

GLASSLAB (Graphics Learning and Smart Sensing Lab)

Data Science

Is something big happening?

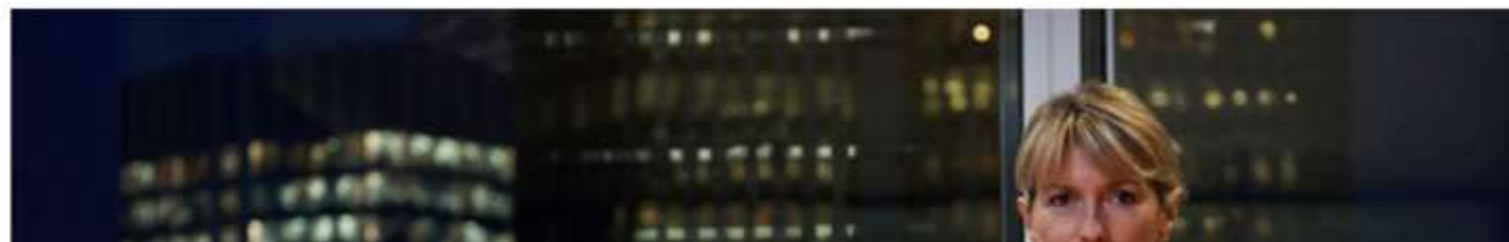
FEATURE

5 tips for working with data science



What the C-Suite Needs To Know About Data Science

October 20, 2015 by Vin Vashishta — Leave a Comment



Forbes

/ Leadership

OCT 6, 2015 @ 04:44 AM

Data Science



FORTUNE

DATA SCIENCE

The White House has plan to help society through data analysis

Big Data Sector

TECH TIMES

PERSONAL TECH BIZ TECH



The HP ElitePad Rugged Tablet

Limited by the Environment



Agricultural Revolution

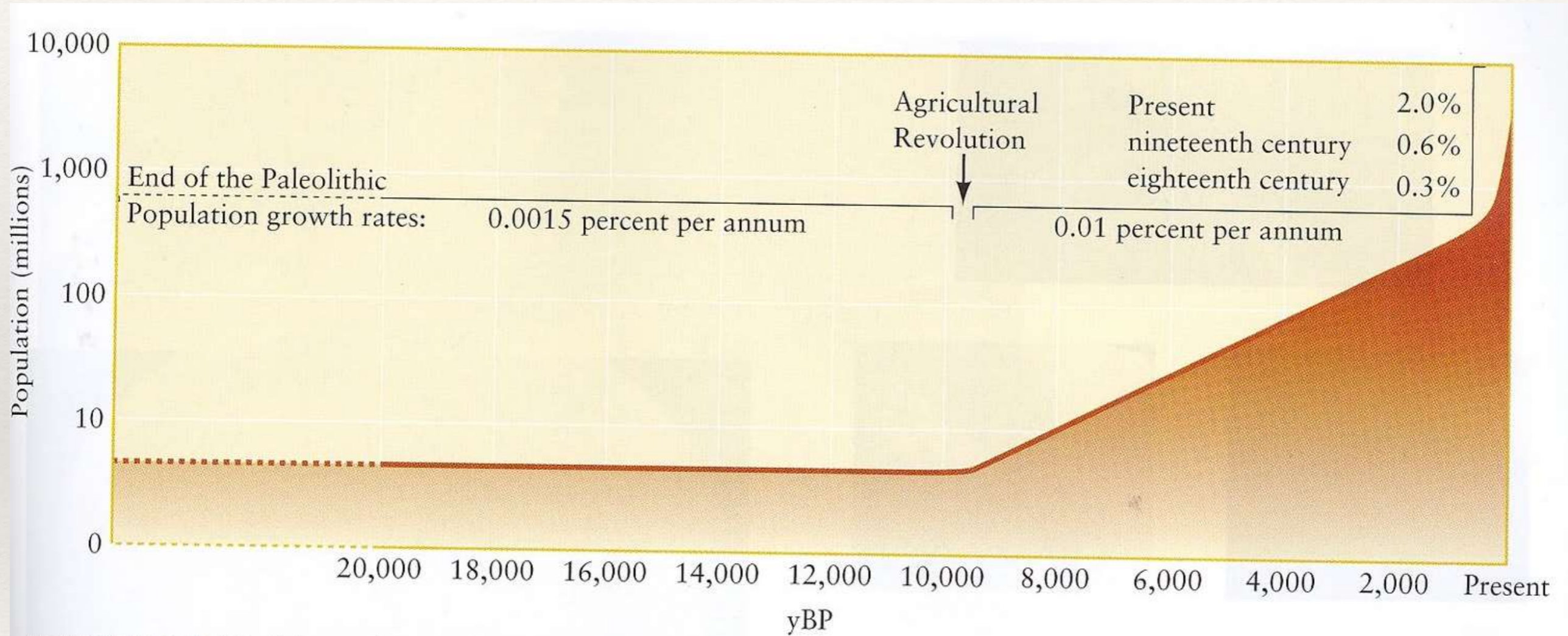


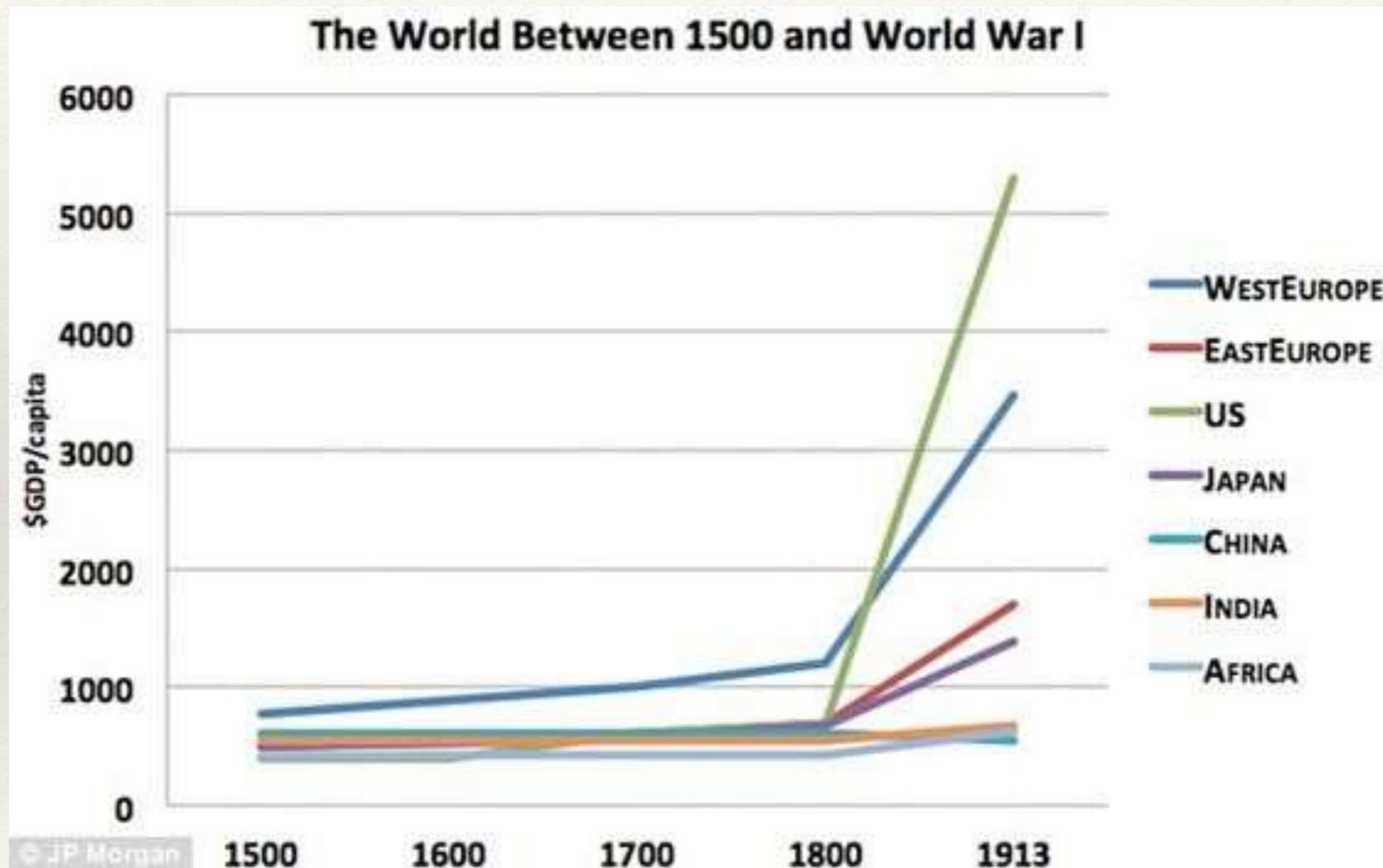
FIGURE 12.9 ■ **World Population Size**

This graph shows the trend in population size on a global scale. Until 10,000 yBP and the advent of agriculture, population remained constant, numbering less than 10,000,000 people. After the agricultural revolution, however, population skyrocketed.

Limited by our muscles



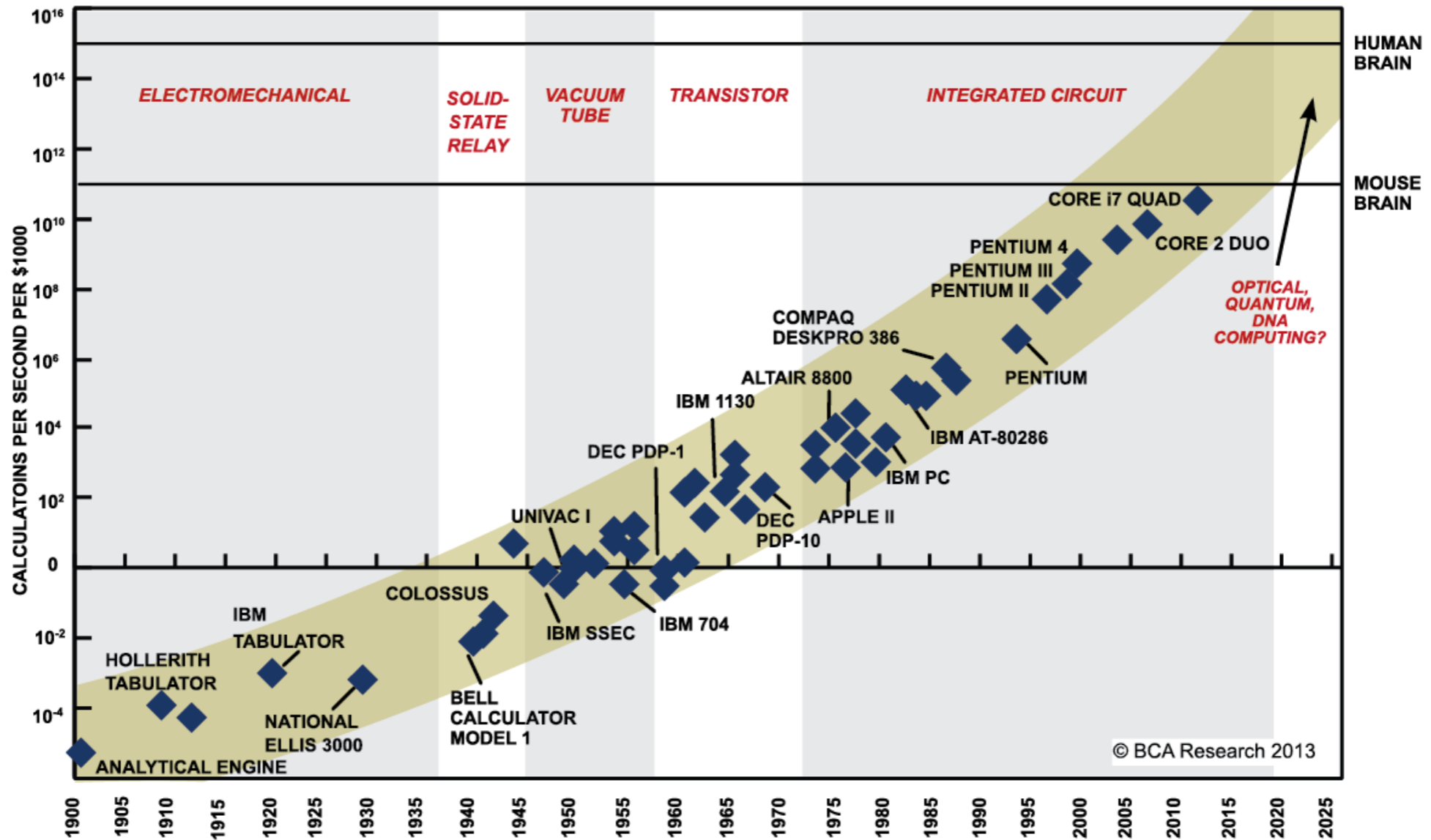
Industrial Revolution



Limited by what we can calculate



Computational Revolution



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, THE VIKING PRESS, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

Limited by what we can understand



An aerial photograph of a city at night, with lights reflecting on the water. The text "Data Revolution" is overlaid in the center in a red, sans-serif font.

Data Revolution

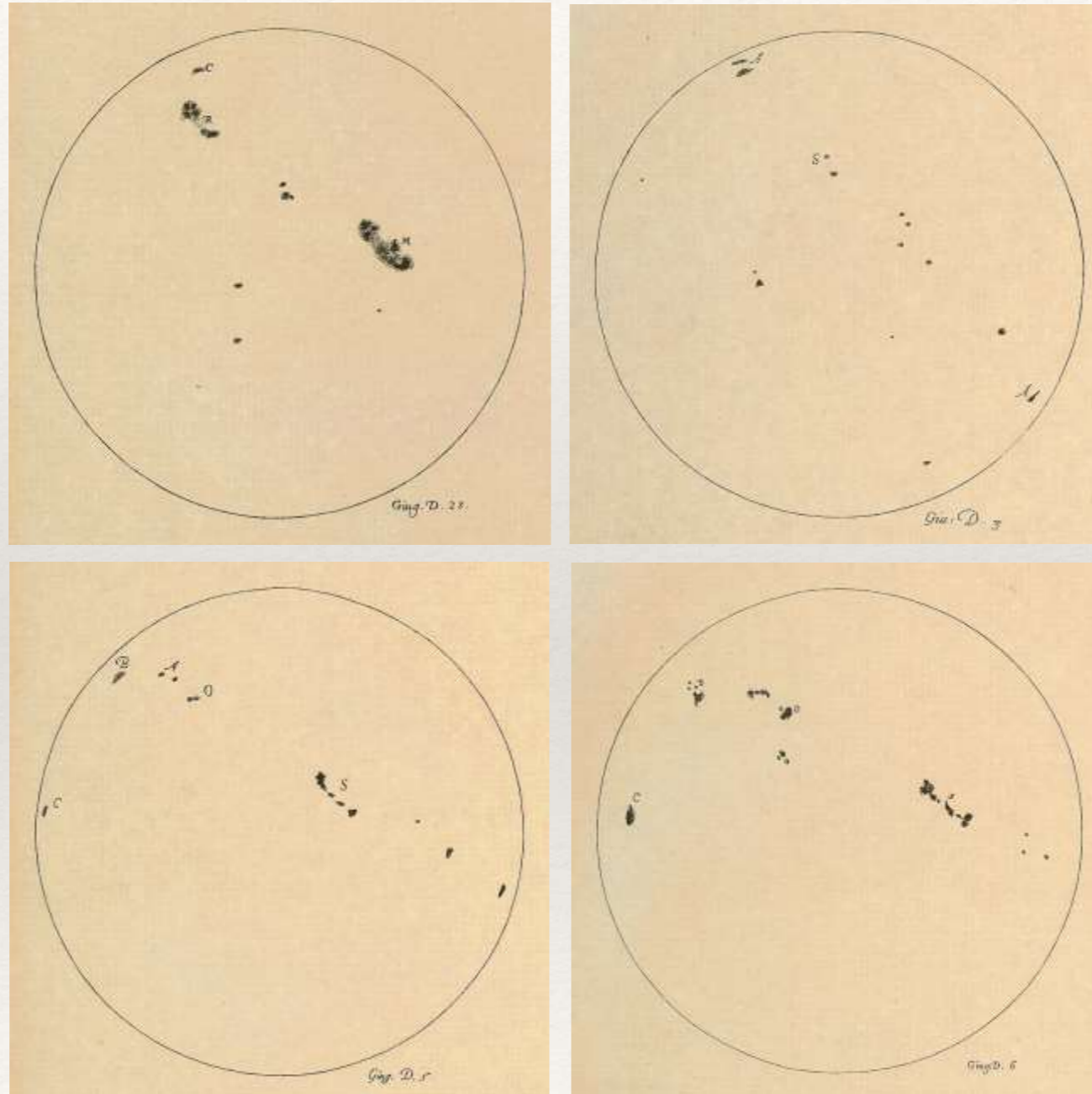
Where is this coming from?

Throughout Human History



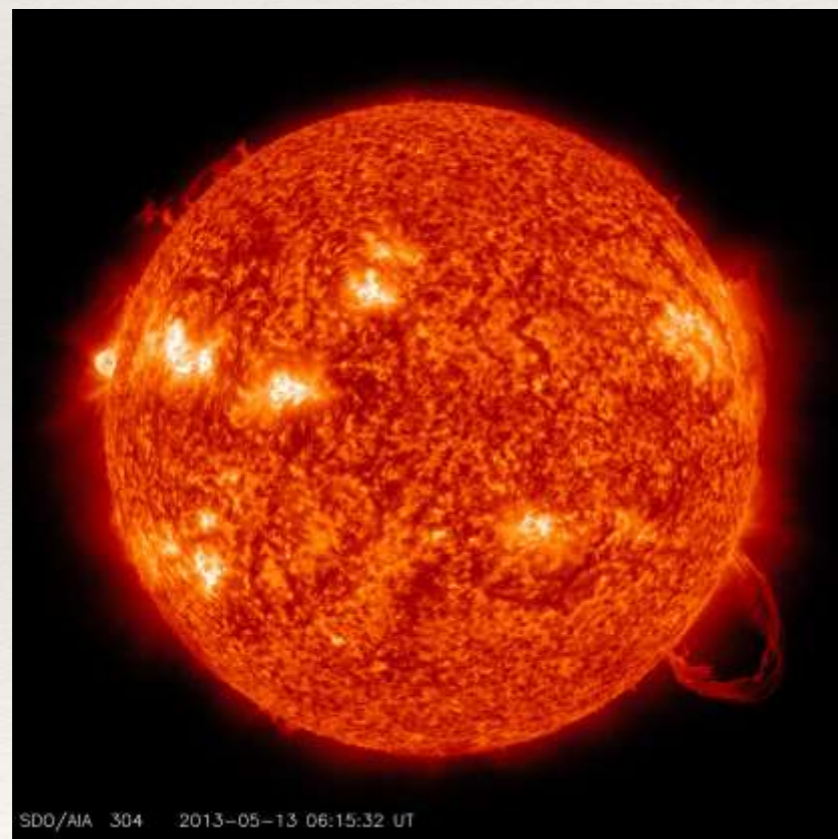
Data difficult to collect and store

Galileo's Hand-drawn Sunspots



vs.

UltraHD video of the Sun 24-7



From NASA SDO



Wilcox Solar Observatory

Flood of Data

GOES 9,10,12
NOAA-15,16,17,18
LandSat 5,7
DMSP F13,14,15,16
Meteosat 6,7,8,9
CBERS-2,2B
SPOT-2,4,5
ENVISAT
Resourcesat 1
CARTOSAT-1,2,2A
RADARSAT-1,2
KOMPSAT-1
THEO-1
GOMs
GMS-5
METEOR-3
OKEAN
Feng-Yun



1 Sensor (MODIS) = 125 GB/DAY

Flood of Data

Large Hadron Collider

37 TB of data/day

Network of Sensors

NASA Earth Observatories



Model/Simulation Output



Social Network Data



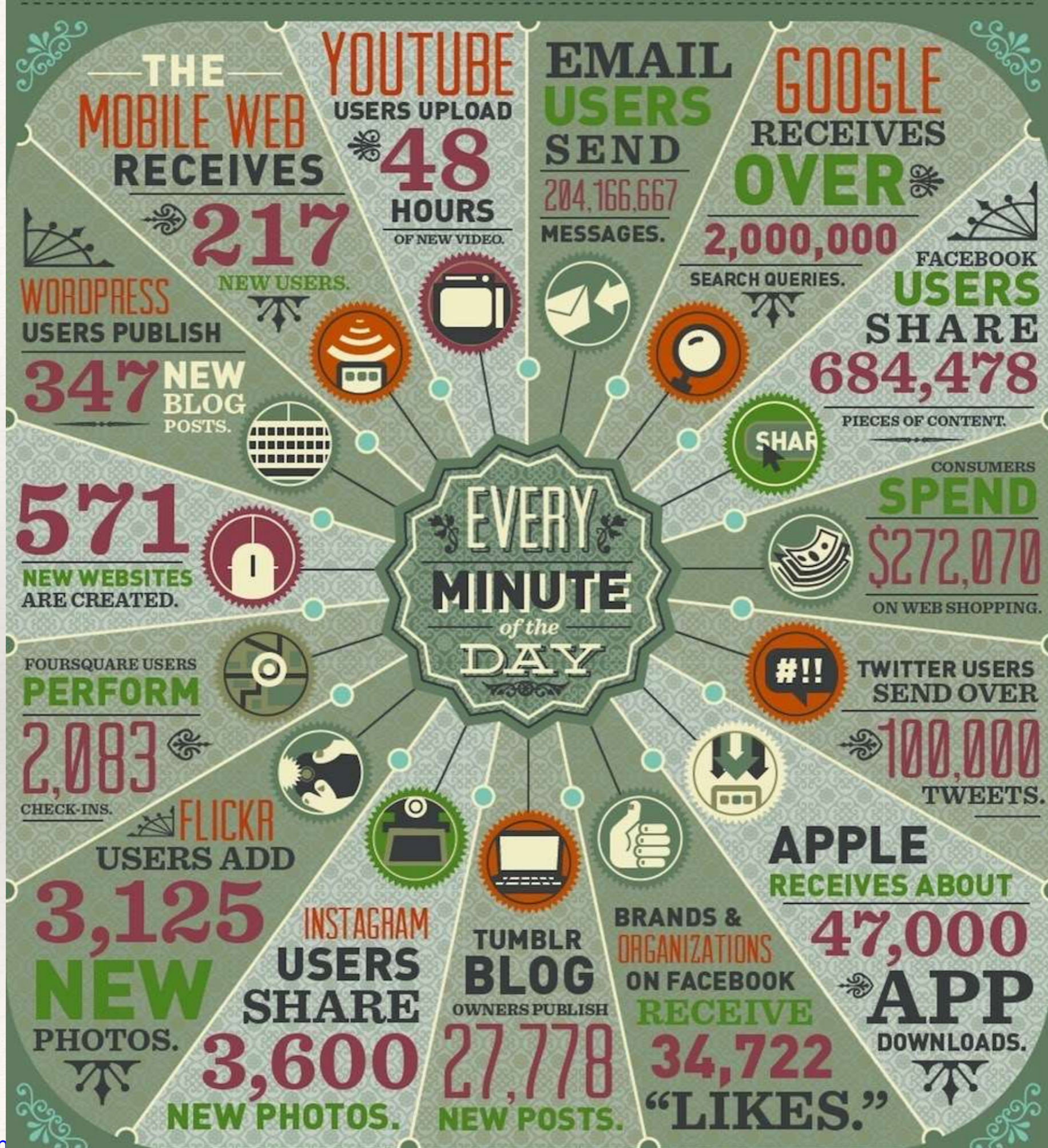
Health Data



Financial Data



Security Data



Every two days now
we create as much
information as we did
from the dawn of
civilization up until
2003

Eric Schmidt, former CEO
Google, 2010



Perfect Storm: Big Data + Machine Learning

Old AI: IBM 1997



We know the rules.

19 Dec 2013

Playing Atari with Deep Reinforcement Learning

Volodymyr Mnih Koray Kavukcuoglu David Silver Alex Graves Ioannis Antonoglou

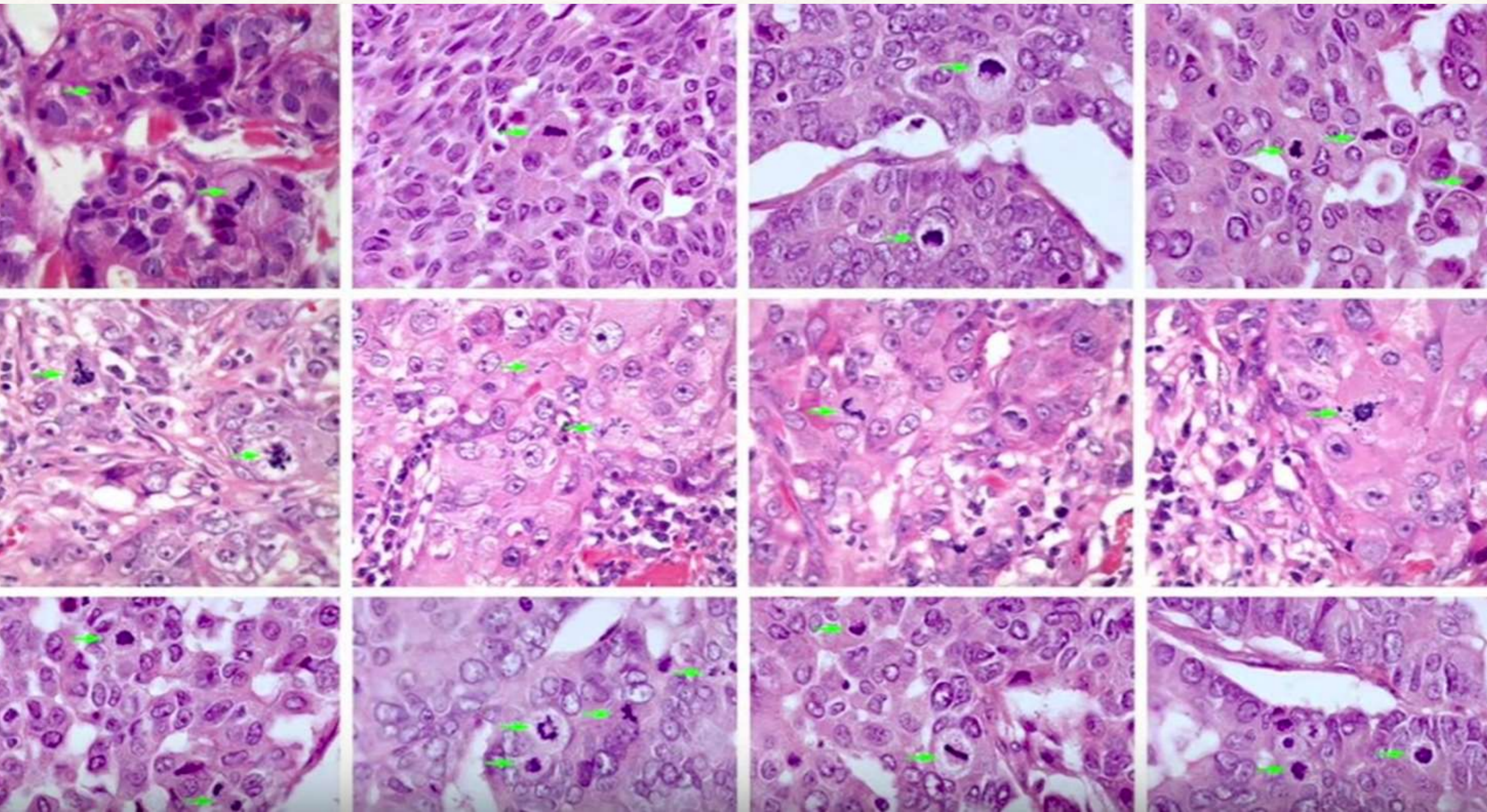
Daan Wierstra Martin Riedmiller

DeepMind Technologies



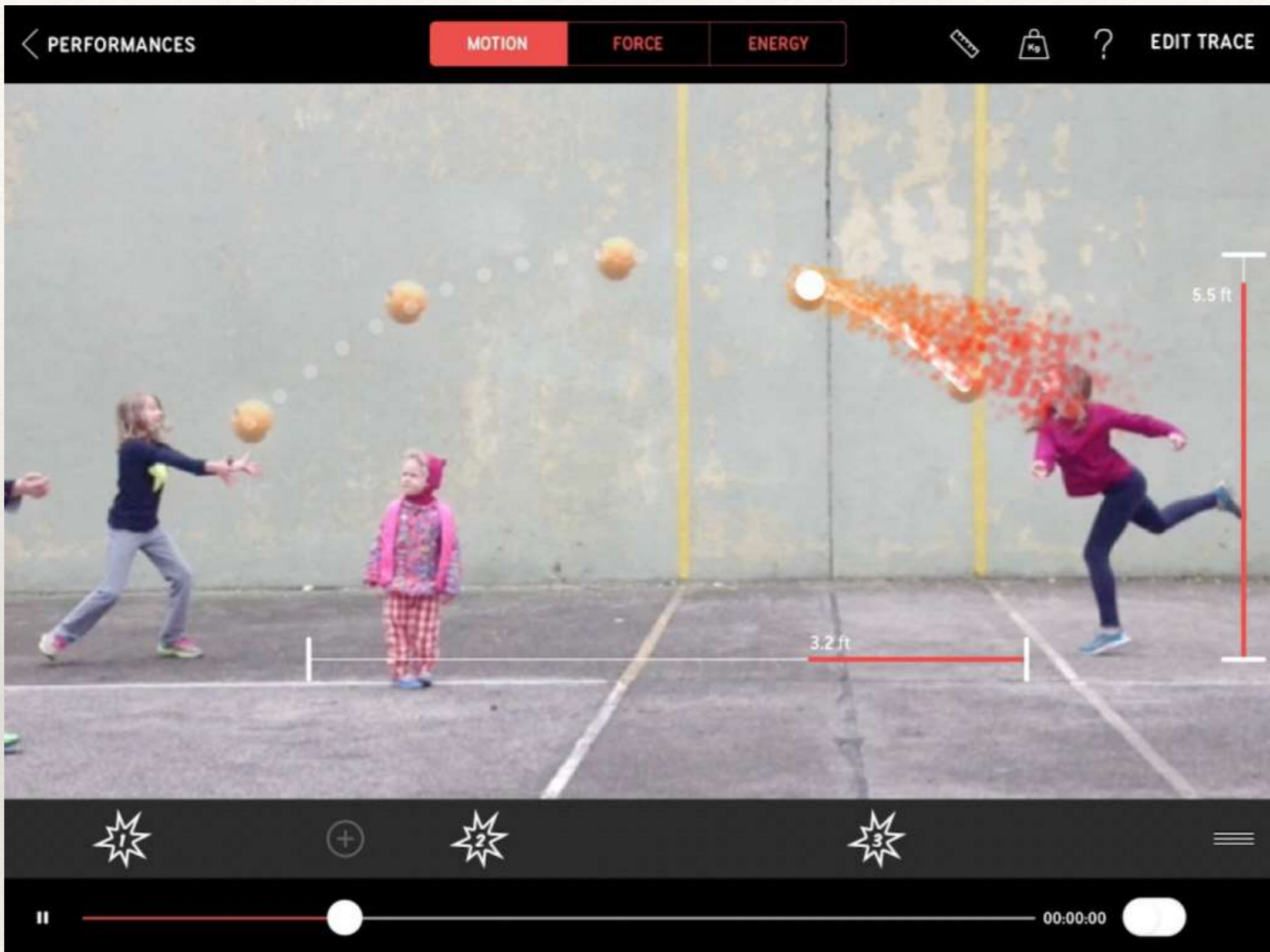
Figure 1: Screen shots from five Atari 2600 Games: (Left-to-right) Pong, Breakout, Space Invaders, Seaquest, Beam Rider

It learns the rules.



What rules?

Is this important for science
and engineering?



NYC Playground Physics

All: Driven by Data Science

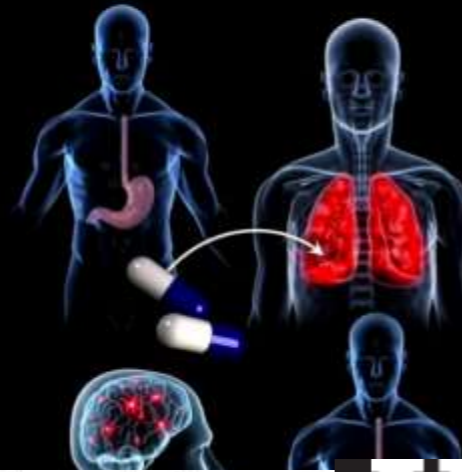
Internet of Things

How Data Science-Driven Software is Eating the Connected World

Sarah Aerni, Principal Data Scientist
Pivotal
@itweetsarah

Hadoop Summit, San Jose, CA
June 10th

Data Science in Drug Discovery

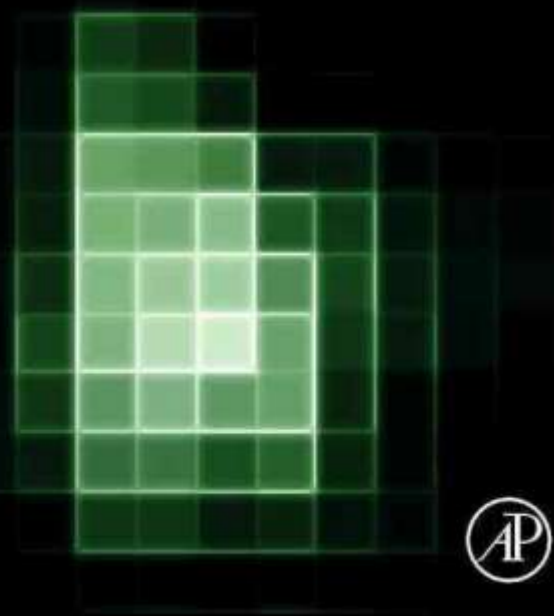


Marina Sirota, PhD
Assistant Professor
Institute for Computational
Health Sciences

October 22, 2015

Energy Storage for Smart Grids

Planning and Optimization for Renewable and Variable Energy Resources (VERs)
Pengwei Du and Ning Lu



Robot learns skills through trial and error, like you do



VB

Now mobile analytics startups are hiring data scientists. Here's why



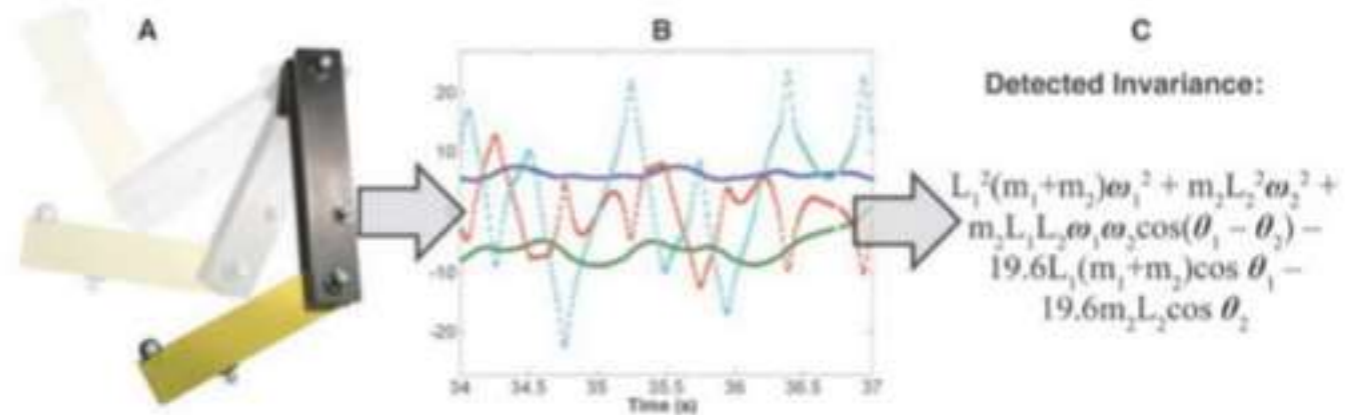
THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

BRANDON KEIM SCIENCE 04.02.09 3:09 PM



Illustration: Marian Bantjes

COMPUTER PROGRAM SELF-DISCOVERS LAWS OF PHYSICS



IN JUST OVER a day, a powerful computer program accomplished a feat that took physicists centuries to



Completed • \$40,000

Merck Molecular Activity Challenge

Thu 16 Aug 2012 – Tue 16 Oct 2012 (3 years ago)

Help develop safe and effective medicines by predicting molecular activity.

Help enable the development of safe, effective medicines.

Deep Learning How I Did It:
Merck 1st place interview

Geoffrey Hinton

Computer scientist



Short Term: Scramble for Talent



"data science", "data scientist"

Find Trends

Find Jobs

what: job title, keywords or company

Job Trends

Job Trends

[Job Postings Per Capita](#)

[Job Market Competition](#)

[Industry Employment Trends](#)

"data science", "data scientist" Job Trends

Scale: **Absolute** - [Relative](#)

[▶ Email to a friend](#)

Job Trends from Indeed.com

— "data science" — "data scientist"

[▶ Post on your blog/v](#)



Top Job Trends

1. [HTML5](#)
2. [MongoDB](#)
3. [iOS](#)
4. [Android](#)
5. [Mobile app](#)
6. [Puppet](#)
7. [Hadoop](#)
8. [jQuery](#)

The 11 hottest jobs on Wall Street in 2015

by Paul Clarke 2 April 2015

PERSONAL FINANCE: SPREAD THE WEALTH

The Top 5 Highest-Paying Tech Jobs in 2015

Which titles to target if you want to make over \$100K in one of the world's fastest-growing industries.

Want better work-life balance? Here are jobs for you

By Sally French
Published: Oct 20, 2015 4:56 p.m. ET

227 179 22

Tech, social media jobs rank high

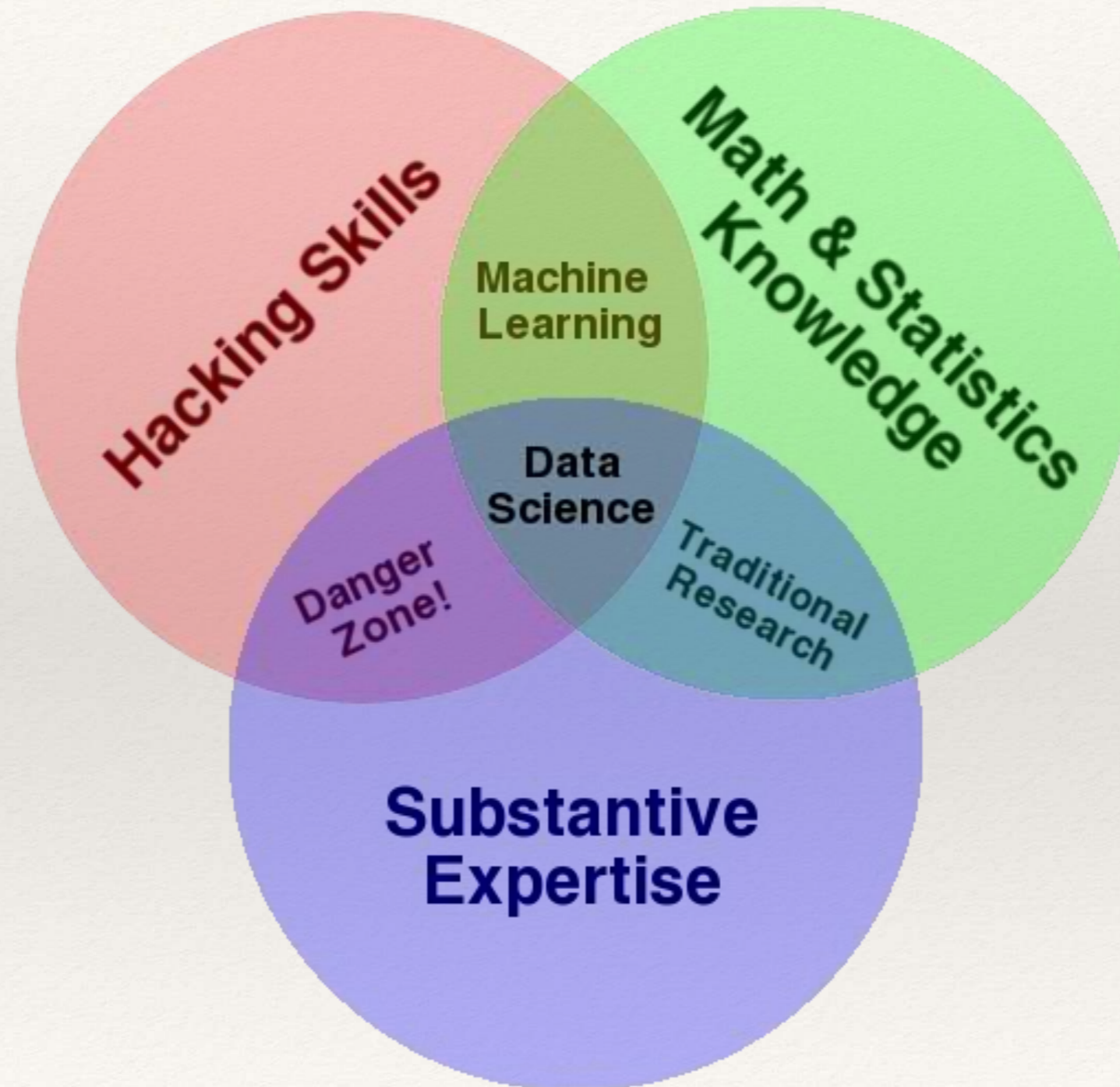
Rank	Job Title	Work-Life Balance
1	Data Scientist	4.2



11. Data scientists

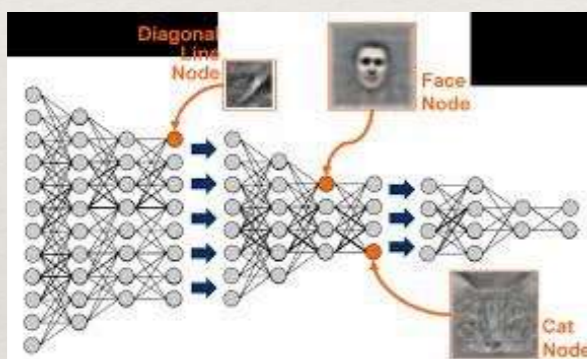
Big data is central to all the hot technology jobs outside of Wall Street currently, but it's also becoming increasingly important for financial institutions to get to grips with. Whether that's hedge funds grappling huge amounts of data on Twitter to make trading decisions or investment banks shaking up their data analysis tools, data scientists are gradually becoming hot property in the financial sector, according to one New York IT recruiter who did not wish to be named. Start boning up on [Hadoop and its related products](#).

What is Data Science?



Coping with Data

Algorithms



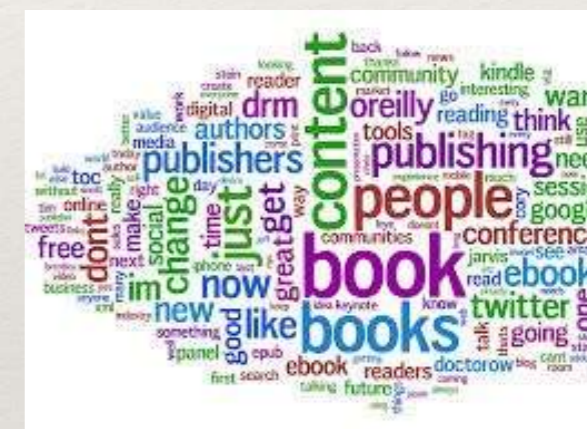
+

Computation

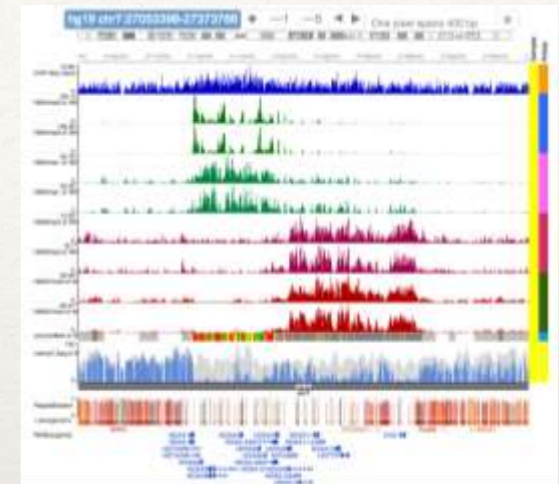
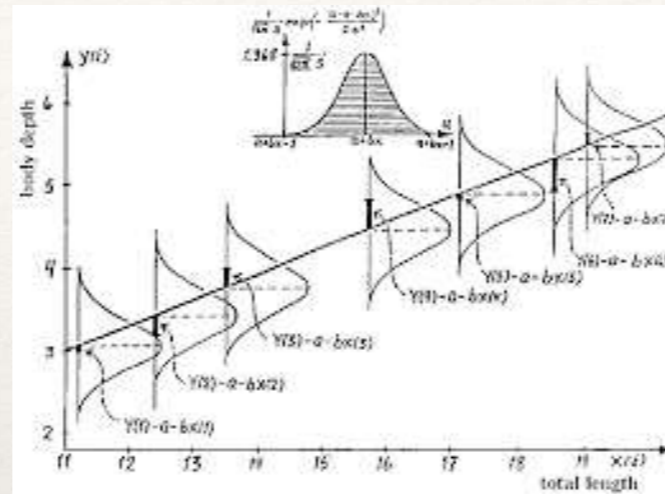
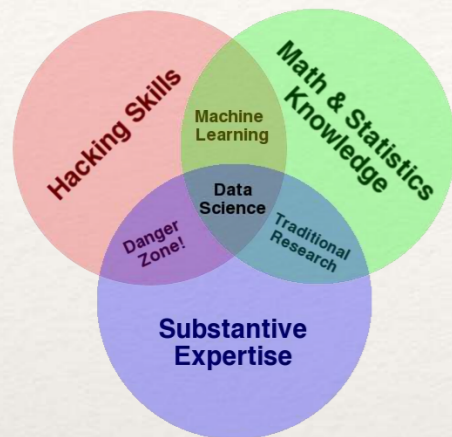


+

Visualization



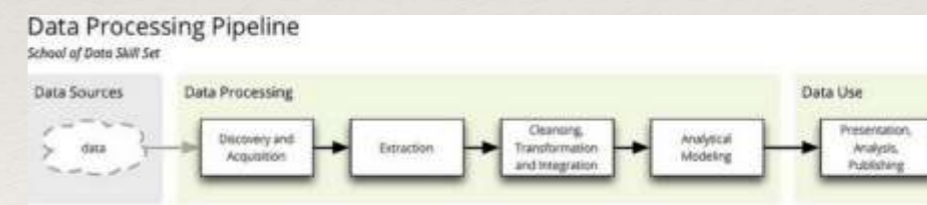
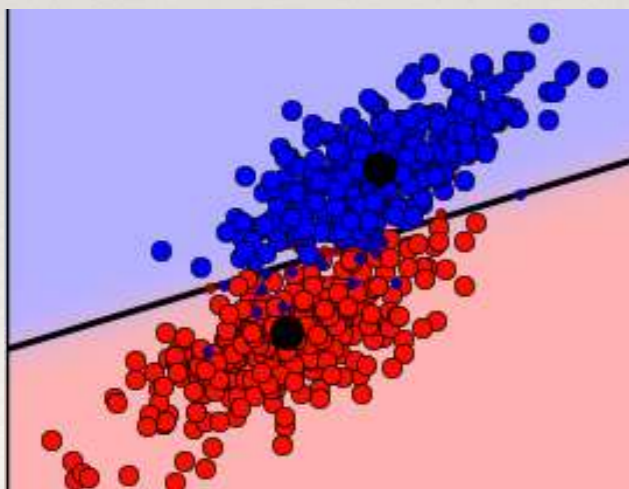
Core Subjects



Intro to Data Science

Applied Stats

Data Visualization



Machine Learning

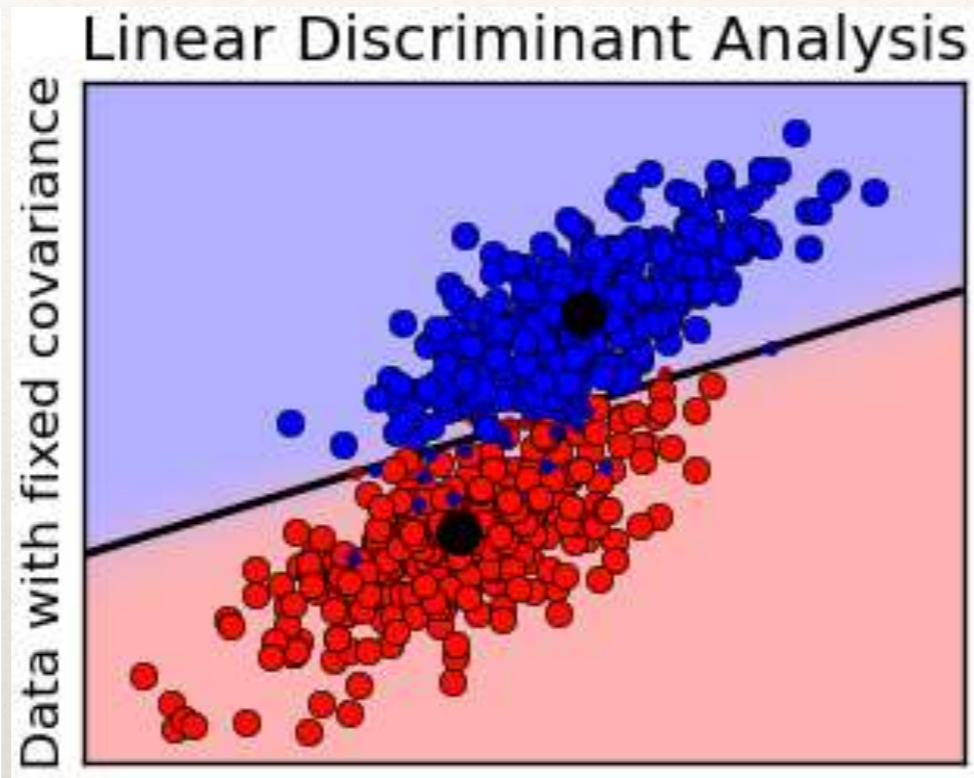
Big Data

Data Engineering

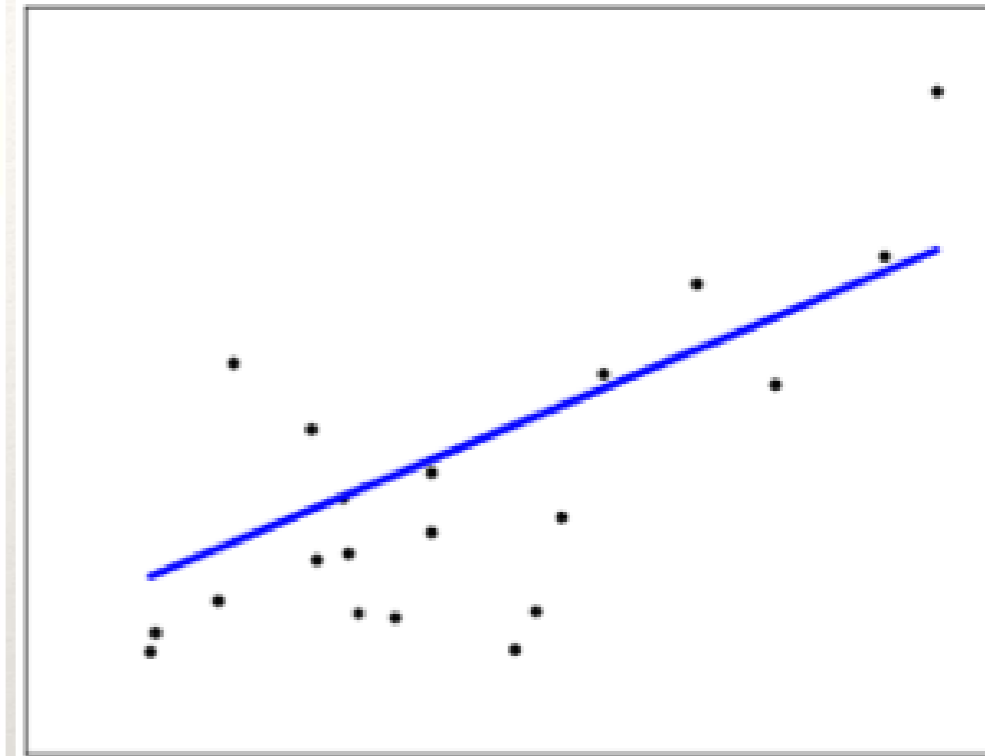
Machine Learning



Clustering Problem



Classification Problem

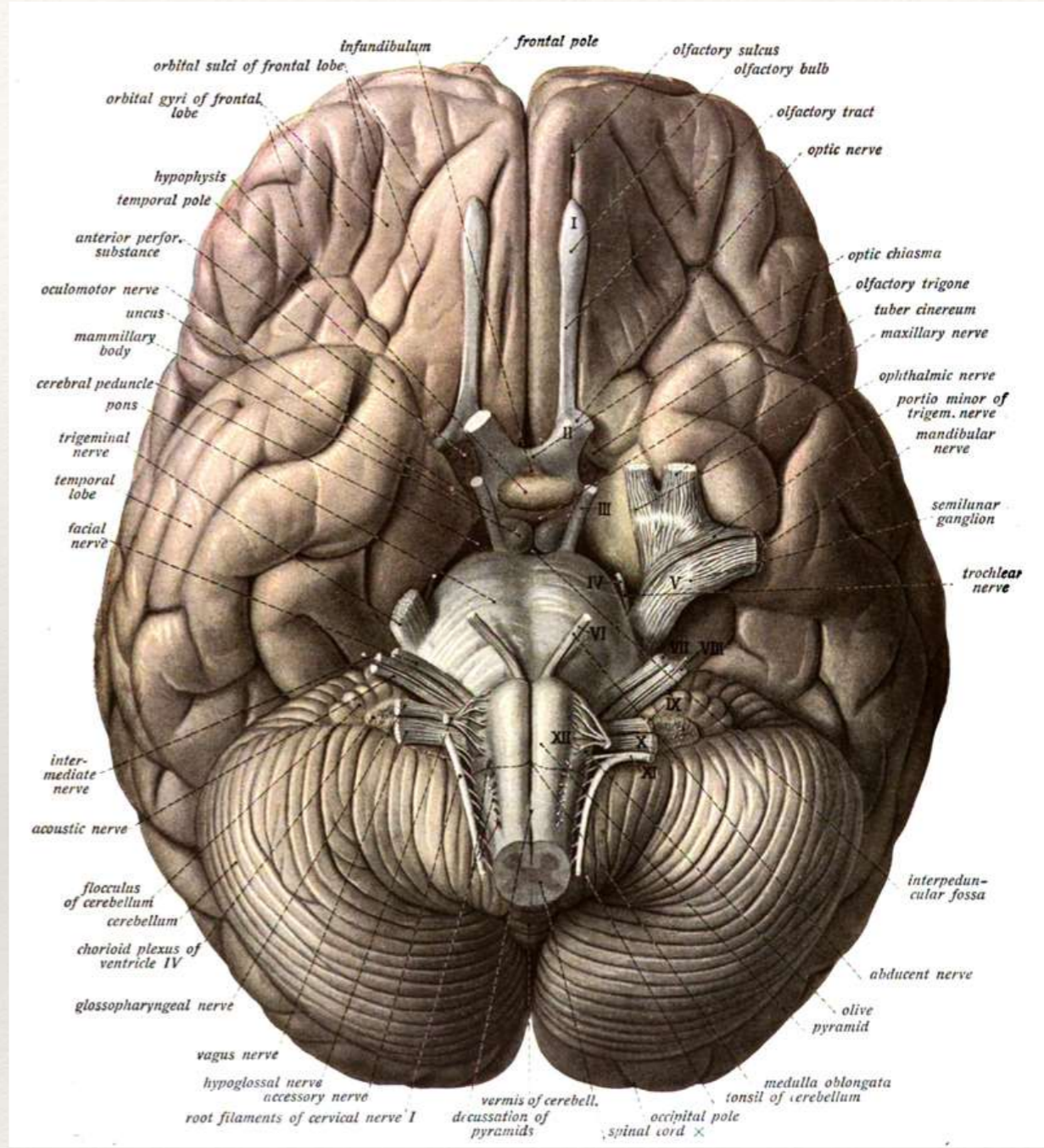
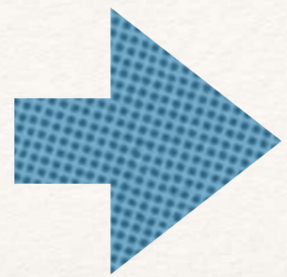


Regression Problem

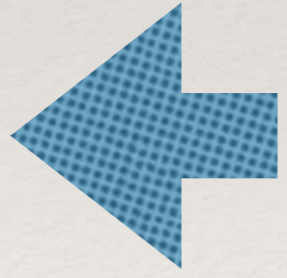
Goal: Explore structure of data and make predictions

Data Visualization

Information

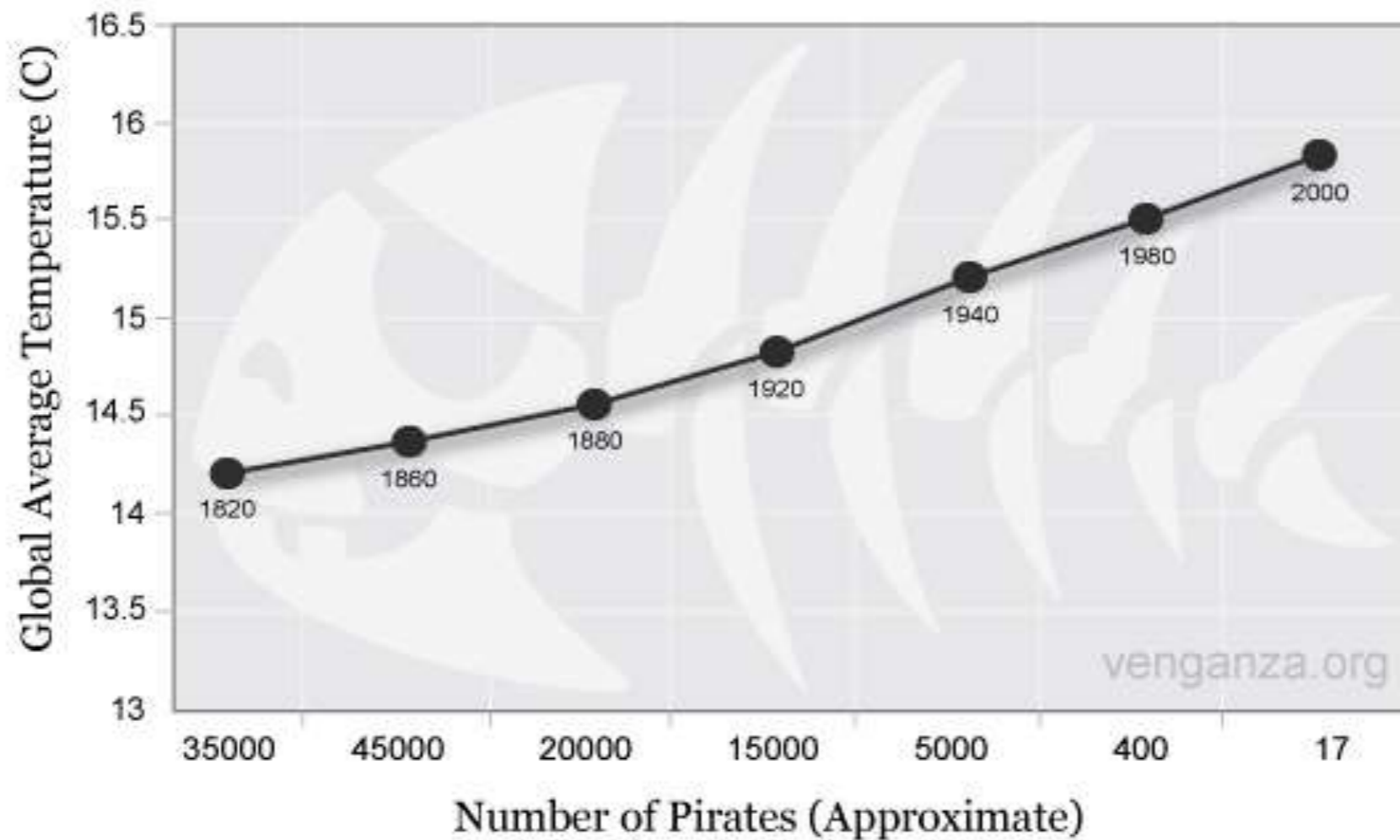


Understanding



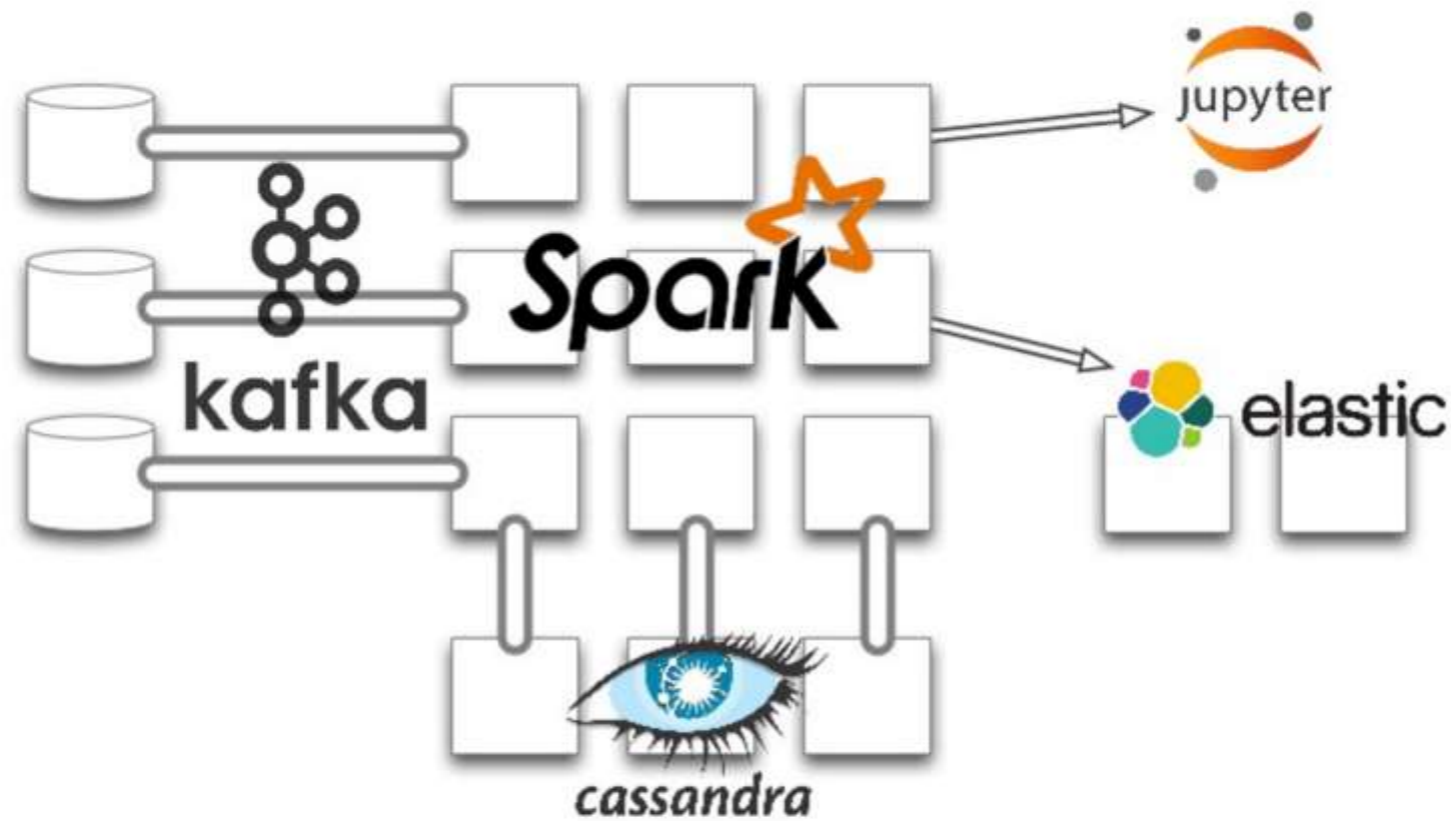
Statistics

Global Average Temperature Vs. Number of Pirates

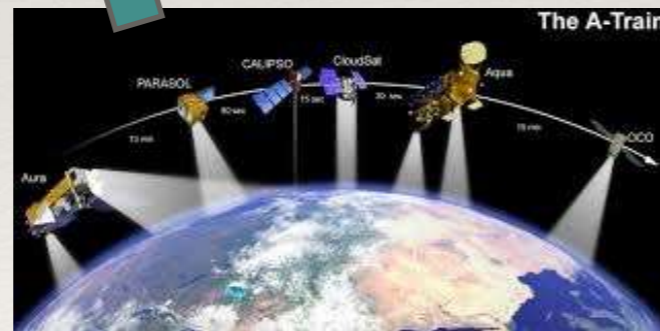
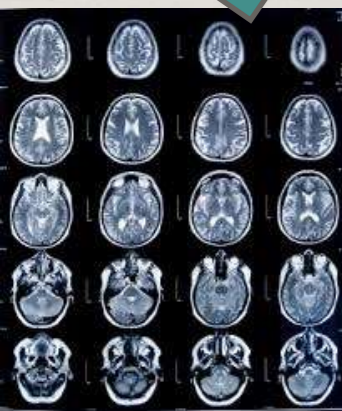
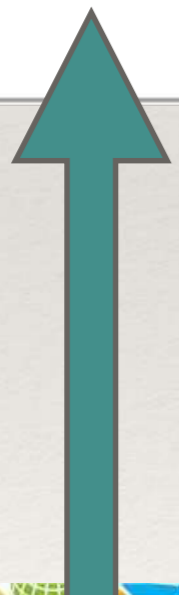
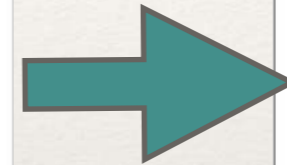
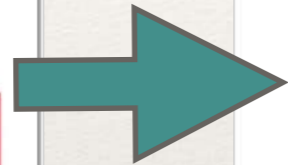
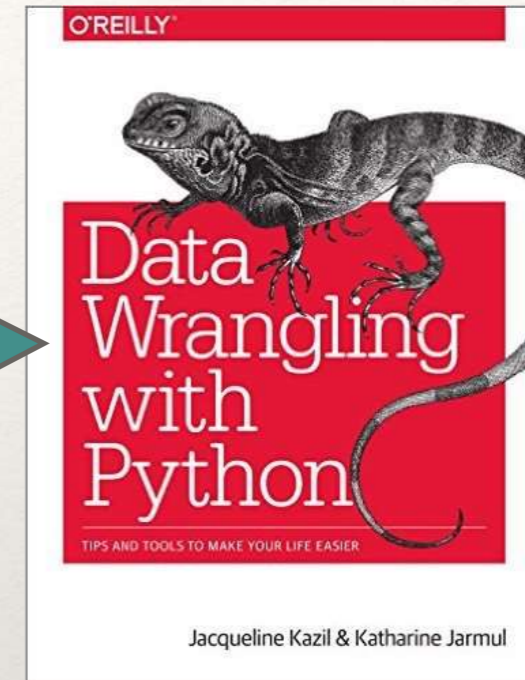
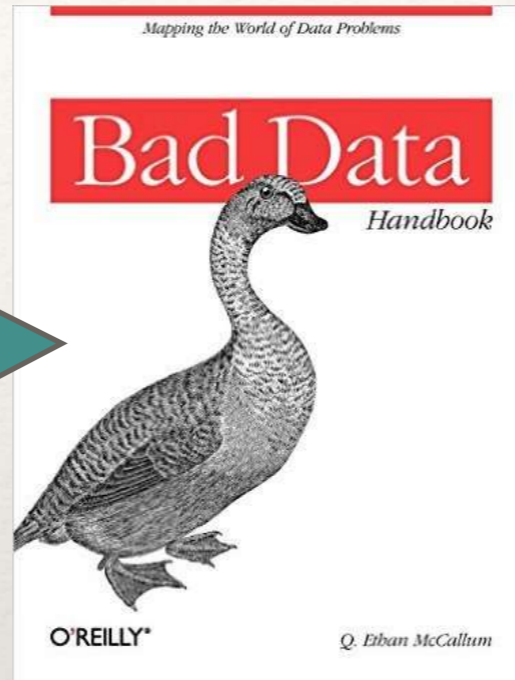
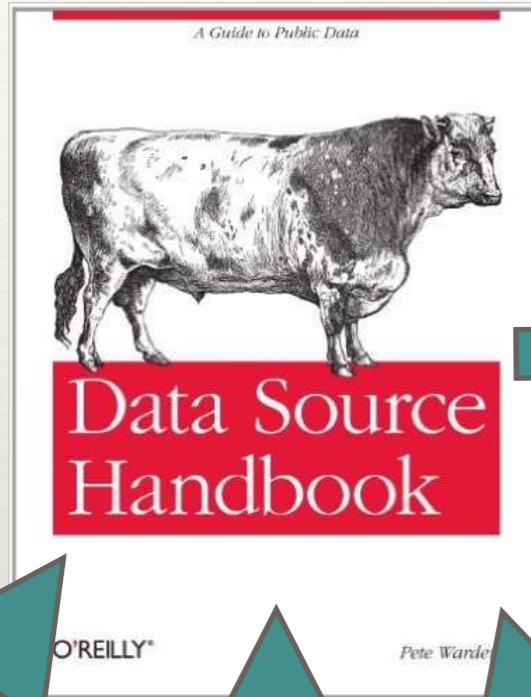


Infrastructure for Big Data

Design Patterns



Data Engineering



Data Cleaning

Data Transformation
Signal/Image Processing

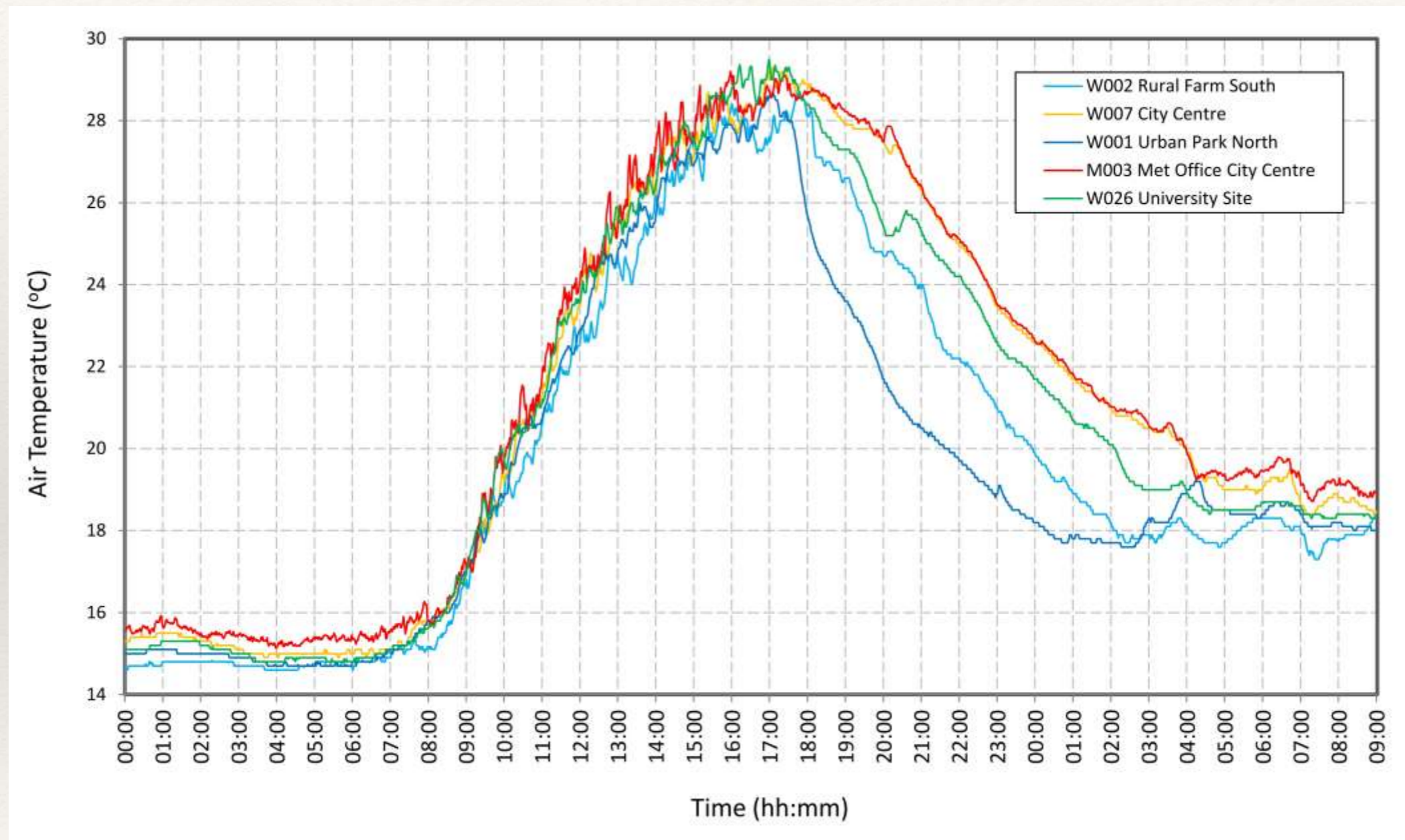


Cities/Towns form a Network



Who is closer to whom?

Similarities in Temp and Rainfall



Who is closer to whom?