

DATA201

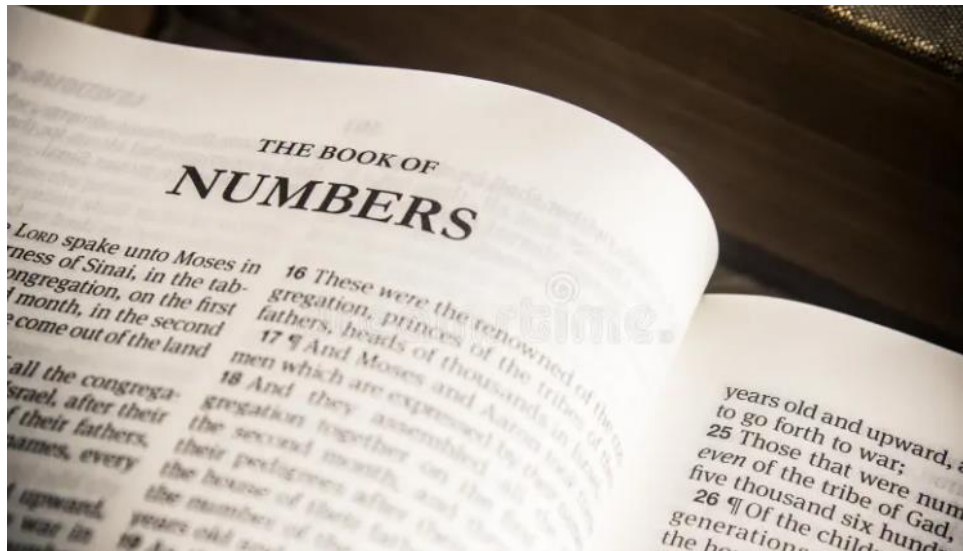
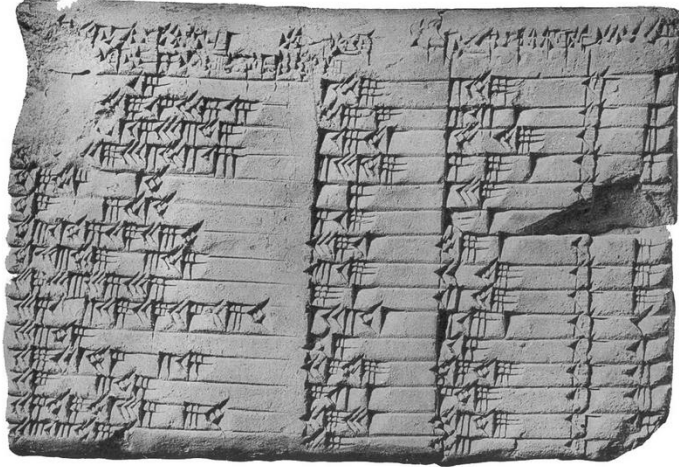
Week 11 - Lecture 1

What is data?

DATA201 – Practical Data Science

- In this course, we have focused on teaching you the practical skills for organising and wrangling data so you can use it for analysis, visualisation, modelling ++
- There are further considerations that you should understand as a “data practitioner” (statistician, data analyst, scientist, engineer etc.)
- Let’s start by understanding what data is...

Data Practices are Inherently Human



OBSERVATIONS

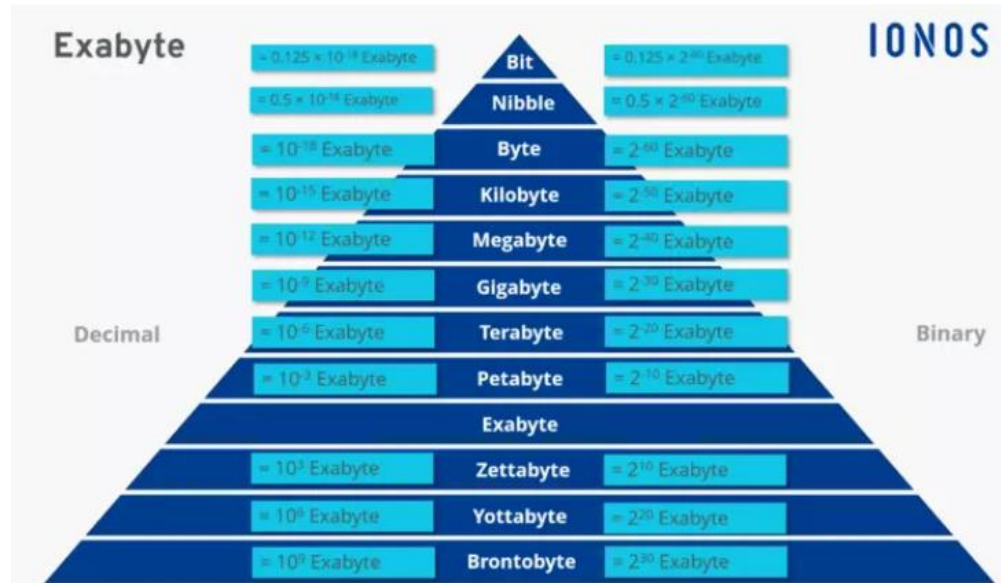
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

By JOHN GRAUNT,
Citizen of
LONDON.

With reference to the Government, Religion, Trade,
Growth, Age, Diseases, and the several Changes of the
said CITY.

— Non, ne ut mirare Turis, labore.
Continetur parvis Libellus —

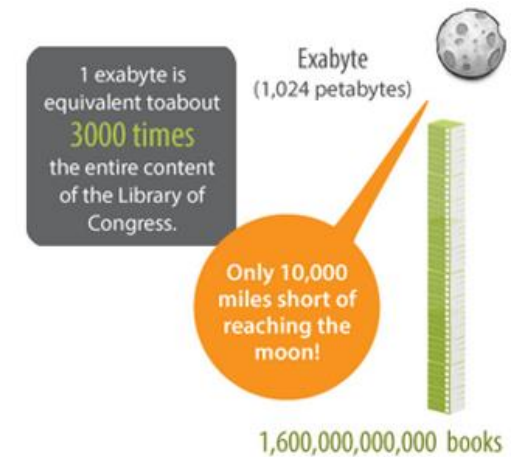
Data Scale Reflects Our Modern World



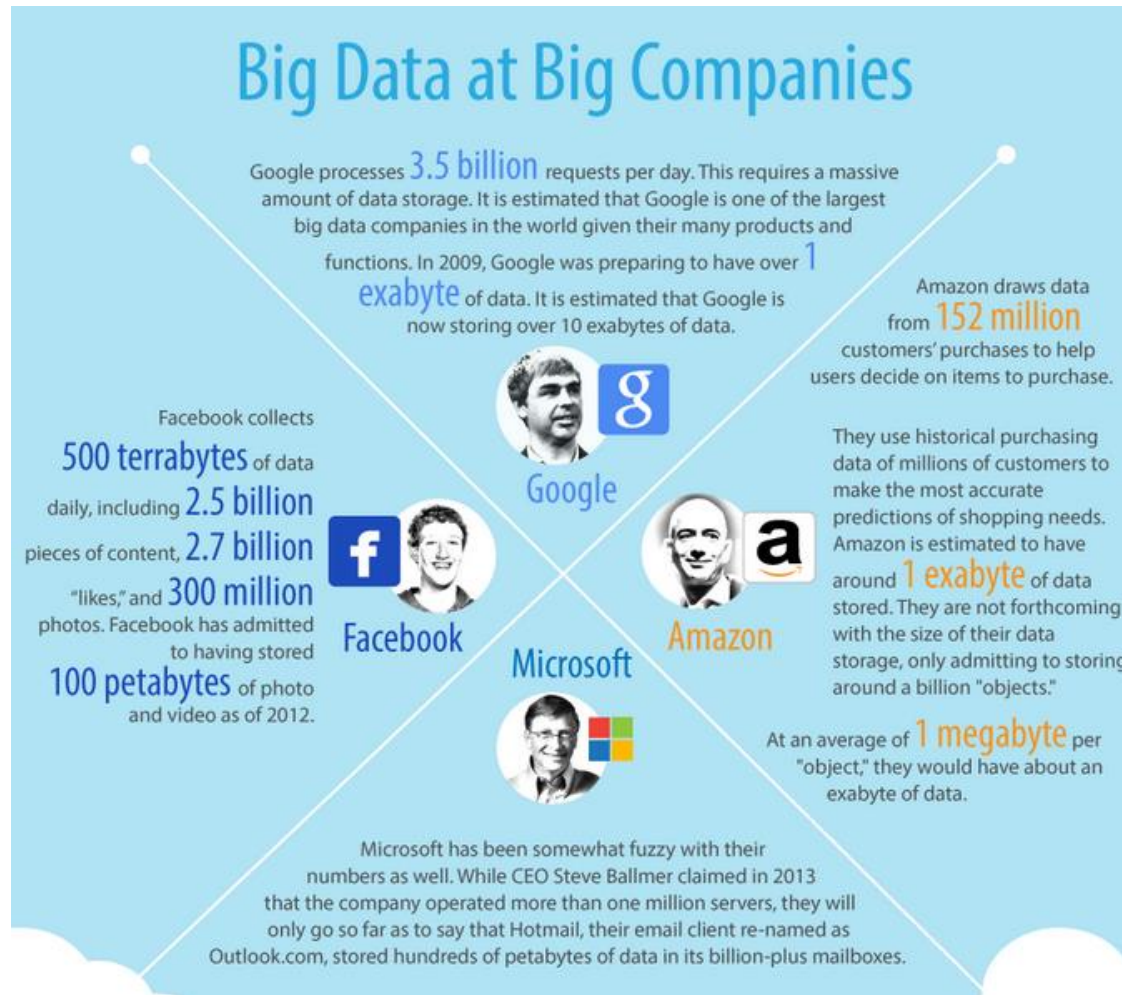
Unit of Measurement	Data Generated
Zettabytes	0.33
Exabytes	328.77
Petabytes	328,767.12
Terabytes	328.77 million
Gigabytes	328.77 billion
Megabytes	328.77 trillion
Kilobytes	328.77 quadrillion
Bytes	328.77 quintillion

Top Data Created Stats

- Approximately **328.77 million terabytes** of data are created each day
- Around **120 zettabytes** of data will be generated this year
- **181 zettabytes** of data will be generated in 2025
- Videos account for **over half** of internet data traffic
- The US has **over 2,700** data centers



Data is BIG Business!



Rank	Name	Market Cap	Price	Today	Price (30 days)	Country
1	Apple AAPL	\$2.796 T	\$177.79	▲ 0.03%		USA
2	Microsoft MSFT	\$2.386 T	\$321.01	▼ 0.59%		USA
3	Saudi Aramco 2222.SR	\$2.193 T	\$9.07	▲ 2.56%		S. Arabia
4	Alphabet (Google) GOOG	\$1.650 T	\$130.17	▼ 0.03%		USA
5	Amazon AMZN	\$1.420 T	\$138.41	▼ 0.11%		USA
6	NVIDIA NVDA	\$1.009 T	\$408.55	▼ 3.62%		USA
7	Berkshire Hathaway BRK-B	\$783.09 B	\$358.35	▲ 0.38%		USA
8	Meta Platforms (Facebook) META	\$773.02 B	\$301.64	▼ 1.34%		USA
9	Tesla TSLA	\$769.07 B	\$242.65	▼ 1.10%		USA
10	Visa V	\$502.73 B	\$240.04	▼ 0.05%		USA



THE ECONOMIST – THE WORLD'S MOST VALUABLE RESOURCE IS NO LONGER OIL, BUT DATA

Data Drives New Technologies



Data Drives Decision-Making



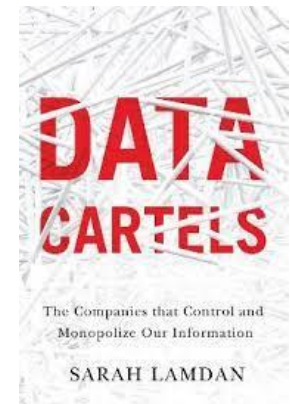
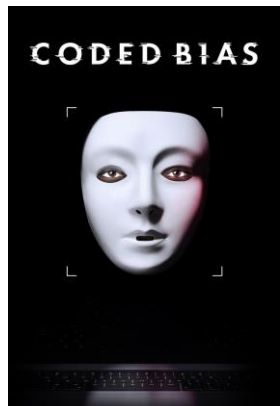
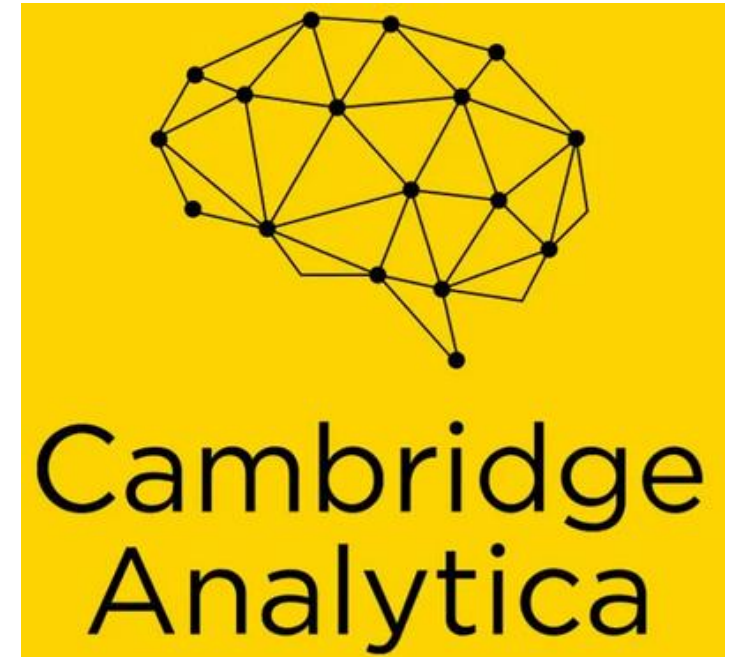
Data Can Cause and Perpetuate Injustices



Source: ProPublica



Machine Learning | CMU - Carnegie Mellon University
Amazon Scraps Secret AI Recruiting ...



So...What is Data?

- Data – what is it in DATAX201 (and other DATAX courses)
 - A collection of numerical or non-numerical observations with some contextual meaning
 - Pulse rates, preferred president polls, opinions on a referendum
 - Digital media – photos, videos, reports, snapchats, tiktoks, blogs etc.

So...What is Data?

- Data – what is it in DATAX201 (and other DATAX courses)
 - A collection of numerical or non-numerical observations with some contextual meaning
 - Mussels (A1), light fixtures (A2), political polls, etc.
 - Digital media – photos, videos, reports, snapchats, tiktoks, blogs etc.
- Data is **Power**!
 - Data is the raw material – data output (e.g statistics) is knowledge gained from the information
 - Having the ability to understand data, its strengths and limitations, and having the ability to extract knowledge from data is power!

Data is Power!

We sit in a powerful and privileged position as data practitioners!

- As data practitioners, we need to understand
 - Data is powerful and used for big decisions – therefore we need to ensure integrity in all aspects of the data ‘process’, from collection to analysis and interpretation
 - Data is power, and we must ensure that we are responsible to those who the data is about, and those who have a stake in the data
 - Striking the balance! Maximising benefits, minimising harms.