Welcome to Applied Data Science
Let's turn data into stories!

Teaching Unit: COMS30050
Assessment Unit: COMS30051 (Year 3); COMSM0055 (Year 4)

Nirav Ajmeri and Seth Bullock

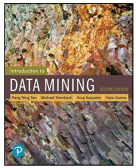
School of Computer Science
University of Bristol

TB2 2025/26

Outline

Resources to look at . . .

- Book: Doing Data Science, Cathy O'Neil and Rachel Schutt (Ch. 1 and 2)[a]
- eBook: Data Science: An Introduction, wikibooks[b]
- Web: Kdnuggets, Kaggle, Data Science Central
- Book: Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar[c]



[a] eBook available through the Library
[b] `https://en.wikibooks.org/wiki/Data_Science:_An_Introduction`
[c] copy available in the Library

- Meetups: PyData, Data Science & ML Study Group, Tech Ethics, AI West

PyData Bristol

BRISTOL DATA SCIENCE AND MACHINE LEARNING STUDY GROUP

- What is data science?
- What will you learn?

Data

Can you think of examples of structured or unstructured data?

- Numerical, categorical, or binary
- Text: emails, ~~tweets~~ skeet, articles
- Records: user-level data, timestamped event data, log files
- Geo-based location data
- Network data
- Sensor data
- Image and video
- Audio and music

Big Data

Numbers...

- 376 Billion – The number of e-mail messages sent each day; up to 80% are spam
- Over 500 – Hours of video were uploaded to YouTube every minute, resulting in nearly 3.4 years of content every hour
- 1 Billion – Stories posted on Meta platforms every day
- 90% – Percentage of the world's data created in the last 2 years

Big Data and Small Data

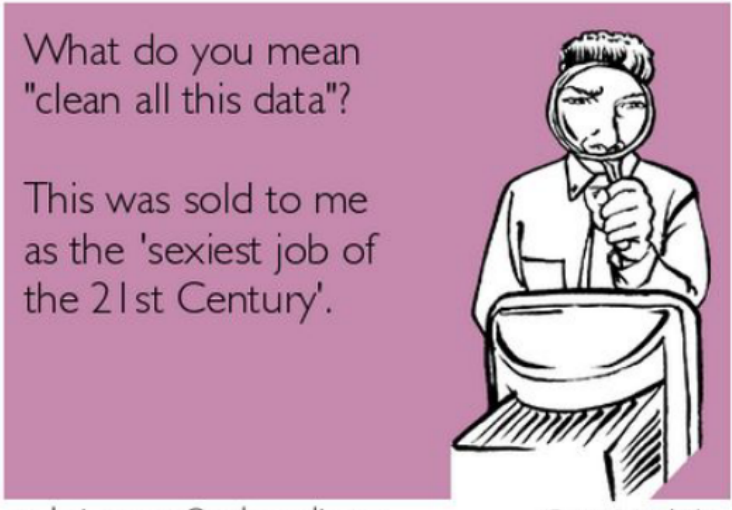
What is **large**?

- Depends on the context

Large text dataset?

1967 1,000,000 words

2006 1,000,000,000,000 words



[1]https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

What is Data Science?

A definition... by *Zico Kolter, Professor, CMU*

Data science is the application of computational and statistical techniques to address or gain insight into some problem in the real world.

- "computational"... involves some sort of algorithmic methods written in code,
- "statistical"... statistical inference lets us build the predictions that we make, and
- "real world"... deriving insight not into some artificial process, but into some "truth" in the real world

Data science = statistics + data collection + data preprocessing + machine learning + visualisation + business insights + scientific hypotheses + big data + (etc)

What is Data Science?

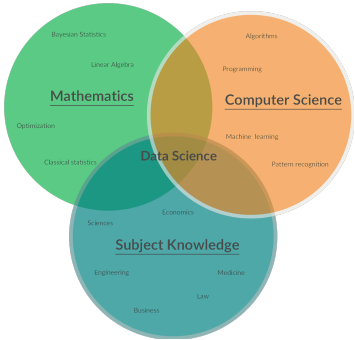
Broad...

Data Science is broad ... not a single discipline; encompasses several disciplines

What is Data Science?

Interdisciplinary. . .

"Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data. These insights can be used to guide decision making and strategic planning."[2]



[2] IBM: What is Data Science? https://www.ibm.com/topics/data-science

[3] Data Science by Megan Kress on Creatly.com
https://i.pinimg.com/originals/fc/61/40/fc61405983965a5ce1bea0e8a353d6f2.png

What is Data Science?

Insight-focussed. . .

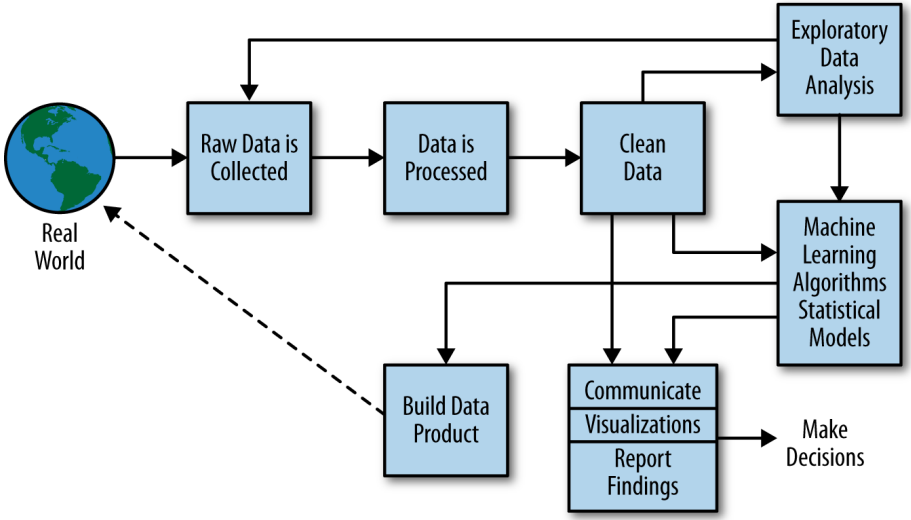
. . . finding insights in data and leveraging them to make an informed decision

Applications of Data Science

Applications are across domain...

- Retail: Customer segmentation, personalised recommendation, market analysis ...
- Finance: Risk assessment, fraud detection, algorithmic trading ...
- Healthcare: Disease prediction, drug discovery, medical image analysis ...
- Food: Waste reduction, demand forecasting ...
- Agriculture: Soil health prediction, precision farming, pest detection ...
- Energy and Logistics: Demand forecasting, customer behaviour analysis ...
- ...

How do we do data science?



[4] Doing data science, Cathy O'Neil and Rachel Schutt

Summary. . . so far. . .

- Applications of Data Science: high-impact, diverse
- Challenges: computational/information complexity

Course Structure

The unit is split in two halves: Learn and Apply

Weeks 13–17. Learn to do data science

- Face-to-face learning via lectures and labs

Weeks 19–23. Apply the learning of data science

- Supervised group coursework

Week	Topic	Lecturers	Tuesday Lecture (Queens 1.15 SLT; 9am–11am)	Thursday Lab (Queens 1.80; 9am–11am)
13 (w/c 19-Jan)	Introduction and Data Ingress	Nirav	Welcome; Data Ingress; Web Scraping	Web Scraping with Beautiful Soup and REST API
14 (w/c 26-Jan)	Data Ethics and Data Privacy	Nirav	Data Privacy; Data Ethics	Familiarising with coursework datasets and exploring data privacy and ethics
15 (w/c 2-Feb)	Data Management and Data Wrangling	Seth	Data Management; Data Wrangling	Graph database and Neo4j
16 (w/c 9-Feb)	Data Fusion and Data Exploration	Seth	Data Exploration; Data Fusion	Dimensionality Reduction Lab
17 (w/c 16-Feb)	Data Visualisation and Data Science in Production	Nirav; Seth	Data Visualisation; Data Science in Production	Visualisation with Matplotlib, Seaborn, and Plotly.js Lab
18 (w/c 23-Feb)		Computer Science Consolidation Week		

Week	Topic	Lecturers	Tuesday Show and Tell and Q&A (Queens 1.15 SLT; 10am–11am)	Thursday Supervision (Various Meeting Rooms; 9am–11am)
19 (w/c 2-Mar)	CW Starts	Nirav; Seth	CW Intro and Q&A	Supervision Meeting
20 (w/c 9-Mar)	CW Continues	Guest Speaker	Show and Tell	Supervision Meeting
21 (w/c 16-Mar)	CW Continues	Nirav; Seth	CW Q&A	Supervision Meeting
(w/c 23-Mar)	**SPRING VACATION**			
(w/c 30-Mar)	**SPRING VACATION**			
(w/c 6-Apr)	**SPRING VACATION**			
22 (w/c 13-Apr)	CW Continues	Guest Speaker	Show and Tell	**CW Presentations** (IVY GATE BLDG 1.01) Thursday 16-Apr, 9am–12pm
23 (w/c 20-Apr)	Supervision Ends	Nirav; Seth	Coursework Q&A	Supervision Meeting (Queens 1.80)
24 (w/c 27-Apr)	**COURSEWORK DUE** *1PM on Tuesday, 28-Apr*			

Lectures and Labs

Weeks 13–17: Before Computer Science Consolidation Week

- Introduction and Data Ingress
- Data Ethics and Data Privacy
- Data Management and Data Wrangling
- Data Fusion and Data Exploration
- Data Visualisation and Data Science in Production

Group Coursework

Weeks 19–23: After Computer Science Consolidation Week

- Supervised group project (5–6 students, same year across programmes)
- Timeline
  - Week 14: Sample list of datasets released before the lab
  - Week 15: Complete list of datasets available
  - Week 16–17: Team registration
  - Week 19–23: Project execution

Deliverables:

- Coursework "work-in-progress" presentation (*9AM–12PM on Thu, 16 Apr 2026*)
- Group report and individual reflection (*Due: 1PM on Tue, 28 Apr 2026*)

People: Teaching Team

Nirav Ajmeri

Seth Bullock

People: Teaching Assistants

- Omar Emara
- Kal Roberts
- Panagiotis Soustas
- Prajwal Gatti
- Saptarshi Sinha
- Siddhant Bansal
- Tianye Wang

People: Diverse Group of Students!

- Year 3 and Year 4
  - Computer Science
  - Computer Science with Innovation
  - Mathematics & Computer Science
  - Engineering Mathematics
- Incoming Study Abroad Students

Questions or Concerns?

Coursework Team Composition

- Year 3 students work with other Year 3 students
- Year 4 students work with other Year 4 students
- Group of 6 members; maximum of 4 members from one set

Set A	Set B
Computer Science	Engineering Mathematics
Computer Science and Electronics	Engineering Mathematics with ...
Computer Science with Innovation	...
Mathematics and Computer Science	...