

EEPS 440/640

Geospatial Data Science (GDS)

The primary class website for this course is on GitHub

The screenshot shows the GitHub interface for the repository **EEPS_GeospatialData** by user **NoemiVergopolan**. The repository is marked as **Private**. The top navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Security, Insights, and Settings. The repository header shows 1 branch, 0 tags, and 4 commits. The file list includes Assignments, Lectures, Syllabus, and README.md, all recently updated. The README content is partially visible, showing the title **EEPS 440/640 - Geospatial Data Science**. The right sidebar contains sections for About (EEPS 440/460 Geospatial Data Science), Releases (No releases published), and Packages (No packages published).

Navigation: <> Code | Issues | Pull requests | Actions | Projects | Security | Insights | Settings

Repository: **EEPS_GeospatialData** (Private) | Unwatch 1 | Fork 0 | Star 0

Branches: main | 1 Branch | 0 Tags | Go to file | Add file | <> Code

Commits: NoemiVergopolan setup main page (9c5ac61 · now) | 4 Commits

File	Commit	Time
Assignments	setup main page	10 minutes ago
Lectures	setup main page	now
Syllabus	setup main page	10 minutes ago
README.md	setup main page	5 minutes ago

README

EEPS 440/640 - Geospatial Data Science

About
EEPS 440/460 Geospatial Data Science
Readme | Activity | 0 stars | 1 watching | 0 forks

Releases
No releases published
[Create a new release](#)

Packages
No packages published

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

You have your own repository

The screenshot shows a GitHub repository page for 'GDS25_nvr' owned by 'NoemiVergopolan'. The repository is private and has 1 watcher, 0 stars, and 0 forks. The main branch is 'main'. A recent commit 'add HW0 file' by 'Noemi Vergopolan Da Rocha' is shown, with a file 'HW0.ipynb' added 43 minutes ago. The README section prompts the user to 'Add a README' with a button labeled 'Add a README'. The right sidebar shows the 'About' section with no description, and sections for 'Releases' (no releases published), 'Packages' (no packages published), and 'Languages' (Jupyter Notebook 100.0%).

Navigation bar: NoemiVergopolan / GDS25_nvr

Repository: GDS25_nvr (Private)

Buttons: Unwatch (1), Fork (0), Star (0)

Branch: main

Commit: Noemi Vergopolan Da Rocha add HW0 file (a2e2e4a · 43 minutes ago)

File: HW0.ipynb add HW0 file (43 minutes ago)

README: Add a README (Add a README button)

Activity: 0 stars, 1 watching, 0 forks

Releases: No releases published (Create a new release link)

Packages: No packages published (Publish your first package link)

Languages: Jupyter Notebook 100.0%

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 **queue -u USER**

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

```
[I 2025-01-23 11:23:59.287 ServerApp] jupyterlab | extension was successfully linked.
[W 2025-01-23 11:23:59.287 NotebookApp] 'ip' 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] '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 ServerApp. 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 ~]$
```

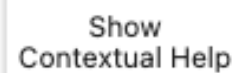
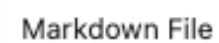
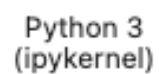
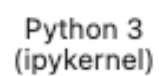
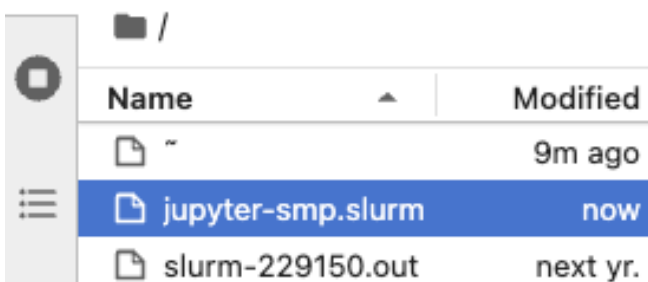

- Open a **NEW** terminal

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

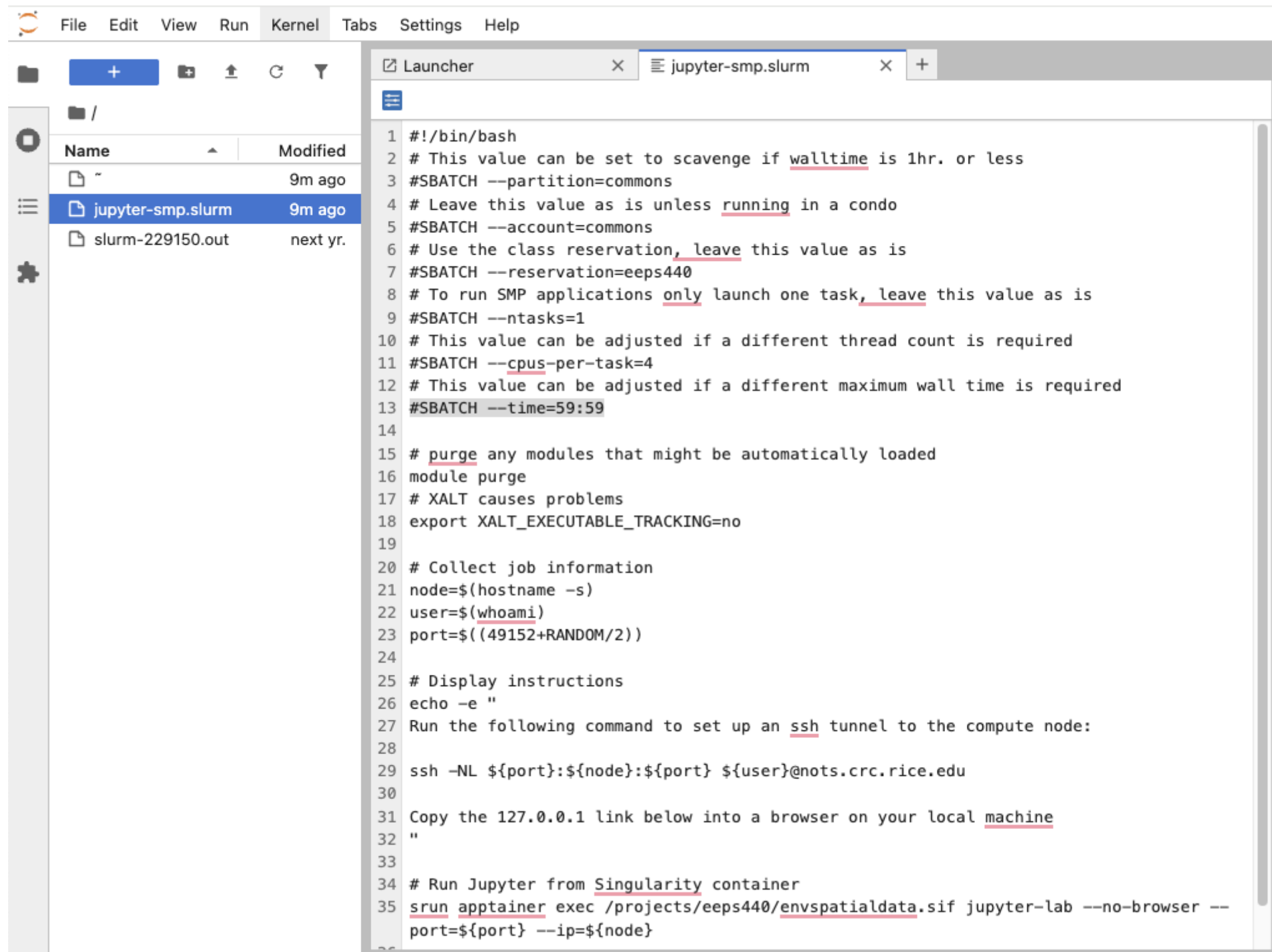
- Type the password and let it be
- Open your internet browser and paste the link with 127.0.....

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

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



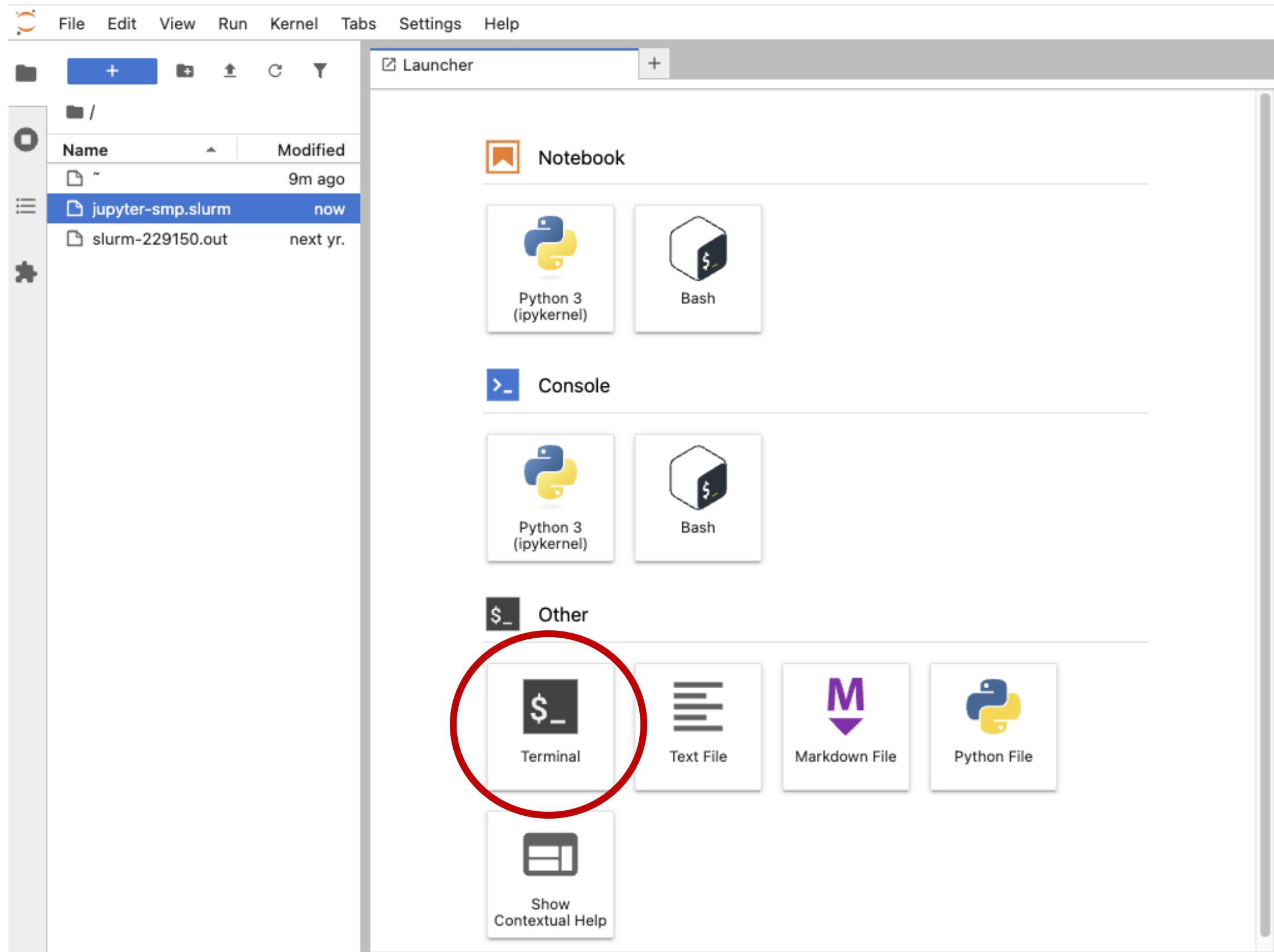
Your job submission script template defaults to 30min run. Let's change it to 59:59 min...



```
1 #!/bin/bash
2 # This value can be set to scavenge if walltime is 1hr. or less
3 #SBATCH --partition=commons
4 # Leave this value as is unless running in a condo
5 #SBATCH --account=commons
6 # Use the class reservation, leave this value as is
7 #SBATCH --reservation=eeps440
8 # To run SMP applications only launch one task, leave this value as is
9 #SBATCH --ntasks=1
10 # This value can be adjusted if a different thread count is required
11 #SBATCH --cpus-per-task=4
12 # This value can be adjusted if a different maximum wall time is required
13 #SBATCH --time=59:59
14
15 # purge any modules that might be automatically loaded
16 module purge
17 # XALT causes problems
18 export XALT_EXECUTABLE_TRACKING=no
19
20 # Collect job information
21 node=$(hostname -s)
22 user=$(whoami)
23 port=$((49152+RANDOM/2))
24
25 # Display instructions
26 echo -e "
27 Run the following command to set up an ssh tunnel to the compute node:
28
29 ssh -NL ${port}:${node}:${port} ${user}@nots.crc.rice.edu
30
31 Copy the 127.0.0.1 link below into a browser on your local machine
32 "
33
34 # Run Jupyter from Singularity container
35 srun apptainer exec /projects/eeps440/envspatialdata.sif jupyter-lab --no-browser --
  port=${port} --ip=${node}
```

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

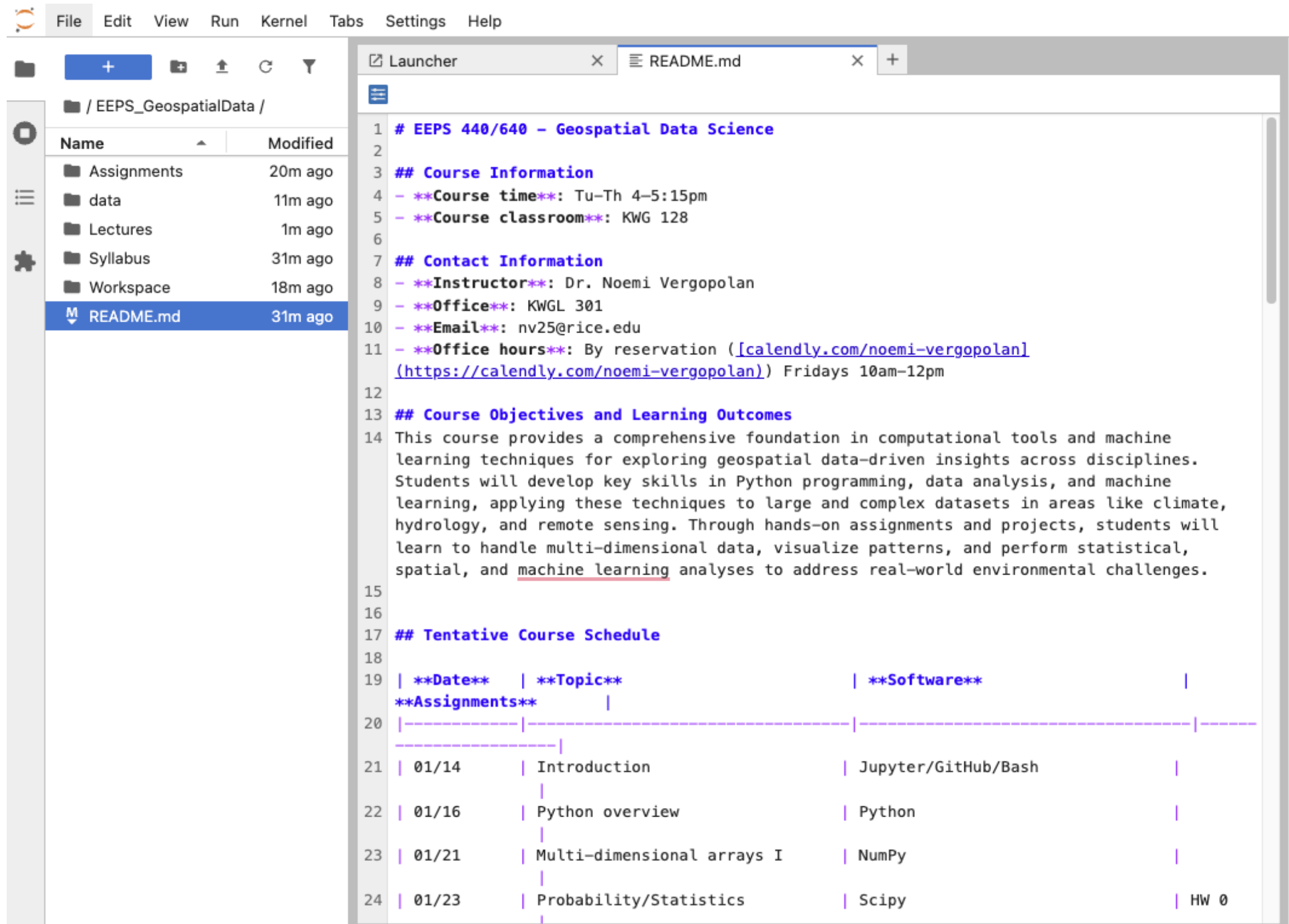
The screenshot shows a JupyterLab environment. On the left is a file browser with a sidebar containing icons for home, recent, and search. The main area lists files and folders:

Name	Modified
/	
EEPS_Geospatial...	8s ago
~	23m ago
jupyter-smp.slurm	5m ago
slurm-229150.out	3m ago

On the right is a terminal window titled "Terminal 2" with the following text:

```
Apptainer> git clone https://github.com/NoemiVergopolan/EEPS_GeospatialData
Cloning into 'EEPS_GeospatialData'...
Username for 'https://github.com':
```

Explore the cloned directory



The screenshot shows a JupyterLab environment. On the left, a file explorer displays the directory structure of a cloned repository named 'EEPS_GeospatialData'. The files listed are 'Assignments', 'data', 'Lectures', 'Syllabus', 'Workspace', and 'README.md'. The 'README.md' file is selected and highlighted in blue.

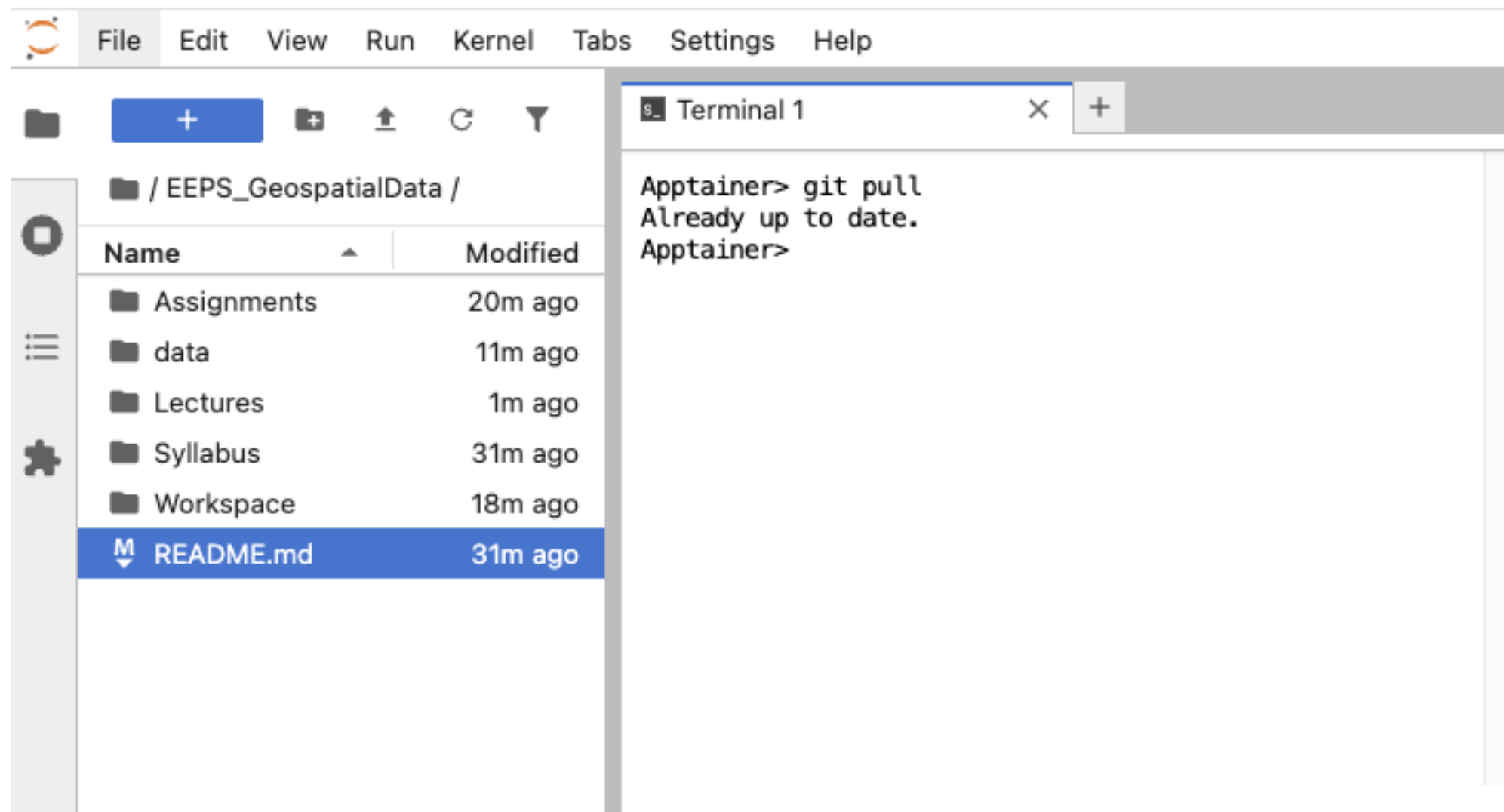
On the right, the code editor displays the content of the 'README.md' file. The content is formatted as a Markdown document with the following sections:

- # EEPS 440/640 - Geospatial Data Science**
- ## Course Information**
 - **Course time**:** Tu–Th 4–5:15pm
 - **Course classroom**:** KWG 128
- ## Contact Information**
 - **Instructor**:** Dr. Noemi Vergopolan
 - **Office**:** KWGL 301
 - **Email**:** nv25@rice.edu
 - **Office hours**:** By reservation ([\[calendly.com/noemi-vergopolan\]](https://calendly.com/noemi-vergopolan)) Fridays 10am–12pm
- ## Course Objectives and Learning Outcomes**

This course provides a comprehensive foundation in computational tools and machine learning techniques for exploring geospatial data-driven insights across disciplines. Students will develop key skills in Python programming, data analysis, and machine learning, applying these techniques to large and complex datasets in areas like climate, hydrology, and remote sensing. Through hands-on assignments and projects, students will learn to handle multi-dimensional data, visualize patterns, and perform statistical, spatial, and machine learning analyses to address real-world environmental challenges.
- ## Tentative Course Schedule**

Date	**Topic**	**Software**	
			Assignments
01/14	Introduction	Jupyter/GitHub/Bash	
01/16	Python overview	Python	
01/21	Multi-dimensional arrays I	NumPy	
01/23	Probability/Statistics	Scipy	HW 0

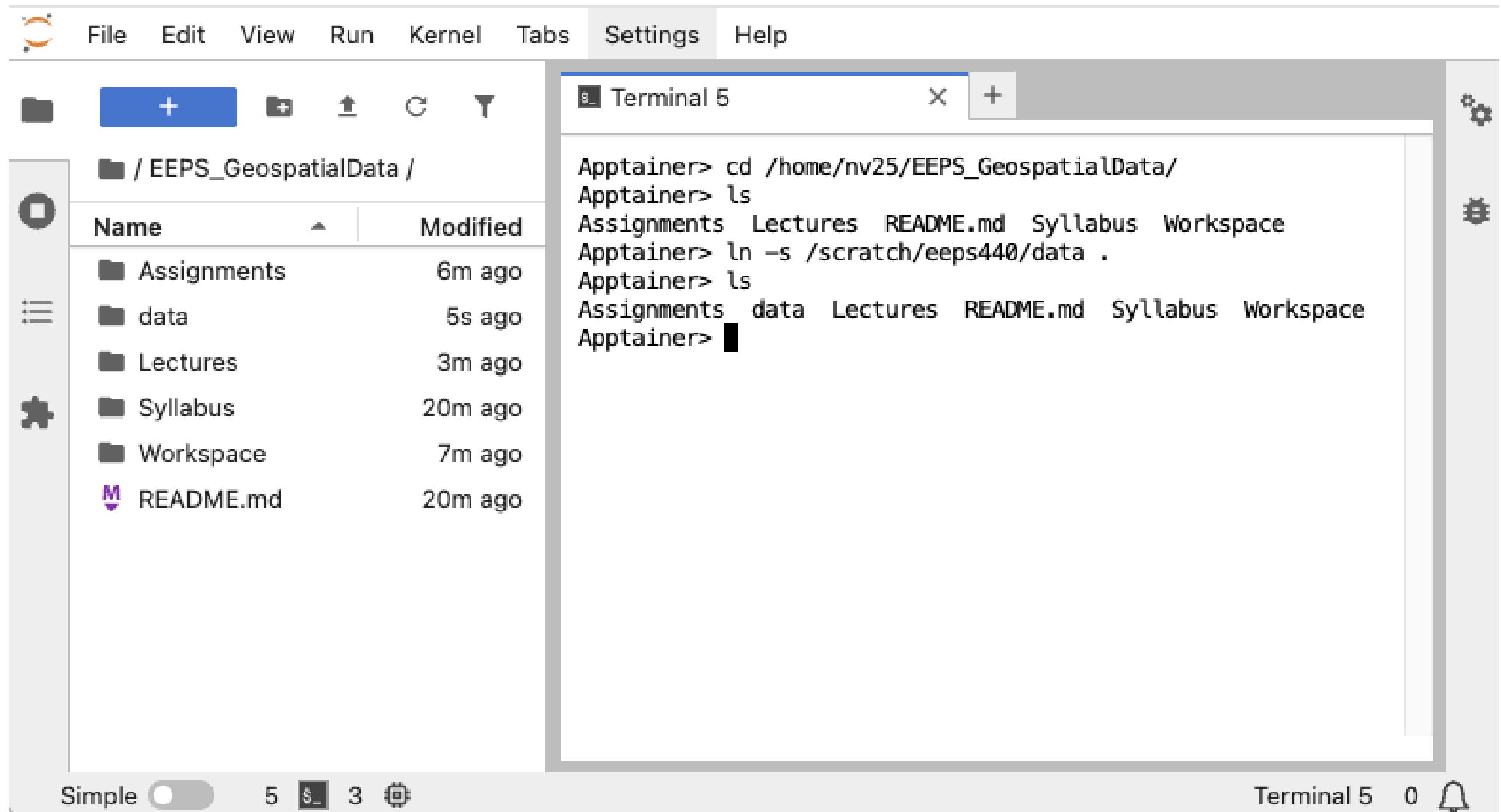
Git pull



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 .`



The screenshot displays the JupyterLab interface. On the left, the file browser shows the directory `/ EEPS_GeospatialData /` with a table of files and folders:

Name	Modified
Assignments	6m ago
data	5s ago
Lectures	3m ago
Syllabus	20m ago
Workspace	7m ago
README.md	20m ago

On the right, the 'Terminal 5' window shows the following commands and output:

```
Apptainer> cd /home/nv25/EEPS_GeospatialData/  
Apptainer> ls  
Assignments Lectures README.md Syllabus Workspace  
Apptainer> ln -s /scratch/eeps440/data .  
Apptainer> ls  
Assignments data Lectures README.md Syllabus Workspace  
Apptainer> █
```

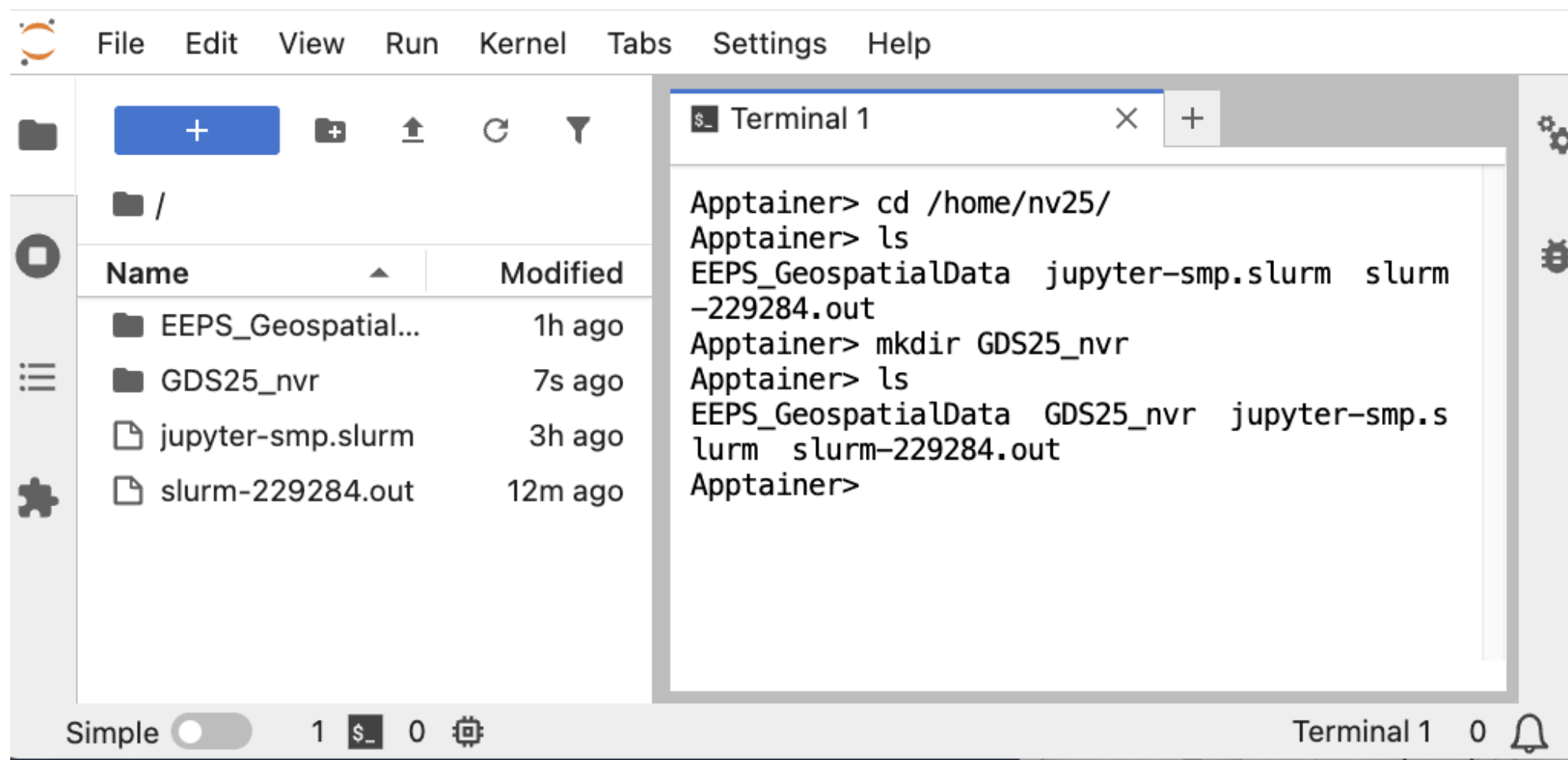
The bottom status bar indicates 'Simple' mode, 5 files, 3 terminals, and the active terminal is 'Terminal 5'.

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 repository

```
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.


 **GDS25_nvr** Private


Unwatch 1 Fork 0 Star 0

main Go to file + <> Code

 **Noemi Vergopolan Da Rocha** add HW0 file a2e2e4a · 22 minutes ago

 HW0.ipynb add HW0 file 22 minutes ago

 **README**



Add a README

Add a README with an overview of your project.

Add a README

About

No description, website, or topics provided.

Activity

0 stars

1 watching

0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

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.