

EEPS 440/640

# Geospatial Data Science (GDS)

# Introduction

- Instructor: Dr. Noemi Vergopolan

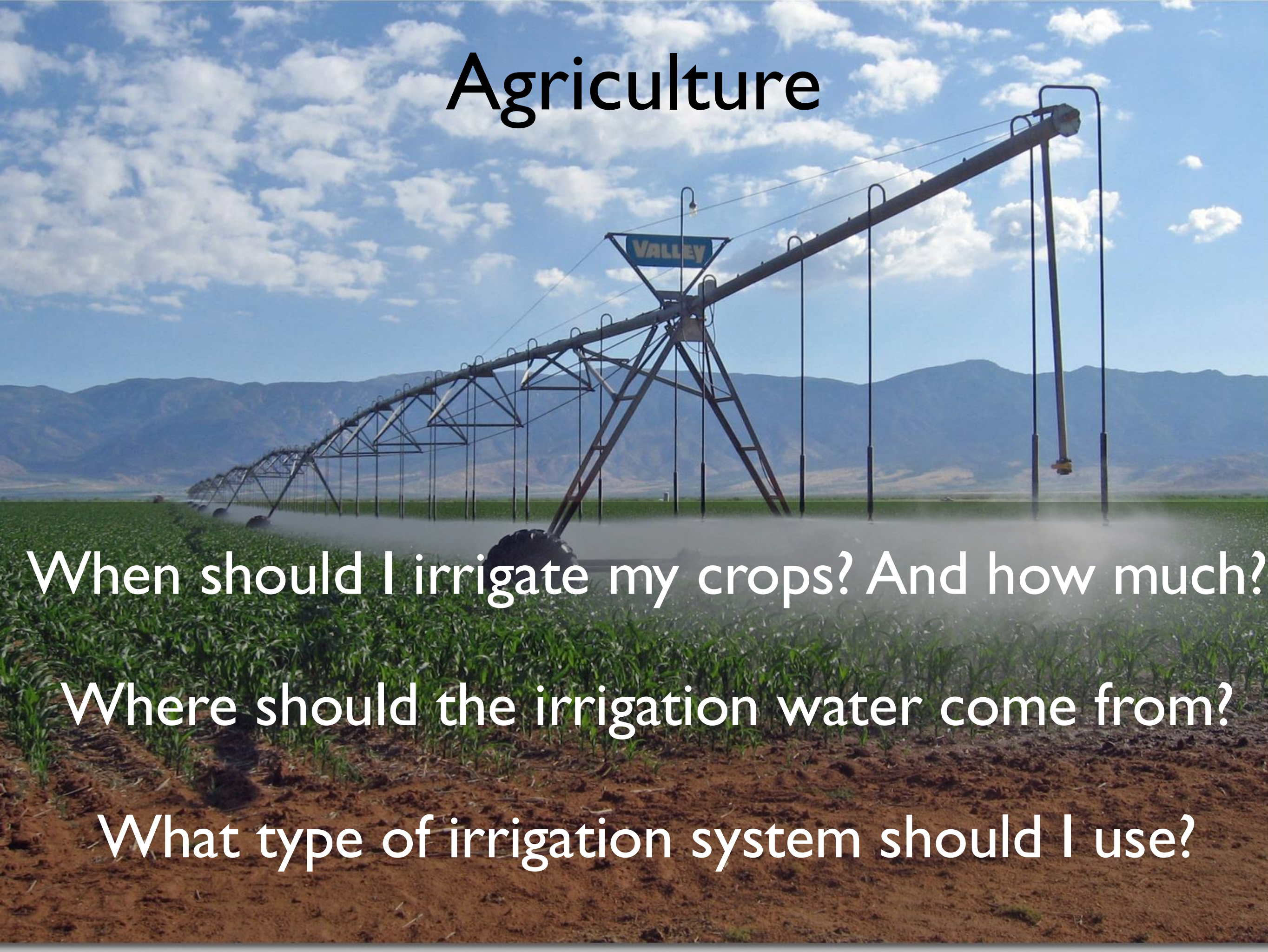
# Motivation

There are big environmental challenges that we want to take on...

Here are some examples from Hydrology



# Agriculture



When should I irrigate my crops? And how much?

Where should the irrigation water come from?

What type of irrigation system should I use?



# Floods

An aerial photograph showing a residential neighborhood completely inundated with floodwater. The water is dark and still, reflecting the sky. Several houses with dark roofs are partially submerged. A large green tree stands in the water on the left. A road with yellow lane markings runs through the center, with a utility pole and power lines crossing it. In the background, a red building is visible. The overall scene depicts the severe impact of flooding on a community.

Where will it flood during a hurricane?

How can we mitigate the impacts of floods?

What is the risk that my house is flooded?



# Water management

An aerial photograph of a large concrete dam. Water is held back on the left side of the dam, forming a reservoir. On the right side, water is cascading over the dam's spillways, creating white rapids and a waterfall. The surrounding landscape is arid and rocky, with some sparse green vegetation. A small building with a green roof is visible near the dam's base on the left.

Where should I place a reservoir?

How much water should I release?

How will a changing climate impact water inflows  
into existing reservoirs?



# Wildfires

A dramatic night scene of a house engulfed in flames. A massive fireball is visible in the sky above the house, which has several windows glowing with light. Bare tree branches are silhouetted against the bright fire.

Which conditions trigger wildfires?

Are wildfires caused by climate change?

What is the risk that my house is burned?

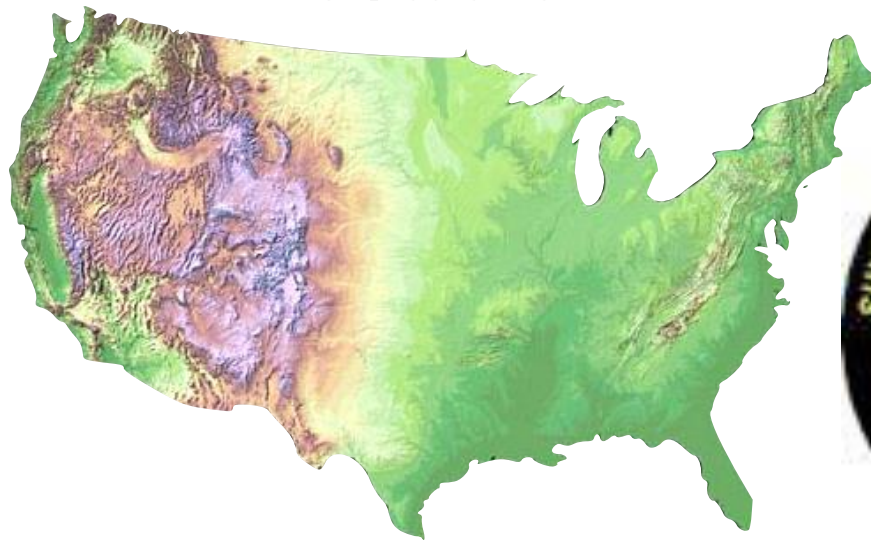
How do we address those  
questions?

Data, data, and more data



# I. Environmental Data

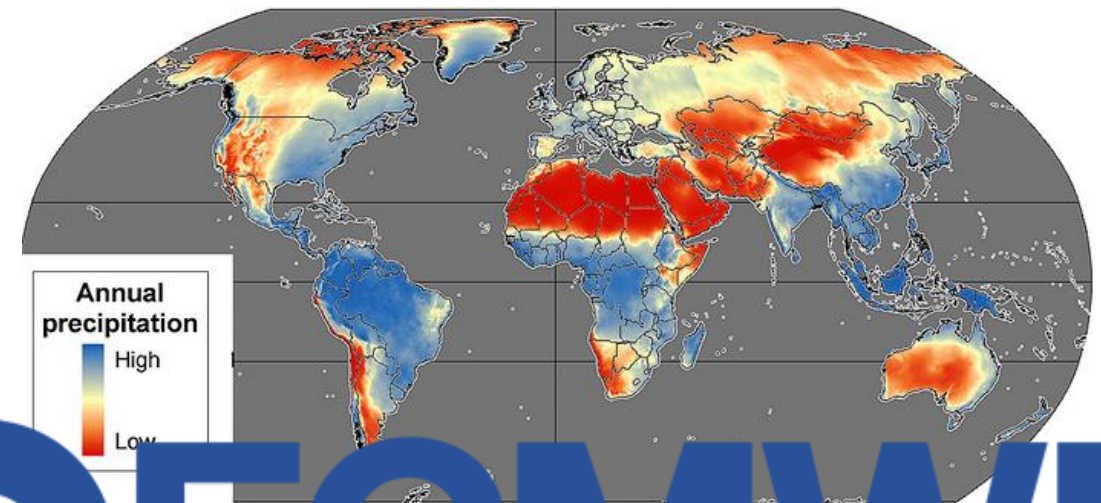
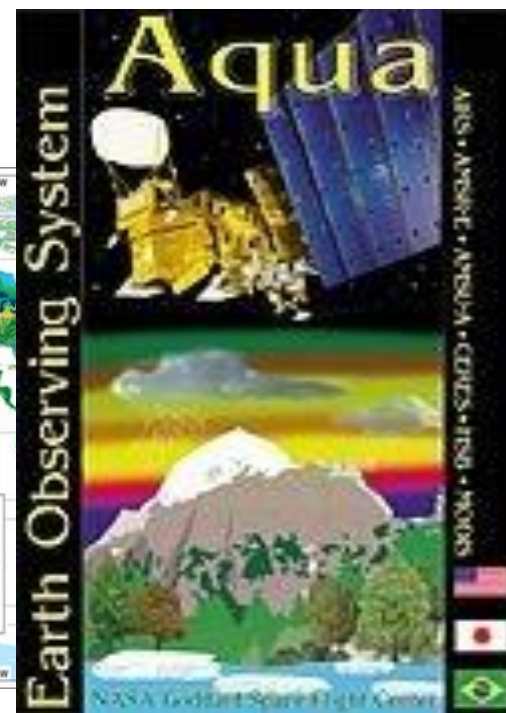
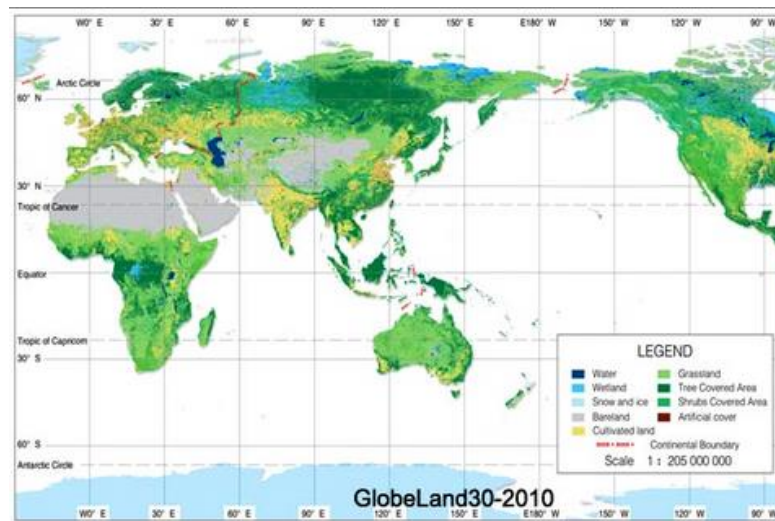
## Elevation



## Soil



## Land Cover





# I. Environmental Data

Elevation

CZO

CRITICAL ZONE OBSERVATORIES

Soil

CRDC

LANDSAT

GOES-R

“Age of Data”

Land Cover

Climate

WATCH  
Water and Global Change

THE LONG TERM  
ECOLOGICAL  
RESEARCH  
NETWORK

EUROPEAN CENTRE FOR MEDIUM RANGE WEATHER FORECASTS



# What Petabytes of data feels like

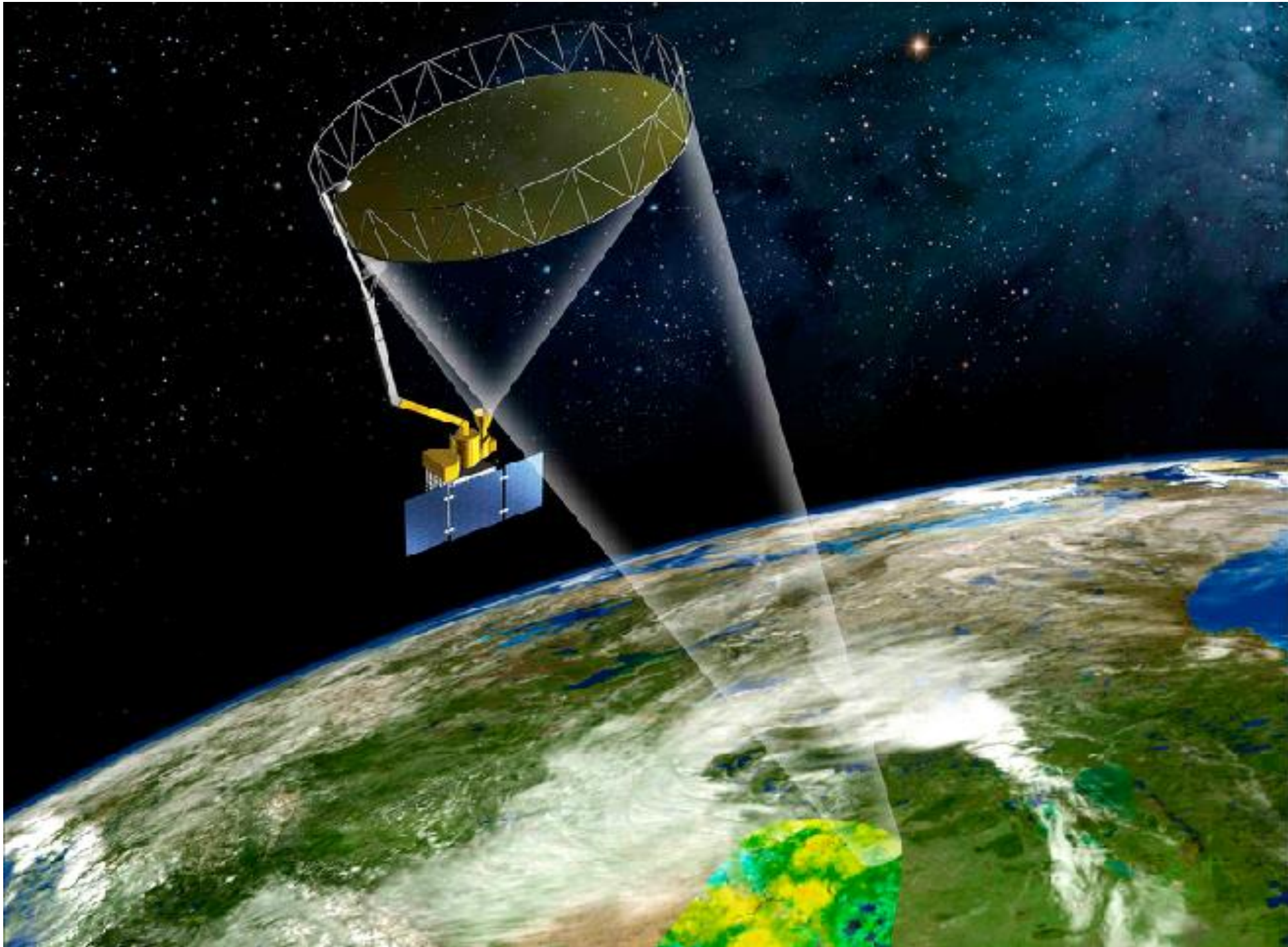




And the problem is even much worse with geospatial data

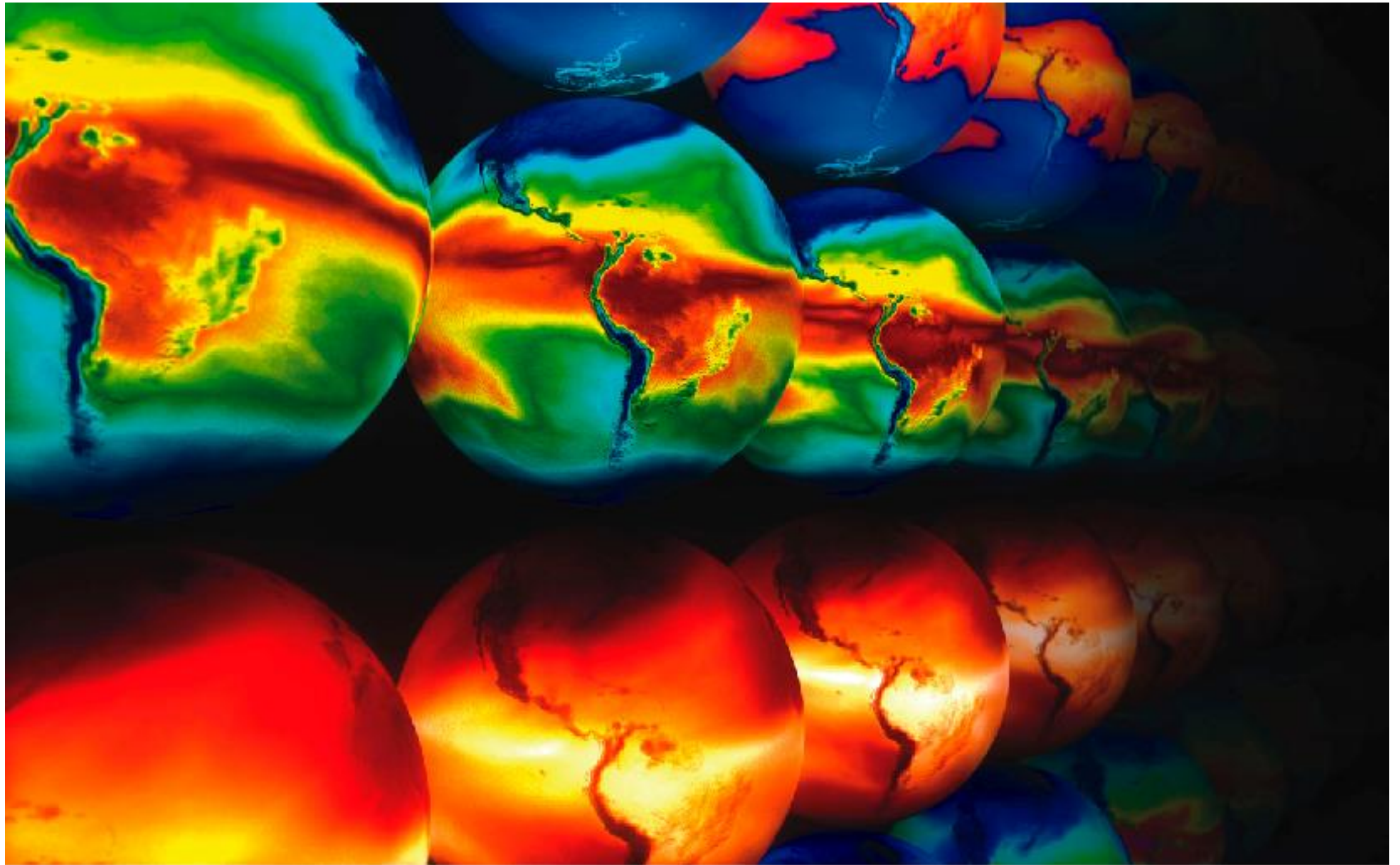


# Example: Satellite remote sensing





# Example: Climate model output





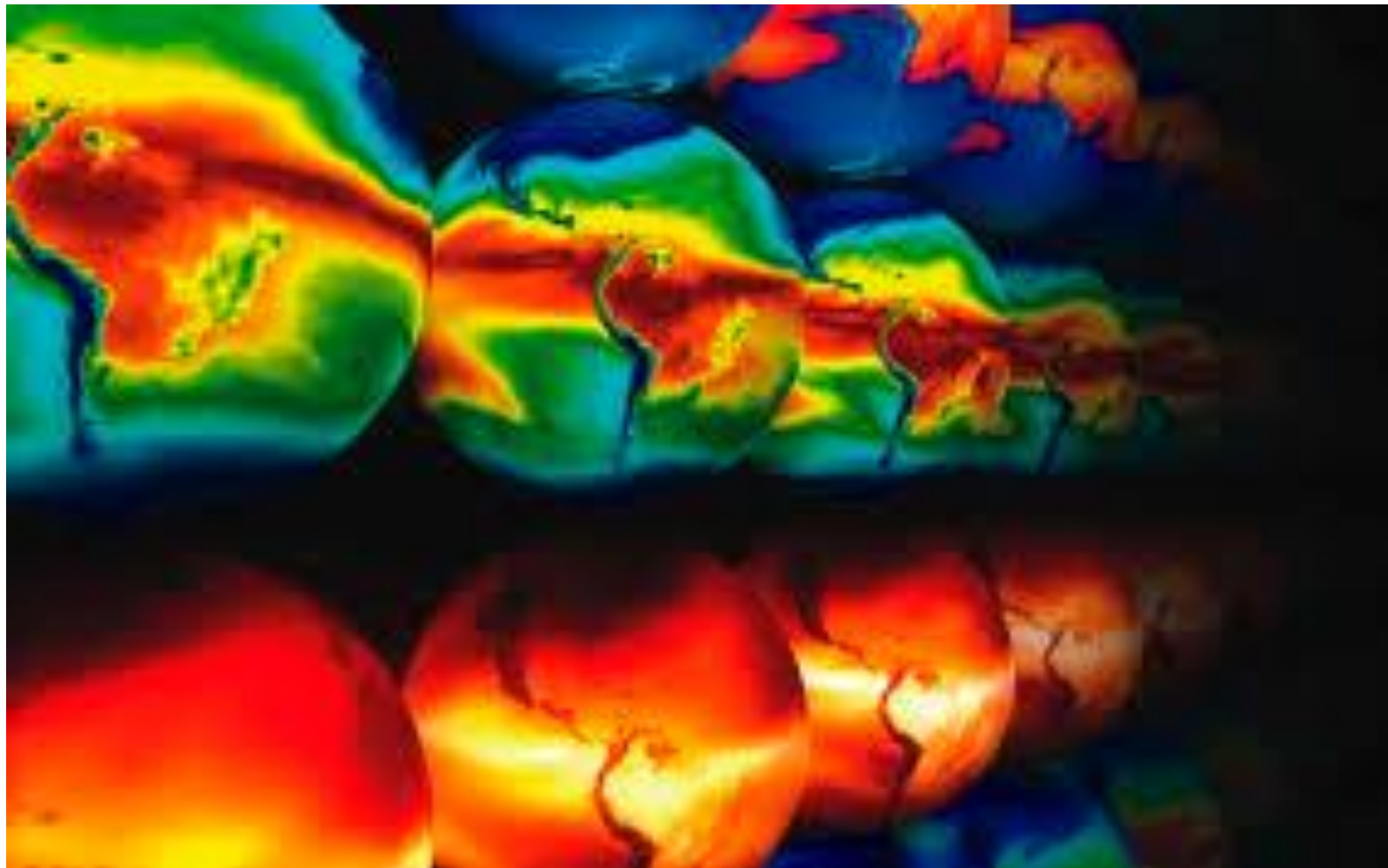
This course aims to give you  
the tools and knowledge on  
how to tackle these data



**What topics will we cover?**



# Topics: Analyze spatial environmental data



How can we analyze terabytes/petabytes of climate model output and/or satellite data?



# Topics: Map projections



Robinson



Plate Carrée



Winkel Tripel



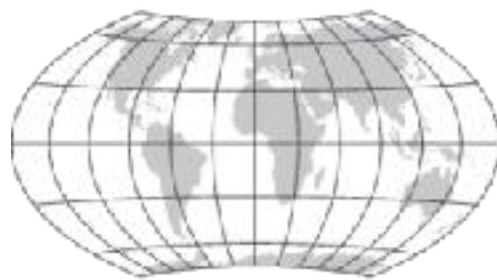
Eckert IV



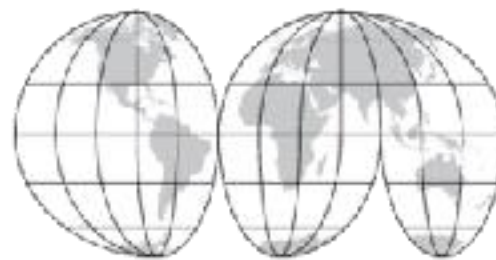
Mollweide



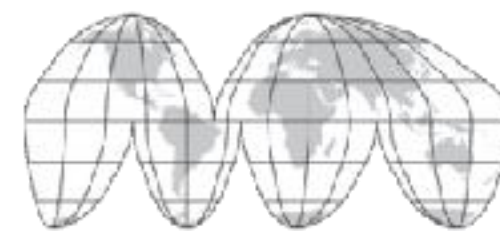
Mercator



Wagner VII



Interrupted Mollweide

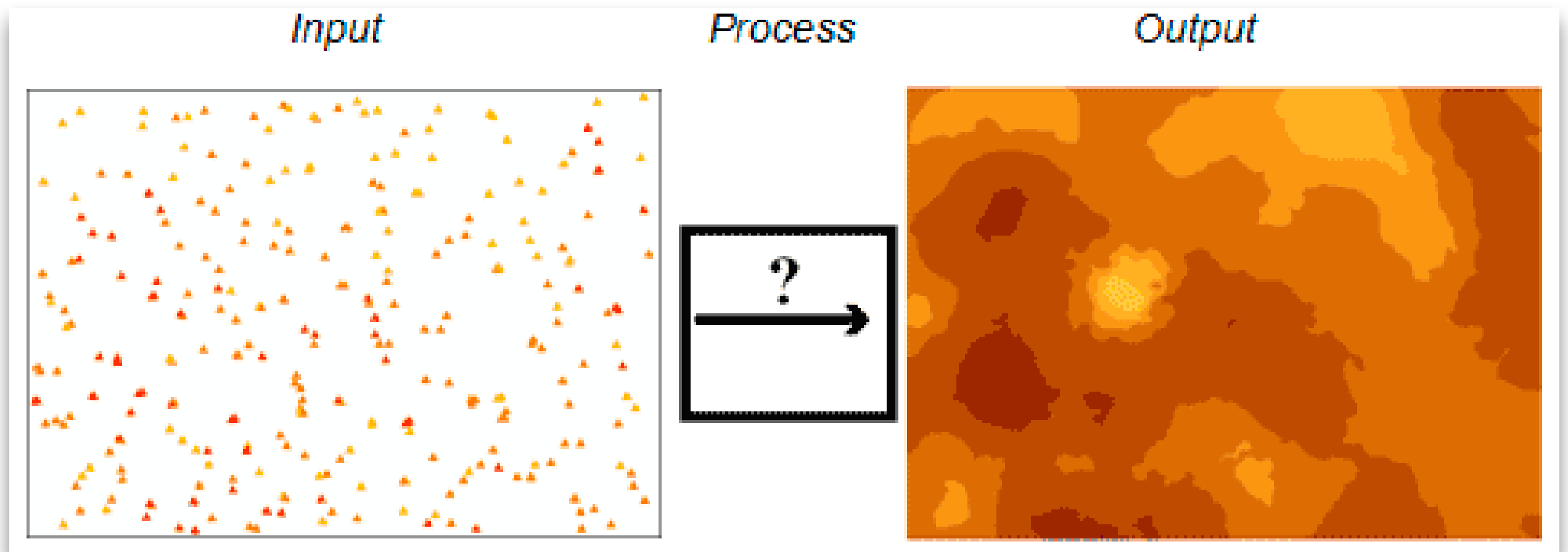


Goode Homolosine

How can we transform a map to a different projection?



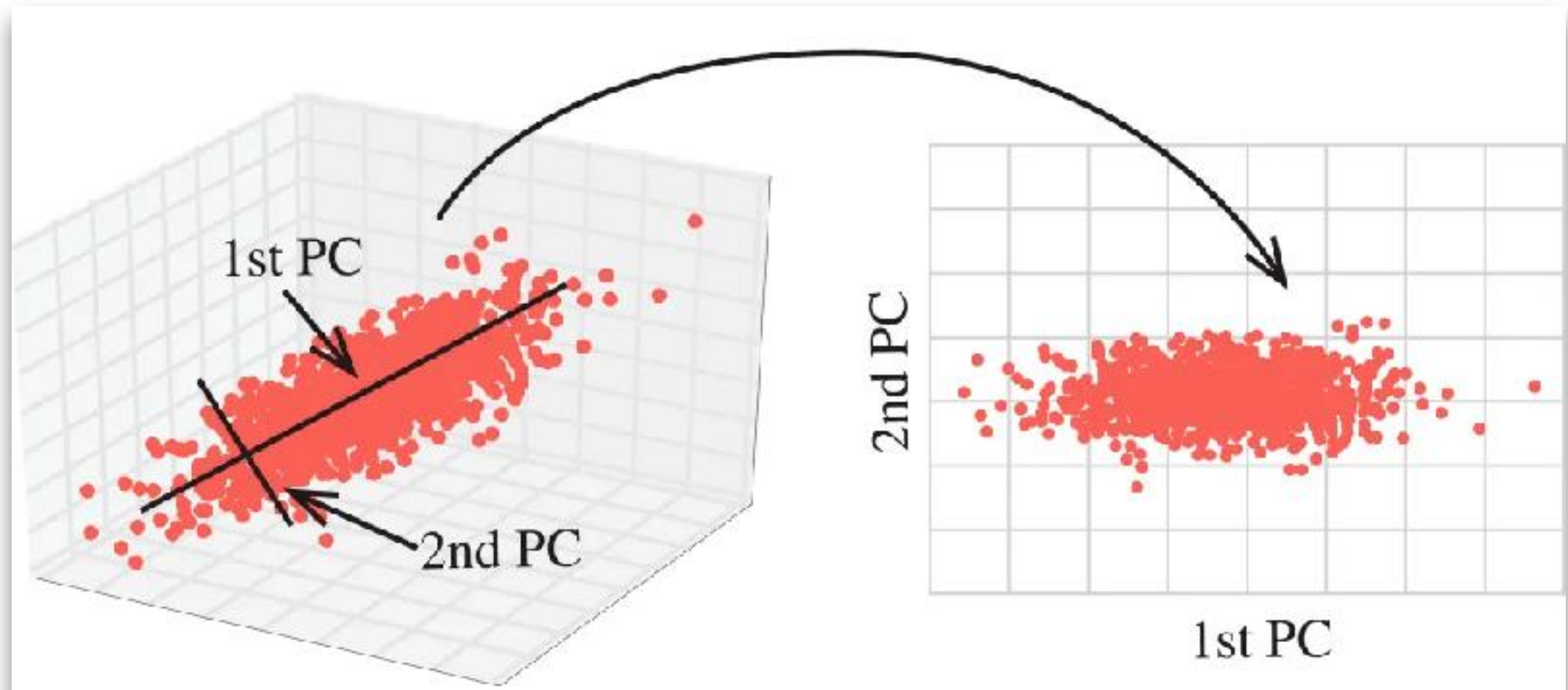
# Topics: Spatial interpolation/ prediction



How do you create a spatial map from point data?



# Topics: Reduce dimensions



How do we compress data?

What are the most important features in the data?

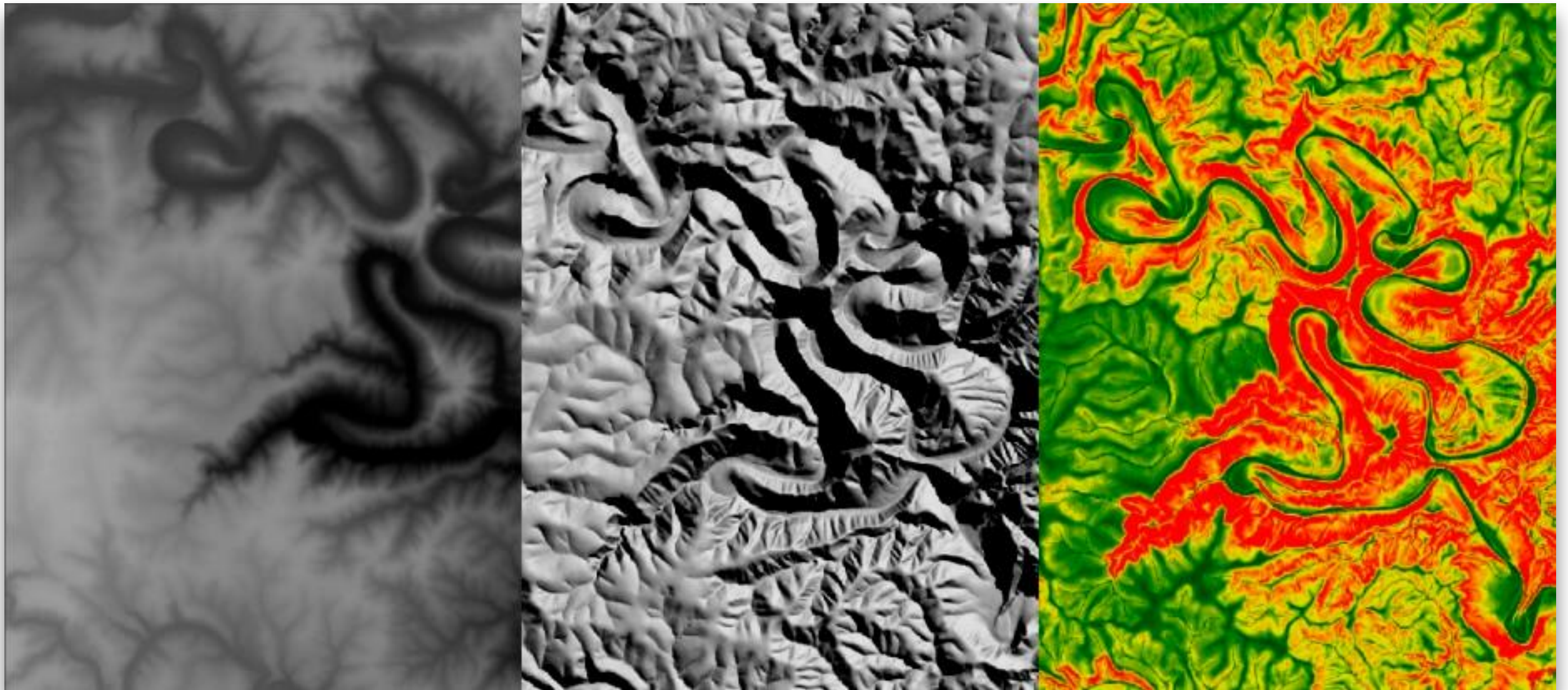


# Topics: Terrain Analysis

Elevation

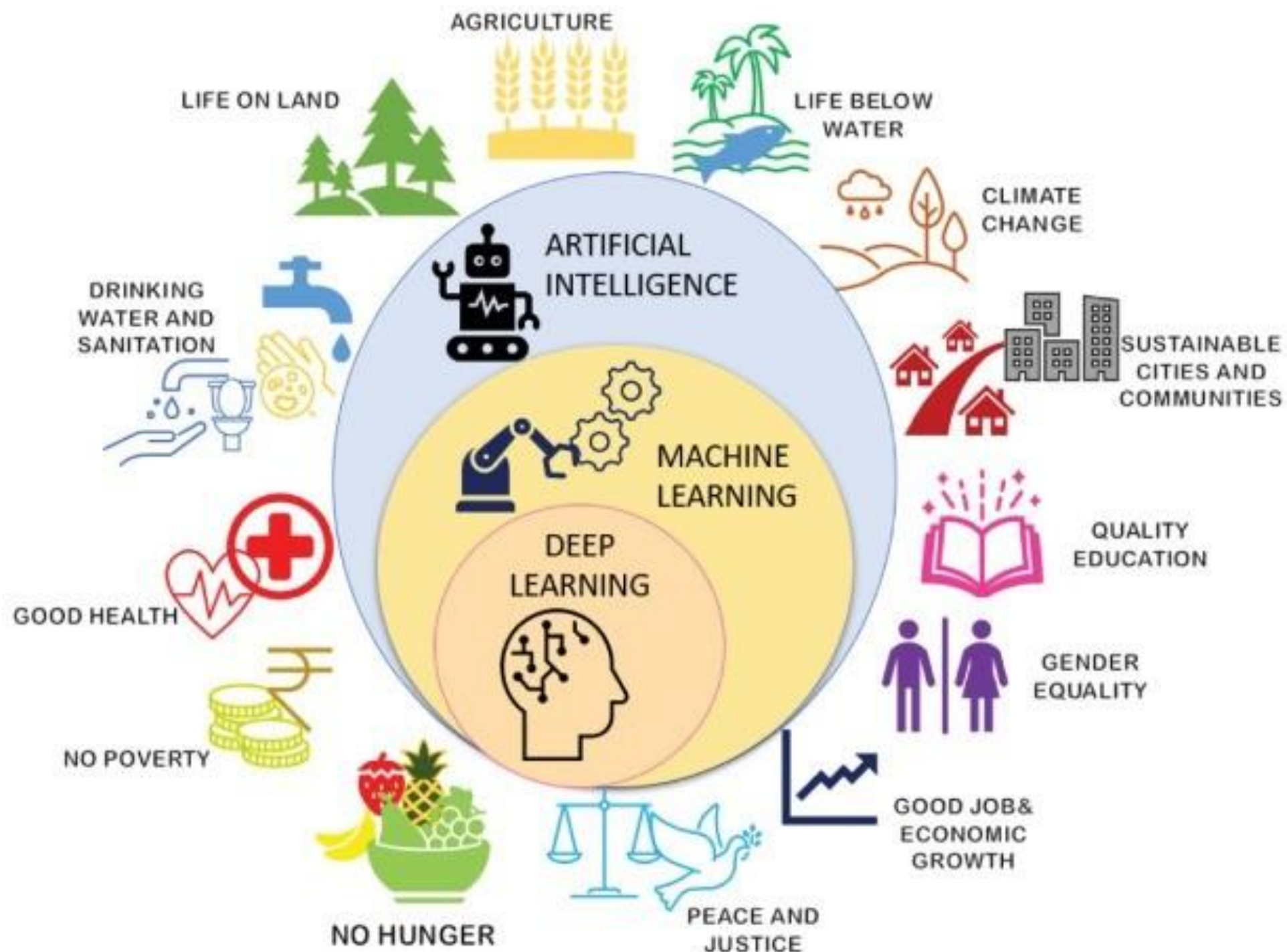
Hillshading

Slope



What can we learn from elevation data?

# Integration of ML methods for geospatial data science





And more...

But what tools will we use to  
learn how to create and analyze  
these large datasets?



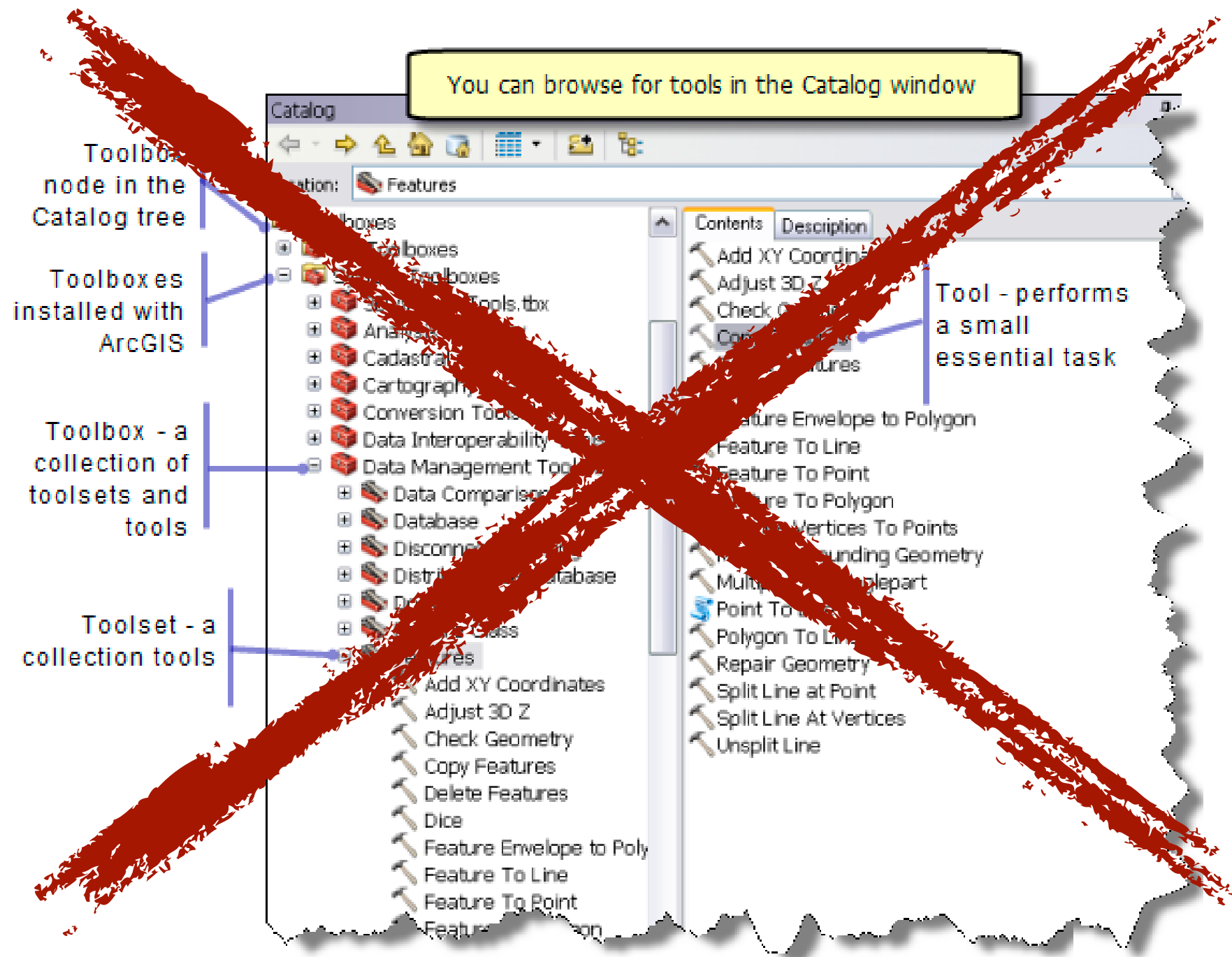
Let's start off with what we will  
NOT be using this semester

This is NOT an ArcGIS course





# This is NOT a “click the button” class



If that was what you were expecting then  
this may not be the class for you

If you want to learn ArcGIS  
Rice has resources for you:

GIS/Data Center (GDC)

<https://library.rice.edu/places/gisdata-center-gdc>

<https://wiki.rice.edu/confluence/x/QIAKAg>



Then what will we use?

# Programming, programming, and more programming

If you are terrified of programming...  
Don't be! It is worth the time investment.

However, if you are not willing to dedicate  
the time and just want to “pass by” then this  
class is probably not a good fit.



Which programming  
languages will we use?

# Primary language: Python



Python level: Intermediate



# Secondary language: Bash



**Bash level: Beginner**

# Don't be afraid...

- We are going to spend the first third of the class learning how to program in Python (and Bash)
- It will get hard, but it will be additive so one can take this class without a background in these programming languages



# Rice Computing Resources

## Python Programming Series

<https://library.rice.edu/services/data-workshops>

Intro to programming – Jan. 24, 1-3 pm

Loops and conditions – Jan. 31, 1-3 pm

Data analysis and Viz w/ Pandas – Feb. 21, 1-2:30 pm

Text analysis – Apr. 8, 1-2:30 pm

## Bash - The World is at your Command...line:

<https://library.rice.edu/courses/world-your-commandline>

Jan. 29, 1-4 pm

# Syllabus



# Course Logistics: General Info

## 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 Logistics: General Info

## Grade Policies

The final grade for **undergraduates** will comprise:

10% Class Participation

40% Homeworks

50% Final Project: Proposal (10%), Presentation (10%), and Paper (30%)

The final grade for **graduate** students will comprise:

10% Class Participation

30% Homeworks

60% Final Project: Proposal (10%), Presentation (10%), and Paper (40%)



# Homework

- Completed assignments will be submitted via each student's private GitHub repository
- Each assignment must be submitted before class on the day listed on the schedule

There are no textbooks required



I am are here to help.

That said, Google and ChatGPT will answer almost all questions much faster (Use it!).



# Participation

- Discussion/questions
- Avoid the one-way lecture approach, which relies on your involvement
- Bring your laptop to class to follow along with the interactive lectures



# Final Project

## Components:

- A 1-page project proposal.
- An ~8-10 page final project paper.
- A final project presentation will be held during the last week of the semester.

## Scope:

- Students select a topic of personal interest or curiosity. While it may relate to their research, the project must be self-contained and utilize the tools introduced in the class or more advanced ones.

## Expectations:

- **Undergraduate Students:** Focus on a well-defined, independent project.
- **Graduate Students:** Address strong scientific research questions with rigorous data analysis using advanced methods from the course. The final report should demonstrate novelty and quality comparable to a peer-reviewed scientific article.
- **Evaluation** will consider the project's originality, quality, and alignment with academic standards.



# Tentative schedule

<b>Timeline</b>	<b>Topic</b>	<b>Software</b>	<b>Assignments</b>
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
01/28	Visualizing data	Matplotlib	
01/30	Data storage	NetCDF/GeoTiff/NetCDF/HDF5/Zarr	
02/04	Probability/Statistics		HW 1
02/06	Bayesian Statistics		

# Tentative schedule

Timeline	Topic	Software	Assignments
02/11	Map Projections I	Cartopy	
02/18	Map Projections II	GDAL	
02/20	Multi-dimensional arrays II	CDO/Xarray	
02/25	Vector Data	OGR/Shapely/GeoPandas	
02/27	Cluster Analysis I	Scikit-Learn	HW 2
03/04	Cluster Analysis II		
03/06	Dimensionality Reduction		
03/11	Decision Trees		
03/13	Random Forests & Boosting		
Spring Break			

# Tentative schedule

Spring Break			
03/25	Artificial Neural Networks		
03/27	Convolutional Neural Networks		HW 3
04/01	Kriging and Semi-variogram		Project Proposal
04/03	Regression Kriging		
04/08	Terrain Analysis		
04/10	Interactive Visualization	Folium/leaflet	
04/15	Parallel computing	Numba/Mpi4py/Dask	HW 4
04/17	BONUS		
04/21	Oral Presentations		
04/23	Oral Presentations		Project Paper



# Leverage Rice University High-Performance Computing (HPC) Infrastructure

- HPC is a group of powerful computers that solve large problems
- A Docker container in the HPC will host our class (virtual machine)
- We will use **Jupyter Lab**, **Notebooks** and the **Terminal** on the Docker





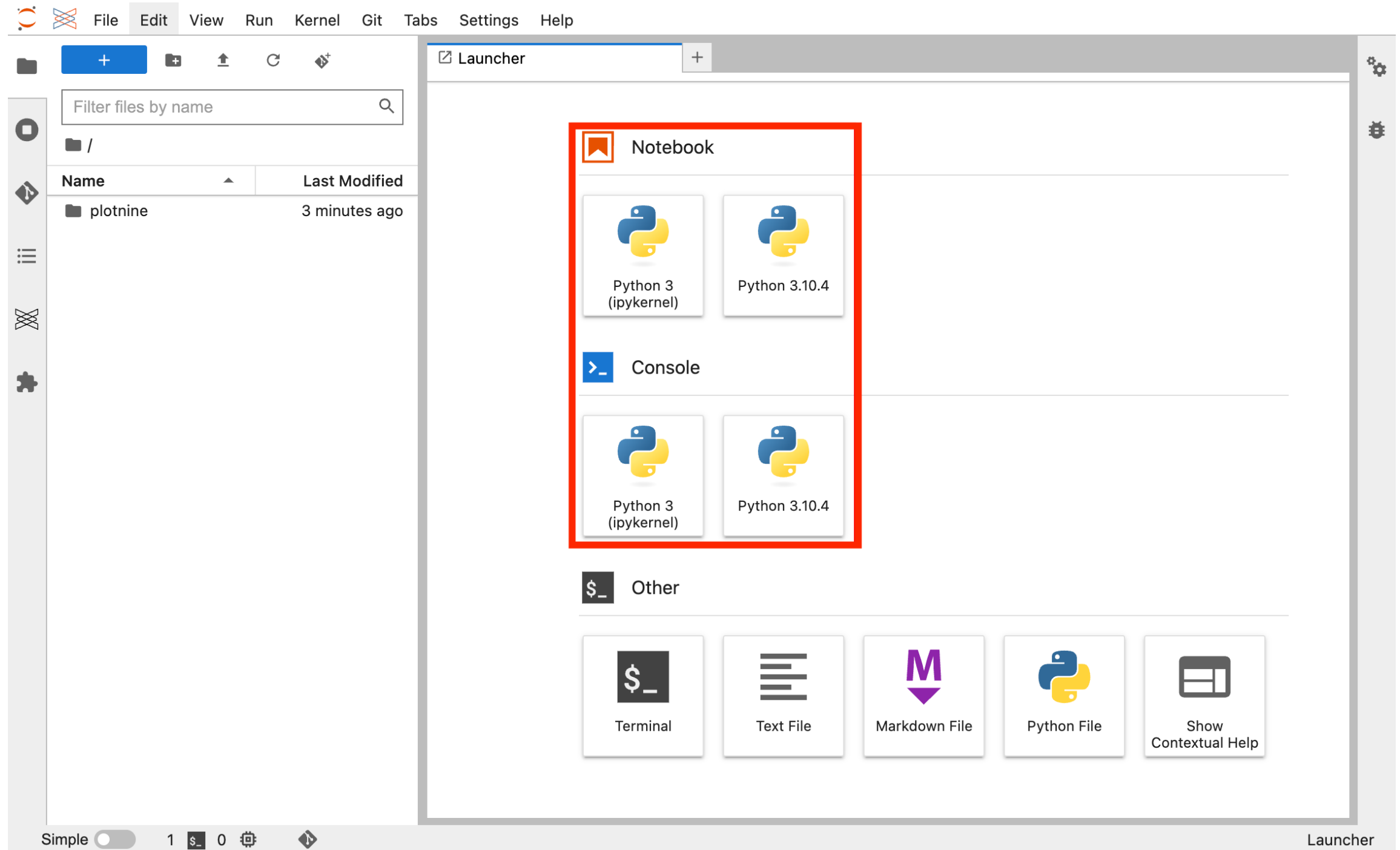
# Leverage Rice University High-Performance Computing (HPC) Infrastructure

Next week we will cover:

- How to access the docker and run the notebooks
- How to access and submit assignments in the docker



# Jupyter Lab



<https://jupyterlab.readthedocs.io/en/stable/>



# Jupyter Notebook



jupyter jupyter-interactive-visualization Last Checkpoint: 3 minutes ago (autosaved)



Quit

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel)



## Interactive Python Visualization Libraries

```
In [ ]: ▶ import numpy as np
import plotly.offline as py
import plotly.figure_factory as ff
from bokeh.models import HoverTool, WheelZoomTool
from bokeh.plotting import figure
from bokeh.io import show, output_notebook
output_notebook()
```

### Plotly

```
In [ ]: ▶ py.init_notebook_mode()
t = np.linspace(-1, 1.2, 2000)
x = (t**3) + (0.3 * np.random.randn(2000))
y = (t**6) + (0.3 * np.random.randn(2000))

colorscale = ['#7A4579', '#D56073', 'rgb(236,158,105)', (1, 1, 0.2), (0.98,0.98,0.98)]

fig = ff.create_2d_density(
    x, y, colorscale=colorscale,
    hist_color='rgb(255, 237, 222)', point_size=3
)

py.iplot(fig, filename='histogram_subplots')
```

[https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what\\_is\\_jupyter.html](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html)

# Terminal



Remains one of the best ways to interact more directly with the computer (flexibility!)

# Interacting with the terminal:

## Bash (Bourne Again SHell)





# Tutorials to learn Bash

<https://linuxconfig.org/bash-scripting-tutorial-for-beginners>

<https://www.codecademy.com/learn/learn-the-command-line>

<https://devhints.io/bash>

<https://hackernoon.com/top-10-bash-file-system-commands-you-cant-live-without-4cd937bd7df1>

# Top bash commands

- 1) **ls** - Lists the folder and file names in the current working directory.
- 2) **cd** - Change Directory to the path specified, for example `cd Lectures`
- 3) **mkdir** - Make directories with this command `mkdir my_folder`
- 4) **mv** - Moves files and folders. The first argument is the file you want to move, and the second is the location to move it to.
- 5) **cp** - Copies files and folders `cp my_file ./projects` . The flag `-r` recursively copies subfolders and files.
- 6) **rm** - Removes files and folders `rm my_folder` . Using `-r` will again recursively delete subfolders, `-f` force deletes, and `-rf` for a recursive force delete.

Let's talk about version control...



# GitHub: Git on the web



<https://www.youtube.com/watch?v=w3jLJU7DT5E>

# Version control: Git



<https://vimeo.com/41027679>

# 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)

Branches: main (1 Branch), Tags: 0 Tags

Search: Go to file

Buttons: Add file, <> Code

Commit history:

Commit	Author	Message	Time
9c5ac61	NoemiVergopolan	setup main page	now
			4 Commits

File list:

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 content:

## EEPS 440/640 - Geospatial Data Science

Right sidebar:

- About**: EEPS 440/460 Geospatial Data Science
- Releases**: No releases published. [Create a new release](#)
- Packages**: No packages published

[https://github.com/NoemiVergopolan/EEPS\\_GeospatialData.git](https://github.com/NoemiVergopolan/EEPS_GeospatialData.git)



# All lectures, assignments, and materials will be on the GitHub repository

The screenshot shows the GitHub interface for the repository 'EEPS\_GeospatialData' by user 'NoemiVergopolan'. The repository is private and has 1 branch and 0 tags. The main branch is 'main'. The repository contains a README.md file and three folders: Assignments, Lectures, and Syllabus. The README file is titled 'EEPS 440/640 - Geospatial Data Science'. The repository has 0 stars, 1 watcher, and 0 forks. The repository is currently being viewed by 1 person.

Navigation bar: NoemiVergopolan / EEPS\_GeospatialData

Search: Type / to search

Actions: Code, Issues, Pull requests, Actions, Projects, Security, Insights, Settings

Repository: EEPS\_GeospatialData (Private)

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

Branches: main (1 Branch), 0 Tags

Go to file

Add file

Code

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

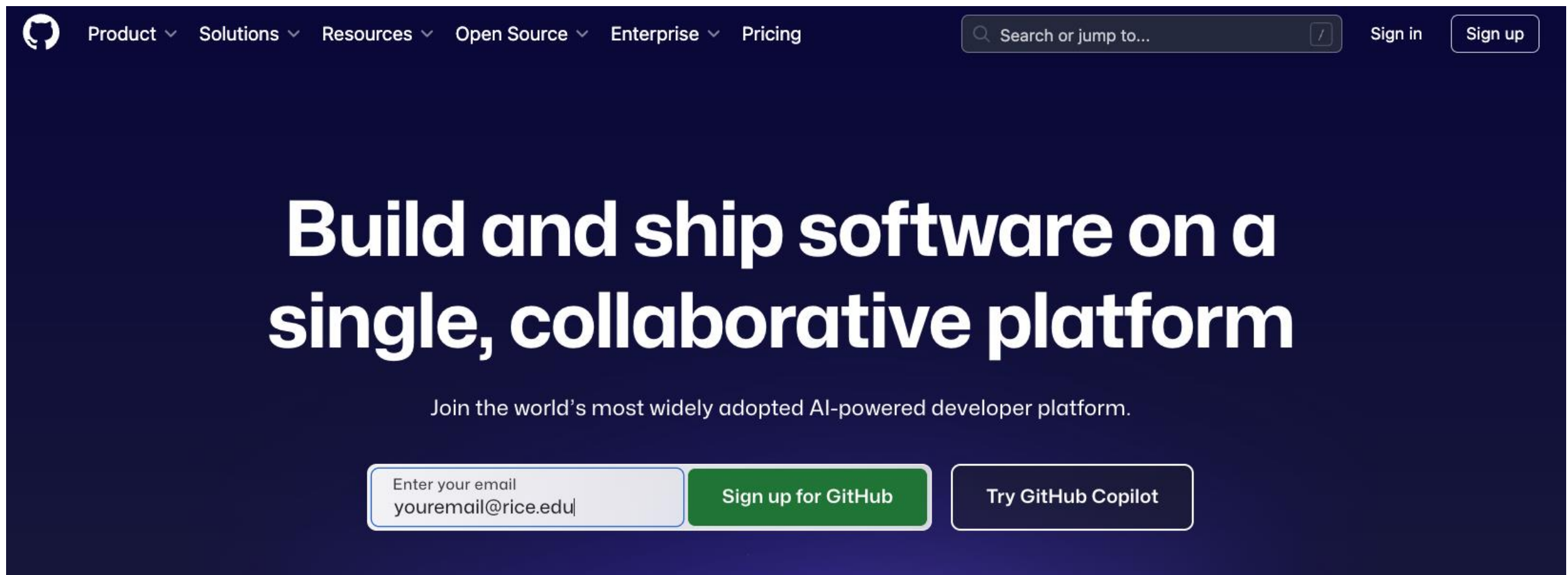
README

EEPS 440/640 - Geospatial Data Science

File/Folder	Commit Message	Commit Hash	Time
Assignments	setup main page	9c5ac61	10 minutes ago
Lectures	setup main page		now
Syllabus	setup main page		10 minutes ago
README.md	setup main page		5 minutes ago

**\*Solutions to assignments will be posted on Canvas**

# Create a personal account on GitHub.com



The image is a screenshot of the GitHub homepage. At the top, there is a dark blue navigation bar. On the left, the GitHub logo is followed by links for Product, Solutions, Resources, Open Source, Enterprise, and Pricing, each with a dropdown arrow. On the right, there is a search bar with the placeholder text "Search or jump to...", a "Sign in" link, and a "Sign up" button. The main content area has a dark blue background. It features a large white heading: "Build and ship software on a single, collaborative platform". Below this heading is a smaller white line of text: "Join the world's most widely adopted AI-powered developer platform." At the bottom, there is a sign-up form. It consists of a light blue input field with the placeholder text "Enter your email" and the example email "youremail@rice.edu". To the right of the input field is a green button with the text "Sign up for GitHub". To the right of the green button is a white button with the text "Try GitHub Copilot".

Product ▾ Solutions ▾ Resources ▾ Open Source ▾ Enterprise ▾ Pricing

Search or jump to... / Sign in Sign up

## Build and ship software on a single, collaborative platform

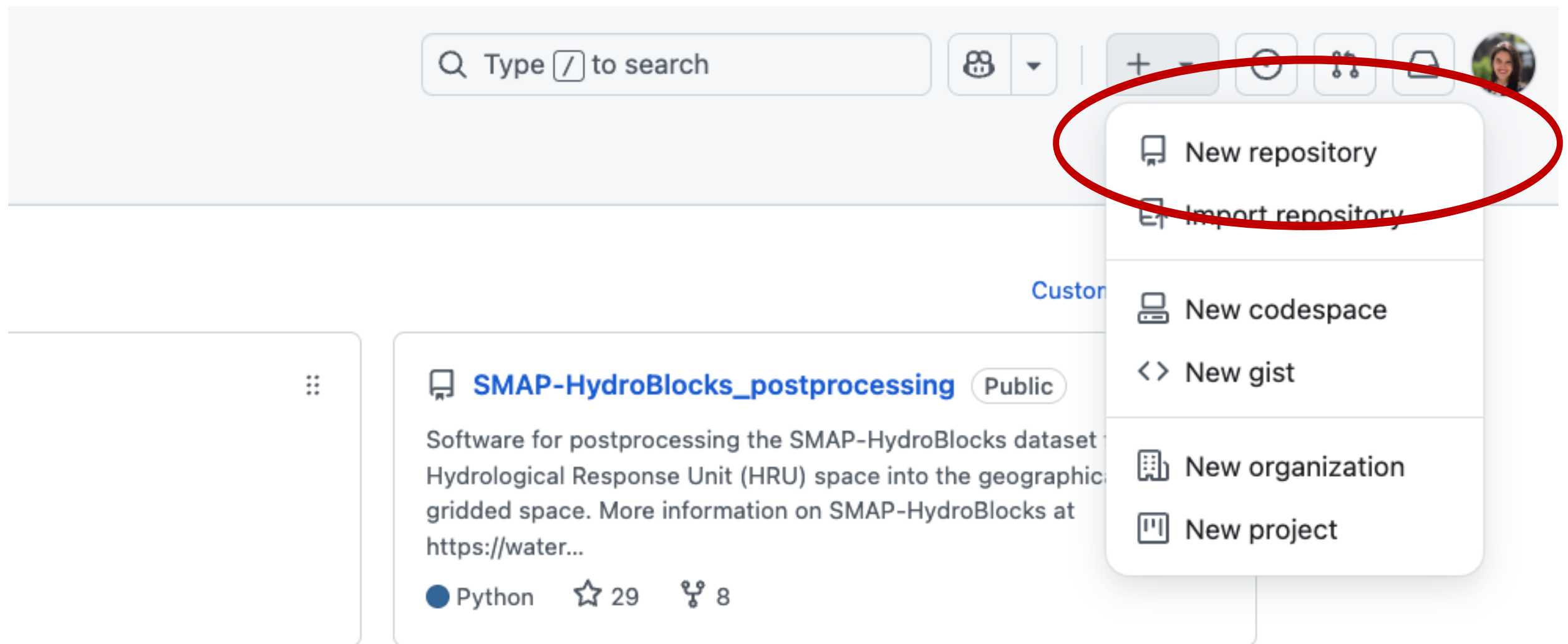
Join the world's most widely adopted AI-powered developer platform.

Enter your email  
youremail@rice.edu

Sign up for GitHub

Try GitHub Copilot

# Create a new repository from your homepage



# Naming your class repository GDS25\_initials

## Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository.](#)

---

*Required fields are marked with an asterisk (\*).*

Owner \*

 NoemiVergopolan ▼

Repository name \*



/ GDS25\_nvr

✓ GDS25\_nvr is available.

Great repository names are short and memorable. Need inspiration? How about **curly-fortnight** ?



# Make private (unless you want everyone to see your completed assignments!)

- ☐  **Public**  
Anyone on the internet can see this repository. You choose who can commit.
- ☒  **Private**  
You choose who can see and commit to this repository.

# Create the repository!

## Initialize this repository with:

☐ Add a README file

This is where you can write a long description for your project. [Learn more about READMEs.](#)

## Add .gitignore

.gitignore template: None ▼

Choose which files not to track from a list of templates. [Learn more about ignoring files.](#)

## Choose a license

License: None ▼

A license tells others what they can and can't do with your code. [Learn more about licenses.](#)

---

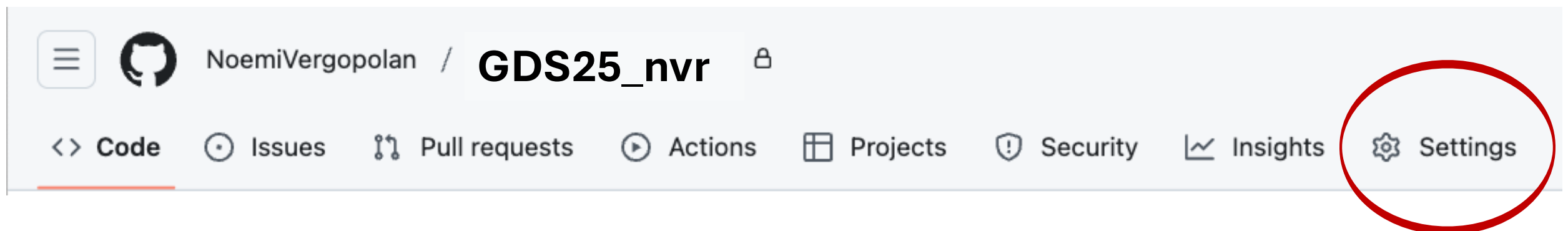
ⓘ You are creating a private repository in your personal account.

---

Create repository

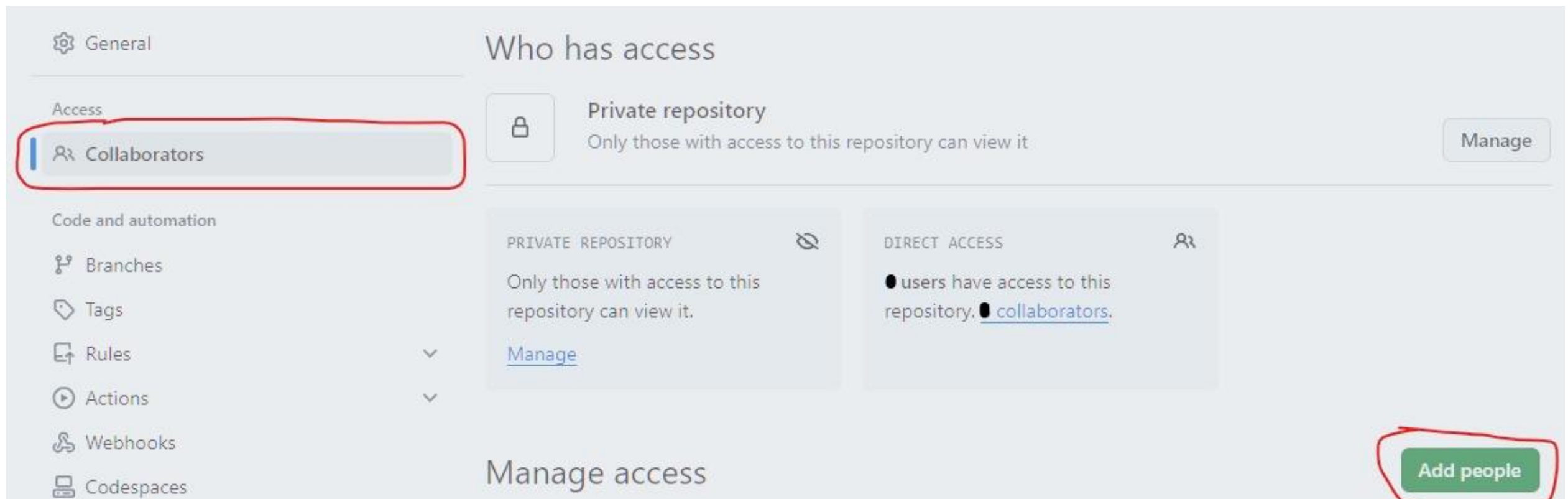
# Last thing you need to do is share your private repository with myself

On your repository, go to:  
“Settings” then “Collaborators” then “Add People”



# Last thing you need to do is share your private repository with myself

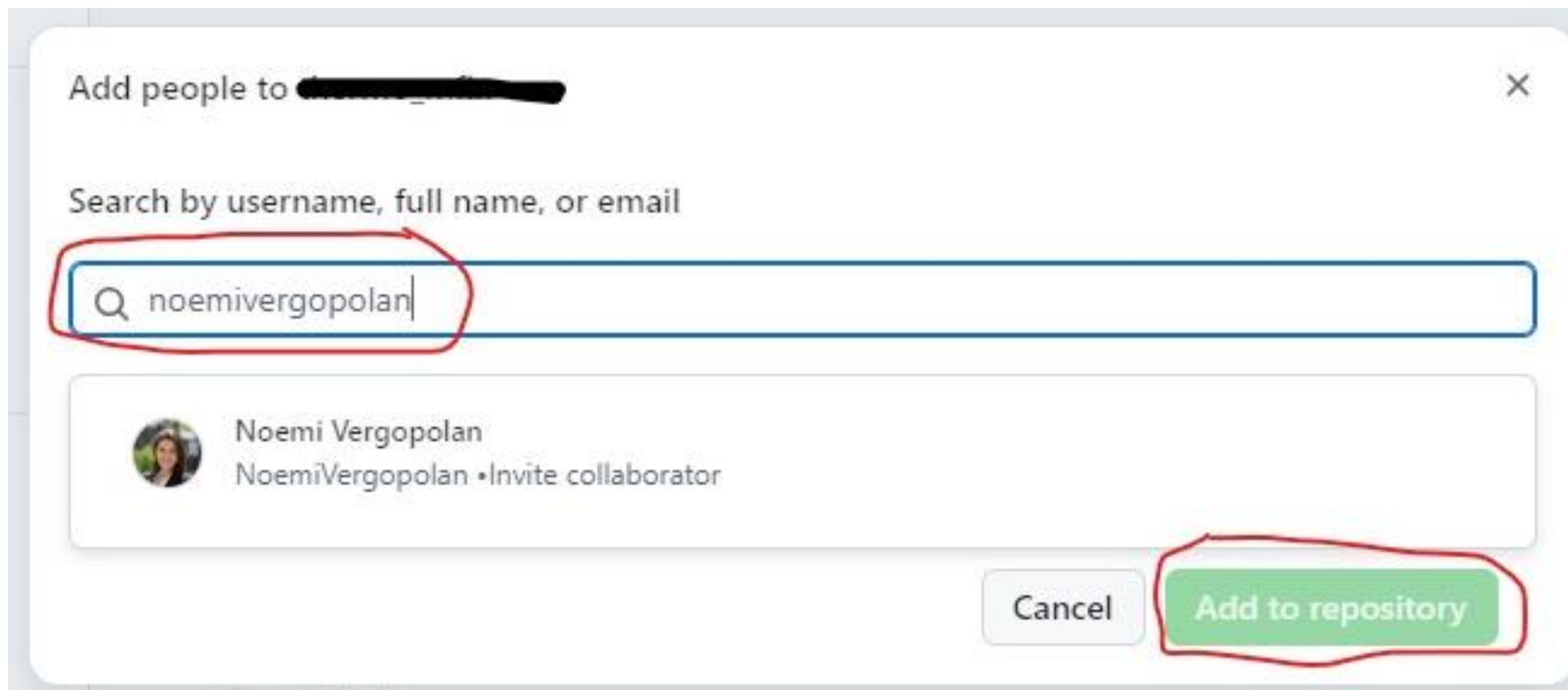
On your repository, go to:  
“Settings” then “Collaborators” then “Add People”





# Last thing you need to do is share your private repository with myself

On your repository, go to:  
“Settings” then “Collaborators” then “Add People”  
**noemivergopolan**



Add people to [redacted] ×

Search by username, full name, or email

Q noemivergopolan


Noemi Vergopolan  
NoemiVergopolan • Invite collaborator


Cancel Add to repository

# https://tinyurl.com/rs2r2xss

**Share with  
me your  
GitHub info  
& fill survey**

## GitHub Info & Survey

noemi.v.rocha@gmail.com [Switch account](#) 

 Not shared

*\* Indicates required question*

Full Name \*

Noemi Vergopolan

GitHub Username \*

NoemiVergopolan

Email used in GitHub account \*

noemi.v.rocha@gmail.com

Repository name (e.g., GDS25\_xxx) \*

GDS\_nvr

**Issue with class starting 15 or 30min earlier?**

**Thursday:**

- Intro to Python Programming

**Next week:**

- How to access the Docker, Github, and Homeworks
- Numpy and Scipy

**Deadlines for the homework will be updated!**