Data types and storage

Learning intentions

We will be learning about the structure of data, specifically to,

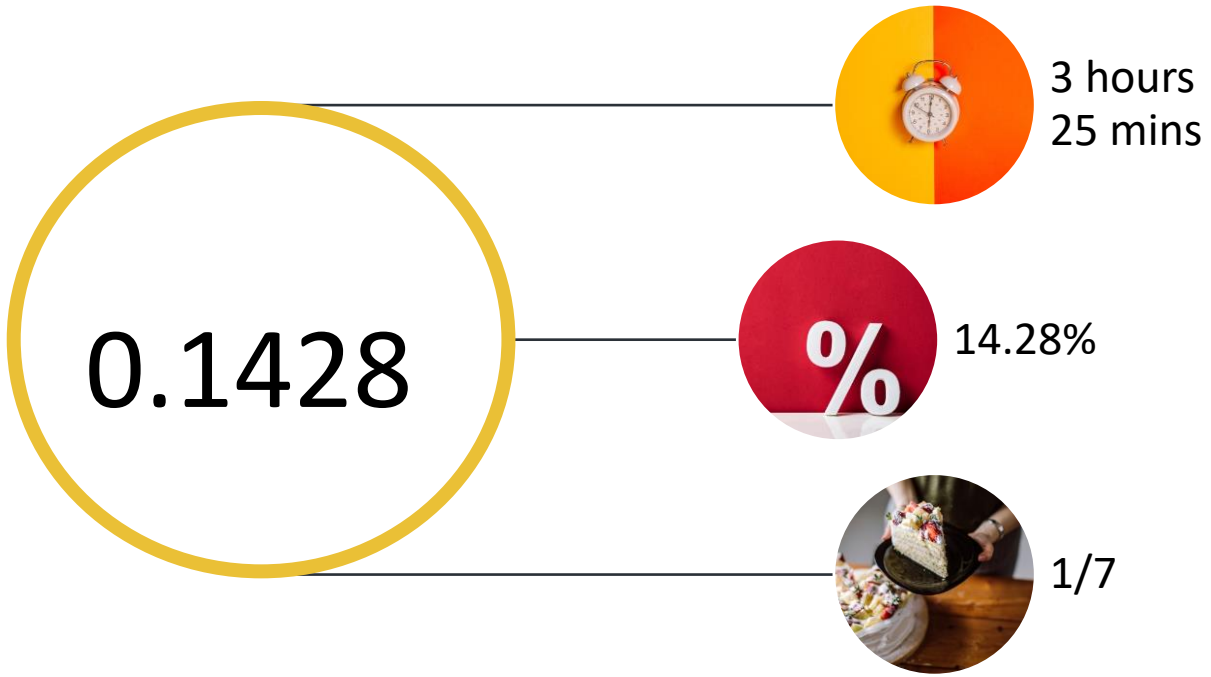
- understand the **different data types** that computers can store
- understand how these **data types can be combined into data structures**
- understand the difference between how data is stored and **how it is displayed**

Background

The way data is displayed to the end user is often **different to how it is stored**.

How the data is held in a computer can impact on the analysis you can perform on it.

In this lesson we are going to look at the **types of data** computers store and the **structures** they hold this data in.



Data types

The data type refers to **how data is stored internally** to the computer.

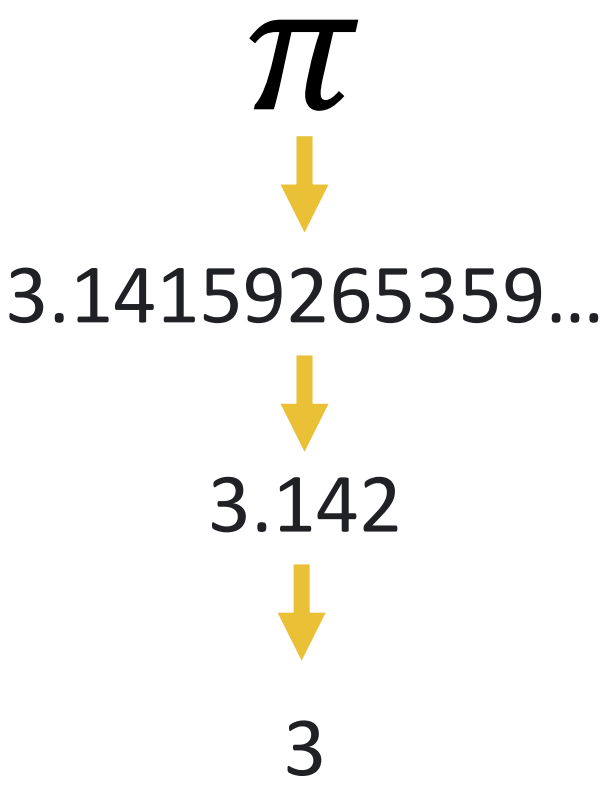
The data type **defines what calculations** the computer can perform on it. (e.g. can you add the data?)

Changing the data type can also affect **the precision** of the value stored.

Data types

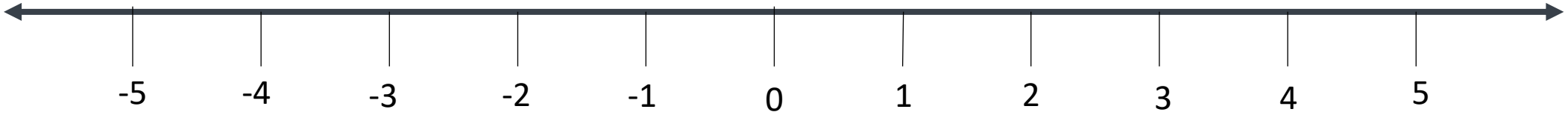
In data science the **precision** of a number relates to the **number of digits stored** in the computer.

By changing the data type, it can change the number of digits the computer stores.



Show me…

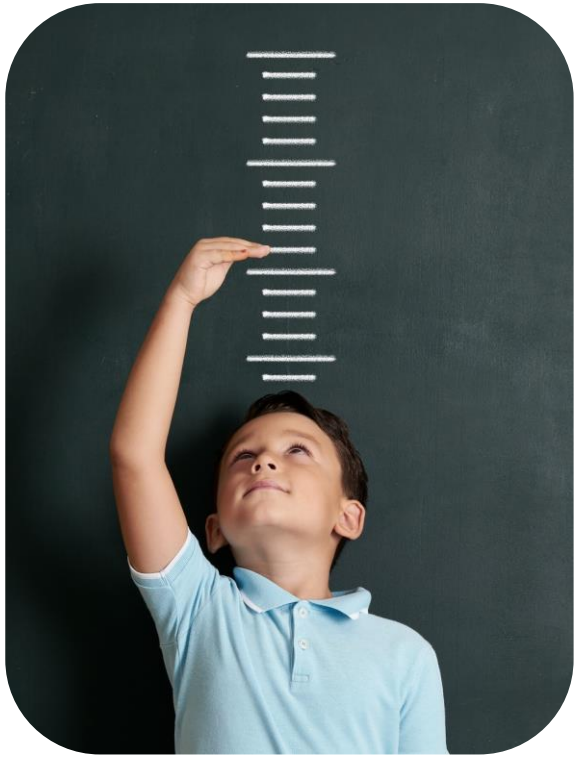
Integer Numbers (positive or negative) with no decimal or fractional parts



Show me…

Floating point numbers that contain a decimal or fractional part

123.15	½
-4.33333	0.15
1.0	¾

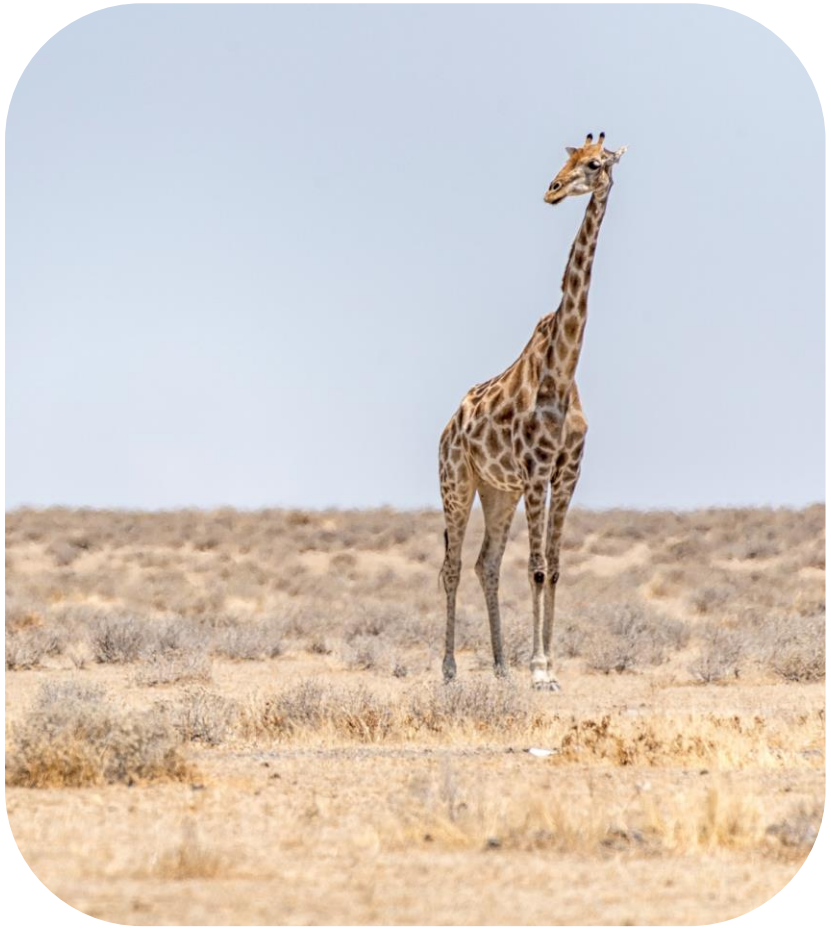


Example

What data type is the height of a giraffe (5.5 metres)?

It is a **floating point.**

Because it is a number that can contain a decimal or fractional part.

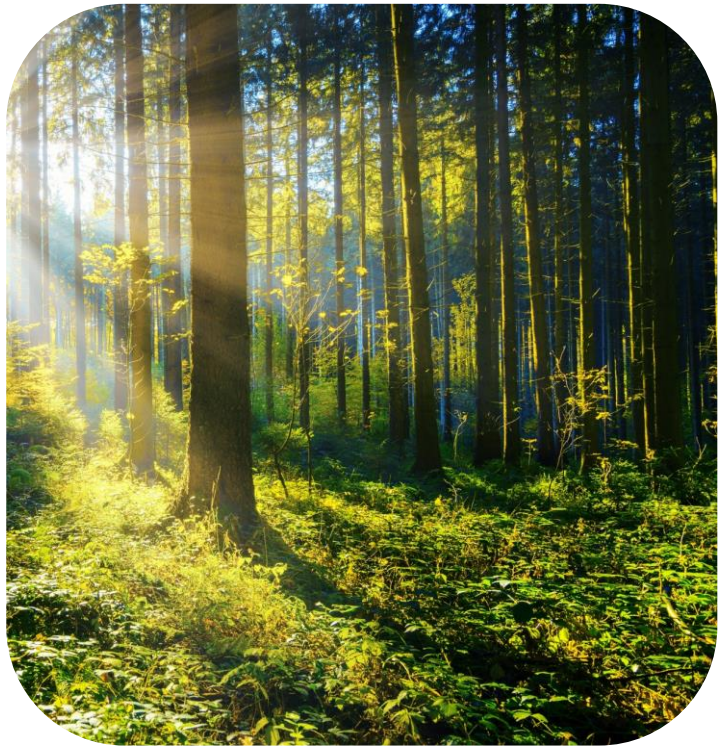


Example

What is the average height(m) of these trees when **using different data types**?

	Floating point (2dp)	**Integer**
tree_1	2.12	2
tree_2	2.49	2
tree_3	5.13	5
Average	**3.25**	**3**

There is **0.25m difference** in the average height of the trees when using the different data types.



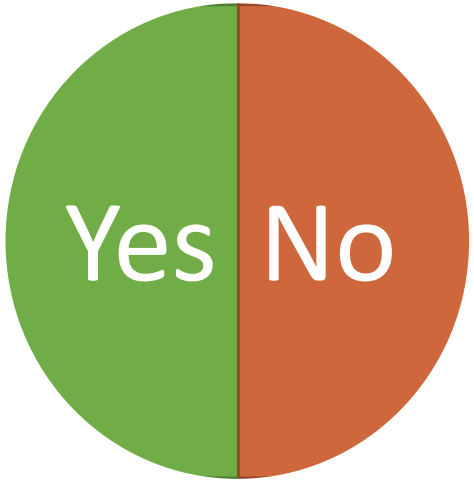
Show me…

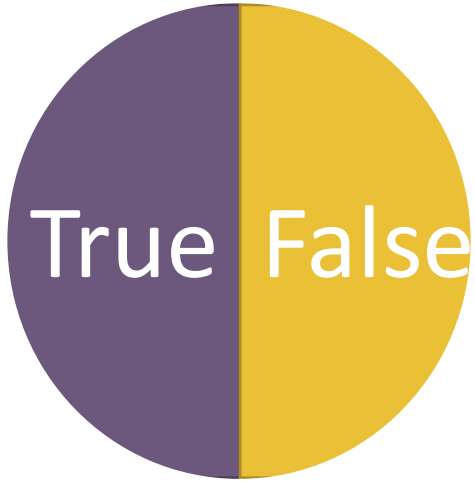
Character a single text character which can be a letter, number or symbol

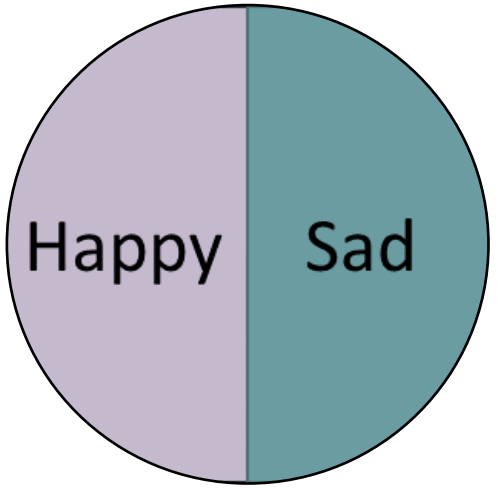
A	?	f
4	£	&

Show me…

Boolean can only take two possible values, such as true/false or yes/no.







Boolean data is often stored as 0/1 where 1 is true or yes

Data types

For computers, interpreting days, months, years, minutes and seconds is complicated.

To overcome this, computers process **date/time as the time passed since the 'epoch' date.**

Different software languages use different 'zero points' to count time from.

Software/language	Epoch date
Unix (e.g. Python)	1 Jan 1970
Microsoft Excel	1 Jan 1900
Windows NT	1 Jan 1601
MATLAB	0 Jan 1 BC
Google Sheets	30 Dec 1899

Show me…

Date and time the number of days or seconds passed since the 'epoch' date

1st Jan 2022 = 44562*

1 hour 30 mins = 0.06

* This example is based on Microsoft Excel.

Example

Date/Time data type
the number of days or seconds passed since the 'epoch' date

What number will Microsoft Excel store when the end user sees the date of **1st January 2030**?

In Excel, the ‘epoch’ date is 1/1/1900.

They are 47,484 days between 1/1/1900 and 1/1/2030.

Therefore, Excel will store the number **47,484**.

Your turn…

You are collecting data on people attending a lesson. What data types are these?

Information	Example	Data type
Date of lesson	15/5/2019	
Age of student (whole numbers)	16	
Arrived on time?	Yes	
Test result (in %)	75%	

Integers: numbers with no decimal or fractional parts

Floating point: numbers that can contain a decimal or fractional part

Character: a single text character which can be a letter, number or symbol

Boolean: can take two possible values such as true/false or yes/no.

Date and time: the number of days or seconds passed since the 'epoch' date, normally 1/1/1970.

Your turn…

You are collecting data on people attending a lesson. What data types are these?

Information	Example	Data type
Date of lesson	15/5/2019	Date
Age of student (whole numbers)	16	Integer
Arrived on time?	Yes	Boolean
Test result (in %)	75%	Floating point

Integers: numbers with no decimal or fractional parts

Floating point: numbers that can contain a decimal or fractional part

Character: a single text character which can be a letter, number or symbol

Boolean: can take two possible values such as true/false or yes/no.

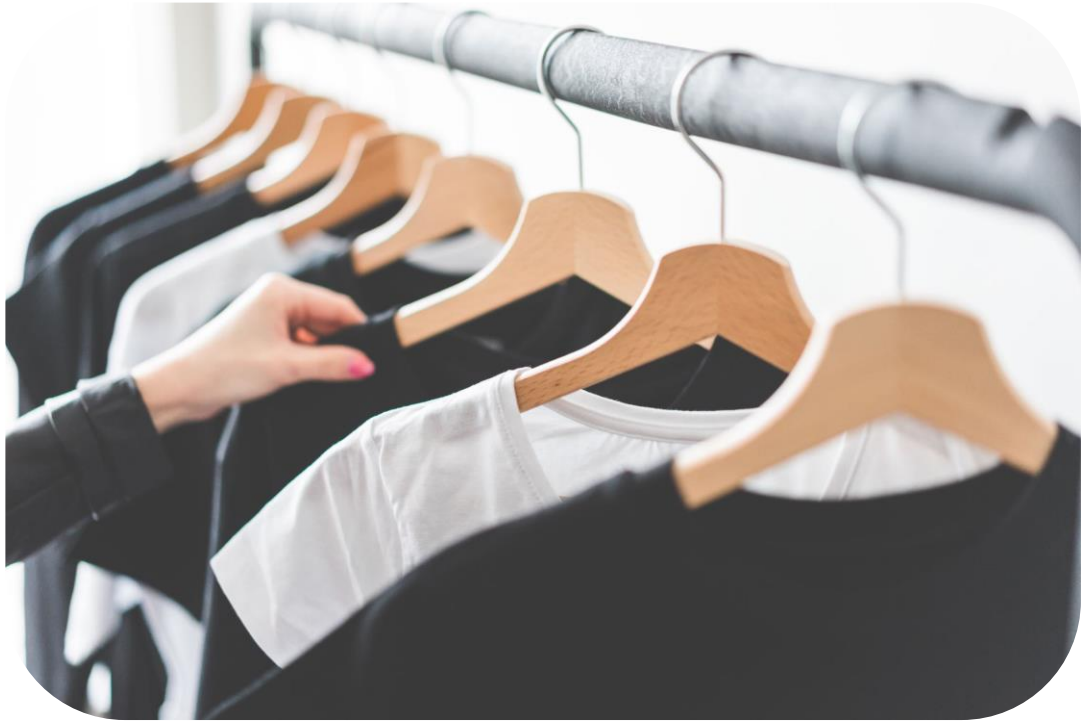
Date and time: the number of days or seconds passed since the 'epoch' date, normally 1/1/1970.

Next steps

Complete **questions 1 to 4**
in **section 1** of the
‘Data types and storage’ workbook.

Data structure

Imagine all your **clothes in a pile on the floor.** How long would it take you to find a specific jumper you want to wear?



Now imagine your **clothes are organised in a structured way** in your wardrobe and drawers. How long would it take to find your jumper now?

For computers to find data efficiently, the **data needs to be held in a structure**.

Definition

Data structure

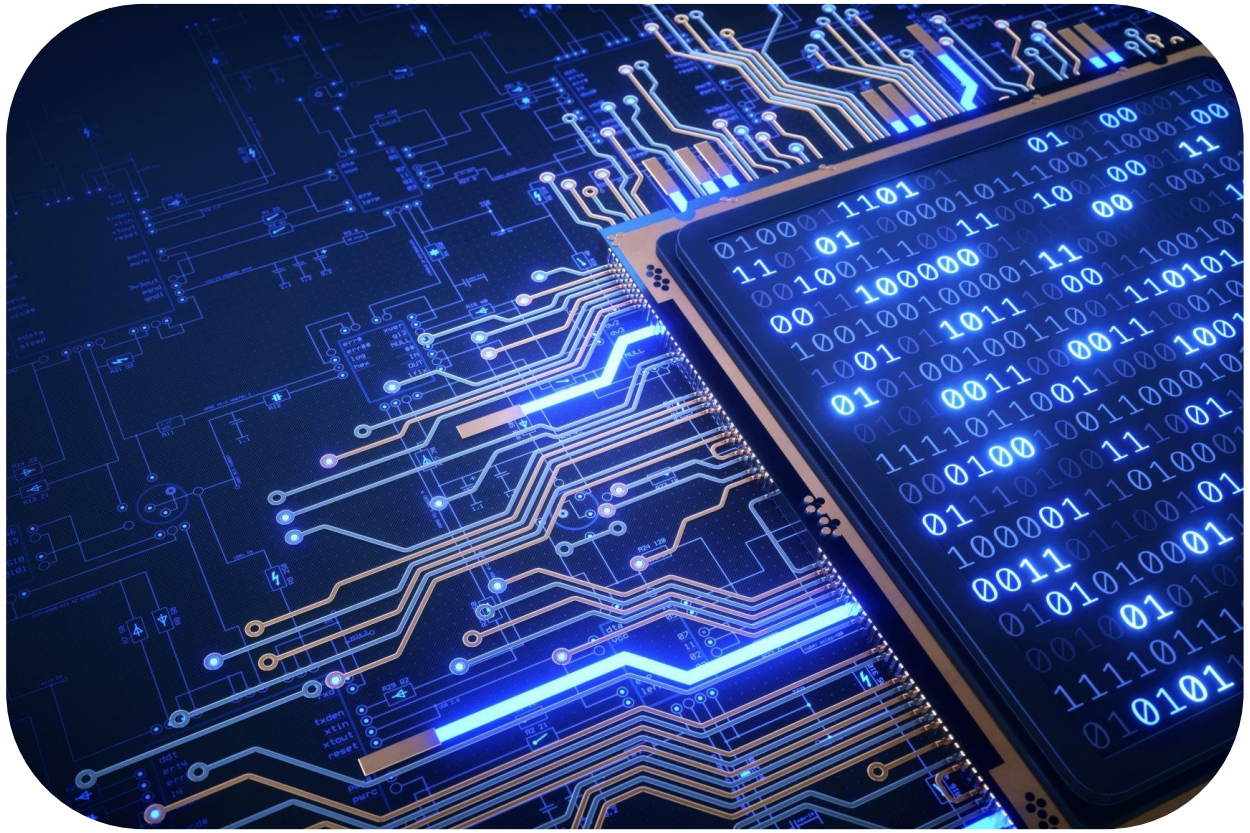
The way data is organised in a computer to make it easier and quicker to use.

Data structures

By using the appropriate data structures, it can allow for faster and easier processing of data.

The structure types normally used to store data are,

- String
- Array
- Dataset



Show me…

Strings a collection of characters combined to create alphanumeric text

BA1542

Edinburgh

#Friday

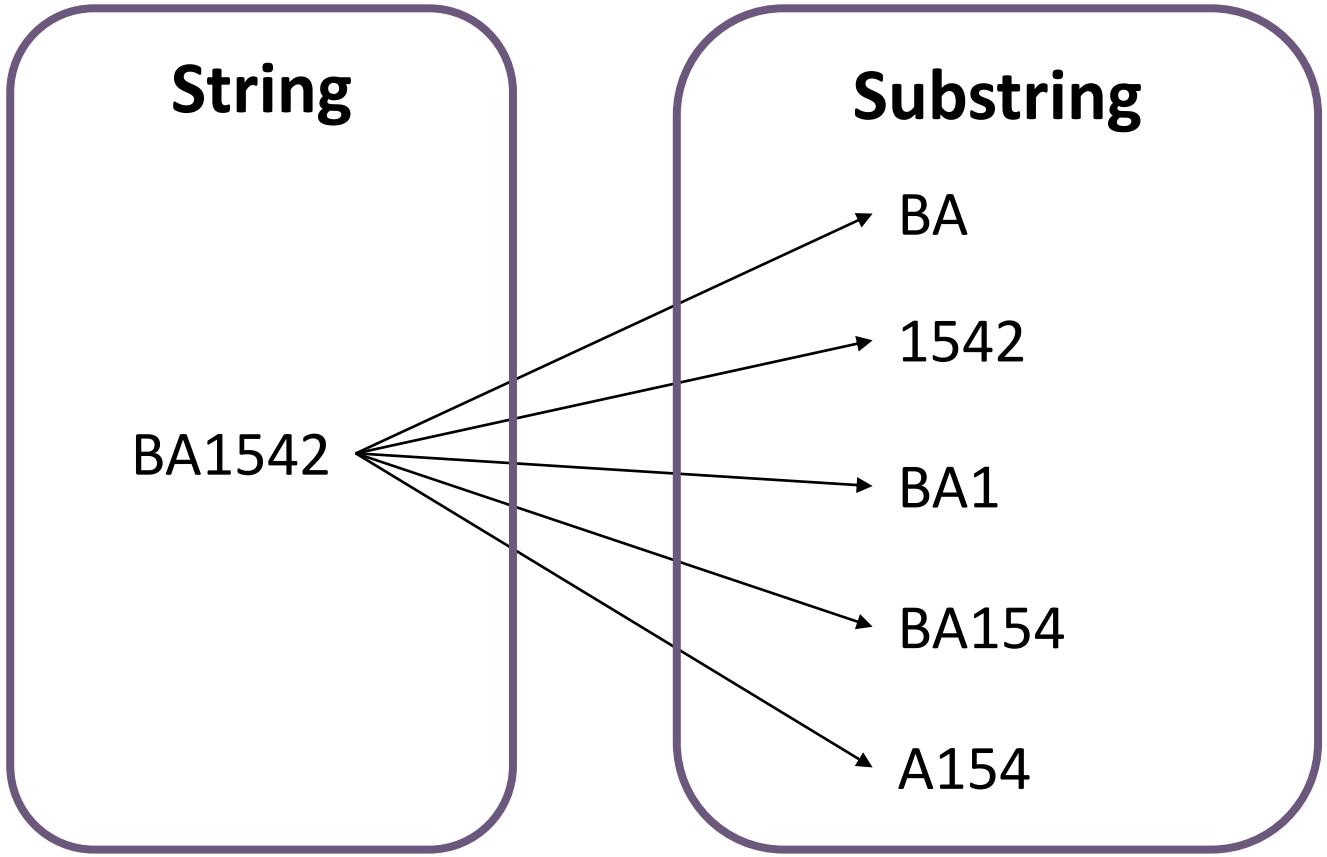
@scotland

AA 12 34 56 B

Your Name

Show me…

Substrings a set of characters that exist inside a string



The substring is any sequence of characters that are inside the original text.

Show me…

Array a structure of a fixed size which holds items of the same data type

Red	Blue
Yellow	White
Purple	Green

4	11	15	57	90

Show me…

Dataset A two-dimensional structure that has rows and columns.

unique_id	age	score	comment
145G45R	15	100.0	Pass
1755FG	43	39.5	Needs retake
LH452d	34	65.0	TBC

Each column can hold different data types

Every cell must contain some information

Your turn…

How could you turn this shopping list into a **dataset** with the following headings, “item”, “colour”, “weight_g”, “quantity”?

To buy

Bread

Pasta (500g)

Blue milk

Red milk

5 apples

Dataset
A two-dimensional structure that has rows and columns.

Your turn…

Turn this shopping list into a dataset.

To buy

Bread

Pasta (500g)

Blue milk

Red milk

5 apples

Each column holds a different type of information

item	colour	weight_g	quantity
Bread	NA	NA	1
Pasta	NA	500	1
Milk	Red	NA	1
Milk	Blue	NA	1
Apples	NA	NA	5

Every cell has some information

Next steps

Complete **questions 1 to 4**
in **section 2** of the
‘Data types and storage’ workbook.

Then complete **questions 1 to 5**
in **section 3** of the
‘Data types and storage’ workbook.

Learning checklist

I can *describe* common stored data types and understand that they can be different from the displayed format.

I can *describe* the structures data can be held in.

How you can use this lesson

You are free to:

- **Share** – copy and redistribute the material in any medium or format
- **Adapt** – remix, transform and build upon the material

Under the following terms:

- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **NonCommercial** — You may not use the material for commercial purposes.
- **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Created by Effini in partnership with Data Education in Schools, The Data Lab and Data Skills for Work, with funding from the Scottish Government.

Effini

THE DATA LAB
value from data

Data Skills FOR WORK

Data Education in Schools