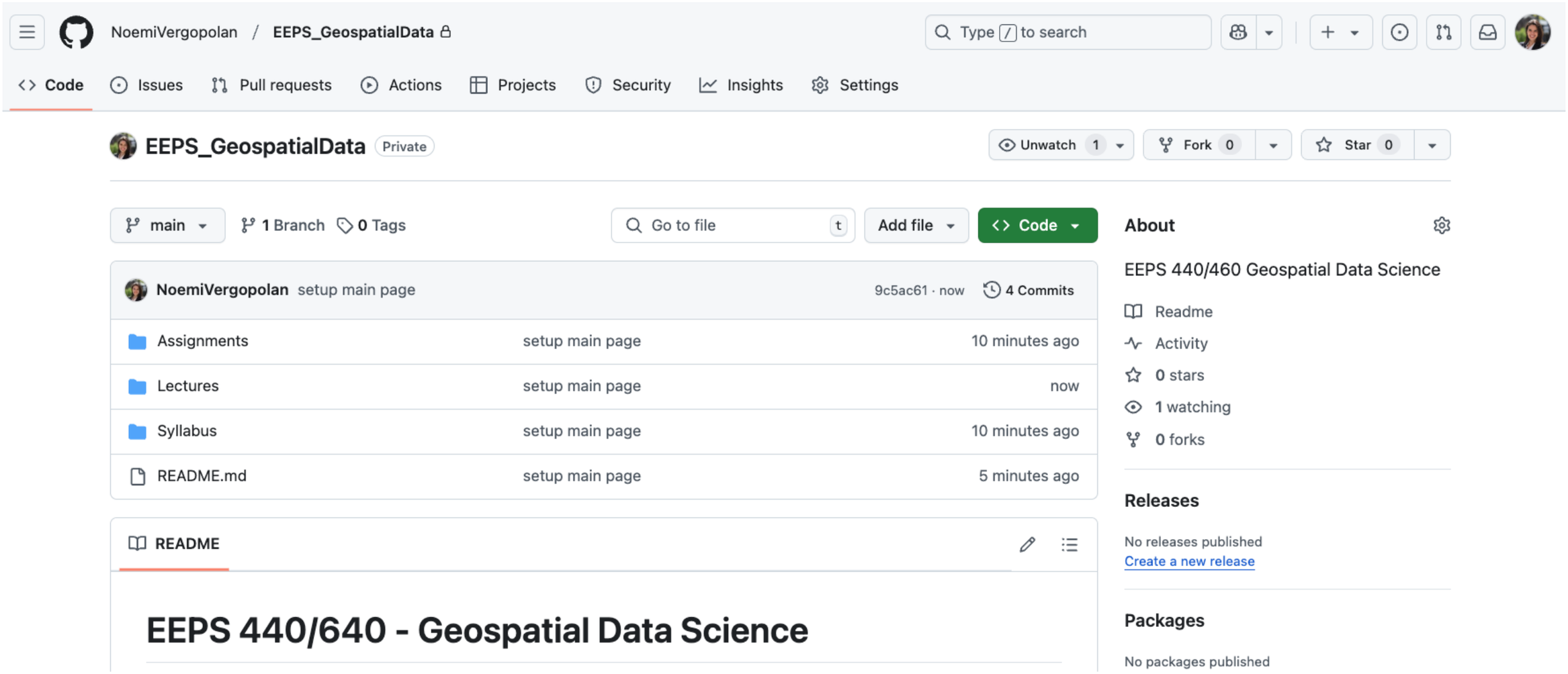
EEPS 440/640

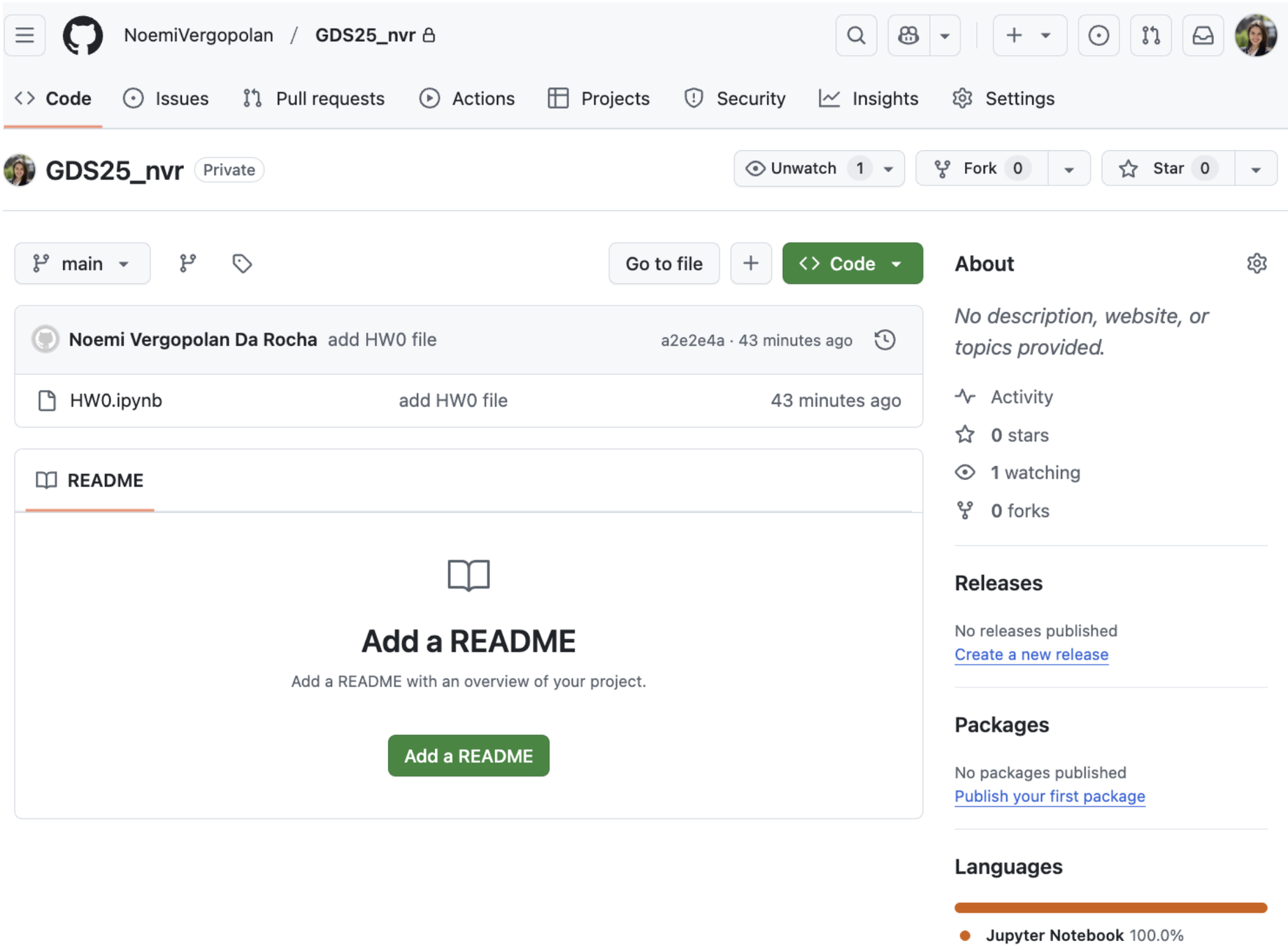
Geospatial Data Science (GDS)

The primary class website for this course is on GitHub



https://github.com/NoemiVergopolan/EEPS_GeospatialData.git

You have your own repository



https://github.com/NoemiVergopolan/GDS25_nvr

Lets create a token password for YOUR repository

Log to your Github Account

https://github.com/settings/personal-access-tokens

- Generate new token
- Give it a name (e.g., GDS)
- Only selected repositories:
  GDS25_nv, EEPS_GeospatialData
- Repository Permissions → Read and write
- Generate Token

Save your token password somewhere safe

How to access the HPC

- Check the HPC account info you received
- Open the terminal on your machine

This week → `ssh nv25@notsx.rice.edu`

Next week onwards → `ssh nv25@nots.rice.edu`

  - Password and login

How to launch a job on the HPC

Copy the job submission script template to your home

```
cp /projects/eeps440/jupyter-smp.slurm .
```

Launch your job

```
[nv25@loginx1 ~]$ sbatch jupyter-smp.slurm
Submitted batch job 229150
```

- See your job in the queue with squeue –u USER

```
[nv25@loginx1 ~]$ squeue -u nv25
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
            229150   commons jupyter-     nv25  R       1:12      1 bb5u26c1
```

- Use ls to look for the job output slurm-JOBID.out

```
[nv25@loginx1 ~]$ ls
~   jupyter-smp.slurm  slurm-229150.out
```

- Use cat slurm-JOBID.out to show what is inside

```
[nv25@loginx1 ~]$ cat slurm-229150.out
```

```
[nv25@loginx1 ~]$ cat slurm-229150.out

Run the following command to set up an ssh tunnel to the compute node:

ssh -NL 60560:bb5u26c1:60560 nv25@nots.crc.rice.edu
```

ssh -NL 60560:bb5u26c1:60560 nv25@nots.crc.rice.edu

```
[W 2025-01-23 11:23:59.287 NotebookApp] 'ip' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. B
e sure to update your config before our next release.
[W 2025-01-23 11:23:59.287 NotebookApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp.
 Be sure to update your config before our next release.
[W 2025-01-23 11:23:59.287 NotebookApp] 'port' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp.
 Be sure to update your config before our next release.
[W 2025-01-23 11:23:59.287 NotebookApp] 'allow_origin' has moved from NotebookApp to ServerApp. This config will be passed to Se
rverApp. Be sure to update your config before our next release.
[I 2025-01-23 11:23:59.288 ServerApp] nbclassic | extension was successfully linked.
[I 2025-01-23 11:23:59.612 ServerApp] notebook_shim | extension was successfully linked.
[I 2025-01-23 11:23:59.612 ServerApp] panel.io.jupyter_server_extension | extension was successfully linked.
[I 2025-01-23 11:23:59.700 ServerApp] notebook_shim | extension was successfully loaded.
[I 2025-01-23 11:23:59.701 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2025-01-23 11:23:59.702 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2025-01-23 11:23:59.705 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.12/site-packages/jupyterlab
[I 2025-01-23 11:23:59.705 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 2025-01-23 11:23:59.705 LabApp] Extension Manager is 'pypi'.
[I 2025-01-23 11:23:59.806 ServerApp] jupyterlab | extension was successfully loaded.
[I 2025-01-23 11:23:59.811 ServerApp] nbclassic | extension was successfully loaded.
[I 2025-01-23 11:23:59.812 ServerApp] panel.io.jupyter_server_extension | extension was successfully loaded.
[I 2025-01-23 11:23:59.812 ServerApp] Serving notebooks from local directory: /home/nv25
[I 2025-01-23 11:23:59.812 ServerApp] Jupyter Server 2.15.0 is running at:
[I 2025-01-23 11:23:59.812 ServerApp] http://bb5u26c1:60560/lab?token=f54ed35c711f6ffee18706435ab6913499c2b86c2fc45d47
[I 2025-01-23 11:23:59.812 ServerApp]     http://127.0.0.1:60560/lab?token=f54ed35c711f6ffee18706435ab6913499c2b86c2fc45d47
[I 2025-01-23 11:23:59.812 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2025-01-23 11:23:59.816 ServerApp]

    To access the server, open this file in a browser:
        file:///home/nv25/.local/share/jupyter/runtime/jpserver-3571747-open.html
    Or copy and paste one of these URLs:
        http://bb5u26c1:60560/lab?token=f54ed35c711f6ffee18706435ab6913499c2b86c2fc45d47
        http://127.0.0.1:60560/lab?token=f54ed35c711f6ffee18706435ab6913499c2b86c2fc45d47
```

http://127.0.0.1:60560/lab?token=f54ed35c711f6ffee18706435ab6913499c2b86c2fc45d47

```
[nv25@loginx1 ~]$
```

- Open a **NEW** terminal

```
ssh -NL 60560:bb5u26c1:60560 nv25@nots.crc.rice.edu
```

- Type the password and let it be
- Open your internet browser and paste the link with 127.0…..

```
http://127.0.0.1:60560/lab?token=f54ed35c711f6ffee18706435ab6913499c2b86c2fc45d47
```

You are running your Jupyter lab on Rice's HPC!

File Edit View Run Kernel Tabs Settings Help

/

Name	Modified
~	9m ago
jupyter-smp.slurm	now
slurm-229150.out	next yr.

Launcher

Notebook

Python 3 (ipykernel)

Bash

Console

Python 3 (ipykernel)

Bash

Other

Terminal

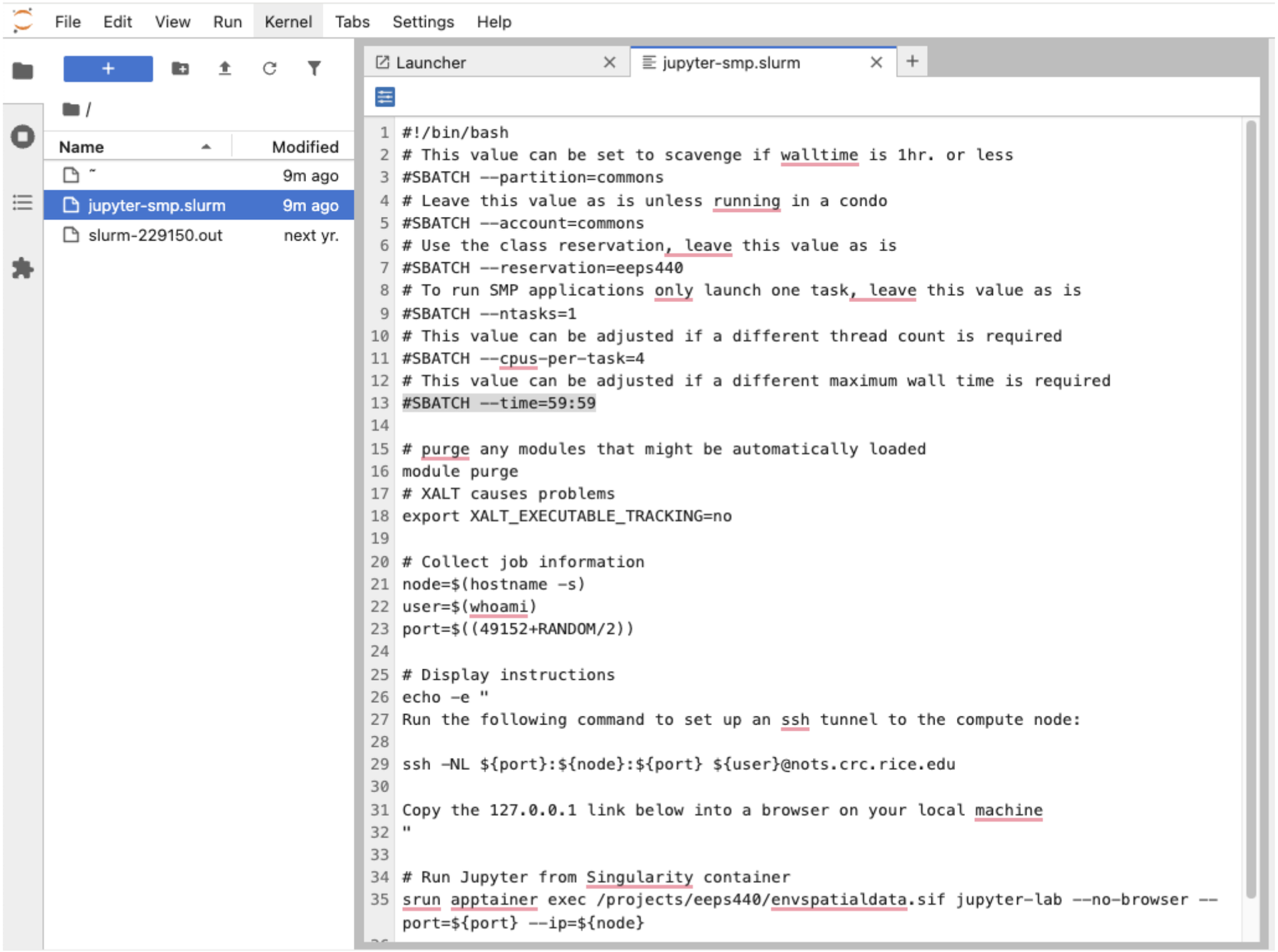
Text File

Markdown File

Python File

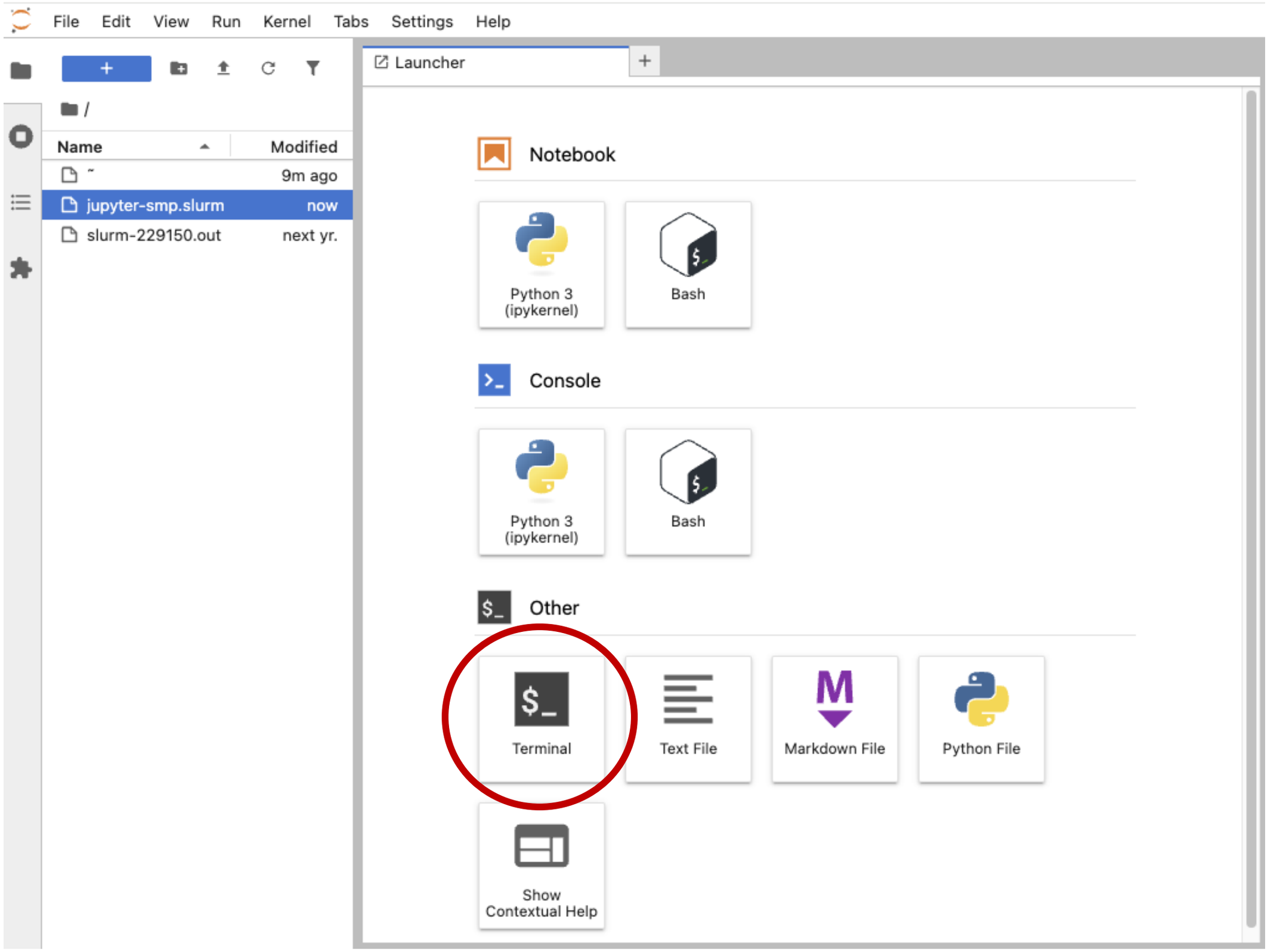
Show Contextual Help

Your job submission script template defaults to 30min run. Let's change it to 59:59 min…



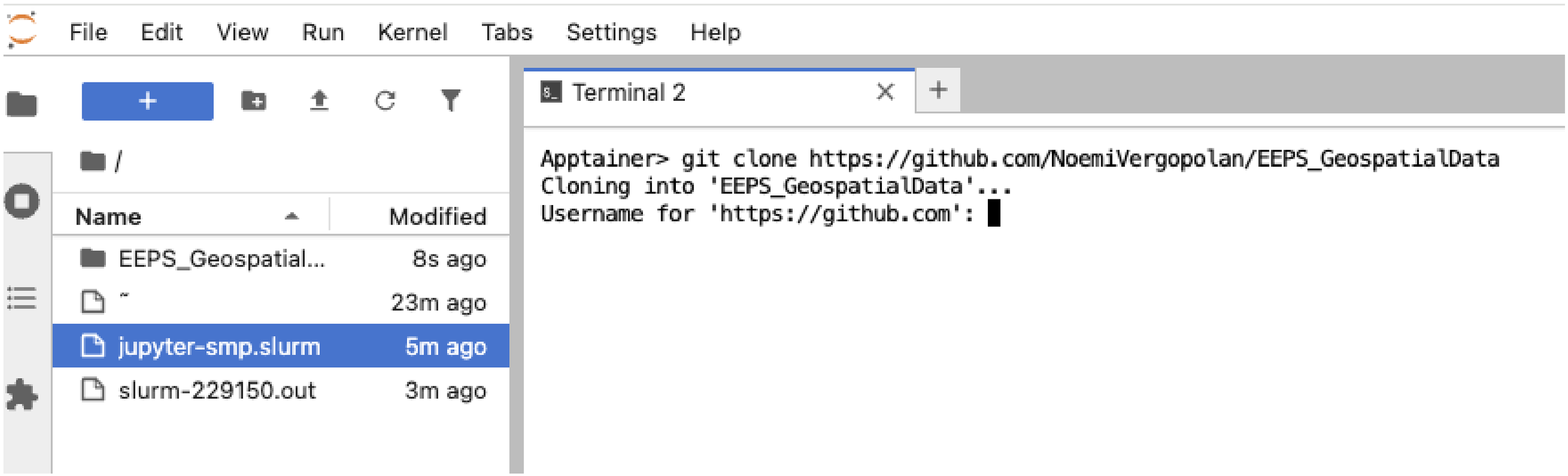
You can ask for more run time, but you will probably sit for longer in the queue…

Open a terminal on Jupyter Lab

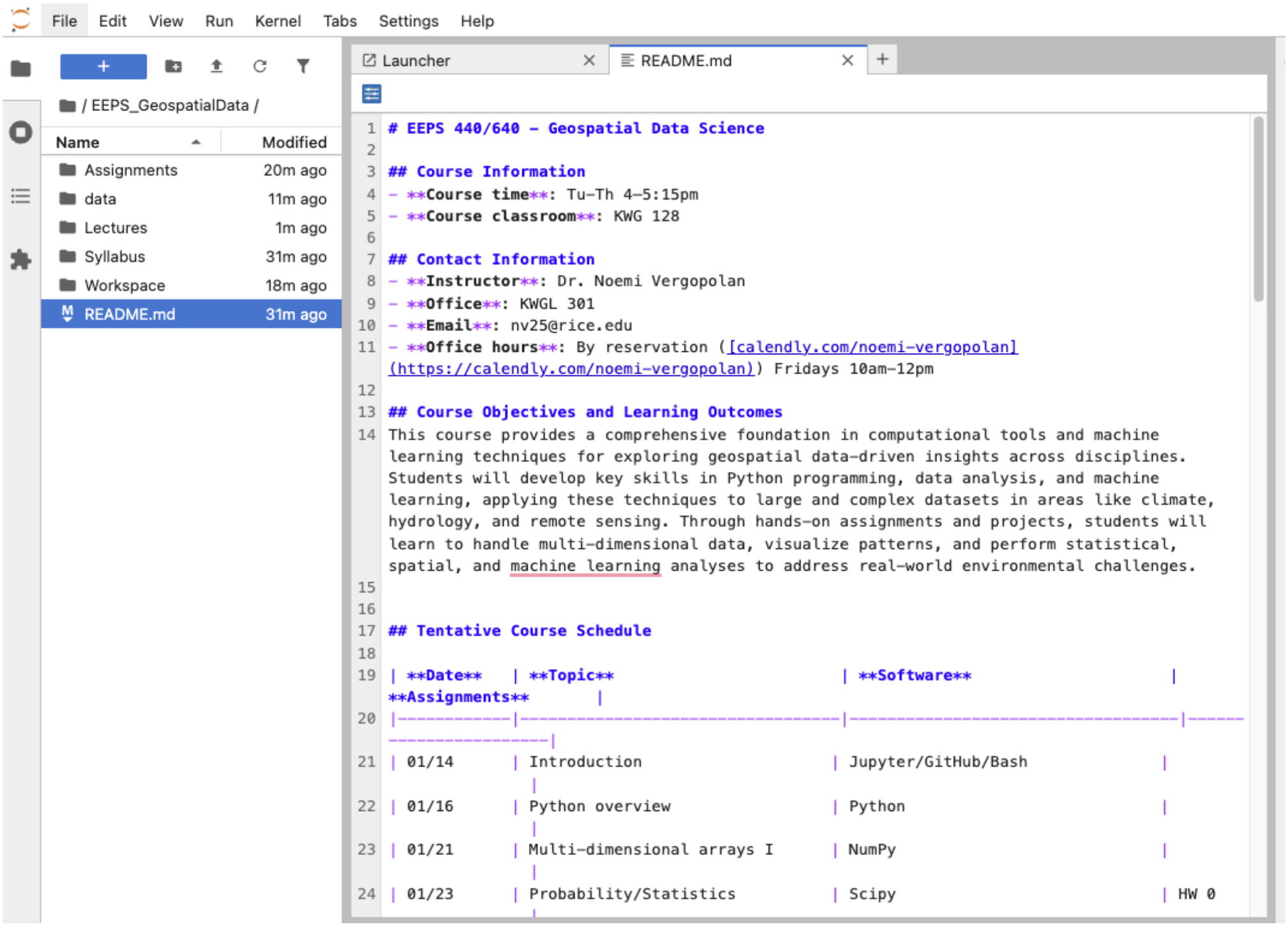


git clone

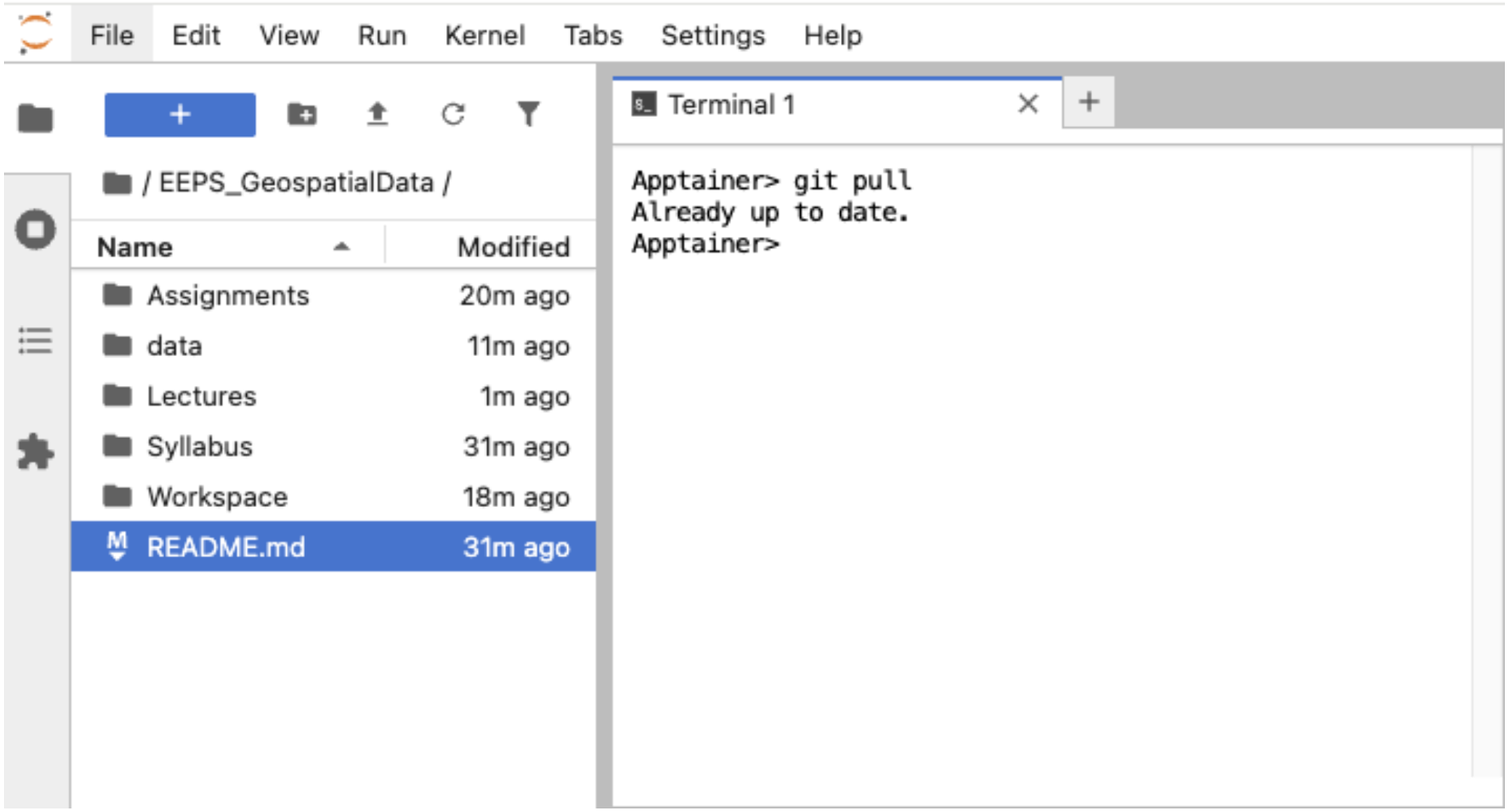
https://github.com/NoemiVergopolan/EEPS_GeospatialData.git



Explore the cloned directory



Git pull



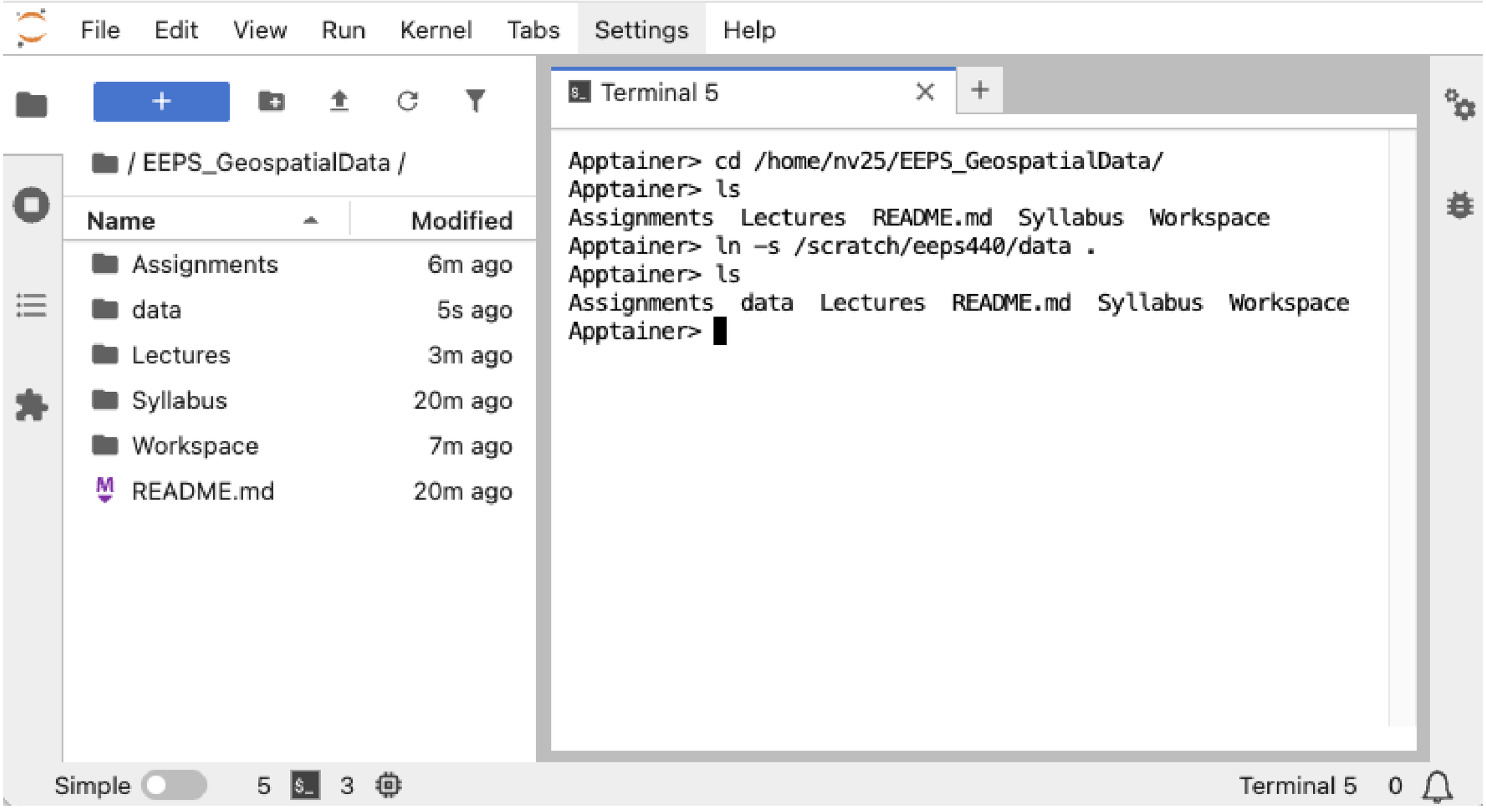
Press enter after writing "git pull"

This is how you will have update the class repository on your system (i.e., as lectures and assignments are added online)

Let's create a symbolic link to our datasets with:

cd /home/USER/EEPS_GeospatialData

ln –s /scratch/eeps440/data .

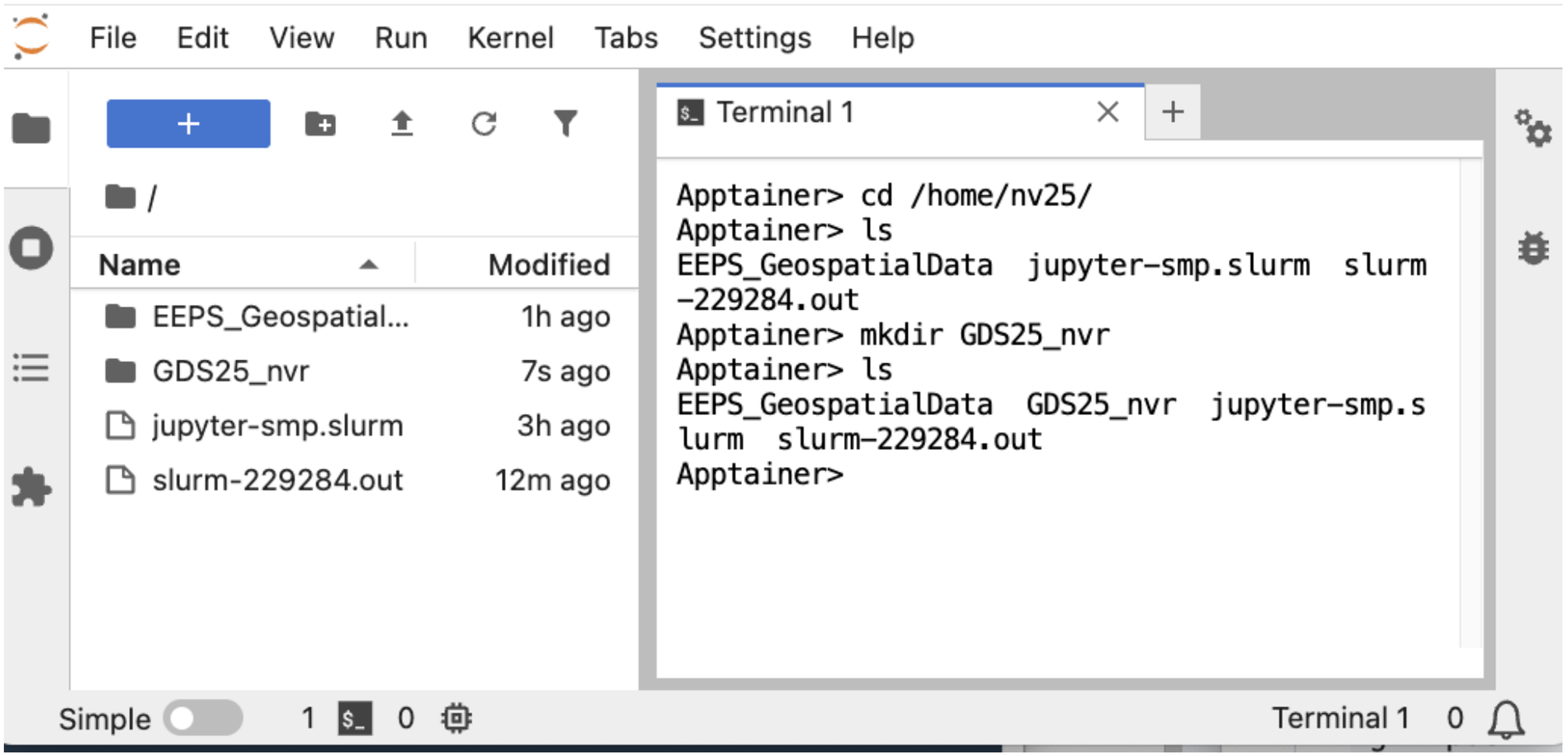


You will be doing all of your work on your container.

But how will you access the assignments and how will you submit them?

Make a GitHub repository to turn in your assignments

cd /home/USER
mkdir GDS25_nvr



Create a directory (use your own initials)

Enter directory

cd GDS25_nvr

Create git repository

git init

Copy over assignment HW0

cp ../EEPS_GeospatialData/Assignments/HW0.ipynb .

Add assignment to your repository

git add HW0.ipynb

Commit changes to repository

git commit -m 'add HW0 file'

Link your folder with your existing epository

git branch -M main

git remote add origin https://github.com/NoemiVergopolan/GDS25_nvr.git

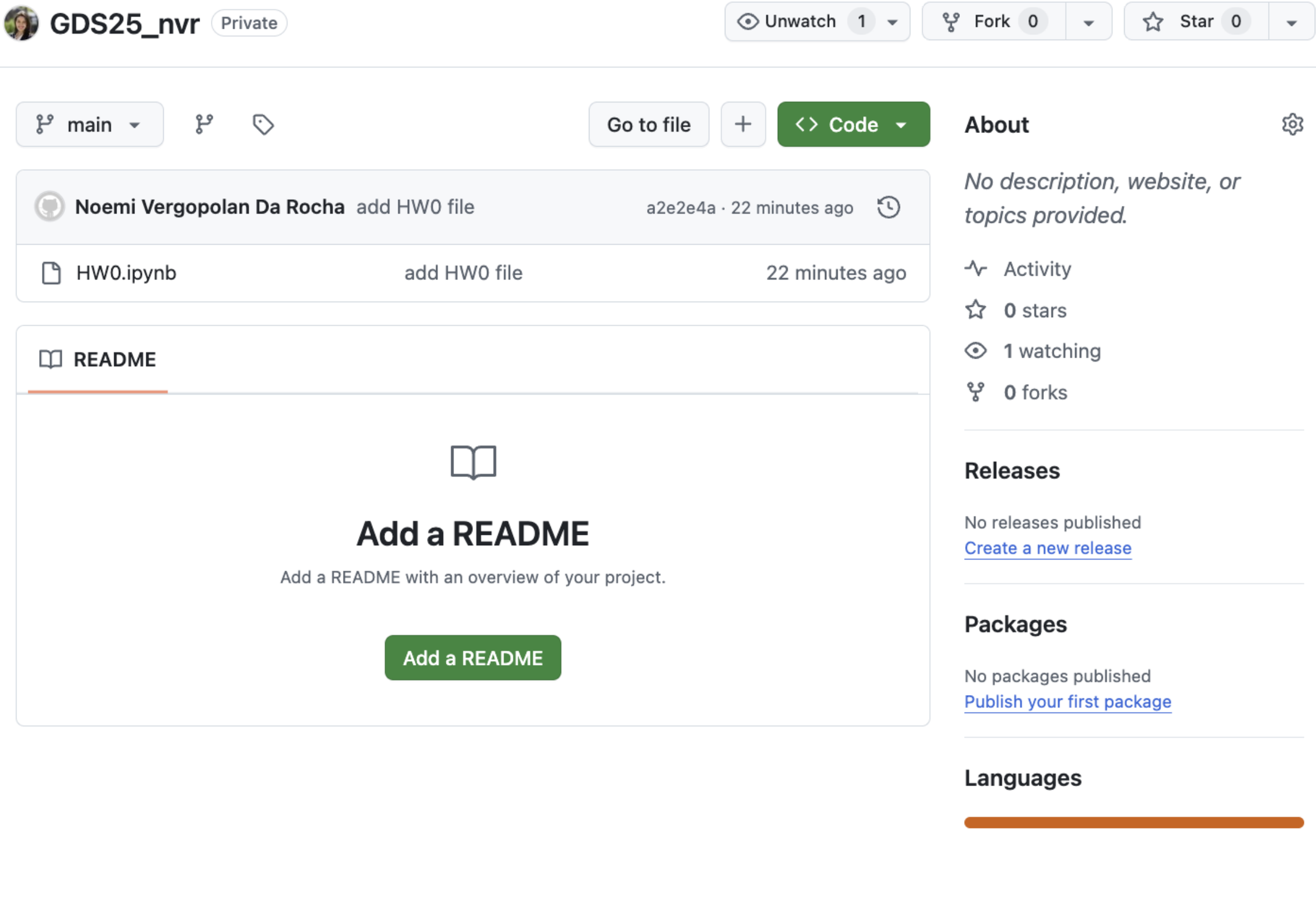
Push your HW to your repository

git push -u origin main

UserName

Token Password

Now they are attached to each other! And the online version has been updated.



With every new assignment:

- Copy the assignment from the class repository
- Add it to your private repository (git add)
- Save and commit the changes (git commit)
- Complete the assignment
- Save and commit the changes (git commit)
- Push the changes to the online repository (git push)

- But why so complicated? Why not just send the completed assignment via email?
- Because using version control is critical to most research and industry data science nowadays.
- Forcing you to use version control throughout the course will ensure you learn how to use it.

Assignments submitted any other way will NOT be accepted.

HW0

Due February 6 (before class)

Get your container up and running, clone the class repository, create and link your private repository, and solve some introductory Python exercises.

This will be hard for many, but if we get this out of the way at the beginning, we will avoid a lot of headaches moving forward.