

DPV Week 2

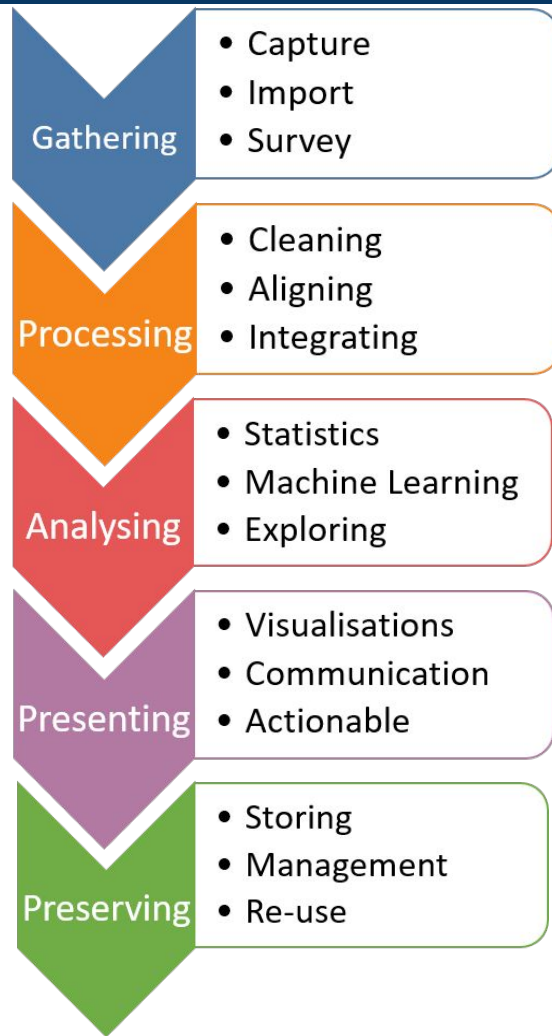
Suzanne Little

Tuesday

- Quick review of Data Analytics Pipeline
 - Case study: understand student housing issues in 2024 for DCU
- Data types
- Assignment 1: fork your skills report repository on gitlab.com/computing

Data Analytics Pipeline

Step 0: Planning



Case Study

Understand student housing issues in 2024 for DCU

Step 0: Planning → Ask important questions

1. Who is my client?
2. What is the question or need?
3. What is the outcome?
 1. DCU management or DCU student union
 2. How has availability and price of accommodation changed over the last 5 years?
 3. A brief report on the data including visualisations

Gathering

What data do you have or do you need?

Where is it? How do you get hold of it?



Gathering

- Capture
- Import
- Survey

Processing



Processing

- Cleaning
- Aligning
- Integrating

Analysing



Analysing

- Statistics
- Machine Learning
- Exploring

Presenting



Presenting

- Visualisations
- Communication
- Actionable

Preserving



Preserving

- Storing
- Management
- Re-use

Your skill record

<https://gitlab.computing.dcu.ie/slittle/csc1158-skills>

Summary: [One sentence summary of what the tool is useful for or what it does]

Data formats in: [where relevant!]

Data formats out: [where relevant!]

Three tips: [one of which must be advanced -- ie, likely to be found in a later chapter or section on datacamp]

- 1.
- 2.
- 3.

Examples of use: [minimal program sample, link to notebook or document or screenshot]

Contribution to data analytics pipeline: [refer to the talk in week 1 for the phases]

Comment on your skill level: [self-rating on this skill and what you'd like to learn to do better or what you'll do next]

References: [links to any sources used or just to helpful reference materials]

Assignment 1: using git

Git project - clone, edit, commit a skill report on Git

Worth 9%

You need to:

1. clone [csc1158-skills](#) repository
2. edit README.md to add your details (ie, not in *my* name!)
3. create git.md from template.md (ie, copy contents to a new file)
4. fill in your first skill report on git
5. commit (with a comment) before **6pm Sep 24th**

Fork the repository?

Makes a copy of **my** repository on **your** personal gitlab.computing account

Rename it appropriately

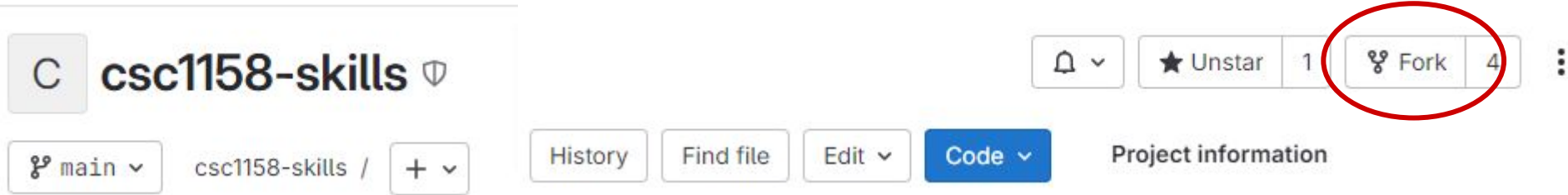
You can then clone and edit **your** version (not mine!)

Push your changes to your repository

If you try to push to mine you won't have permission

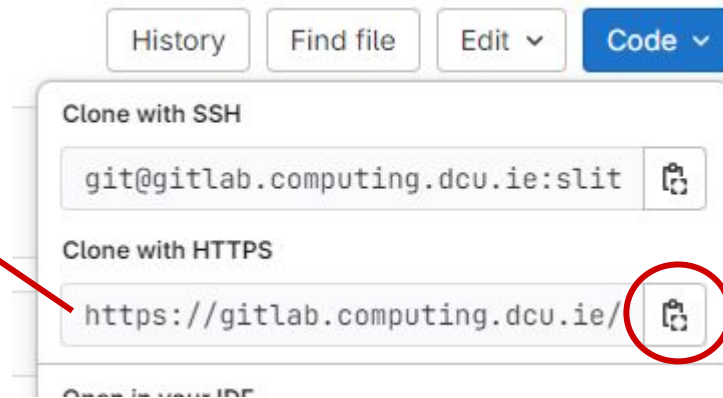
You can link forks so that updates to the parent repo can be imported to the fork

1. Fork this repo <https://gitlab.computing.dcu.ie/slittle/csc1158-skills>



2. Rename your forked repo correctly! (see next slide for instructions)

3. `$ git clone`



Change repo name - 2 steps

Naming, topics, avatar

Update your project name, topics, description, and avatar.

1

Project name

Topics

Topics are publicly visible even on private projects. Do not include spaces or special characters.

2

Advanced

Housekeeping, export, archive, change path, transfer, and delete.

Change path

A project's repository name defines its URL (the one you use to access the project via a browser).

- Be careful. Renaming a project's repository can have unintended side effects.
- You will need to update your local repositories to point to the new location.

Path

https://gitlab.computing.dcu.ie/slitttle/	csc1158-skills
---	----------------



Change path

Thursday

- Finish assignment 1
- Data sources
- Open Data (+ Exercise)
- Files