Mathematics, Data Science and Social Impact

Provocative Debate about their Chemistry if any

(mine from Big Data perspective)

Data

is the sword of the twenty first century, those who wield it well, the Samurai.

– Jonathan Rosenberg, former SVP of product management at Google

Source: Rosenberg, Jonathan. 2009. "From the Height of this Place," Official Google Blog https://googleblog.blogspot.com/2009/02/from- height-of-this-place.html

"Big Data" Sources

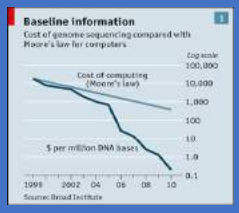
It's All Happening On-line

Every:
Click
Ad impression
Billing event
Fast Forward, pause,…
Server request
Transaction
Network message
Fault
…

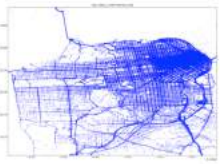
User Generated (Web & Mobile)

Internet of Things / M2M

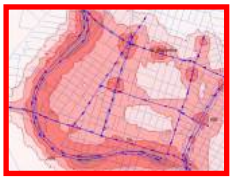
Health/Scientific Computing



What can you do with the data?
One example application:
Traffic Prediction and Earthquake Warning







Crowdsourcing + physical modeling + sensing + data assimilation

to produce:


4

From Alex Bayen, UCB, Director, Institute for Transportation Studies

Big Data: The Good (Bright and Ubiqutious)

Experiments, observations, and numerical simulations in many areas of science and business are currently generating terabytes of data, and are on the verge of generating petabytes and beyond for some applications.

Analyses of the information contained in these data sets have already led to major breakthroughs in fields ranging from genomics to astronomy and high-energy physics and to the development of new information-based industries.

- Frontiers in Massive Data Analysis, National Research Council of the National Academies

Big Data: The Bad (Dark and Scarce)

Given a large mass of data, we can, by judicious selection, construct perfectly plausible unassailable theories—all of which, some of which, or none of which may be right.

- Paul Arnold Srere

Big Data: The Hopeful

The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it — that's going to be **a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids**. Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

- Hal Varian, Google's Chief Economist,
http://www.mckinsey.com/insights/innovation/hal_varian_on_how_the_web_challenges_managers

The goal of this summer school: Getting students to be able to think critically about data.

(with the focused theme of **Math, Data Science and Social Impact**)

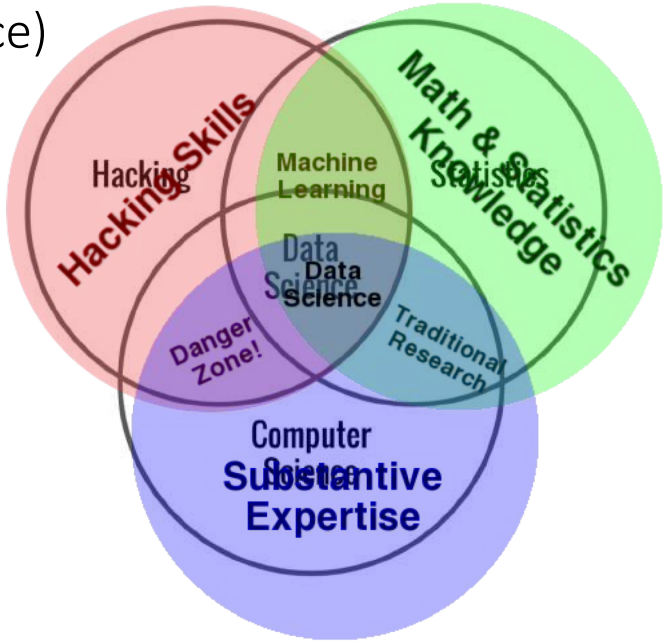
Big Data: The Hopeful

The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it — that's going to be **a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids**. Because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

- Hal Varian, Google's Chief Economist,
http://www.mckinsey.com/insights/innovation/hal_varian_on_how_the_web_challenges_managers

Educate and Learn to be able to think critically about data.

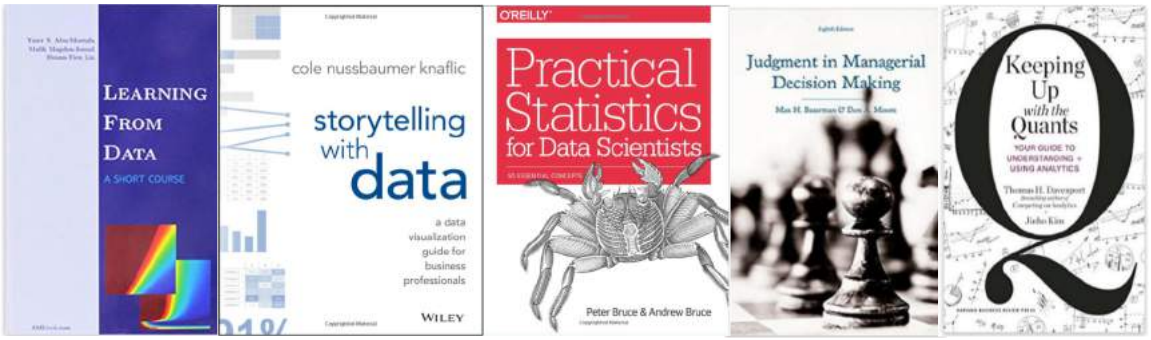
Data Science – A Visual Definition
(Skills for Data Science)



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Business Intelligence v.s. Data Science

Business Intelligence	Data Science
Querying the past	Querying the past present and future



Machine Learning v.s. Data Science

Machine Learning
Develop new (individual) models
Prove mathematical properties of models
Improve/validate on a few, relatively clean, small datasets
Publish a paper

Data Science
Explore many models, build and tune hybrids
Understand empirical properties of models
Develop/use tools that can handle massive datasets
Take action!

Turn data into data products

Mathematics and Data Science: Strong Chemistry

Digital Images and Problems

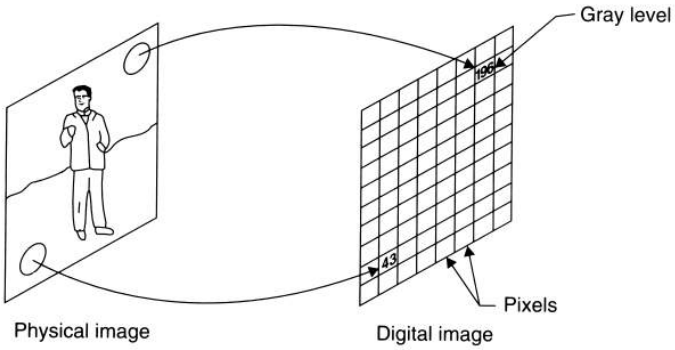


Figure 1–1 A physical image and a corresponding digital image

Mining Images with ML and Deep Neural Networks

- Mathematical techniques (Fourier, wavelets, SVD, etc.)
- Problems from data analysis (mainly image analysis)

What's a Digital Image?

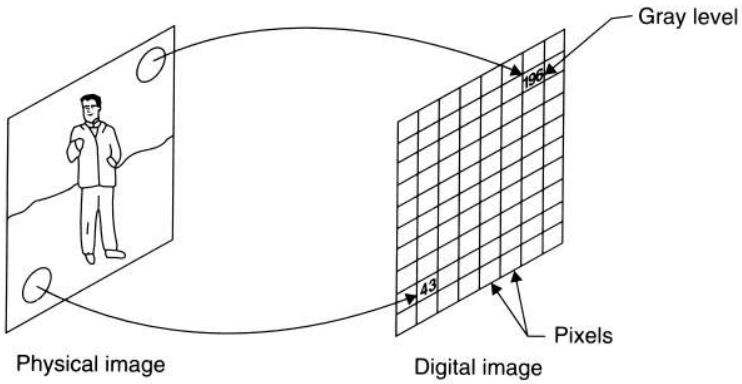
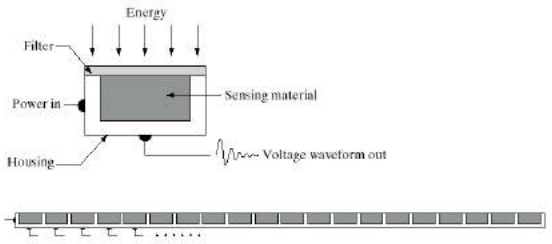


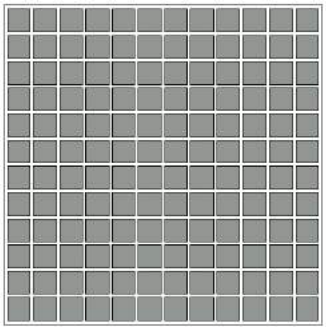
Figure 1–1 A physical image and a corresponding digital image

Mechanism for digitizing

a
b
c

FIGURE 2.12
(a) Single imaging sensor.
(b) Line sensor.
(c) Array sensor.

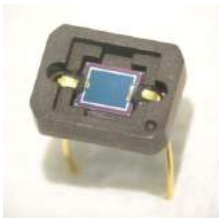




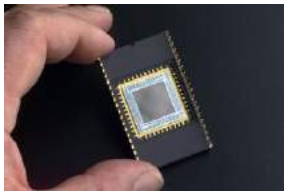
Examples of Sensors

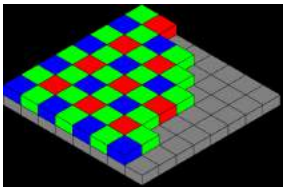
Well known from physics courses…

photodiode



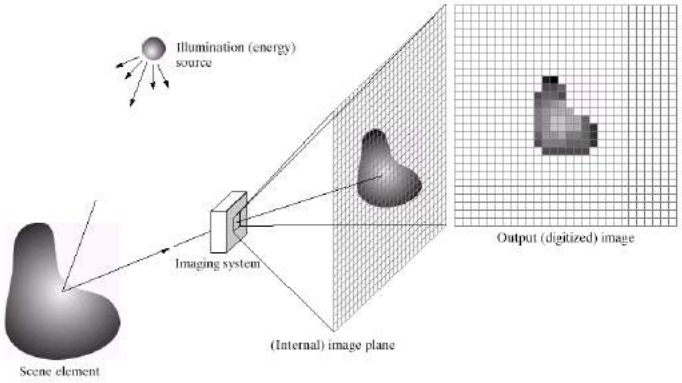
Common in Digital Camera
Charged-Couple Device (CCD)





Digital Image Acquisition

Sampling and Quantization



a b c d e

FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

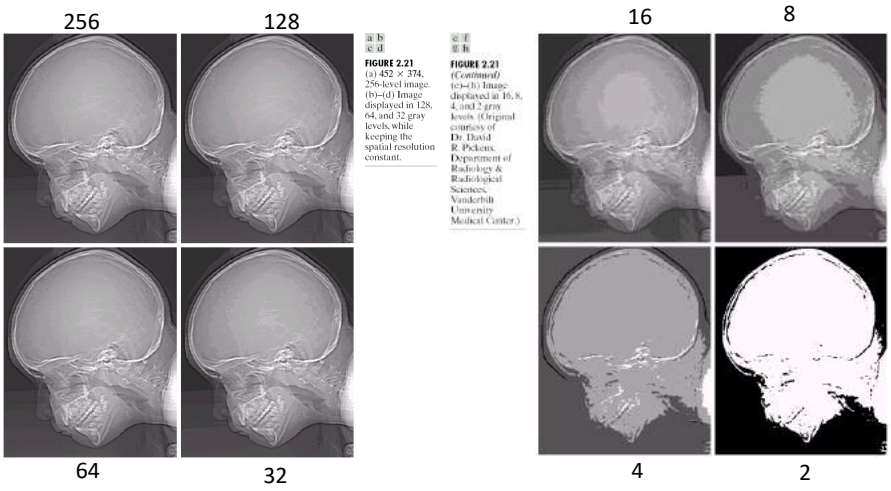
Introduction to the Mathematics of Image and Data Analysis by Gilad Lerman@umn.edu

Problem: Effect of Compression

- Color image of 600x800 pixels
  - Without compression
    - (600*800 pixels) * (24 bits/pixel)
      = 11.52M bits = 1.44M bytes
  - After JPEG compression (popularly used on web)
    - only 89K bytes
    - compression ratio ~ 16:1
- Movie
  - 720x480 per frame,
  - 30 frames/sec,
  - 24 bits/pixel
  - Raw video ~ 243M bits/sec
  - DVD ~ about 5M bits/sec
  - Compression ratio ~ 48:1

"Library of Congress" by M.Wu (600x800)
Based on slides by W. Trappe

Problem: Effect of Quantization



a b
c d

FIGURE 2.21 (a) 452 × 374, 256-level image. (b)–(d) Image displayed in 128, 64, and 32 gray levels, while keeping the spatial resolution constant.

e f
g h

FIGURE 2.21 *(Continued)* (e)–(h) Image displayed in 16, 8, 4, and 2 gray levels. (Original courtesy of Dr. David R. Pickens, Department of Radiology & Radiological Sciences, Vanderbilt University Medical Center.)

Introduction to the Mathematics of Image and Data Analysis by Gilad Lerman@umn.edu

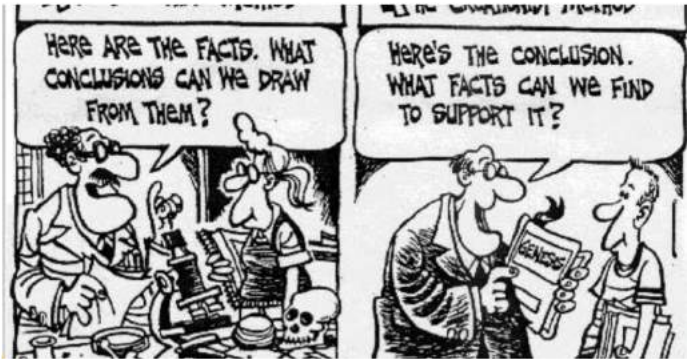
Data Science and Social Impact

Decision Making Biases

Selective Perception & A specific case

- Selective Perception
  - See things from our own personal perspective
  - Organize and interpret events/information based on this perception
  - Influences what we pay attention to and the problems we identify, and the alternatives we develop or consider.

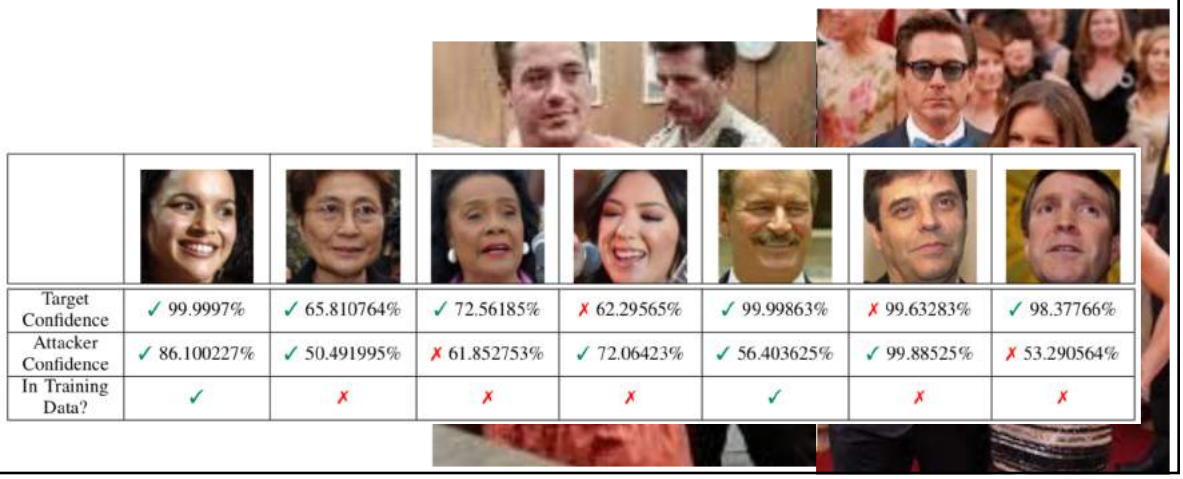
Confirmation Bias:



Stacey Truex and Ling Liu. Countering Membership Inference Attacks with Differentially Private DNN Training

Overconfidence Bias

- When we are given factual questions and asked to judge the probability that our answers are correct, we tend to be far too optimistic.



Target Confidence	✓ 99.9997%	✓ 65.810764%	✓ 72.56185%	✗ 62.29565%	✓ 99.99863%	✗ 99.63283%	✓ 98.37766%
Attacker Confidence	✓ 86.100227%	✓ 50.491995%	✗ 61.852753%	✓ 72.06423%	✓ 56.403625%	✓ 99.88525%	✗ 53.290564%
In Training Data?	✓	✗	✗	✗	✓	✗	✗

Most Importantly...

BE AWARE!

Neither human nor machine can escape these biases but if you are **aware of them** and **challenge them**, you will be become a stronger critical thinker and better decision maker overall!

Mathematics + Social Impact

Two sides of the big data sword