# DATAX201 Week 11 - Lecture 1

What is data?

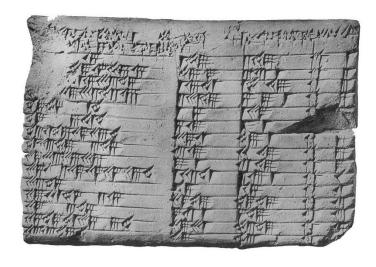
#### DATAX201 — Practical Data Science

 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 ++

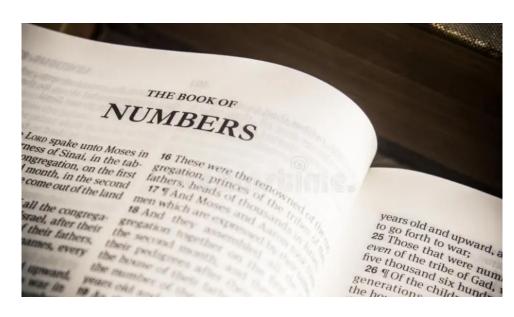
• There are further considerations that you should understand as a "data practitioner" (statistician, data analyst, scientist, engineer etc.)

Let's start by understanding what data is...

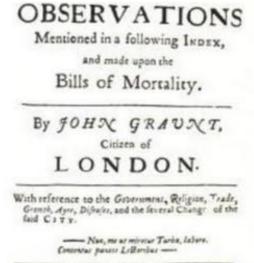
### Data Practices are Inherently Human



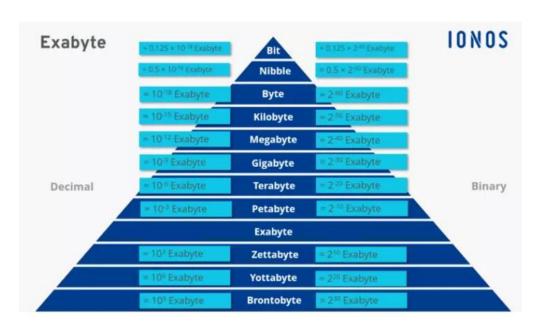








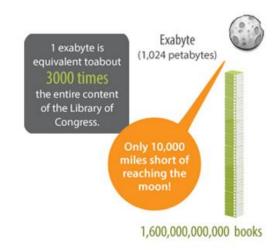
#### Data Scale Reflects Our Modern World



Top Data	Created	<b>Stats</b>
----------	---------	--------------

- 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

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



#### Data is **BIG** Business!

#### Big Data at Big Companies

Google processes 3.5 billion requests per day. This requires a massive amount of data storage. It is estimated that Google is one of the largest big data companies in the world given their many products and

functions. In 2009, Google was preparing to have over

**exabyte** of data. It is estimated that Google is now storing over 10 exabytes of data.

Facebook collects

and video as of 2012.

500 terrabytes of data daily, including 2.5 billion pieces of content, 2.7 billion "likes," and 300 million photos. Facebook has admitted to having stored 100 petabytes of photo



Facebook







Amazon draws data from 152 million

customers' purchases to help users decide on items to purchase.

They use historical purchasing data of millions of customers to make the most accurate predictions of shopping needs. Amazon is estimated to have around 1 exabyte of data stored. They are not forthcoming with the size of their data storage, only admitting to storing around a billion "objects."

At an average of 1 megabyte per "object," they would have about an exabyte of data.

Microsoft has been somewhat fuzzy with their numbers as well. While CEO Steve Ballmer claimed in 2013 that the company operated more than one million servers, they will only go so far as to say that Hotmail, their email client re-named as Outlook.com, stored hundreds of petabytes of data in its billion-plus mailboxes.

Rank	*	Name	9	\$ Market	t Cap \$	Price \$	Today	Price (30 days)	Country
1	l	Ć	Apple		\$2.796 T	\$177.79	▲ 0.03%	~~~	■ USA
2	2		Microsoft MSFT		\$2.386 T	\$321.01	<b>▼</b> 0.59%	~~~~	■ USA
3	3	*	Saudi Aramco		\$2.193 T	\$9.07	<b>2.56%</b>		S. Arabia
4	ļ	G	Alphabet (Google)		\$1.650 T	\$130.17	<b>→</b> 0.03%	~~~	■ USA
5	5	<u>a</u>	Amazon AMZN		\$1.420 T	\$138.41	<b>→</b> 0.11%	~~~	■ USA
6	5	<b>©</b>	NVIDIA NVDA		\$1.009 T	\$408.55	<b>→</b> 3.62%		■ USA
^1 7	,	B	Berkshire Hathaway	9	\$783.09 B	\$358.35	▲ 0.38%		■ USA
<b>v</b> 1 8	3	$\infty$	Meta Platforms (Facebook)	9	\$773.02 B	\$301.64	<b>→</b> 1.34%		■ USA
9	)	Ŷ	Tesla TSLA	\$	\$769.07 B	\$242.65	<b>▼</b> 1.10%	~~~	■ USA
10	)	VISA	Visa	9	502.73 B	\$240.04	<b>→</b> 0.05%	My	■ USA



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

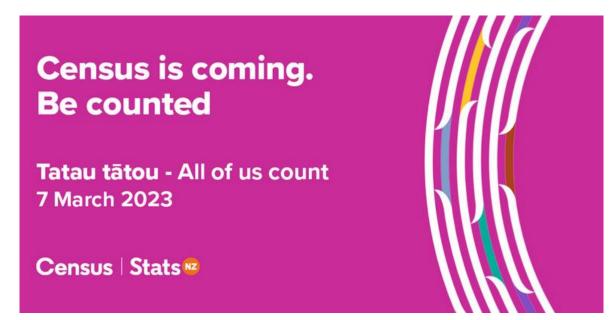
## Data Drives New Technologies







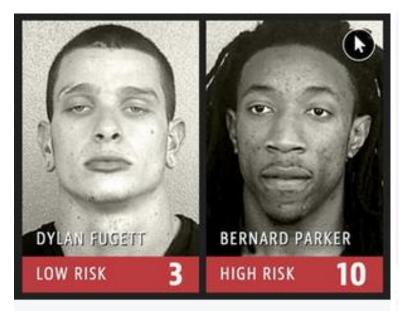
### Data Drives Decision-Making



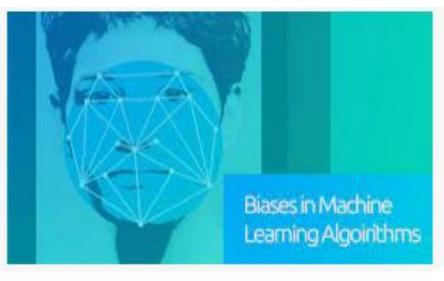




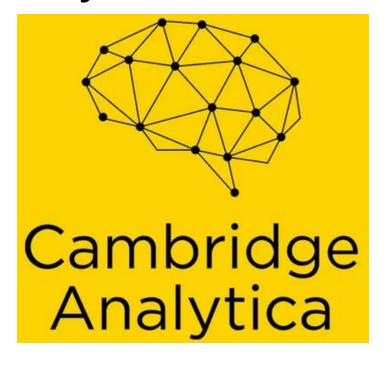
### Data Can Cause and Perpetuate Injustices

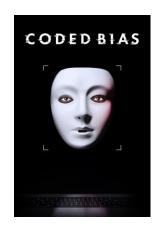




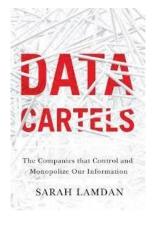


Machine Learning | CMU - Carnegie Mellon University
Amazon Scraps Secret Al Recruiting ...









#### 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 <u>Power!</u>

- 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.