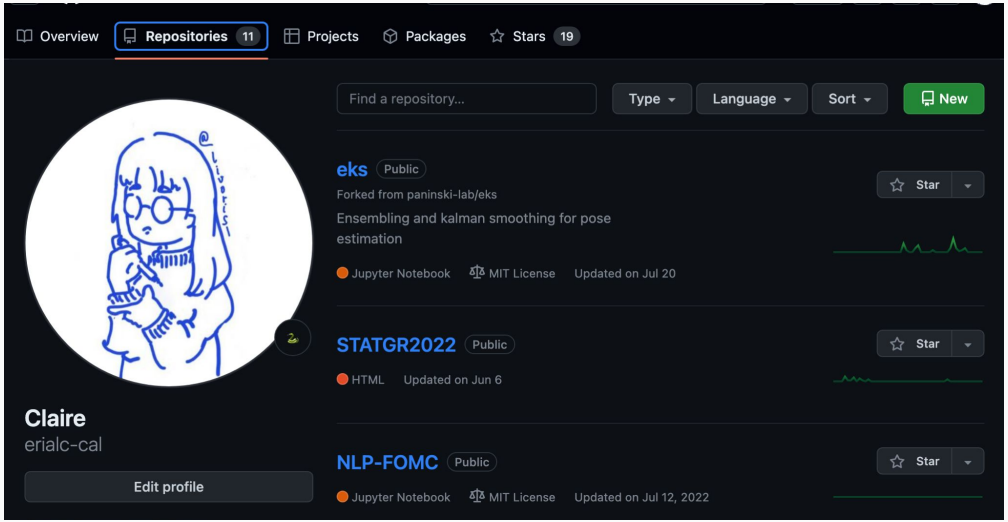
Tutorial for project 1 and introduction to GitHub

Applied Data Science Fall 2023

Getting started with GitHub

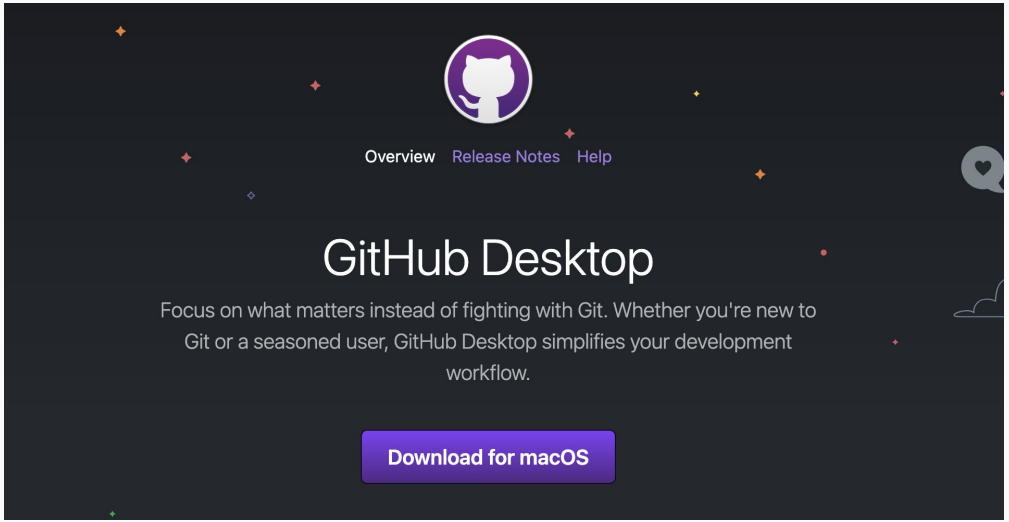
Create a GitHub account and start observing the options on GitHub website:

- creating new repository
- forking repository from other people



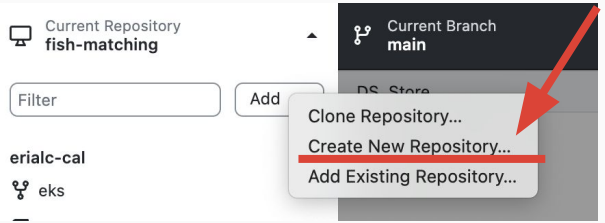
Getting started with GitHub

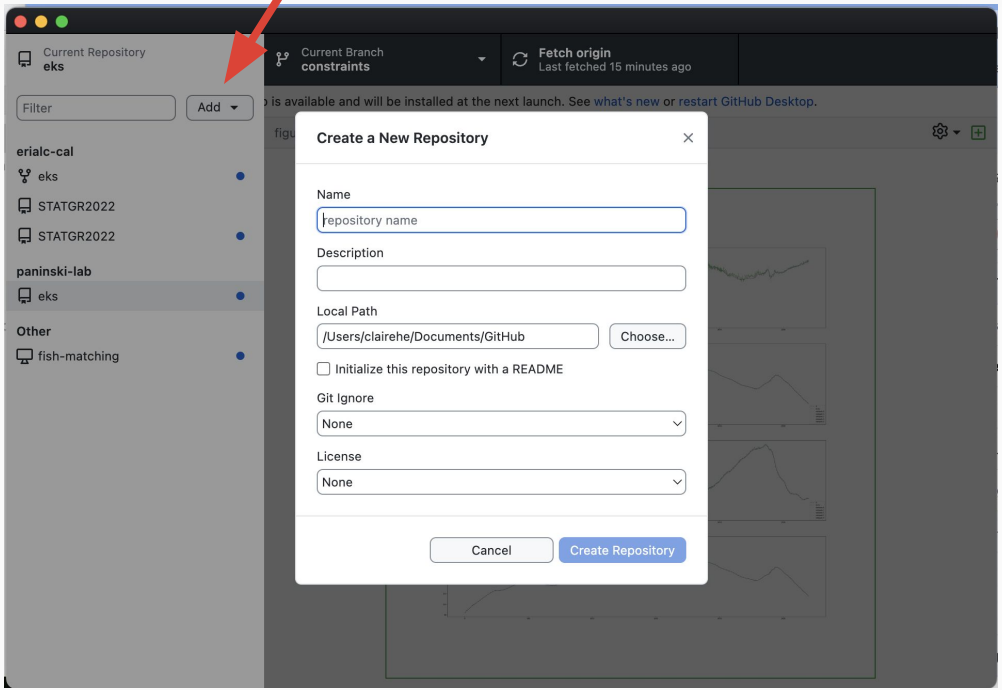
Download and install the GitHub desktop app corresponding to your distribution (MacOS, Linux, Windows) on https://desktop.github.com/ and sign in your GitHub account. Any project forked or created on GitHub website can be now cloned onto local (your computer).



Getting started with GitHub

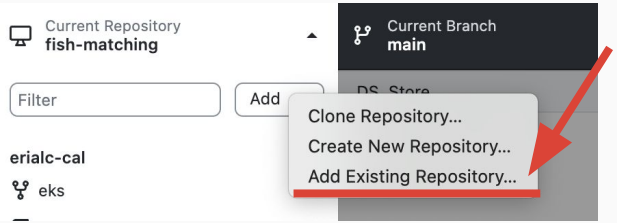
Create a repository 1: from GitHub desktop

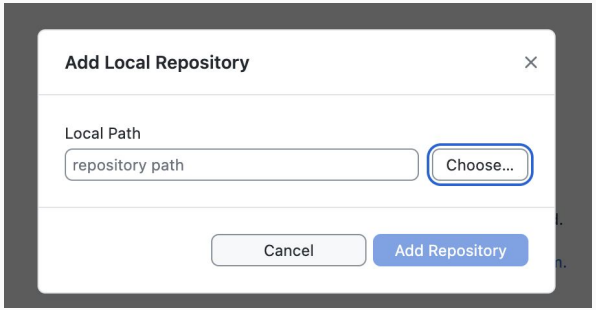


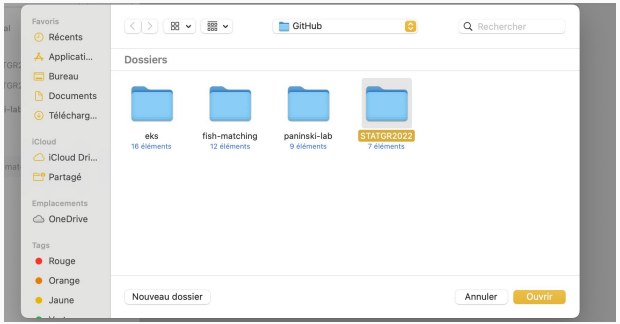


Getting started with GitHub

Create a repository 2: from existing folder



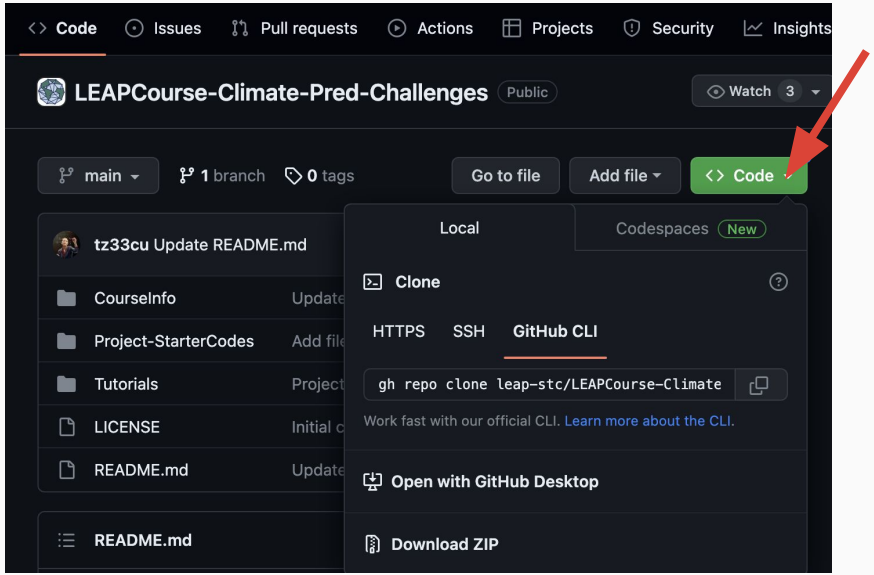




Getting started with GitHub

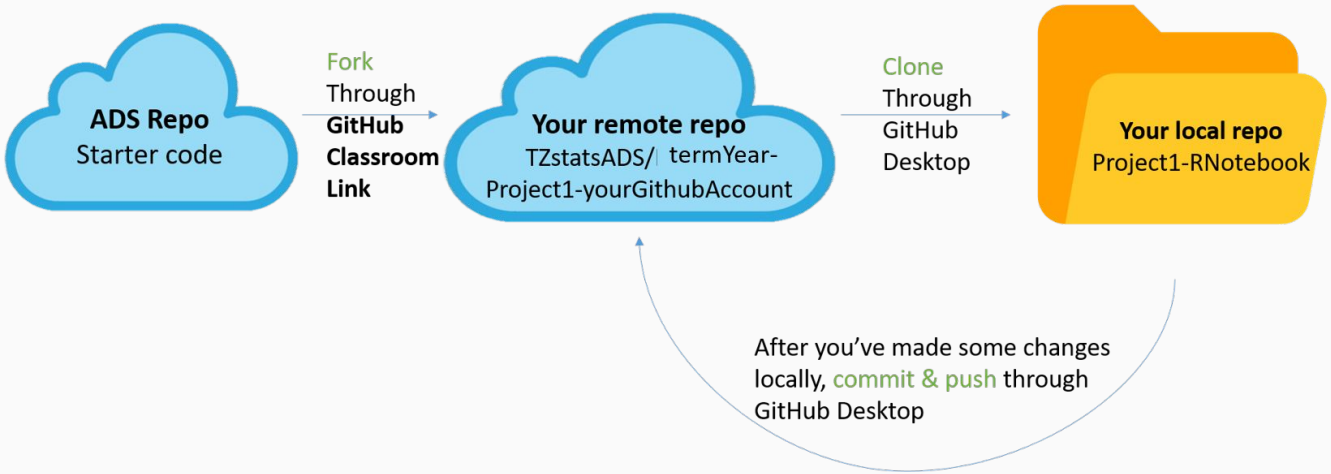
Create a repository 3: cloning an existing repository from website

We will see later how to use branching for collaborative work.



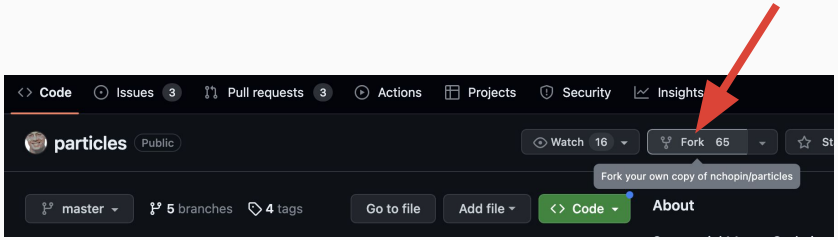
Getting started with project 1

Outline of project 1



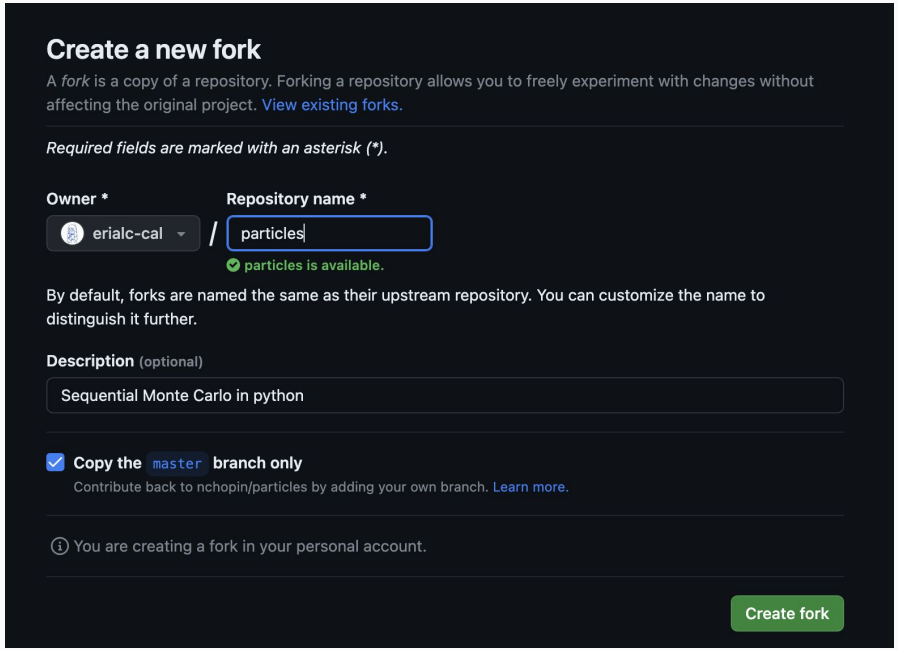
Getting started with project 1

Fork: github copy of project



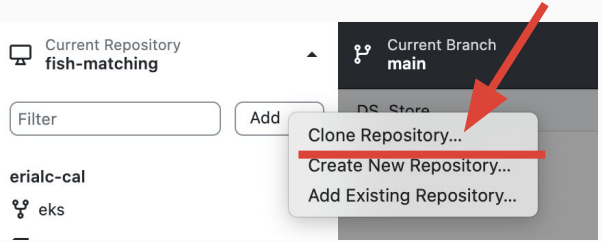
Project is now in your repositories.

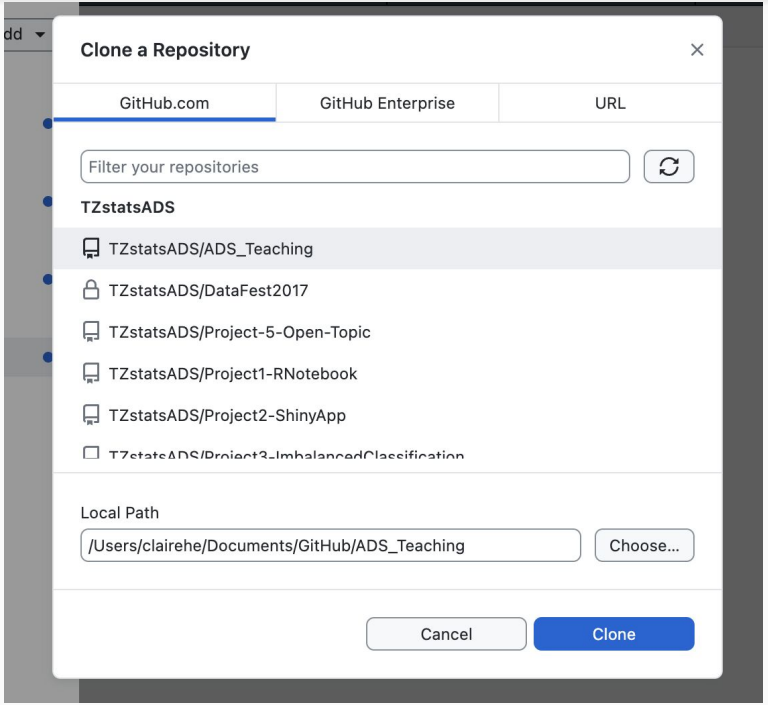
Next: clone to local to work on your laptop.



Getting started with project 1

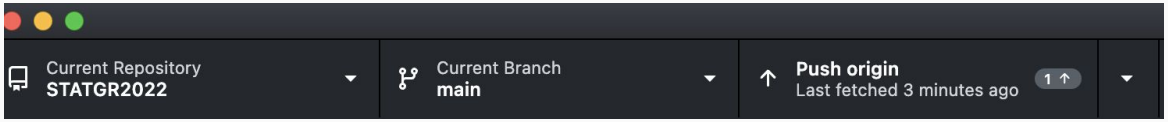
Clone to local from GitHub profile.



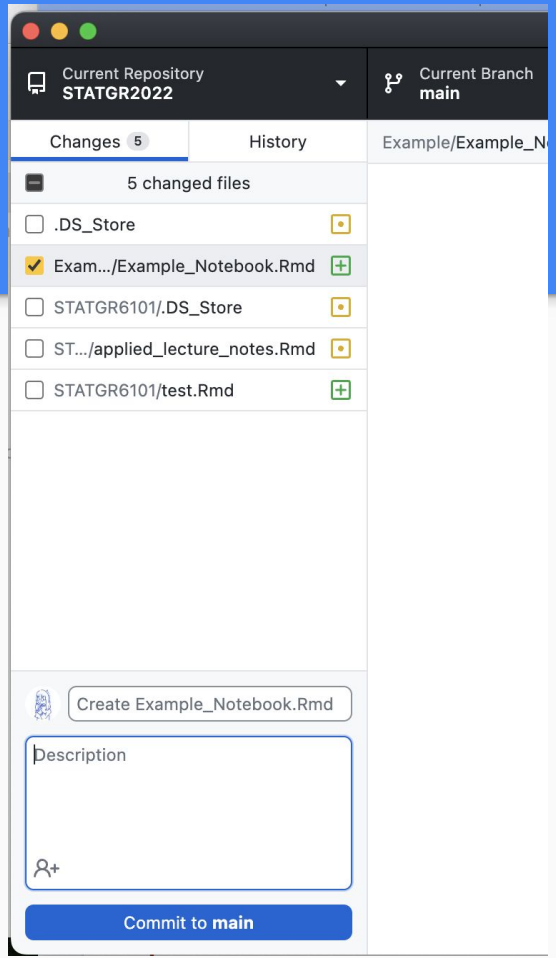


Getting started with project 1

Create a notebook (R or Python) and commit.



Once commit is done, you can Push to origin.
GitHub website will show the changes now.



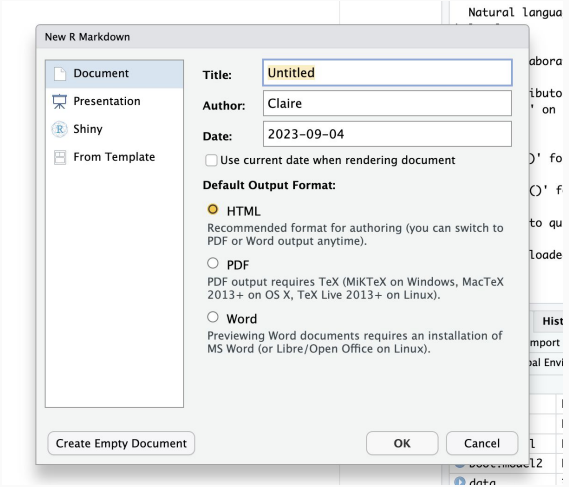
Creating a project notebook

If you choose R

- Install Rstudio corresponding to your distribution
- Create .rmd file: default is html, to knit pdf you will need to download files supporting LaTeX

If you choose Python

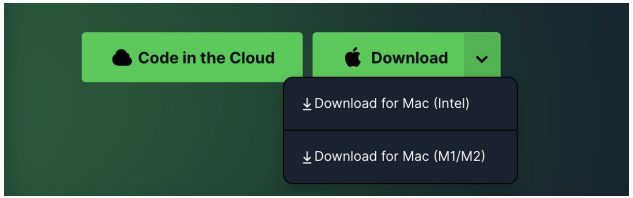
- Recommend installing Anaconda and launching Jupyter Notebook
- Alternatively you can use Google Colab notebooks



Jupyter Notebooks

If you choose Python

- Recommend installing Anaconda and launching Jupyter Notebook



- Alternatively you can use Google Colab notebooks

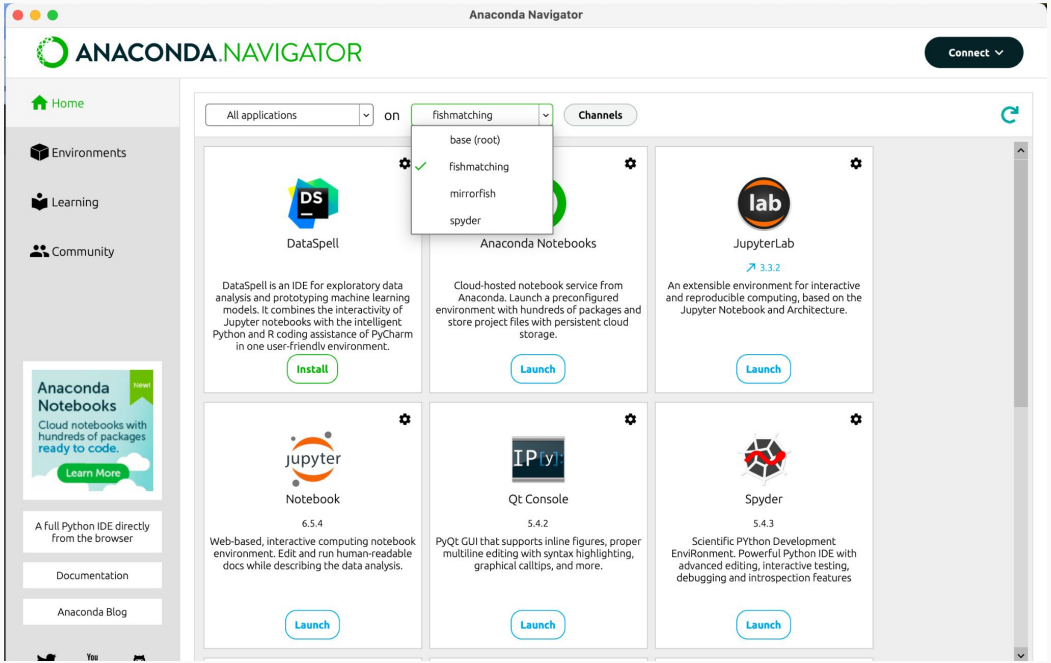
  Tutorial https://colab.research.google.com/#

Getting started with Anaconda

Anaconda is a distribution that includes R and Python.

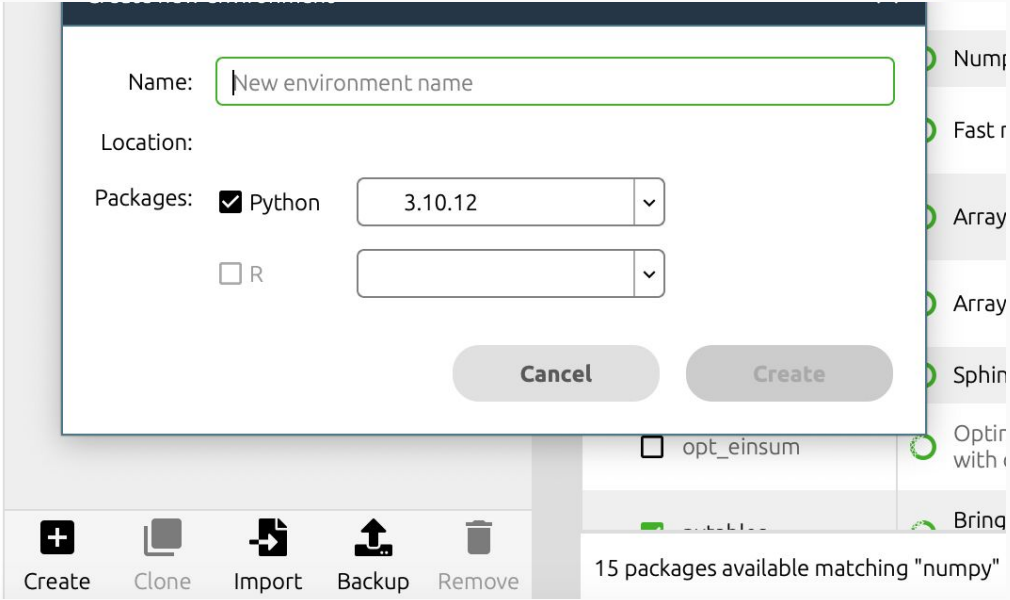
For installation: https://docs.anaconda.com/free/anaconda/install/index.html

For notebooks : launch either *JupyterLab, JupyterNotebook*



Setting up an environment

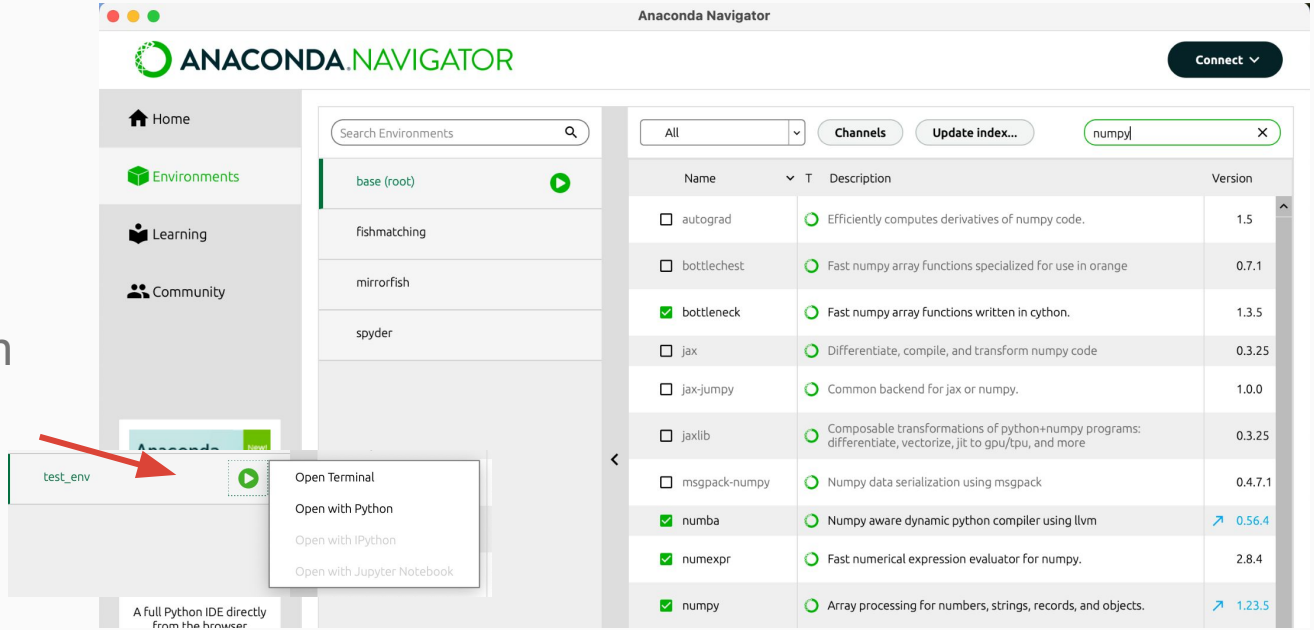
Create your environment using the base packages that Anaconda loads.



Setting up an environment

Manage package installation using the search bar

Note some packages that are not hosted on conda will need to be installed via console



Good practices for notebooks (not an exhaustive list)

- Be careful in variable name attribution between cells
- Do not overload chunks, you can still code functions to not make the notebook too heavy and import it as module
- Be careful when installing packages: always check if package is available on your environment base, if not use terminal install (to avoid package conflicts)
- Find the import line in error messages and use stack overflow, stack exchange etc.
- Google a lot ! You can also get inspired by projects available on GitHub.

Good practices for notebooks (not an exhaustive list)

- Definitely make use of the “Markdown Cells” to make the report readable.
- To have a more polished report: here's something to refer to : https://chris-said.io/2016/02/13/how-to-make-polished-jupyter-presentations-with-optional-code-visibility/
- We will open a thread on Piazza for everyone to share their helpful resources.