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

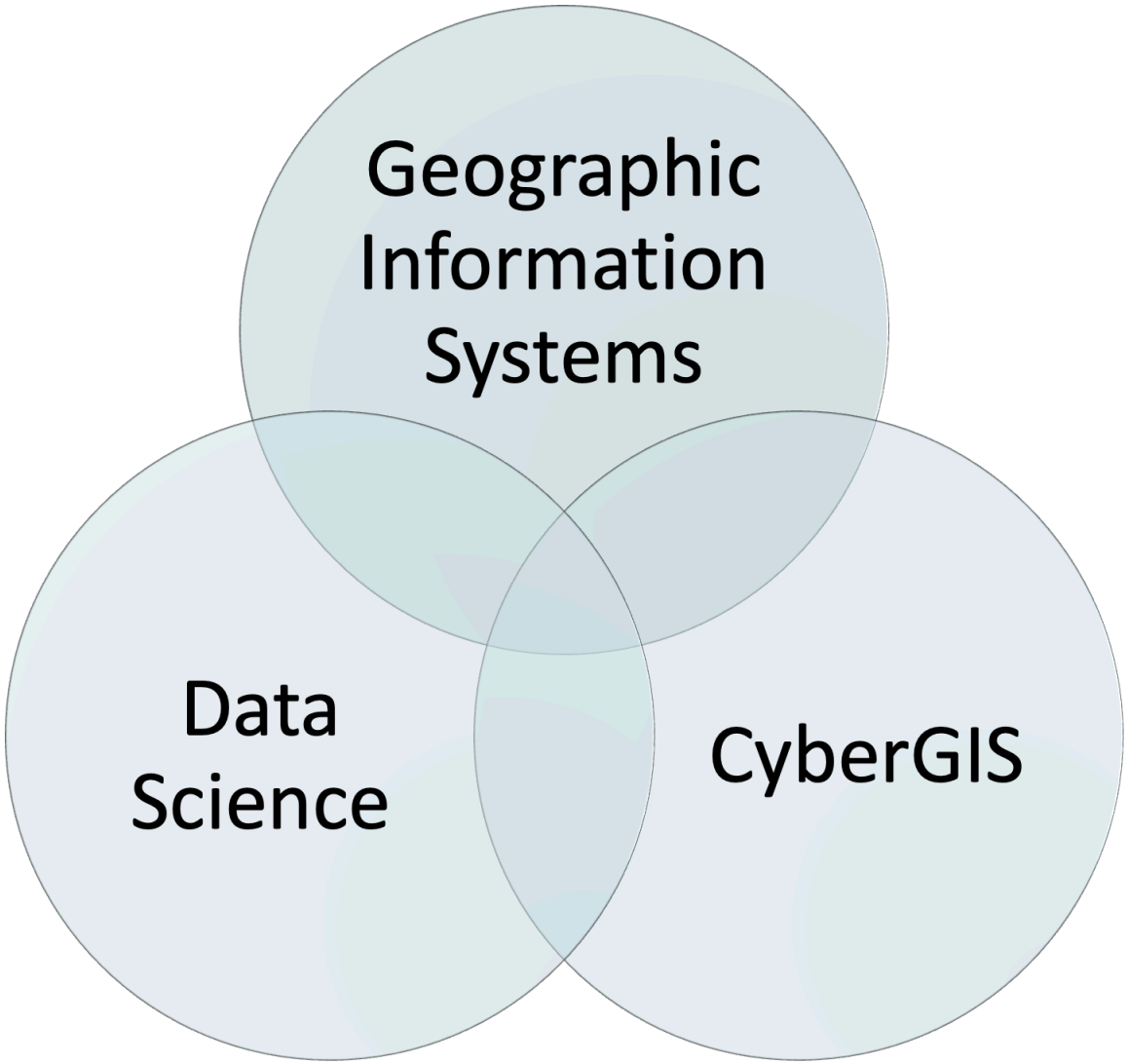
1. This course is like a gym subsciption
2. Principles over technology
3. Collaborate, do not copy

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

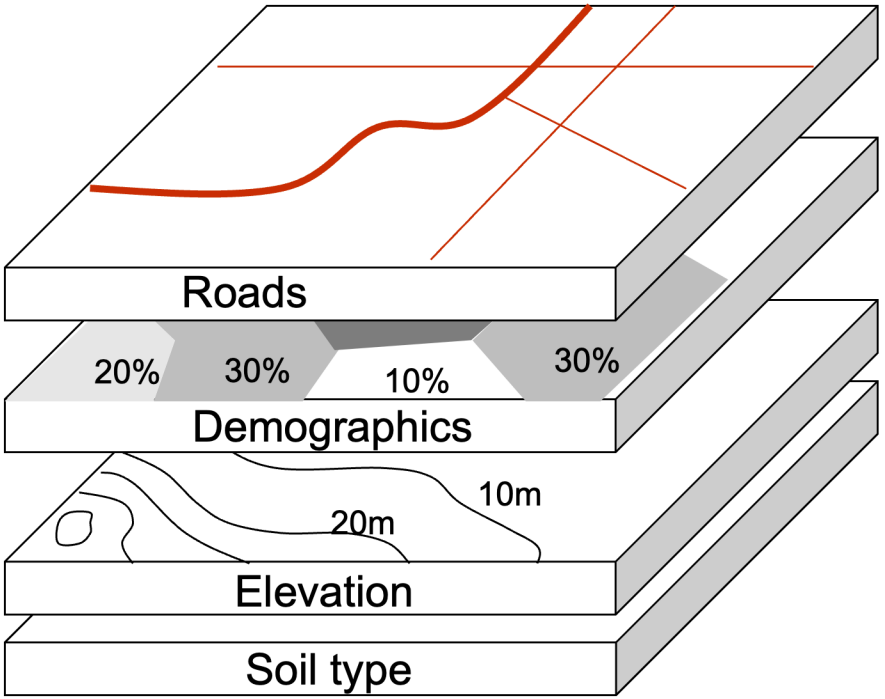
Agriculture

Healthcare

Urban planning

And more…

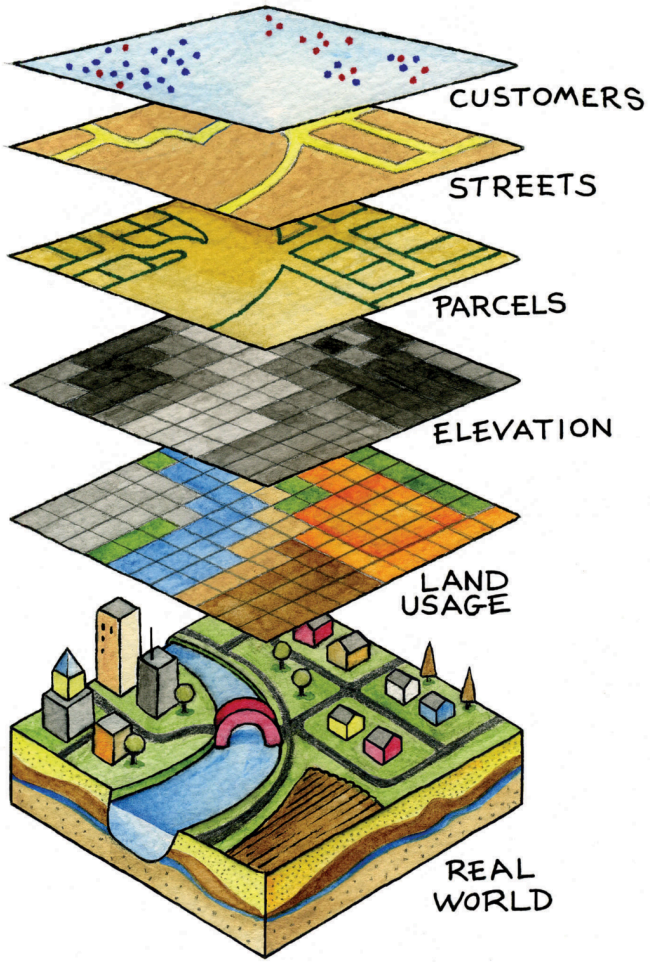
GIS



Layers - Image - Data

lsoa11nm	nsity (pp per hecta	residents	area (hectar)
Liverpool 017D	53.5	2040	38
Liverpool 017E	44.7	1424	32
Liverpool 02...	27.5	1899	69
Liverpool 02...	41.3	1422	34
Liverpool 02...	32.7	2175	66
Liverpool 001A	54.9	1642	30
Liverpool 00...	63.2	1505	24

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.

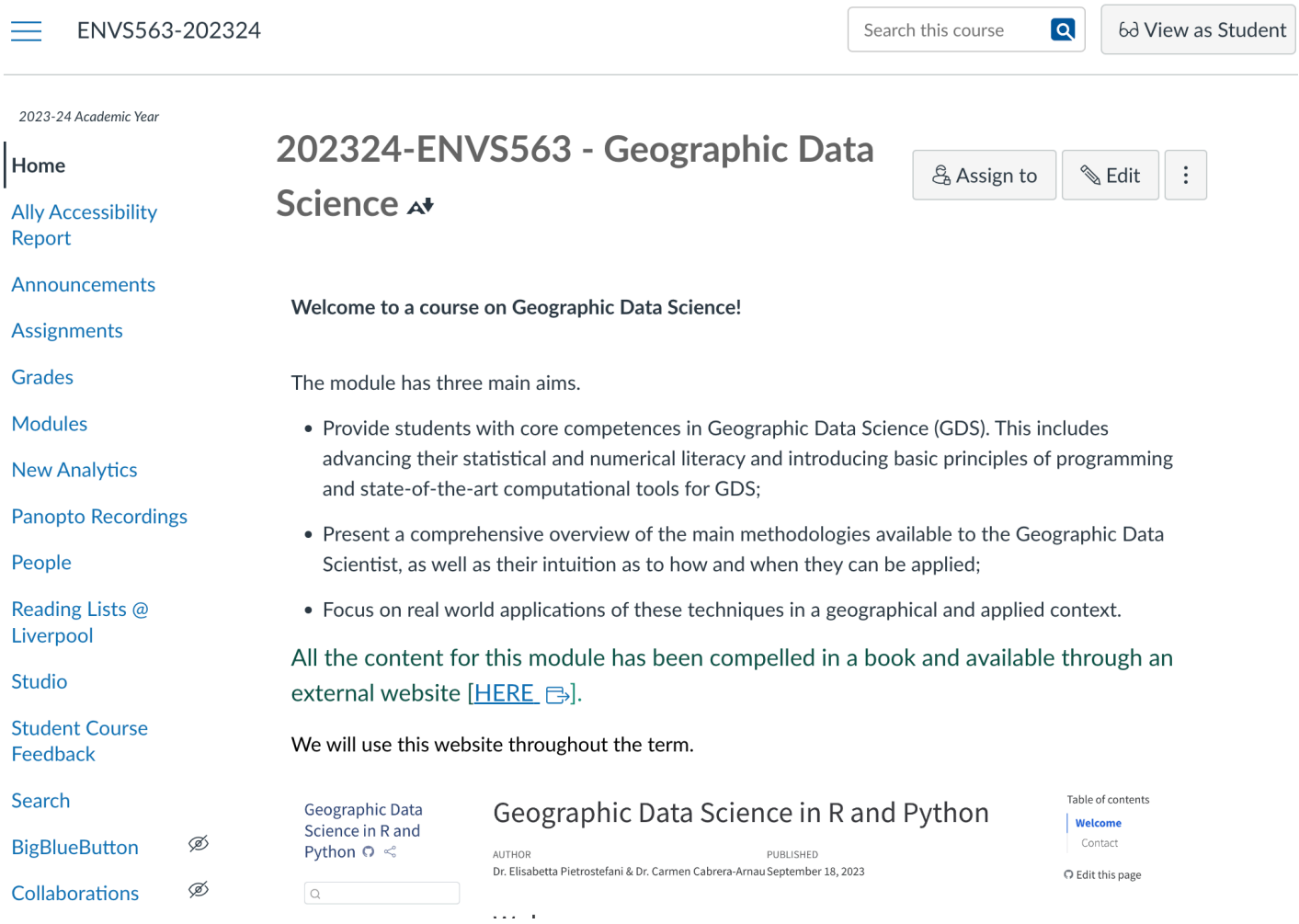
Spatial Databases: The advent of spatial databases empowers us to store and manipulate manageable subsets within the vast sea of geographic data.

Logistics

Sessions

- Lectures *and* labs
- **Monday 12-1pm** (Lecture 1h)
- **Monday 3-5pm** (Lab 2h approx)
- Keep in touch on Teams!

Canvas


ENVS563-202324
Search this course
View as Student
2023-24 Academic Year
Home
Ally Accessibility Report
Announcements
Assignments
Grades
Modules
New Analytics
Panopto Recordings
People
Reading Lists @ Liverpool
Studio
Student Course Feedback
Search
BigBlueButton
Collaborations
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

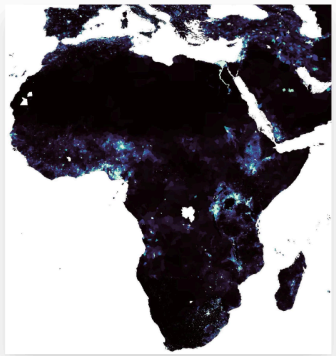
A course in Geographic Data Science

AUTHOR
Dr. Elisabetta Pietrostefani & Dr. Carmen Cabrera-Arnau

PUBLISHED
August 28, 2024

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.

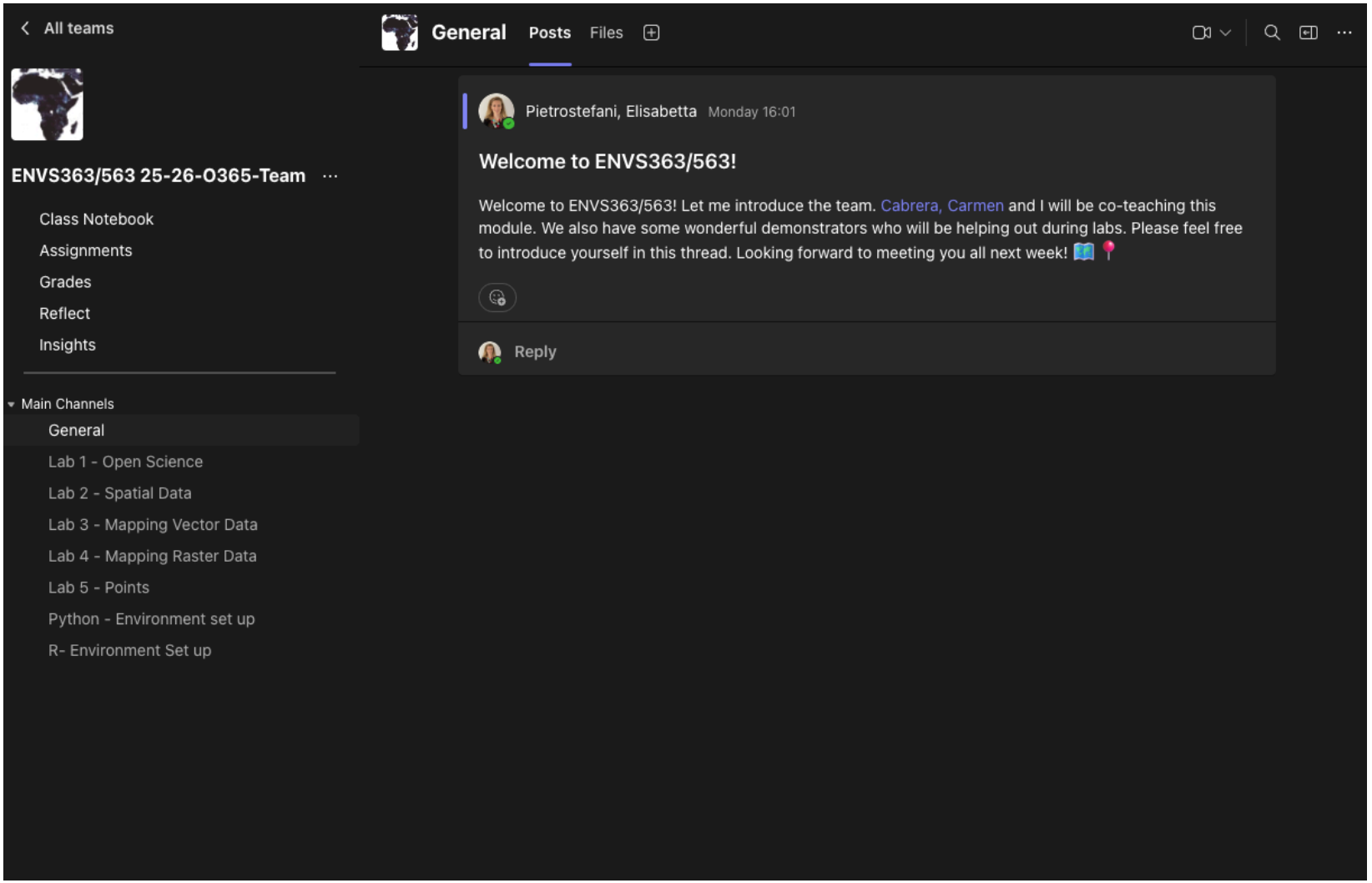
Core to the set of employable skills to be taught in this course is an introduction to programming tools. Students will be able to whether to develop their skills in either R or Python in Lab sessions.

 A compilation of this web course is hosted as a GitHub repository that you can access:

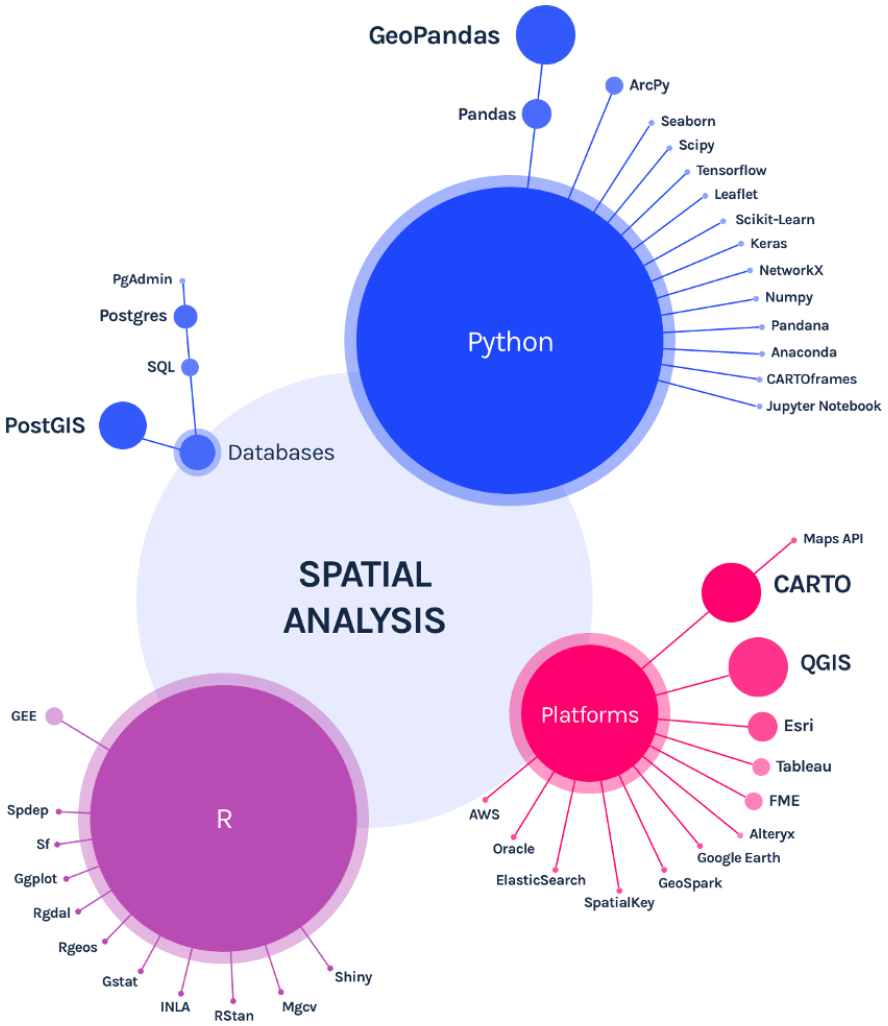
- As an html website.
- As a GitHub repository.

https://gdsl-ul.github.io/gds/

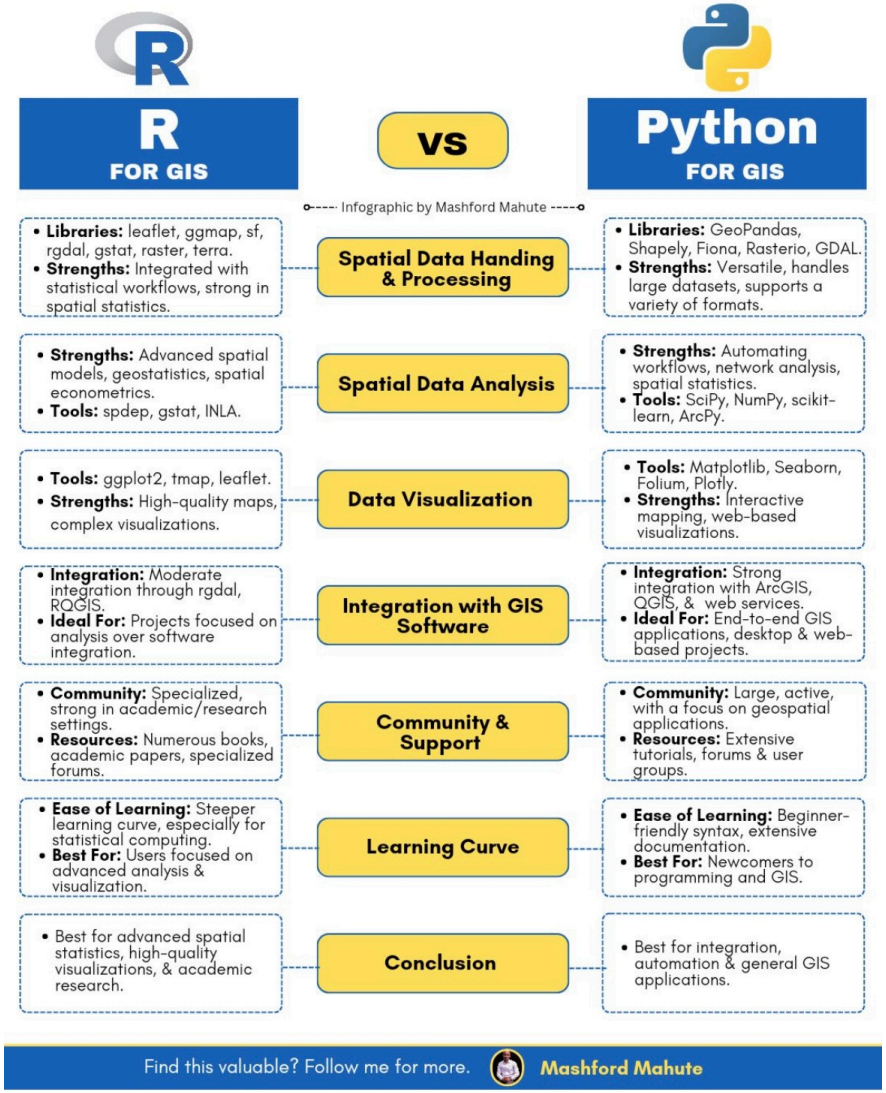
Teams


All teams
General
Posts
Files
ENVS363/563 25-26-O365-Team
Class Notebook
Assignments
Grades
Reflect
Insights
Main Channels
General
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
R- Environment Set up
Pietrostefani, Elisabetta Monday 16:01
Welcome to ENVS363/563!
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!
Reply

Code


GeoPandas
Pandas
ArcPy
Seaborn
Scipy
Tensorflow
Leaflet
Scikit-Learn
Keras
NetworkX
Numpy
Pandana
Anaconda
CARTOframes
Jupyter Notebook
Python
PgAdmin
Postgres
SQL
PostGIS
Databases
SPATIAL ANALYSIS
Maps API
CARTO
QGIS
Esri
Tableau
FME
Alteryx
Google Earth
GeoSpark
SpatialKey
ElasticSearch
Oracle
AWS
Platforms
GEE
Spdep
Sf
Ggplot
Rgdal
Rgeos
Gstat
INLA
RStan
Mgcv
Shiny
R

Code


R
FOR GIS
VS
Python
FOR GIS
Infographic by Mashford Mahute
• Libraries: leaflet, ggmap, sf, rgdal, gstat, raster, terra.
• Strengths: Integrated with statistical workflows, strong in spatial statistics.
Spatial Data Handing & Processing
• Libraries: GeoPandas, Shapely, Fiona, Rasterio, GDAL.
• Strengths: Versatile, handles large datasets, supports a variety of formats.
• Strengths: Advanced spatial models, geostatistics, spatial econometrics.
• Tools: spdep, gstat, INLA.
Spatial Data Analysis
• Strengths: Automating workflows, network analysis, spatial statistics.
• Tools: SciPy, NumPy, scikit-learn, ArcPy.
• Tools: ggplot2, tmap, leaflet.
• Strengths: High-quality maps, complex visualizations.
Data Visualization
• Tools: Matplotlib, Seaborn, Folium, Plotly.
• Strengths: Interactive mapping, web-based visualizations.
• Integration: Moderate integration through rgdal, RQGIS.
• Ideal For: Projects focused on analysis over software integration.
Integration with GIS Software
• Integration: Strong integration with ArcGIS, QGIS, & web services.
• Ideal For: End-to-end GIS applications, desktop & web-based projects.
• Community: Specialized, strong in academic/research settings.
• Resources: Numerous books, academic papers, specialized forums.
Community & Support
• Community: Large, active, with a focus on geospatial applications.
• Resources: Extensive tutorials, forums & user groups.
• Ease of Learning: Steeper learning curve, especially for statistical computing.
• Best For: Users focused on advanced analysis & visualization.
Learning Curve
• Ease of Learning: Beginner-friendly syntax, extensive documentation.
• Best For: Newcomers to programming and GIS.
• Best for advanced spatial statistics, high-quality visualizations, & academic research.
Conclusion
• Best for integration, automation & general GIS applications.
Find this valuable? Follow me for more.
Mashford Mahute

Code

Illustrating having funR: Artwork by '@'allison_horst

Code


Rmarkdown
TEXT. CODE. OUTPUT.
(GET IT TOGETHER, PEOPLE.)
#HEADERS
CODE
TEXT
PANDOC
KNITR
FAIRY DUST
RECIPE
1. ADD TEXT
2. ADD CODE
3. KNIT
4. (magic)
5. CELEBRATE PERCEIVED WIZARDRY
rmarkdown
knitr
pandoc

Code

Website

- Syllabus
- Assessments
- Environment

Let’s have a look

Assignments

Assignment I

Title: Programmed Map

Due date: 30th October 2025 (week 6)

40% of the final mark

Assignment II

Title: Computational Essay

Due date: 4th December 2025 (week 11)

60% of the final mark

AI

GAI Boilerplate

- *You are reminded that the inappropriate use of Generative Artificial Intelligence Tools in the preparation of this assignment is strictly prohibited.*
- *This assignment should be prepared using your own words. All use of AI translation tools should be properly acknowledged. Extensive use of AI proof-reading tools is prohibited. Whilst you may use spelling/grammar checks typically found in word-processing packages, using AI tools to change words/sentence structure may incur an Academic Integrity penalty.*
- *If your assessment is referred for an Academic Integrity Investigation, you may be asked to demonstrate that the work you have submitted is your own. Therefore, it is advised that you keep hold of earlier files, drafts, notes and other relevant preparatory materials that you have used.*

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

Download R before Lab

If using your own laptop

Environment

This course can be followed by anyone with access to a bit of technical infrastructure. This section details the set of local and online requirements you will need to be able to follow along, as well as instructions or pointers to get set up on your own. This is a centralised section that lists *everything* you will require.

Coding Languages

In this course, you have the option to follow along using either `R` or `Python`, depending on your past experience with these programming languages and preference. Please choose **one** language to focus on and stick to it throughout.

- If you want to follow the course in `R`, you can find instructions to set up your environment here.
- If you want to follow the course in `Python`, you can find instructions to set up your environment here.

This course has two assignments and you will be required to submit both assignments in the same programming languages. The next two sections will guide you through the process of setting up your development environment in `R` or `Python`, so you can get started with the course smoothly.

Questions