

# Geographic Data Science

Introduction

Elisabetta Pietrostefani & Carmen Cabrera

# (Self-)Quiz

- Have you ever used data to make decisions in your life?
- Have you ever heard the term “Data Science ”?
- Have you ever written a line of computer code?

# Philosophy

- (Lots of) **methods** and techniques
  - General overview
  - Intuition
  - Very little math
  - Lots of ways to continue on your own
- Emphasis on the **application** and **use**
- Close connection to “**real world**” applications

# Philosophy

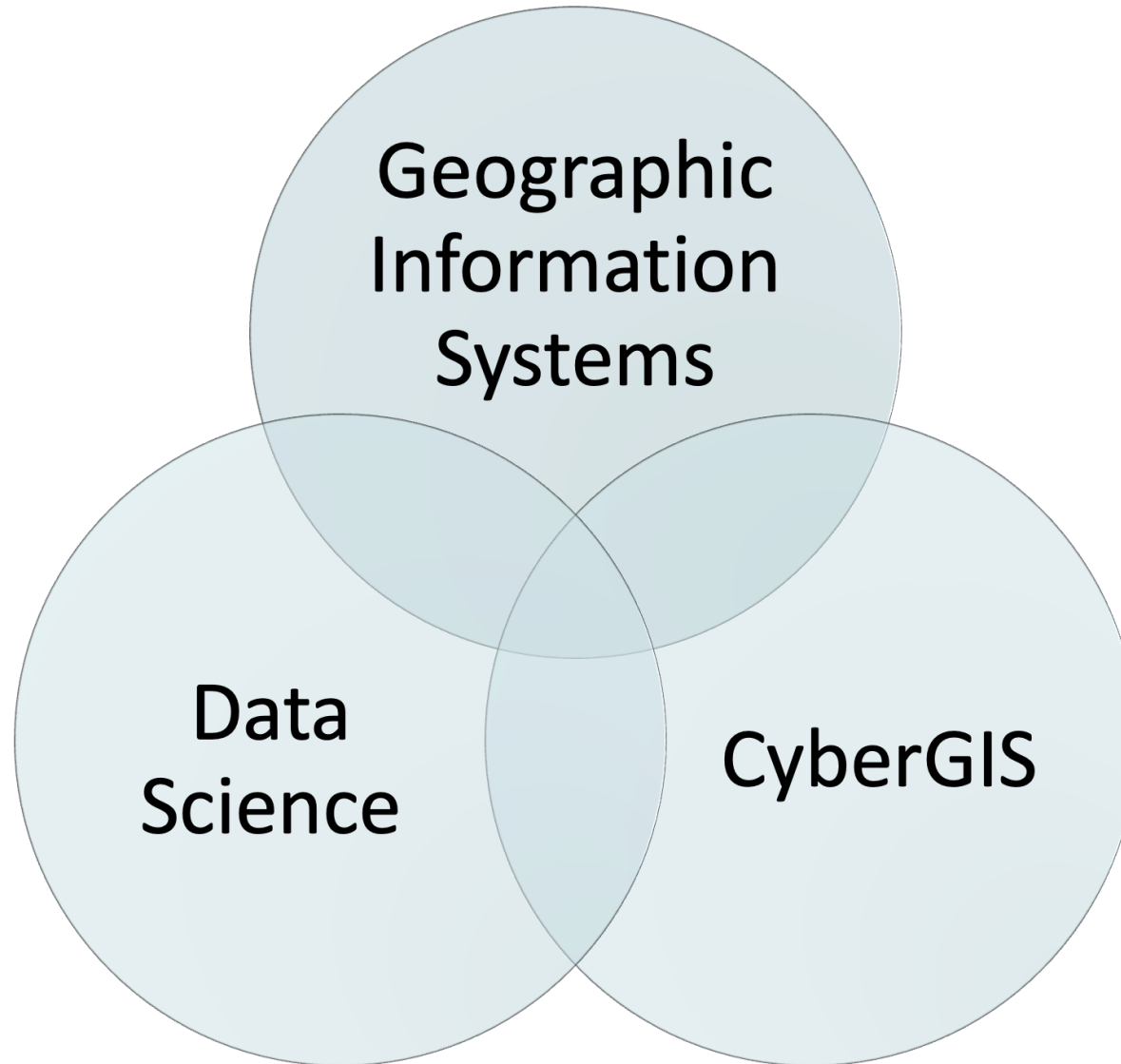
1. This course is like a gym subscription
2. Principles over technology
3. Collaborate, do not copy

# Format

- **Concepts:** lectures (website + slides), readings, videos
- **Hands-on:** concepts in (interactive) action
- **Do-It-Yourself:** practical material to do on your own

# **What is Geographic Data Science?**

# Geographic Data Science



# Geographic Data Science

- **Analyse and extract** insights from geospatial data
- Work with **real-world data** on a number of domains and problems
- Acquire key **data science skills** and important tools to answer spatial questions

**It is in very high demand in industry.**

# ***Philosophy of Geographic Data Science***

Statistician George Box :

*All models are wrong, but some are useful In a similar fashion.*

Geographer Keith Ord :

*All maps are wrong, but some are useful.*

# In what fields is it useful?

Housing

Transportation

Insurance

Telecommunications

Energy

Retail

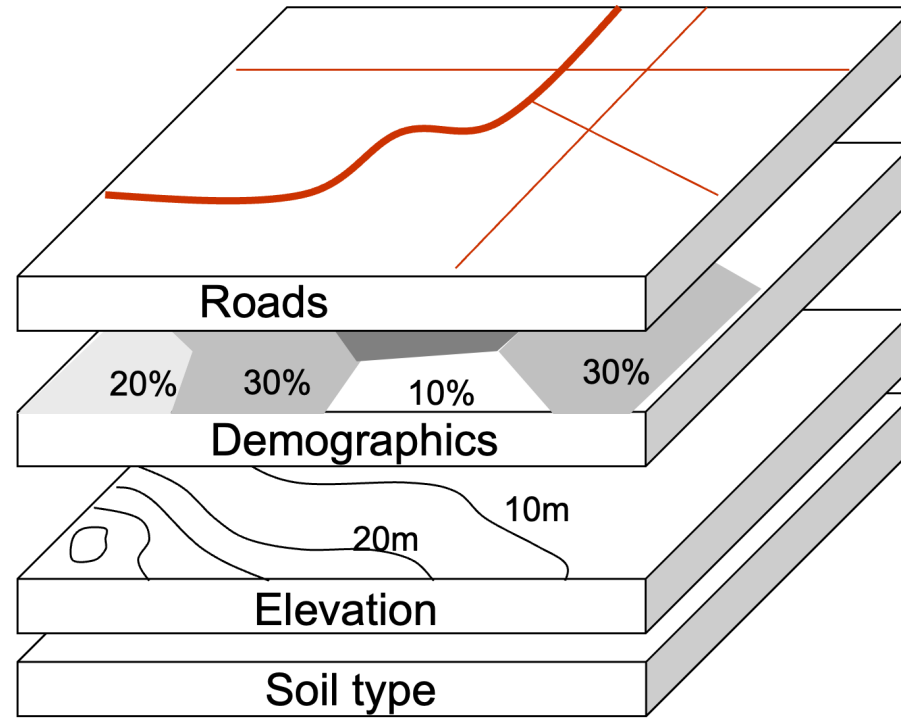
Agriculture

Healthcare

Urban planning

And more...

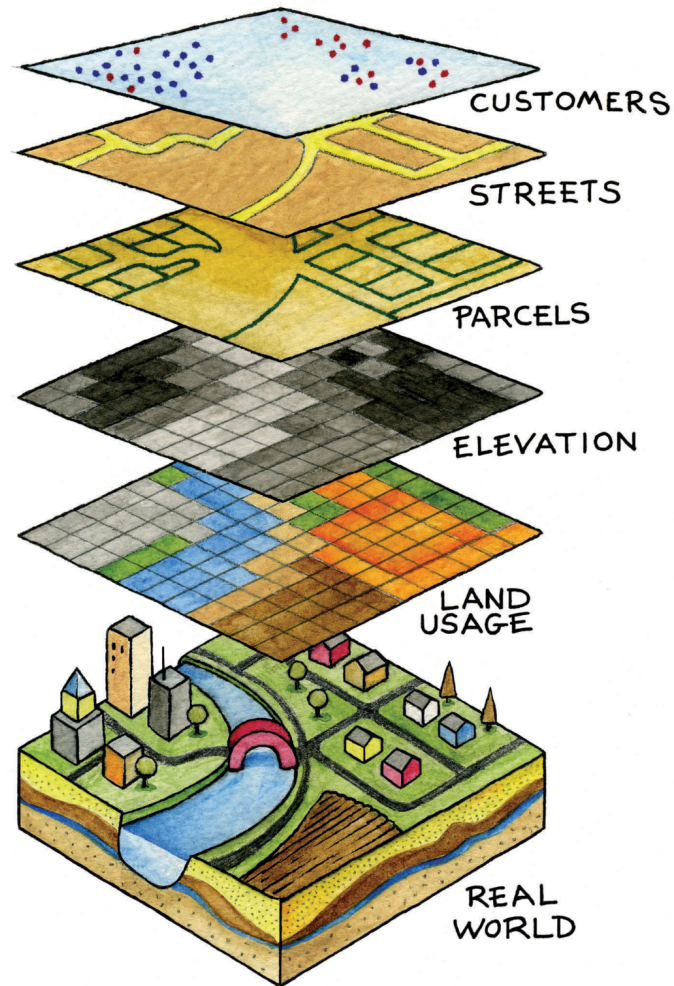
# GIS



# Layers - Image - Data



# GIS world vs. Real World



# Skills

**Hard Skills** - Programming Language - Transparency and Reproducibility - Version control

**Soft Skills** - Communication - Storytelling - Geospatial analytics acumen - Ethical skills

# Open Science

# Command line interface

## Graphical User Interfaces (GUIs)

Open source Geographic Information Systems (GIS), exemplified by software like QGIS, have revolutionized the accessibility of geographic analysis on a global scale. However, they inadvertently introduce a challenge to reproducibility.

## Command Line Interfaces (CLIs)

Command Line Interfaces (CLIs) offer a solution to the reproducibility challenge in GIS.

# The geodata 'revolution'

**Advanced Hardware:** High-performance computer hardware combined with efficient algorithms are driving the geospatial data revolution, allowing us to process vast datasets quickly.

**Scalable Software:** Scalable software solutions are essential for sifting through this data deluge, helping us extract valuable insights from the noise.

**Spatial Databases:** The advent of spatial databases empowers us to store and manipulate manageable subsets within the vast sea of geographic data.

# Logistics

# Sessions

- Lectures *and* labs
- **Tuesdays 10-11am** (Lecture 1h)
- **Tuesdays 1-3pm** (Lab 2h approx)
- Keep in touch on Teams!

## 202324-ENVS563 - Geographic Data Science

Assign to

Edit

### Welcome to a course on Geographic Data Science!

The module has three main aims.

- Provide students with core competences in Geographic Data Science (GDS). This includes advancing their statistical and numerical literacy and introducing basic principles of programming and state-of-the-art computational tools for GDS;
- Present a comprehensive overview of the main methodologies available to the Geographic Data Scientist, as well as their intuition as to how and when they can be applied;
- Focus on real world applications of these techniques in a geographical and applied context.

All the content for this module has been compelled in a book and available through an external website [\[HERE\]](#).

We will use this website throughout the term.

Geographic Data  
Science in R and  
Python

### Geographic Data Science in R and Python

AUTHOR  
Dr. Elisabetta Pietrostefani & Dr. Carmen Cabrera-Arnau  
PUBLISHED  
September 18, 2023

...

Table of contents

Welcome  
Contact

Edit this page

# Website

A course in  
Geographic Data  
Science 🔍 🔗

Welcome

- Syllabus
- Overview
- Assessments
- Environment
  - R
  - Python
- 1 Introduction
  - OpenScience in R
  - OpenScience in Python
  - Do-It-Yourself
- 2 Spatial Data
  - Lab in R
  - Lab in Python
  - Do-It-Yourself
- 3 Mapping Vector Data
  - Lab in R
  - Lab in Python
  - Do-It-Yourself
- 4 Mapping Raster Data
  - Lab in R

## A course in Geographic Data Science

AUTHOR  
Dr. Elisabetta Pietrostefani & Dr. Carmen Cabrera-Arnau August 28, 2024

PUBLISHED

Table of contents

Welcome

Contact

🔗 Edit this page

## Welcome

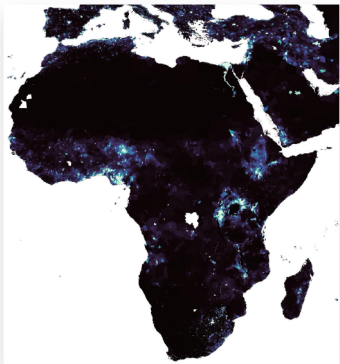
This is the website for the “Geographic Data Science” module **ENVS363/563** at the University of Liverpool. This is course designed and delivered by Dr. Elisabetta Pietrostefani and Dr. Carmen Cabrera-Arnau from the Geographic Data Science Lab at the University of Liverpool, United Kingdom. Much of the course material is inspired by Dani Arribas-Bel’s [course on Geographic Data Science](#).

This module will introduce students to the field of **Geographic Data Science (GDS)**, a discipline established at the intersection between Geographic Information Science (GIS) and Data Science. The course covers how the modern GIS toolkit can be integrated with Data Science tools to solve practical real-world problems.

Core to the set of employable skills to be taught in this course is an introduction to programming tools. Students will be able to whether to develop their skills in either R or Python in Lab sessions.

The website is **free to use** and is licensed under the [Attribution-NonCommercial-NoDerivatives 4.0 International](#). A compilation of this web course is hosted as a GitHub repository that you can access:

- As an [html website](#).
- As a [GitHub repository](#).



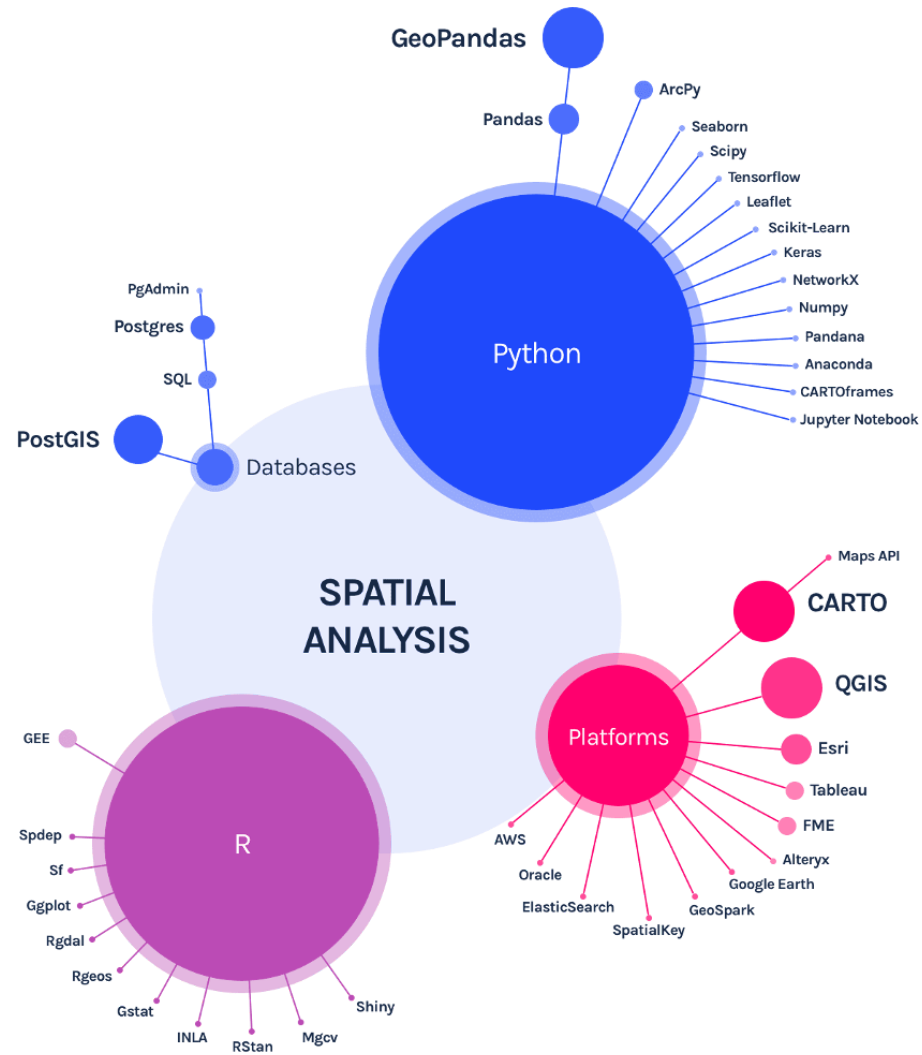
<https://gds-l-ul.github.io/gds/>

# Teams

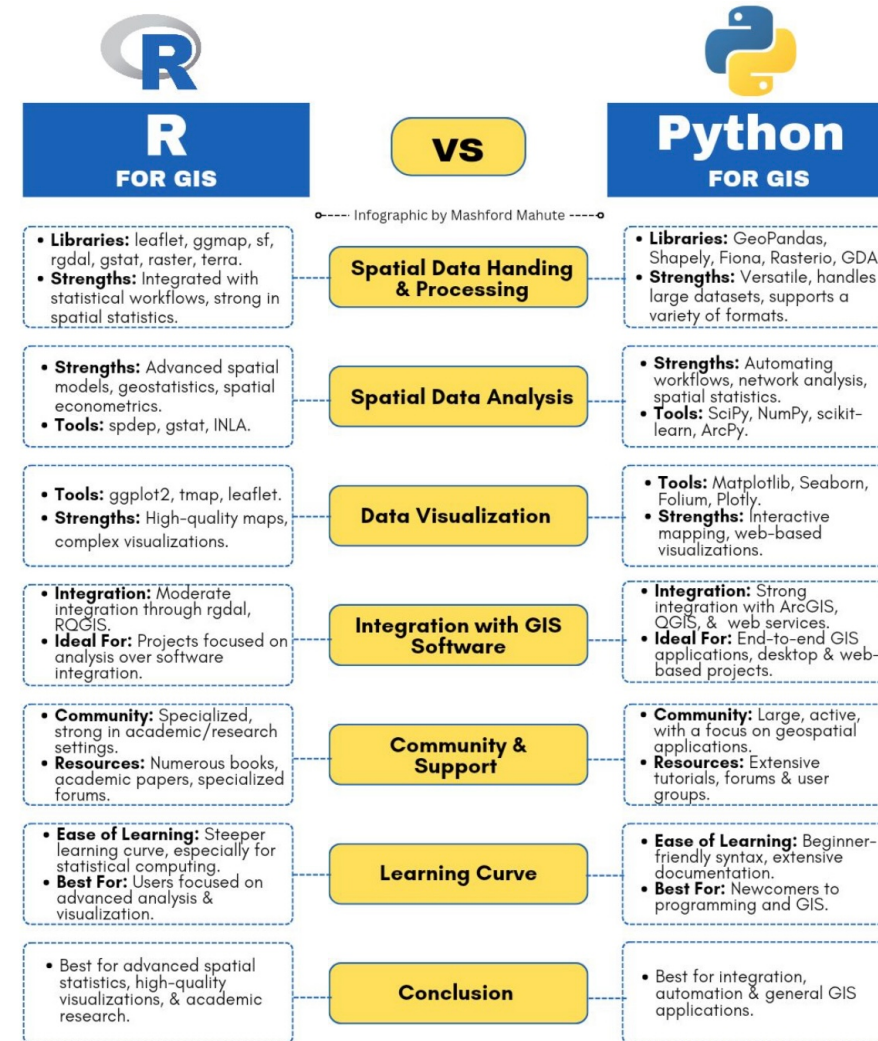
The screenshot shows a Microsoft Teams interface. On the left is a sidebar with a team icon (a map of Africa) and the team name 'ENVS363/563 25-26-O365-Team'. Below the team name are links to 'Class Notebook', 'Assignments', 'Grades', 'Reflect', and 'Insights'. Under a 'Main Channels' section, there is a list of channels: 'General' (highlighted), 'Lab 1 - Open Science', 'Lab 2 - Spatial Data', 'Lab 3 - Mapping Vector Data', 'Lab 4 - Mapping Raster Data', 'Lab 5 - Points', 'Python - Environment set up', and 'R- Environment Set up'.

The main area shows the 'General' channel. At the top, there are tabs for 'General', 'Posts', and 'Files', along with icons for video call, search, and more options. A post by 'Pietrostefani, Elisabetta' from Monday 16:01 is displayed. The post has a title 'Welcome to ENVS363/563!' and a message: 'Welcome to ENVS363/563! Let me introduce the team. [Cabrera, Carmen](#) and I will be co-teaching this module. We also have some wonderful demonstrators who will be helping out during labs. Please feel free to introduce yourself in this thread. Looking forward to meeting you all next week! 📖 📌'. Below the post is a 'Reply' button.

# Code



# Code



Find this valuable? Follow me for more.



Mashford Mahute

# Code

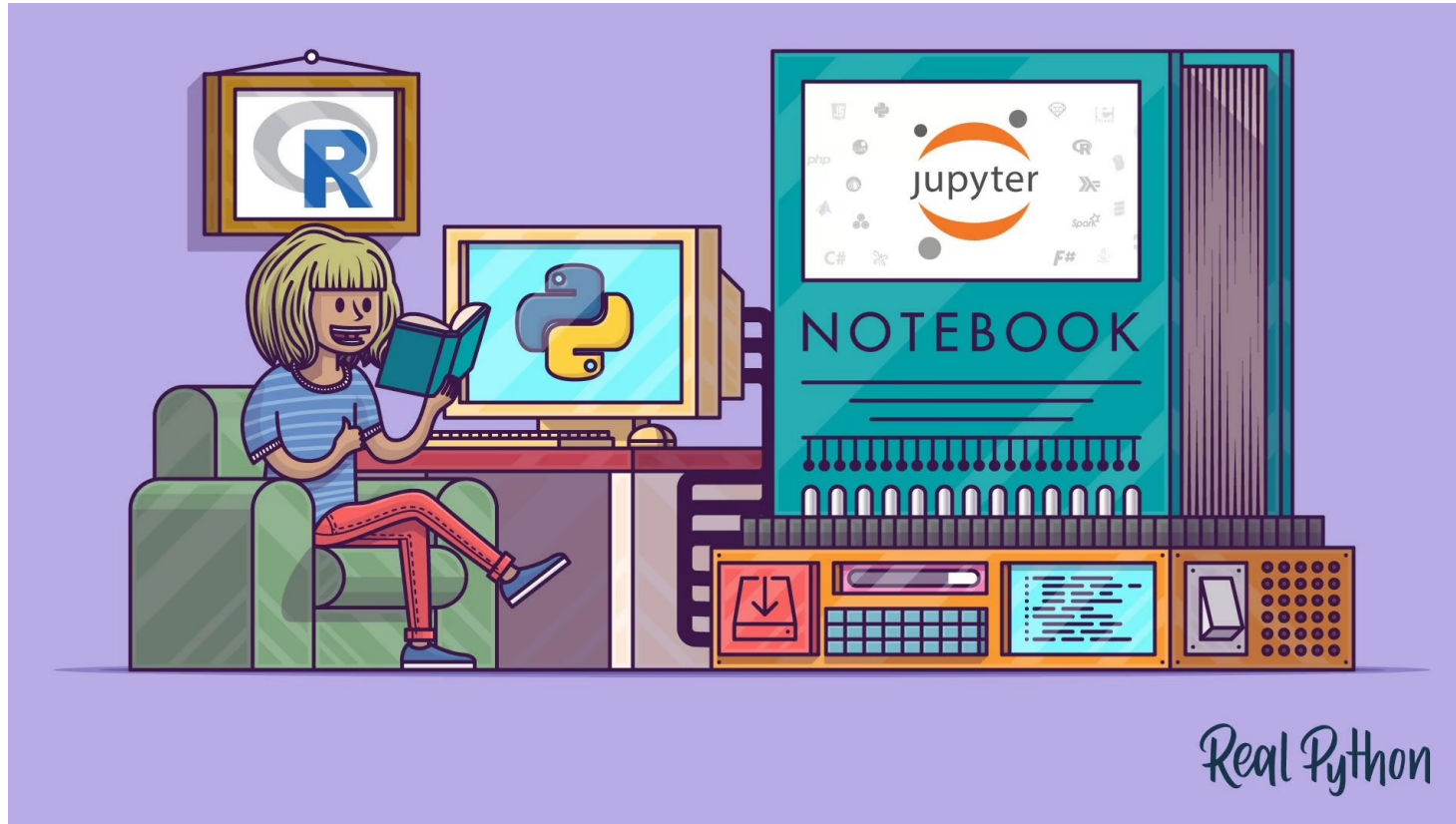


Illustrating having funR: Artwork by '@allison\_horst

# Code



# Code



# Website

- Syllabus
- Assessments
- Environment

Let's have a look

# Assignments

## Assignment I

Title: Programmed Map

Due date: 5th November 2026 (week 6)

40% of the final mark

## Assignment II

Title: Computational Essay

Due date: 10th December 2026 (week 11)

60% of the final mark

# Generative AI

You **may** use Generative AI tools (ChatGPT, Copilot, Claude...) for support.

But **inappropriate use is prohibited** — if you can't explain or justify it yourself, it's an **Academic Integrity concern**.

A **short declaration** is required with every submission: which tool, for which part (narrative/code), and how.

# Generative AI — what to declare

-  **Narrative:** your own words — AI proof-reading that changes wording/structure must be declared
-  **Code:** AI support is fine, but you must be able to explain every line
-  **Undisclosed use found later = penalty**, not just a missing declaration

Full declaration form: [gdsl-ul.github.io/gds/ai-declaration.html](https://gdsl-ul.github.io/gds/ai-declaration.html)

# More Help

This course is much more about *learning to learn* and **problem solving** rather than acquiring specific programming tricks or stats wizardry.

- Learn to ask questions (but don't expect exact answers all the time!!!)
- **Help others** as much as you can (the best way to learn is to teach)
- Search heavily on **Google + Stack Overflow**

# Workflow

come to the **Lectures**

1. Go over the Concepts sections of each week after the lecture
2. Have a look at the Readings and/or videos
3. Record questions and post them on Teams prior to the lab

# Workflow

come to the **Labs**

1. Come work through the code and DIY sections
2. Live answers to questions posted
3. Support from your lecturers and demonstrators
  - Hands on!
  - Collaborate *and* participate

# Questions



Geographic Data Science by [Elisabetta Pietrostefani](#) is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).