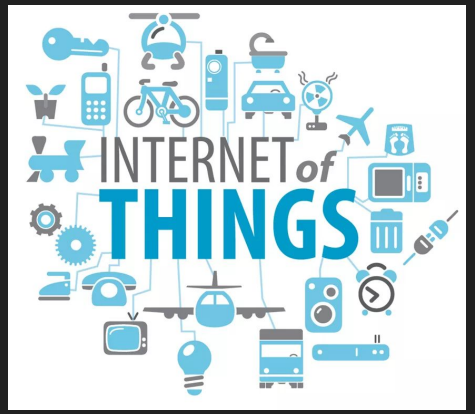
What is the Big Data Age?

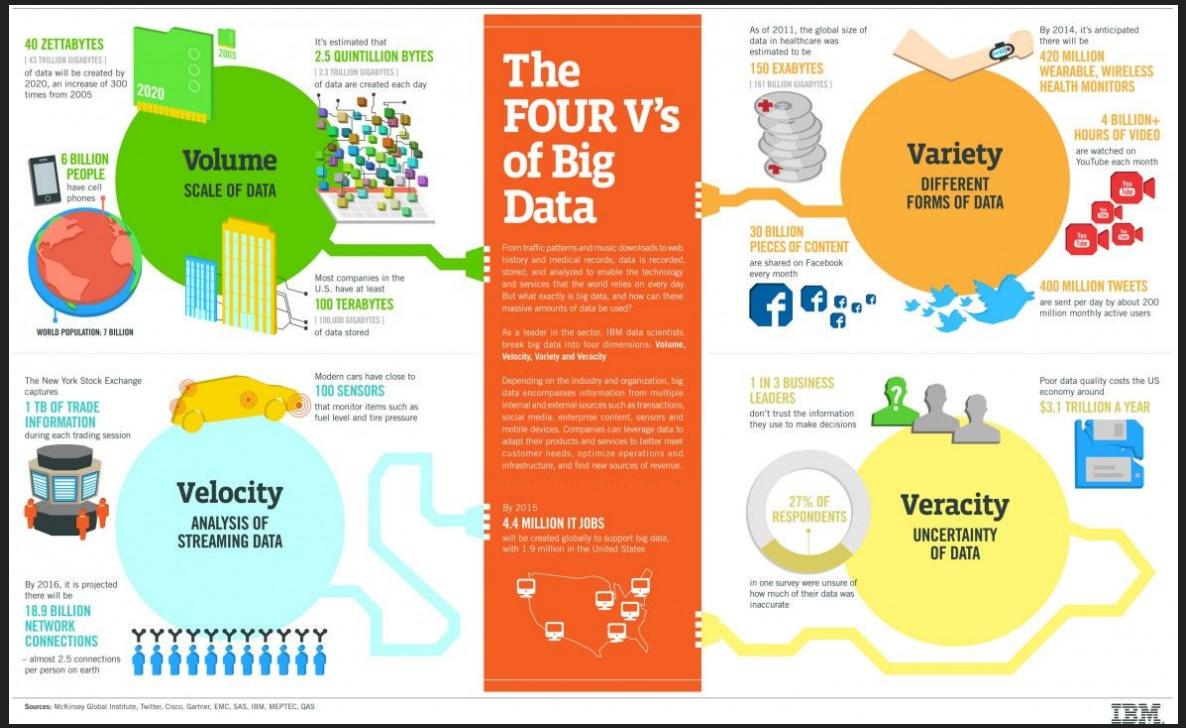
- Volume
- Velocity
- Variety
- Veracity



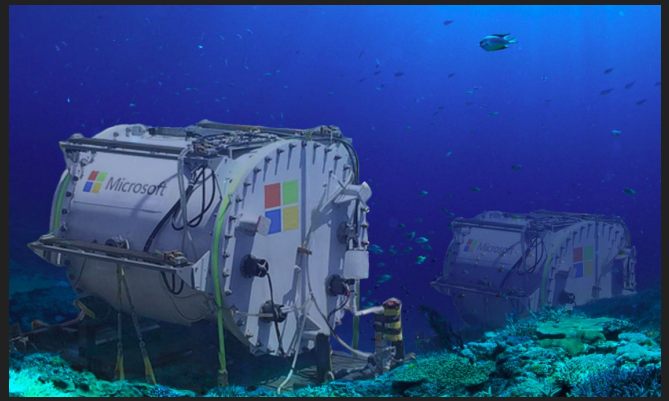
What/Who is producing Big Data?

- people's online activity
- Machines and sensors
- more generally "*things*"

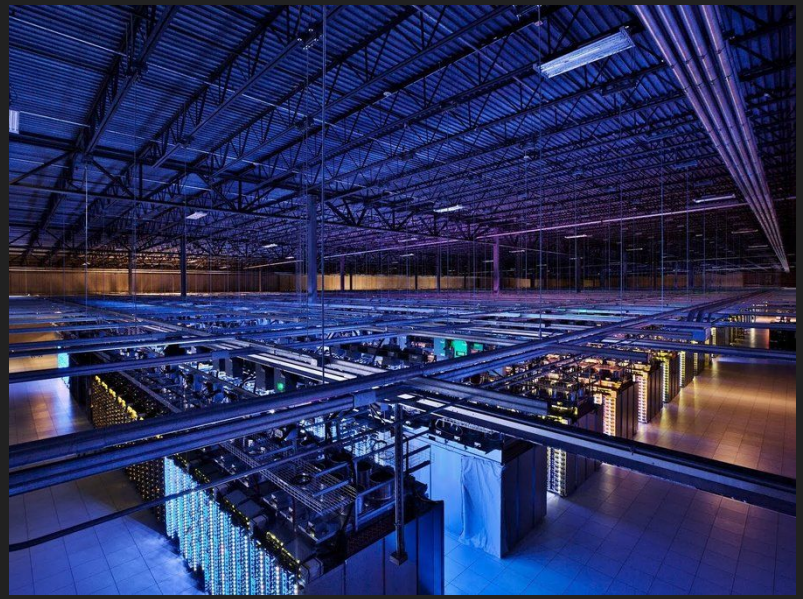




Where do we Store & Compute the Big Data?



- *in huge computer farms*
- “Cloud” Storage and Computing



Why is Data called the “New Oil”?

- improve processes by exploiting patterns in the data - *prediction?*
- How are patterns in data used for prediction?

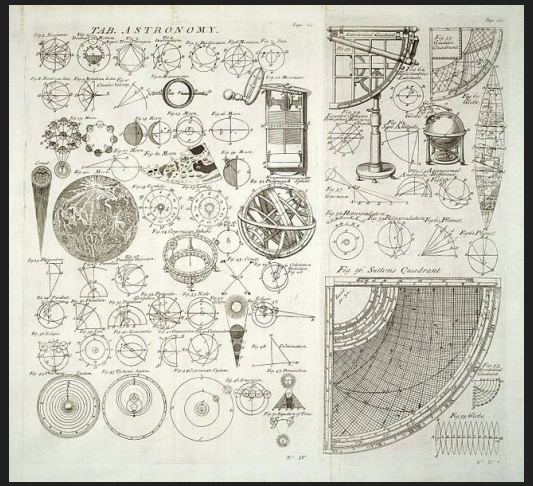
source: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource

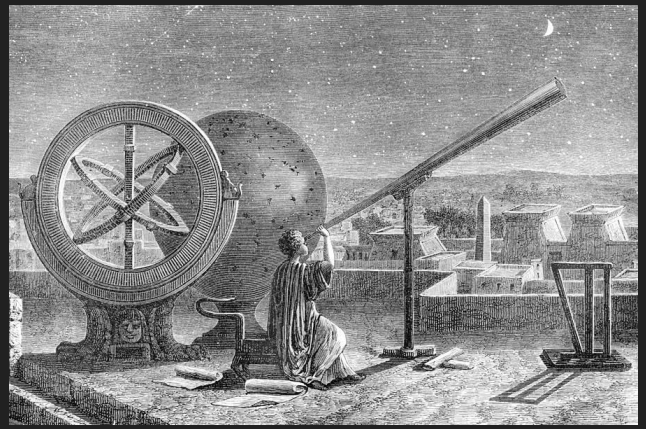
Is making prediction from observation new?

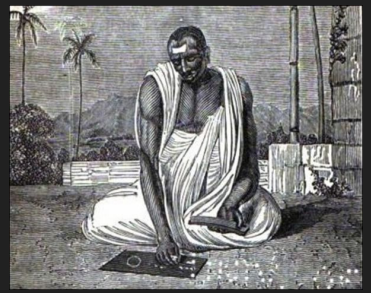
Is making prediction from observation new?

No!

Astronomy is ancient.







How do you make prediction from observation?

How do you make prediction from observation?

The way you can go from observation to prediction is via a MODEL

- Use DATA (past experience) as a feedback to improve your predictive model
- Model can be:
  - Non-mathematical (what we do with our brains all the time)
  - Mathematical (usually implemented as computer algorithm in the cloud)

How do you make prediction from observation?

MODEL

Non-Mathematical

Daily models:
- Cooking
- Household chores
- Worldview / beliefs

Multi-generational models:
- Plant/Animal breeding
- Fishing/Hunting
- ...

Mathematical

Constructive
- Scientific (falsifiable)
- Feedback from data
- Always improving
- Direct measurements

Destructive
- Encode destructive biases and opinions
- Use proxies instead of direct measurements
- No corrective feedback loops with DATA
- Usually not transparent (IP-protected)

“Models are biases & opinions encoded using mathematics in a computer program”, Cathy O’Neil

MODEL

Non-Mathematical

Daily models:
- Cooking
- Household chores
- beliefs - how can we be constructive?

Multi-generational models:
- Plant/Animal breeding
- Fishing/Hunting
- ...

Mathematical

Constructive is “lagom” for society
- Scientific (falsifiable)
- Feedback from data
- Always improving
- Direct measurements

Destructive
- Encode destructive biases and opinions
- Use proxies instead of direct measurements
- No corrective feedback loops with DATA
- Usually not transparent (IP-protected)

“Models are biases & opinions encoded using mathematics in a computer program”, Cathy O’Neil

MODEL

Non-Mathematical

Daily models:
- Cooking
- Household chores
- beliefs - sexism, ageism, racism, etc

Multi-generational models:
- Plant/Animal breeding
- Fishing/Hunting
- ...

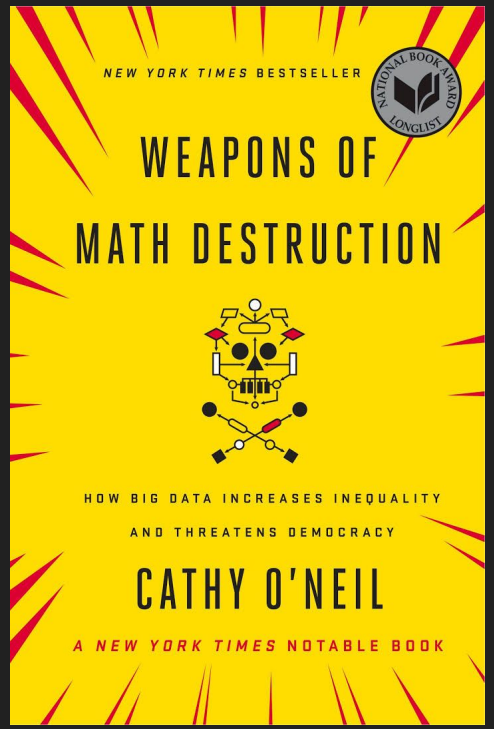
Mathematical

Constructive is “lagom” for society
- Scientific (falsifiable)
- Feedback from data
- Always improving
- Direct measurements

Destructive to some part of society
- Encode destructive biases and opinions
- Use proxies instead of direct measurements
- No corrective feedback loops with DATA
- Usually not transparent (IP-protected)

Models are used to automate life-changing decisions

- Algorithmic Trading in Finance
- Universities rat race for World Rankings
- Online Advertisement
- Getting a Job
- Performance in a Job
- Predictive policing & criminal justice (US)
- Getting credit
- Getting insurance
- Individually *micro-targeted citizen in social media*
- *Let's listen Cathy speak for 10 minutes now:*
  - *https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end#t-786714*



What can we do about it?

Fight back!

Check out Mathematics for Social Activism Workshop at Uppsala June 4-5 2018

We have been working on Project MEP (Meme Evolution Programme)
https://lamastex.github.io/scalable-data-science/sds/research/mep/

To help shed transparent light on

- *How we are possibly micro-targeted as citizens in social media (US Election)?*

Looking for coders who want to help with similar projects around SE Election:

  - *Uppsala Big Data Course Set next period: https://lamastex.github.io/360-in-525/*
  - *Create an "ideological forest of Swedish peoples from their Twitter activity in public streams"*

Bruce Schneier's Fundamental Issue of the Information Age

https://youtu.be/GhWJTWUvc7E?start=1580&end=1760

"And it is not a matter of all surveillance is bad. I think this is a complex issue. This is an issue of designing systems to extract group value from our data while protecting people individually. And I actually think this is a fundamental issue in the information age. Our data has enormous value to us collectively, and our data has enormous value to us each individually. How do we reconcile this?"

Is there a solution? What is the mathematical problem here?

Work has started on mathematical statistical notions of differential privacy.

See Michael Jordan's talk: https://youtu.be/ggq7HiDO0OU?start=158&end=1724

Math-Stats: Minimax Optimal Procedures for Locally Private Estimation, 2017