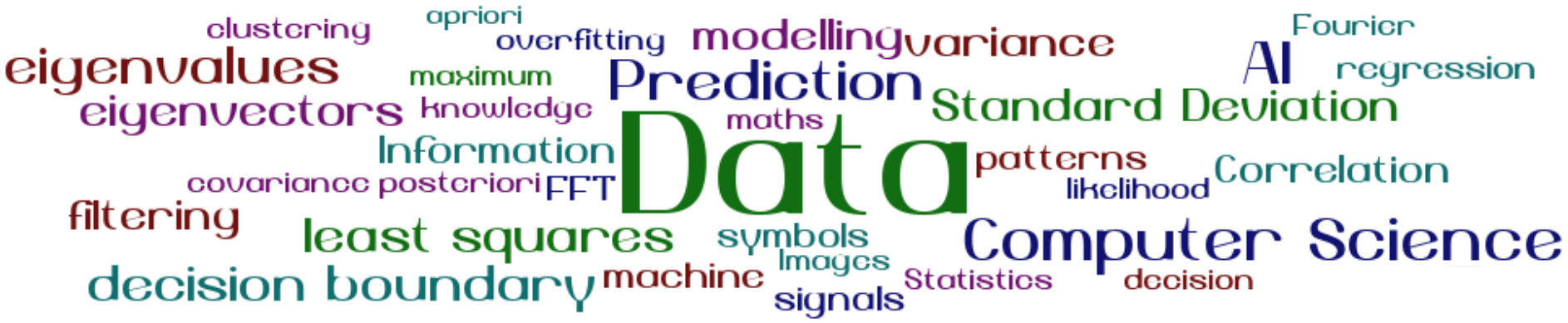
COMS20011 – Data-Driven Computer Science



February 2021

Majid Mirmehdi

Some slides in this lecture are adapted from those authored by **Dima Damen** and **Andrew Calway**

Lecture Video #1

COMS20011 Unit – born this year

- This is a brand new unit in 2020-21 academic year
- Replaces the 20CP COMS20212 (SPS) unit
- Exam materials can be used for revision BUT...
- Use SPS materials with caution...depth, breadth & requirements may differ.

What is Data?

- Data comes in many forms, e.g. symbols, patterns and signals!
- Data: *Structured and Unstructured*
  - Numeric (measurements, finance spreadsheets, ...)
  - Textual (emails, social media, web pages, medical records, ...)
  - Visual (images, video, graphics, animations)
  - Auditory (speech, audio)
  - Signals (GPS signals, accelerometer, heart rate, ...)
  - Many others...

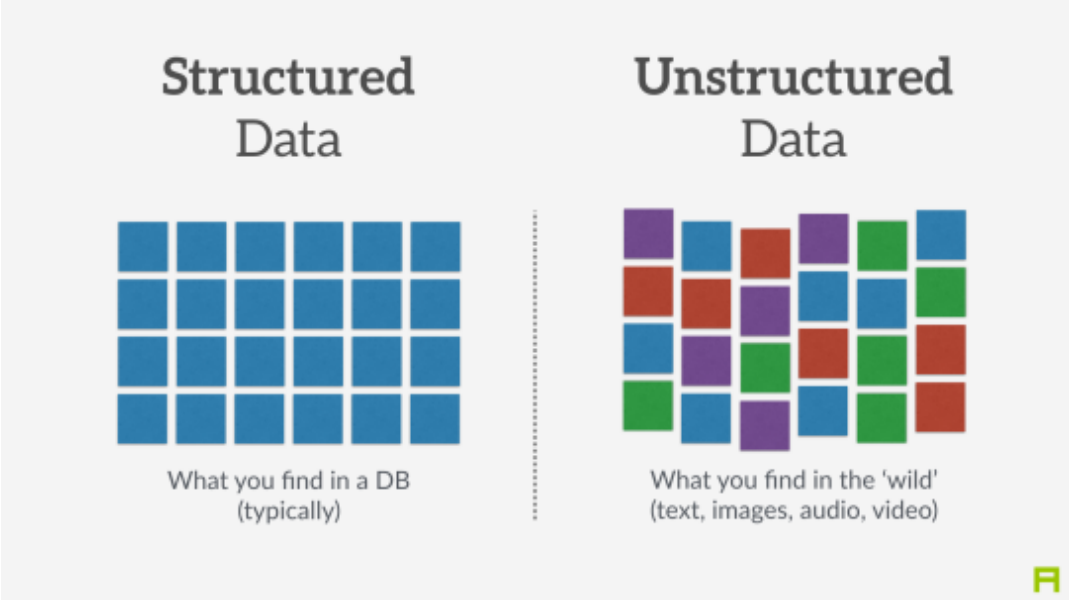


Image from Garrett Hollander

This Unit (adapted from COMS20212: Symbols, Patterns and Symbols)

- This unit is about doing things with data... *but not*
  - storing, shuffling, searching (Algorithms I & II)
  - sending (Computer Systems)
  - compressing or encrypting (Cryptology)

- This unit is about:
  - extracting knowledge from data
  - generating data and making predictions
  - making decisions based on data
  - Often referred to as:

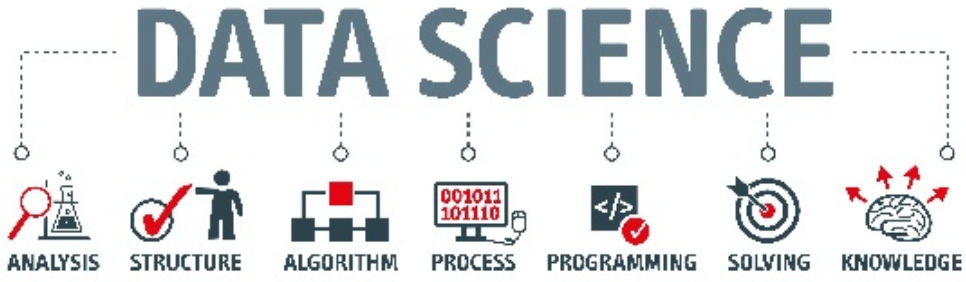


Image from https://www.datanami.com

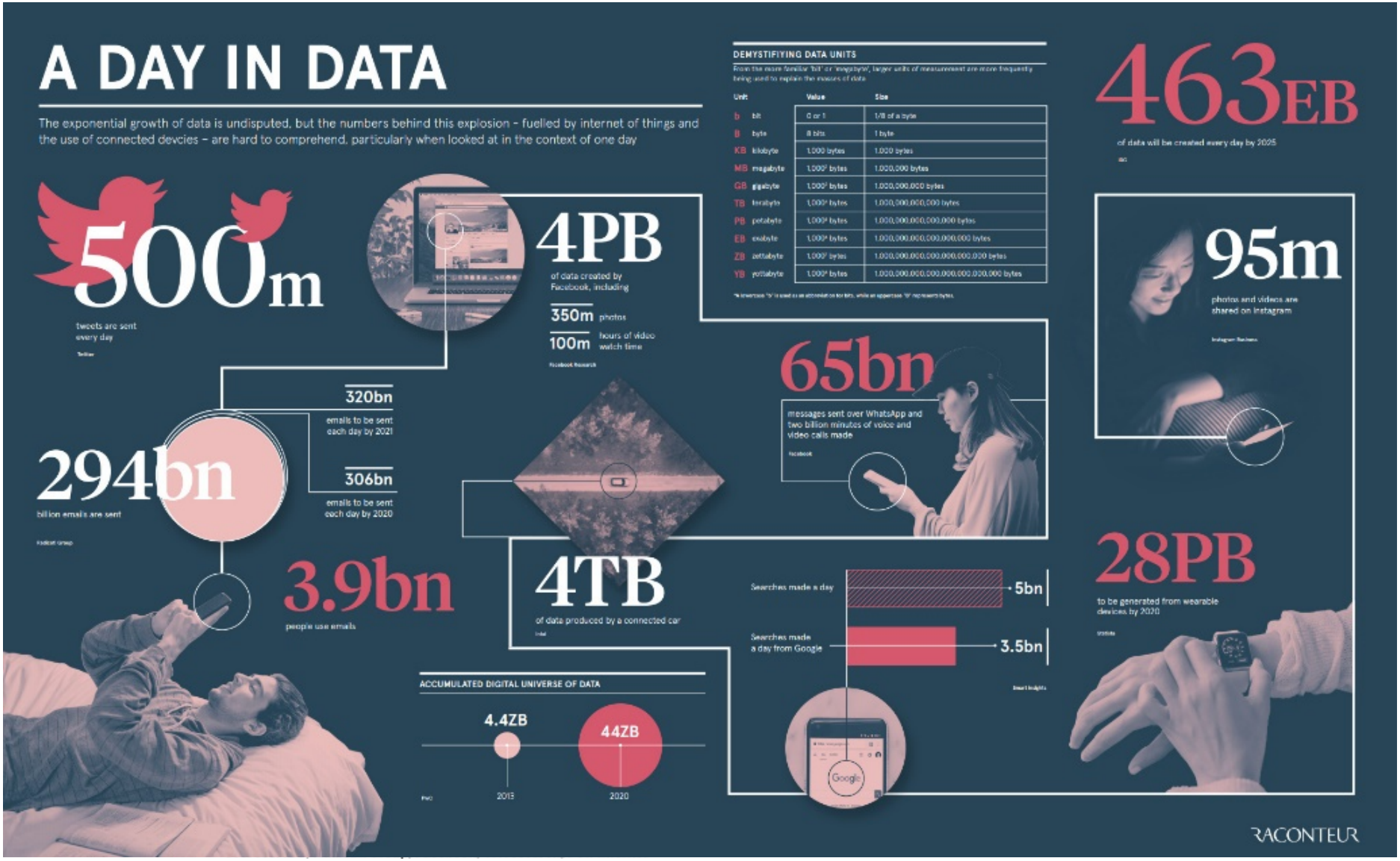
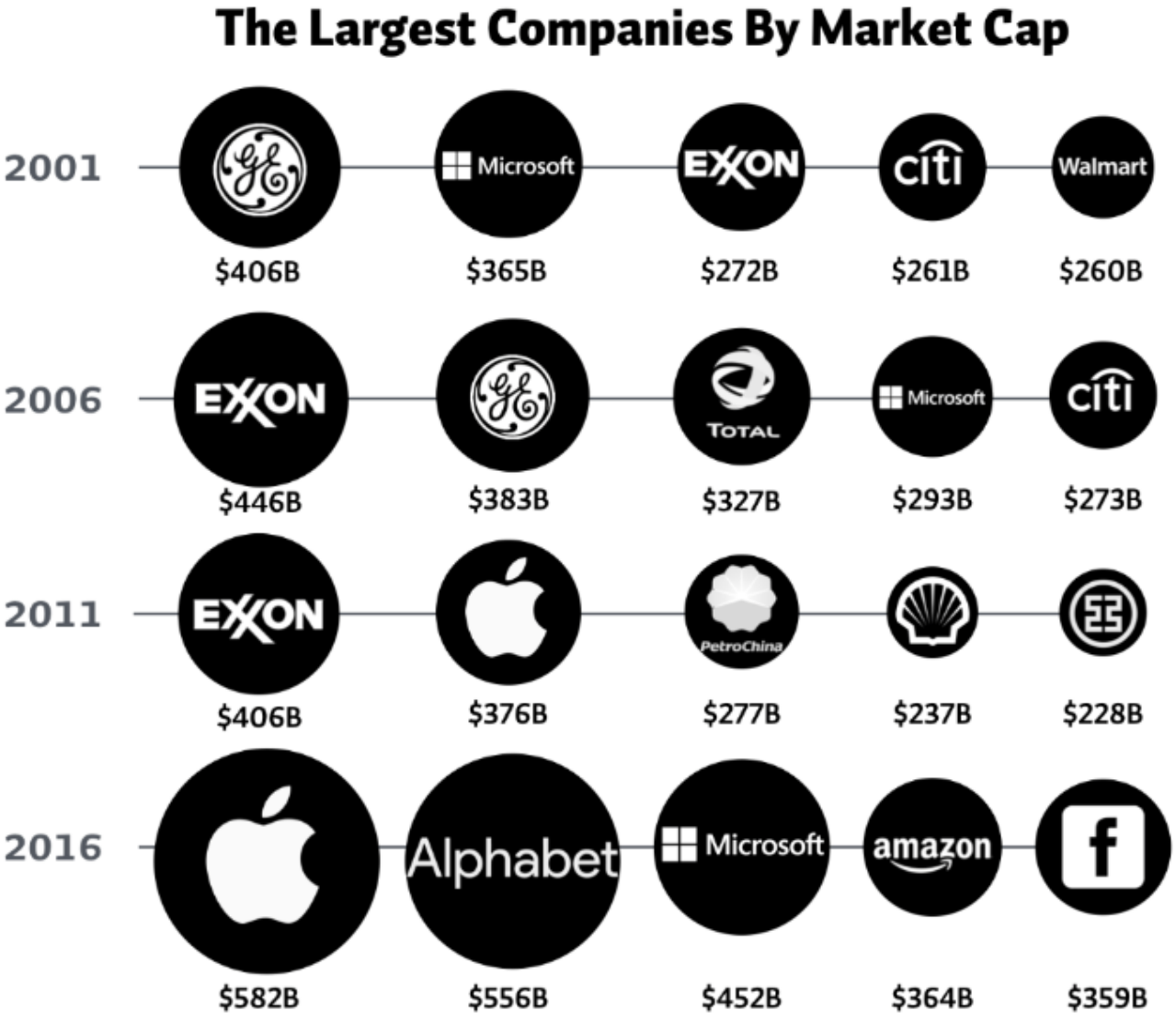
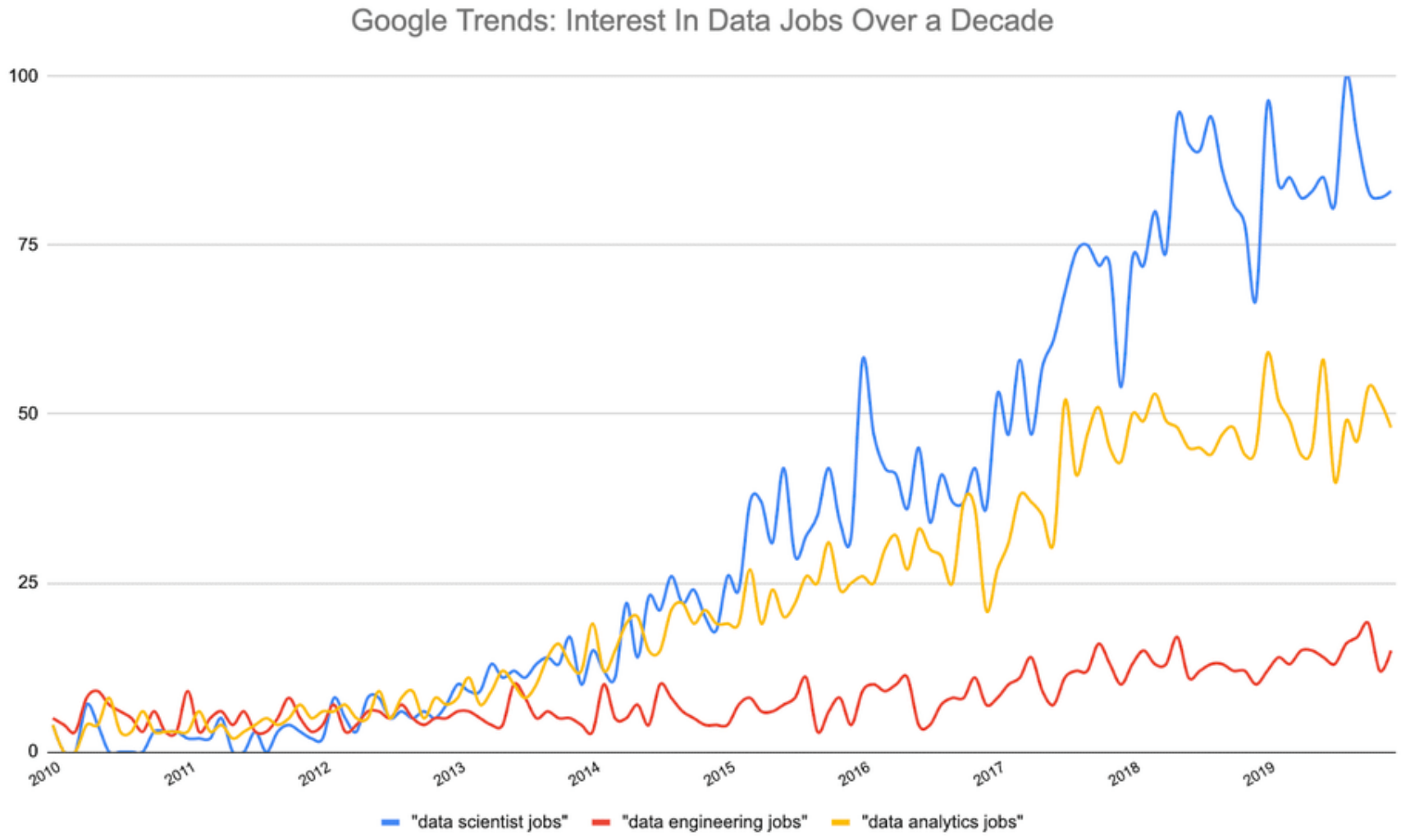


Image from Raconteur https://www.weforum.org

Data is the new Oil



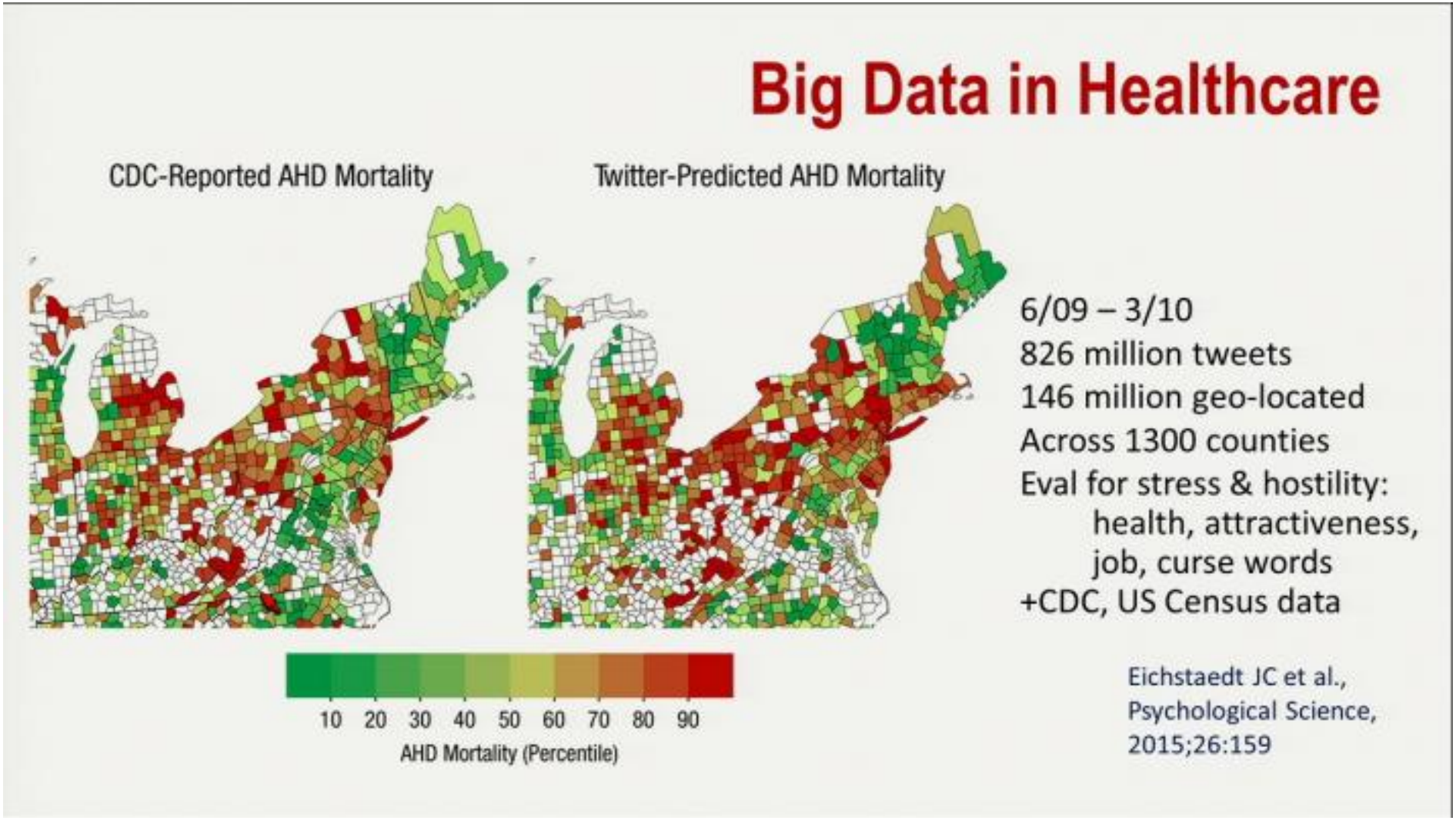
Data Science & Analytics



https://onlinedatasciencemasters.virginia.edu/blog/data-science-vs-data-engineering/

But it's not about the data – it's about the science

Tracking and predicting [disease,mortality,floods,fires, and fun etc.] by Twitter!

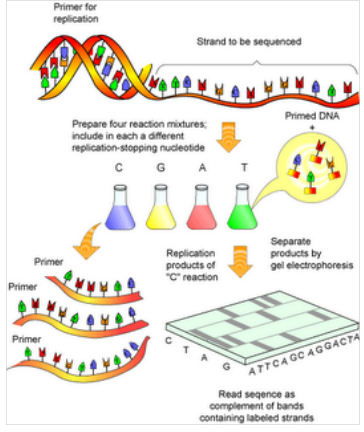


https://www.dicardiology.com/article/understanding-how-big-data-will-change-healthcare

This Unit

Why is it important for Computer Science?

- Fundamental to many application areas:
  - Artificial Intelligence, Machine Learning, Deep Learning
  - Image Processing and Pattern Recognition
  - Graphics, Animation and Virtual Reality
  - Computer Vision and Robotics
  - Speech and Audio Processing.
  - With growing applications in: neuroscience, literature, agriculture, etc.
- Hence, preparation for application units in years 3 and 4.

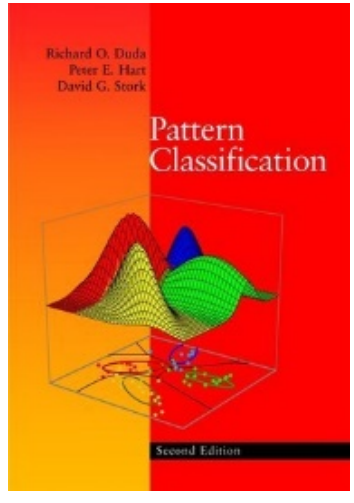


Ex1. A Fishy Problem

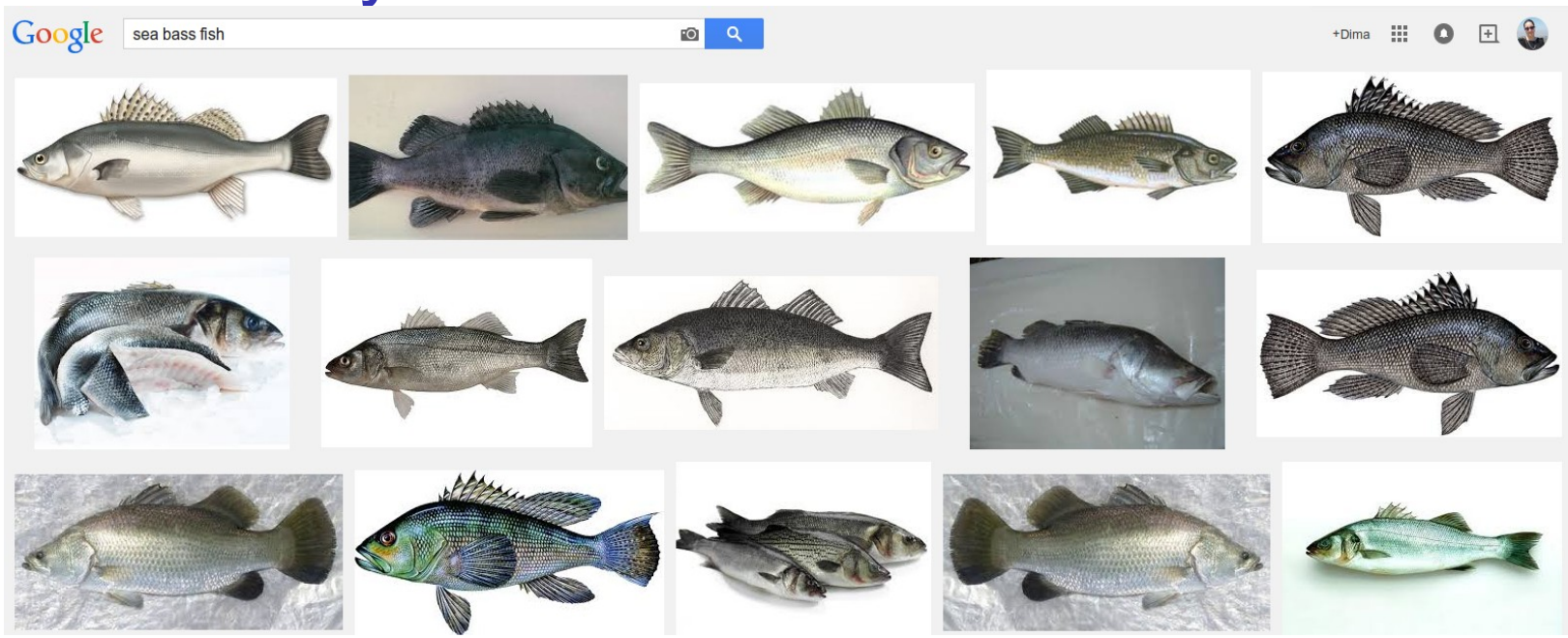
Data: images of fish

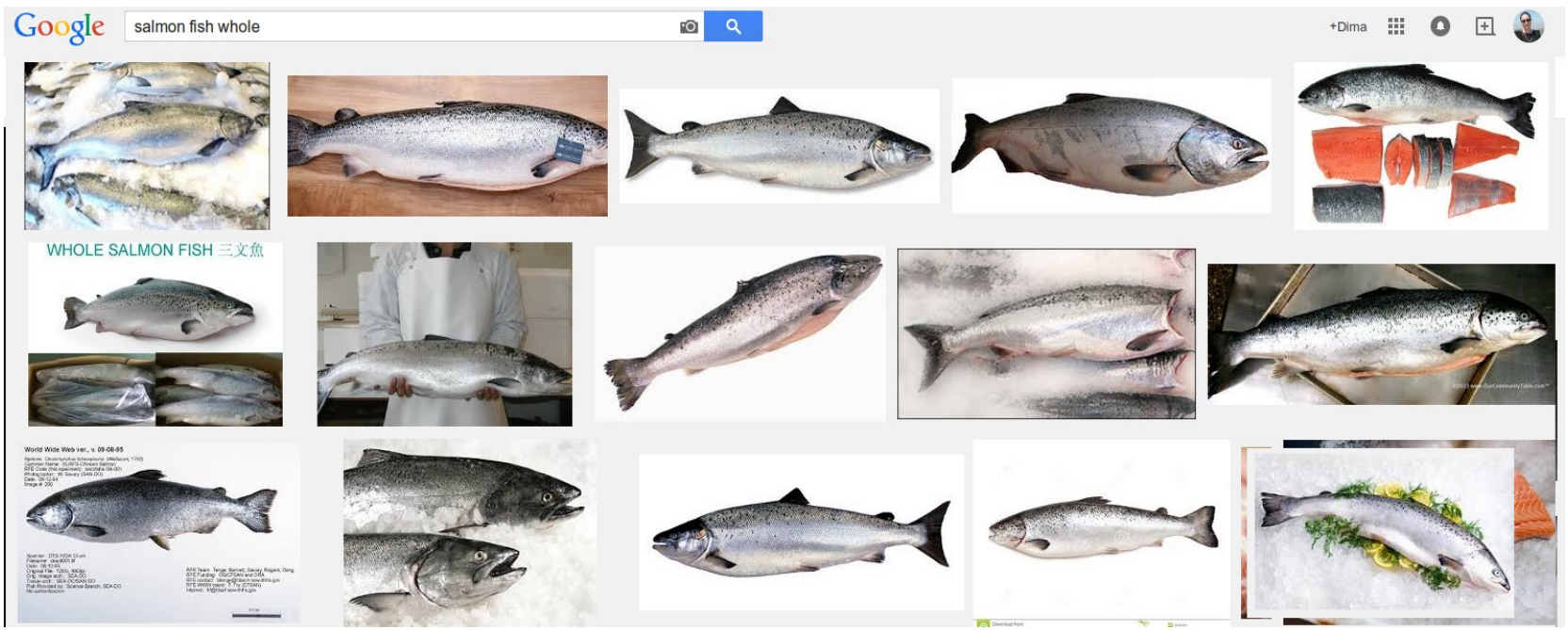
Aim: distinguish between sea bass and salmon

From: Pattern Classification by *Duda, Hart and Stork*, 2nd Edition, Wiley Interscience



Ex1. A Fishy Problem





Fishing for a Solution

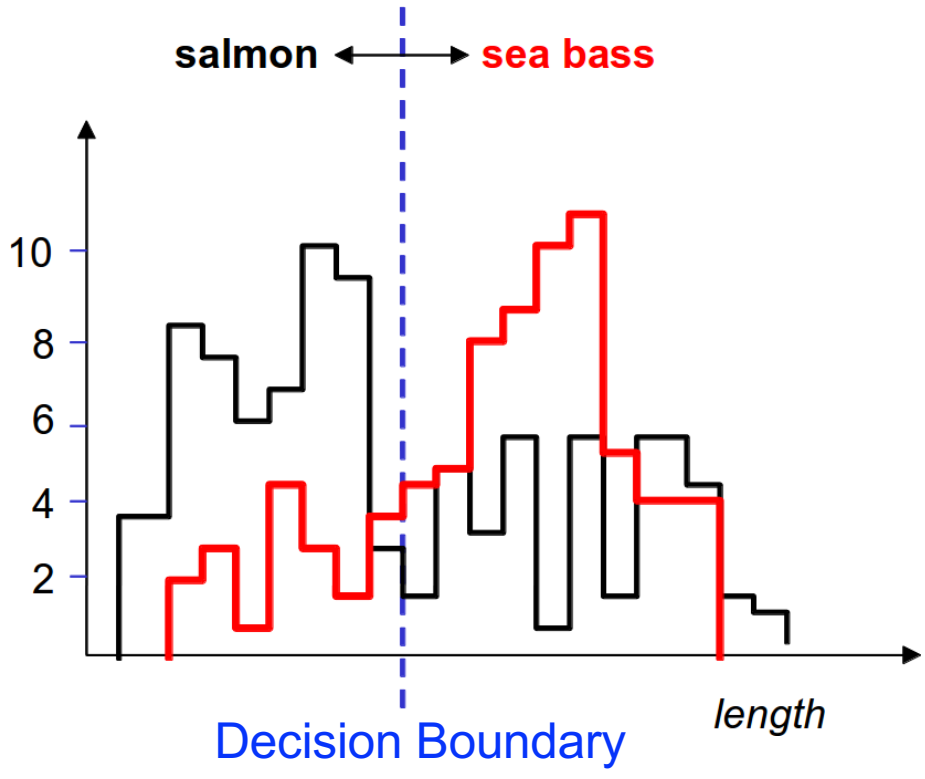
Steps:

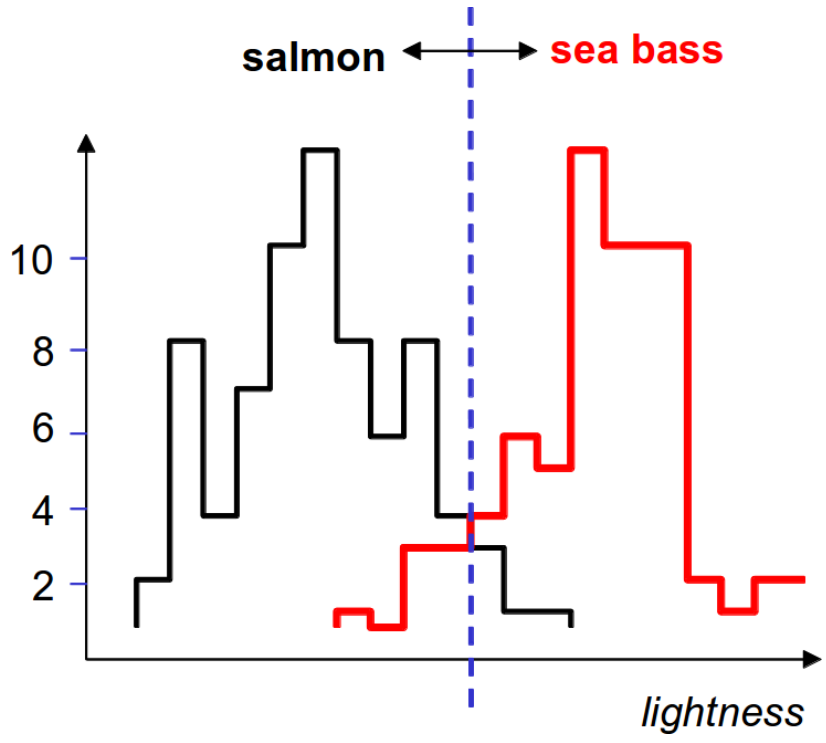
1. Pre-processing e.g. Rotate and align, Segment fish from background
2. Feature Selection e.g. Measure length
3. Classification e.g. Find a threshold

Fishing for a Solution

Steps:

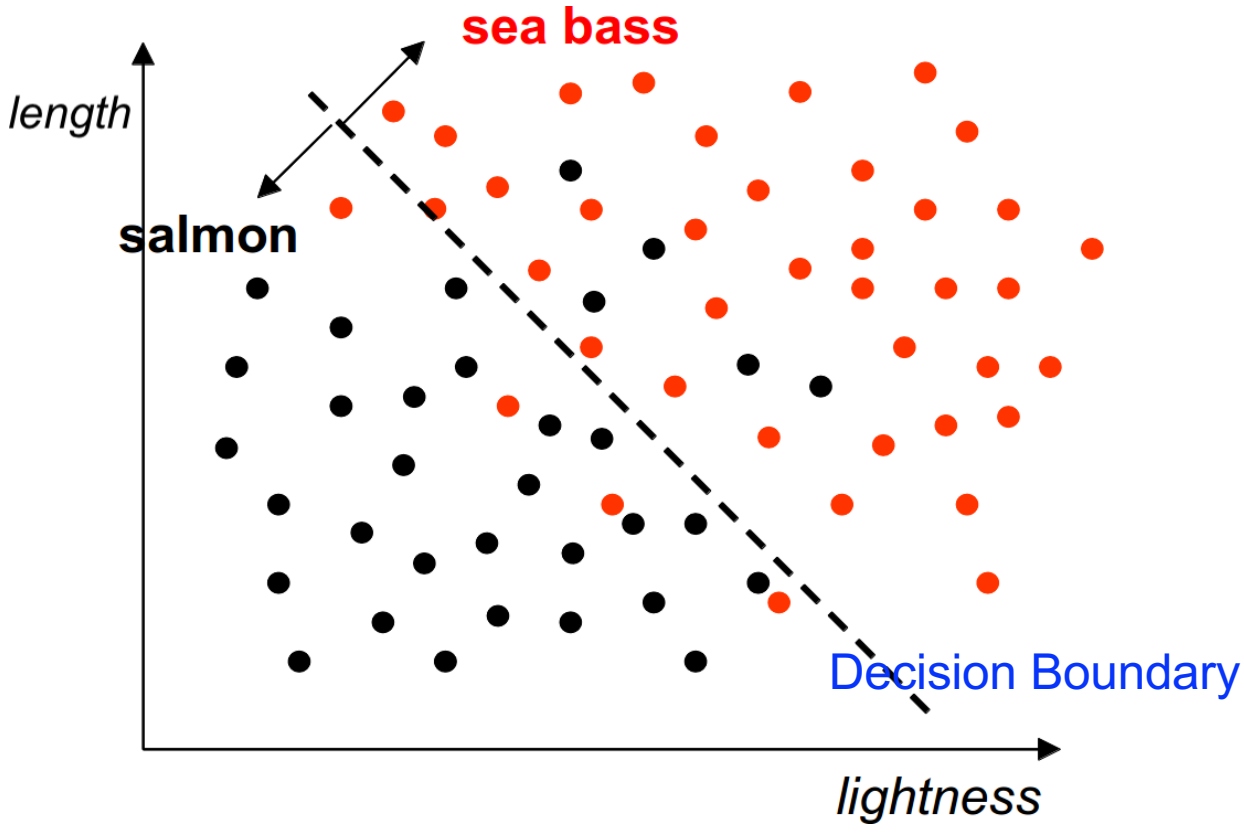
1. Pre-processing e.g. Rotate and align, Segment fish from background
2. Feature Selection e.g. Measure length or lightness
3. Classification e.g. Find a threshold





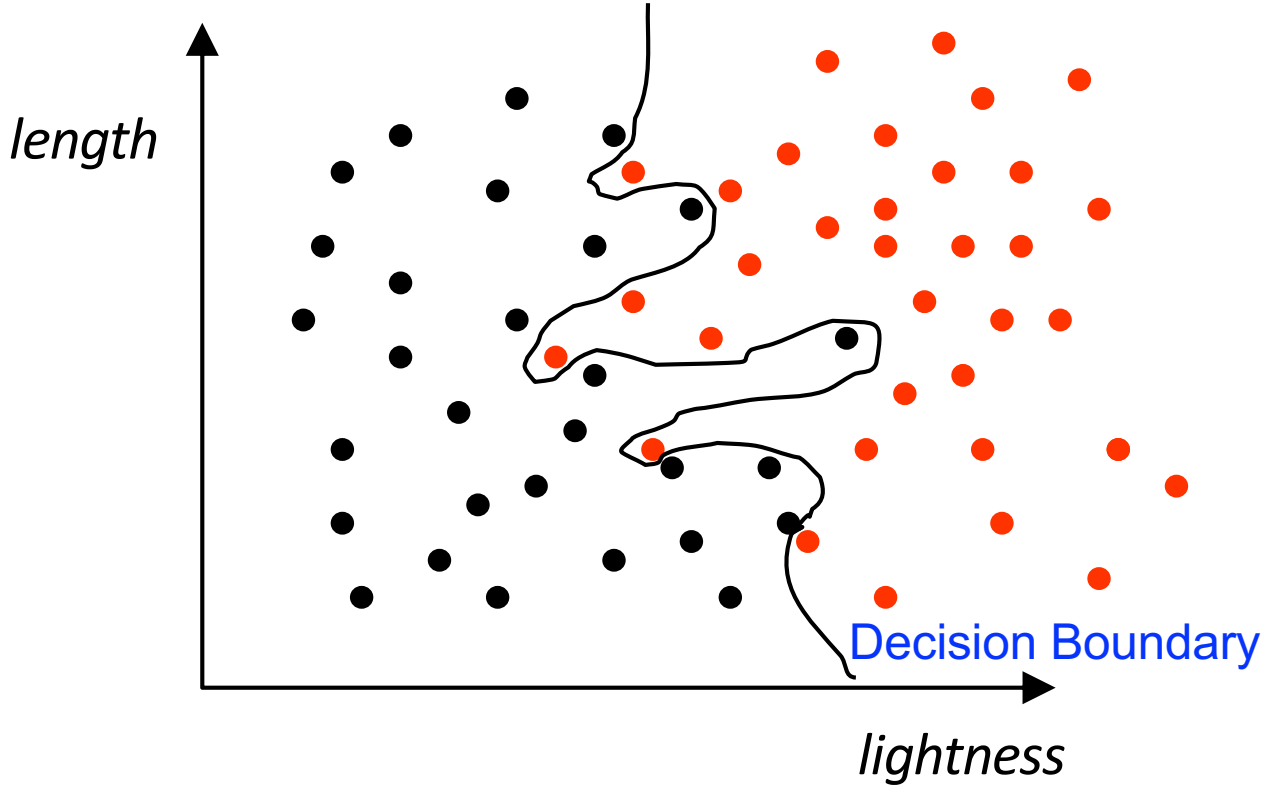
Fishing for a Solution

Multiple features could be selected, resulting in a multi-dimensional feature vector.



Fishing for a Solution

Complex decision model



Typical Data Analysis Problem

Steps:

1. Pre-processing [Unit - Part 1] → Majid Mirmehdi (~10%)
2. Feature Selection [Unit - Part 3] → Majid Mirmehdi (~40%)
3. Modelling & Classification [Unit - Part 2] → Laurence Aitchison **[UD]** (~50%)

Next Video

More example applications...