THE UNIVERSITY *of* EDINBURGH | usher institute

Tutorial 2

Data Types and Structures in Python and R

Supported by

THE UNIVERSITY *of* EDINBURGH | DDI Data-Driven Innovation

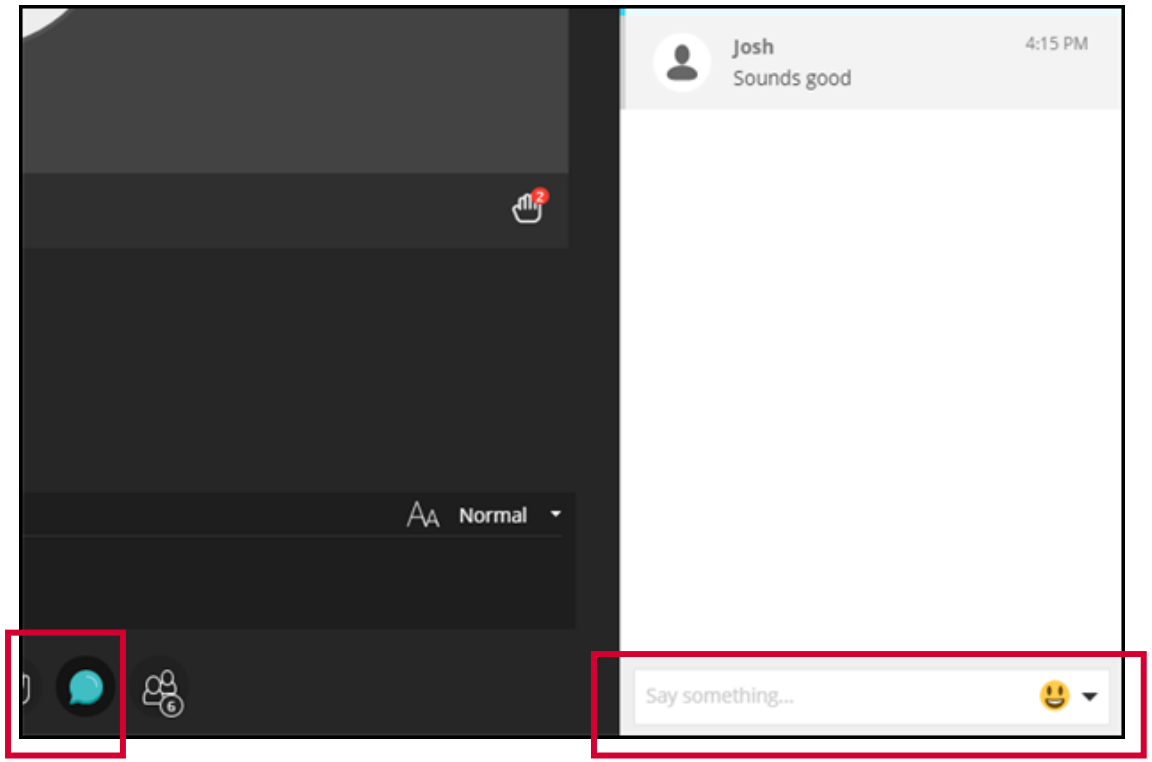
Audio check

Can you hear the presenter talking?

Please type **yes** or **no** in the “Text chat area”

If you can’t hear:

- Check your Audio/Visual settings in the Collaborate Panel
- Try signing out and signing back into the session
- Type into the chat box and a moderator will try to assist you



Recording

This session will now be recorded. Any further information that you provide during a session is optional and in doing so you give us consent to process this information.

These sessions will be stored by the University of Edinburgh for one year and published for 30 days after the event. Schools or Services may use the recordings for up to a year on relevant websites.

By taking part in a session, you give us your consent to process any information you provide during it.

THE UNIVERSITY *of* EDINBURGH | usher institute

Tutorial 2

Data Types and Structures in Python and R

Supported by

THE UNIVERSITY *of* EDINBURGH | DDI Data-Driven Innovation

Agenda

- Course update
- Error of the week
  - Understanding and navigating folder structures in code
- Data types overview
- Data types activity
- Any questions

Special guest lecture on date and time data: 26 April 1pm BST

- Nick Christofidis, Healthcare Analyst at Public Health Scotland
- See Data Dialogue Episode 3 in the Week 2 content folder

Error of the week

Understanding and navigating folder structures in code

How to find your current path

Python

- `OS` module to work with paths independent of operating systems
- `import os`

  `os.getcwd()`
  - to get the current working directory

R

- `getwd()`
  - base R function to get the current working directory
- When using R Projects, your current working directory automatically points to the root folder where that .Rproj file is saved

Navigating folder structures in code

File path = location of a file on a computer's file system structure

Absolute path = specified from the root directory (which is the first or topmost directory)

- AKA "full file paths"
- ~ commonly used to represent user's home directory
- e.g., `C:\Users\bblankin\Teaching\AY2023-24\Data Types and Structures in Python and R\Tutorial Slides` or `~\Teaching\AY2023-24\Data Types and Structures in Python and R\Tutorial Slides`

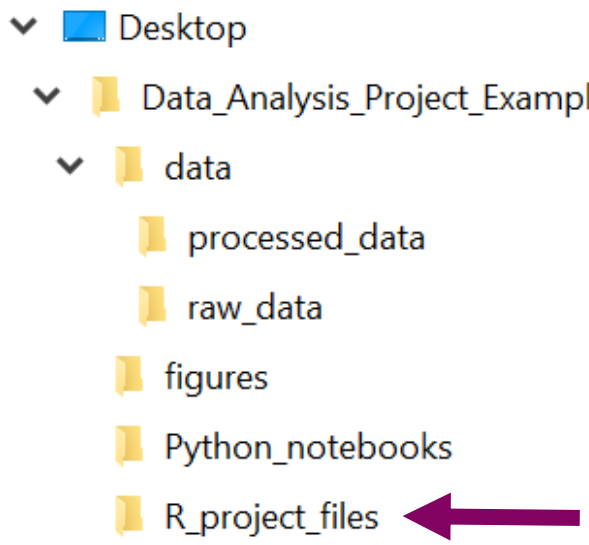
Relative path = path relative to the current directory

- Reproducible if you share your code with someone who has the same file and folder set up!
- Single dot (.) indicates current directory & double dot (..) represents parent directory
- e.g., `.\Data Types and Structures in Python and R\Tutorial Slides`

Path separators formatting note: Mac & Linux use / whereas Windows uses \

- URLs follow a standard format always using forward slash / regardless of operating system

Navigating folder structures in code



In Noteable, your home directory (`\home\jovyan\` or `~\`) = Jupyter Notebook Dashboard (what you see when you open a Standard Notebook (Python 3) server

The absolute path to a RMD file in "R_project_files" is location:

```
~\Desktop\Data_Analysis_Project_Example\
R_project_files\R_file.RMD
```

To navigate **up** the folder tree, use the "`..`" prefix.
To navigate 2 levels up, repeat the up prefix twice "`..\..\`"
For example to go up to figures from R_project_files it would be:

```
..\figures\figure_1.png
```

To get to raw_data from within R_project_files:

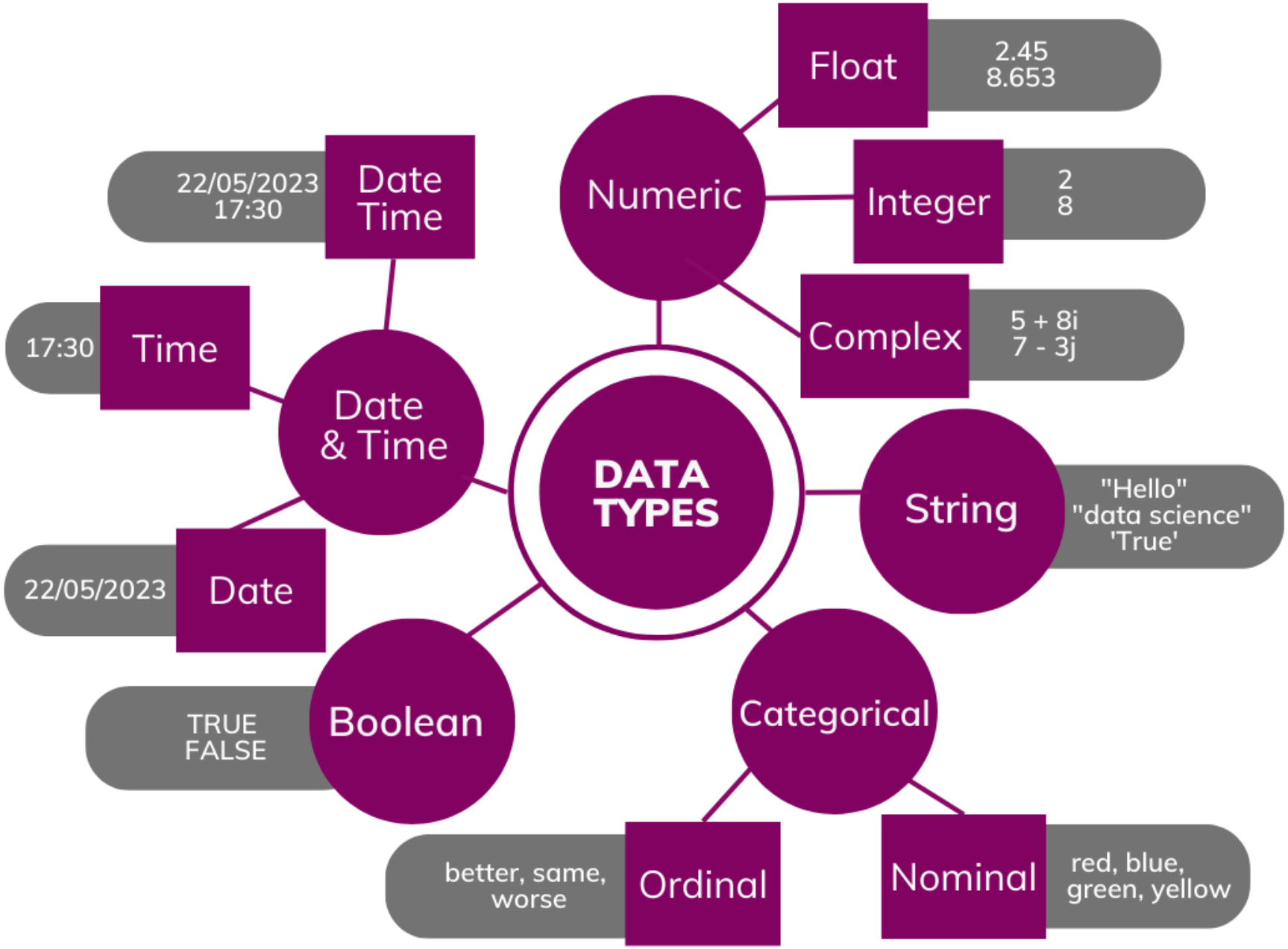
```
..\data\raw_data\raw_data.csv
```

What are data types

- Type of data or values an object contains
  - Internal construct that a programming language uses to understand how to store and manipulate data

- Determines:
  - What kind of mathematical, relational, or logical operation can be applied
  - Which operations can be performed to create, transform, and use the variable in further computation


DATA
TYPES

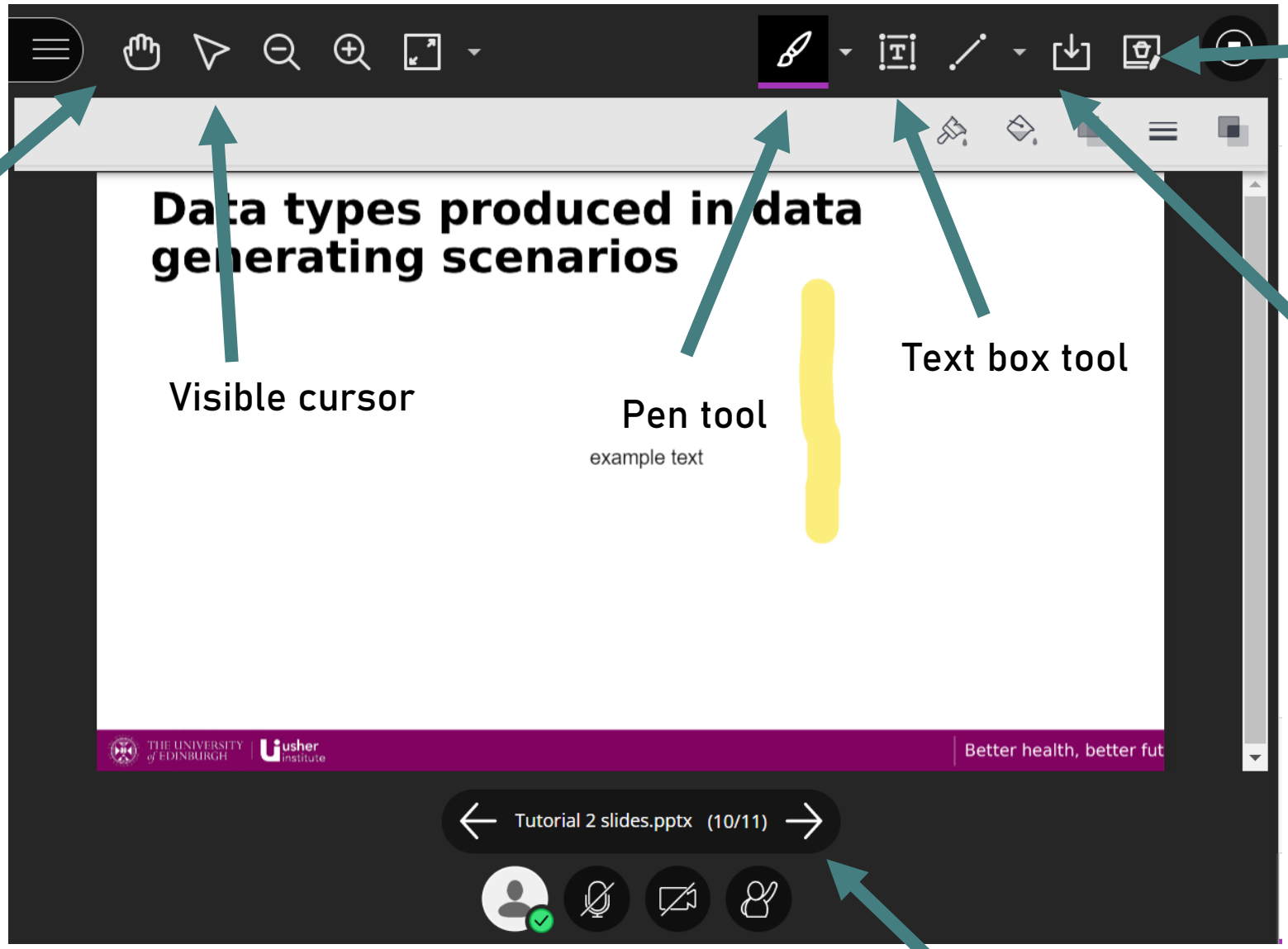

DATA TYPES
Numeric
Float
2.45
8.653
Integer
2
8
Complex
5 + 8i
7 - 3j
String
"Hello"
"data science"
'True'
Categorical
Nominal
red, blue, green, yellow
Ordinal
better, same, worse
Boolean
TRUE
FALSE
Date & Time
Date Time
22/05/2023 17:30
Time
17:30
Date
22/05/2023

Data types produced in data generating scenarios

- Emergency service call outs and transfers (e.g., ambulance)
- Hospital waiting times for non-elective surgeries
- Smoking behaviours in the community
- At-home carers in a local council area (local region)
- Any other health and social care situation you can think of!

Data types produced in data generating scenarios

- 35 min: Discuss the data generated in your selected scenario
  - What types of data they are
  - What are the possible range of values
  - Could the data type vary depending on your analytic use case? If so, how?
- Share with everyone what you spoke about in your groups
- Be sure to save your document so you can post it on the discussion boards after!



Select/ Move annotation items

Visible cursor

Pen tool

Text box tool

Delete annotations (be careful!)

Download slides with Annotations

Move between slides