

# Geographic Data Science

Introduction

Elisabetta Pietrostefani & Carmen Cabrera-Arnau

# (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

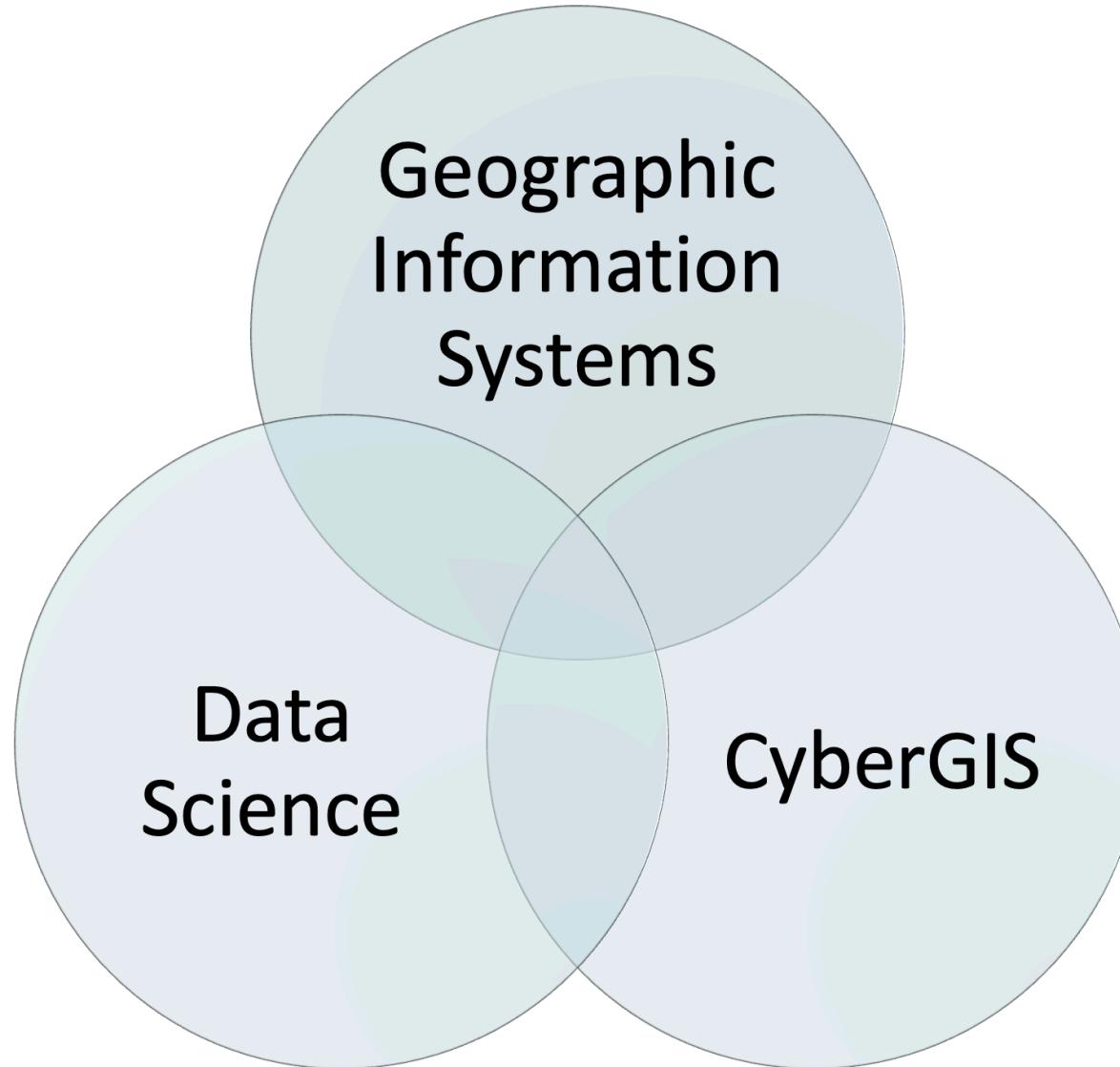
1. This course is like a gym subscription
2. Principles over technology
3. Collaborate, do not copy

# 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

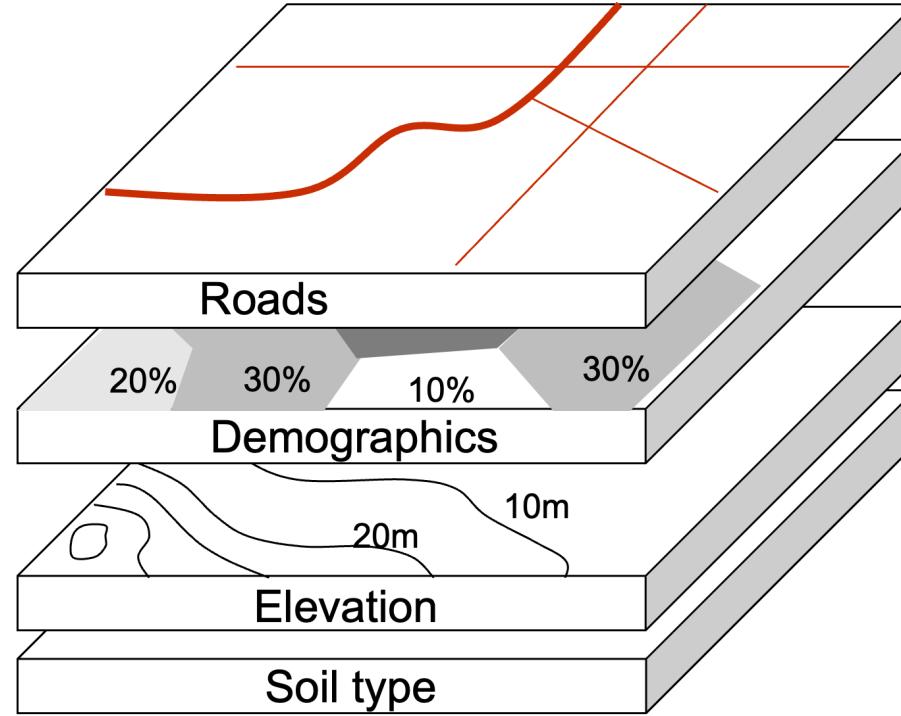
Agriculture

Healthcare

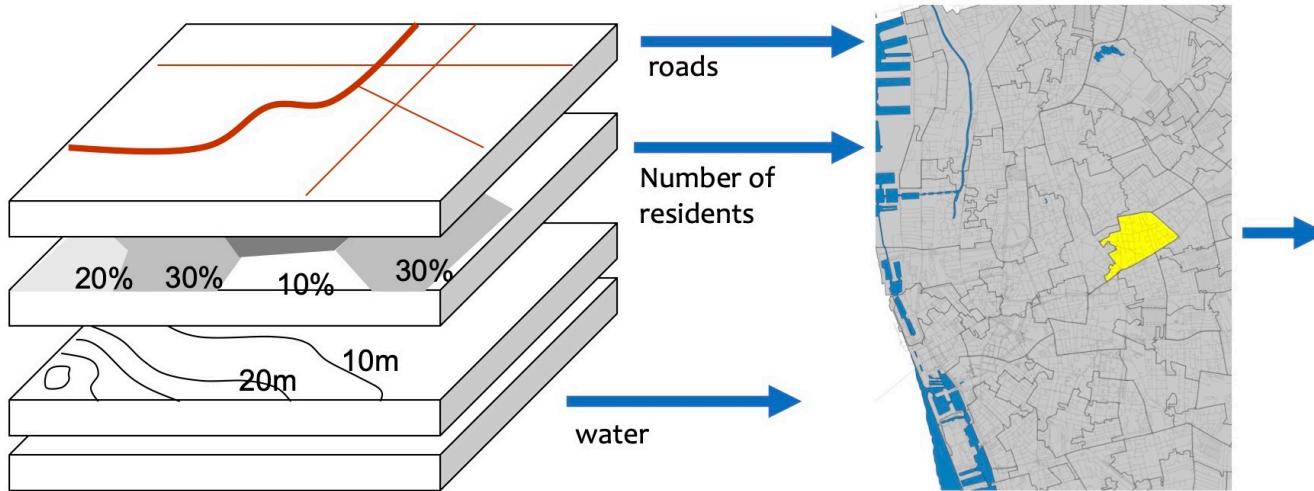
Urban planning

And more...

# GIS

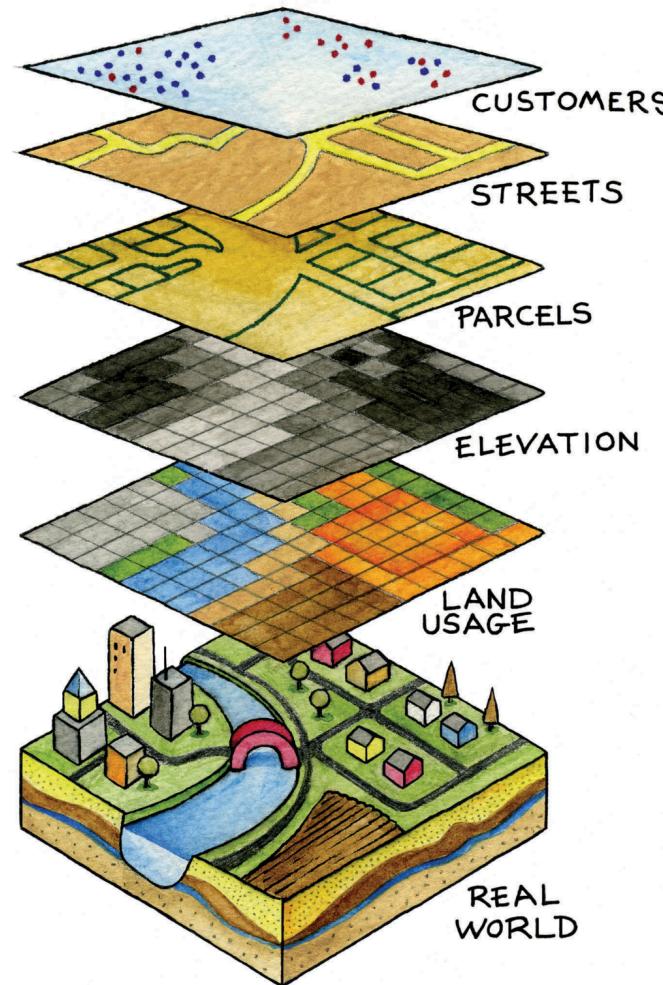


# Layers - Image - Data



Isodata ID	Density (pp per hectare)	Residents	Area (hectares)
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
- Mondays 3-4pm (Lecture 1h)
- Tuesday 3-5pm (Lab 2h approx)
- Keep in touch on Teams!

# 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

### PUBLISHED

August 28, 2024

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 a 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://pietrostefani.github.io/gds/>

# Teams

< > Search ... The University of Liverpool 

< All teams  General Posts Files + Meet ⚙ ⓘ ...

ENVS363/563 23-24-O365... 

Class Notebook  
Assignments  
**Grades**  
Reflect Grades  
Insights

---

Channels  
**General**  
Lab 1 - Open Science

Welcome to ENVS363/563 23-24-O365-Team  
Choose where you want to start

   
[Upload Class Materials](#) [Set up Class Notebook](#)

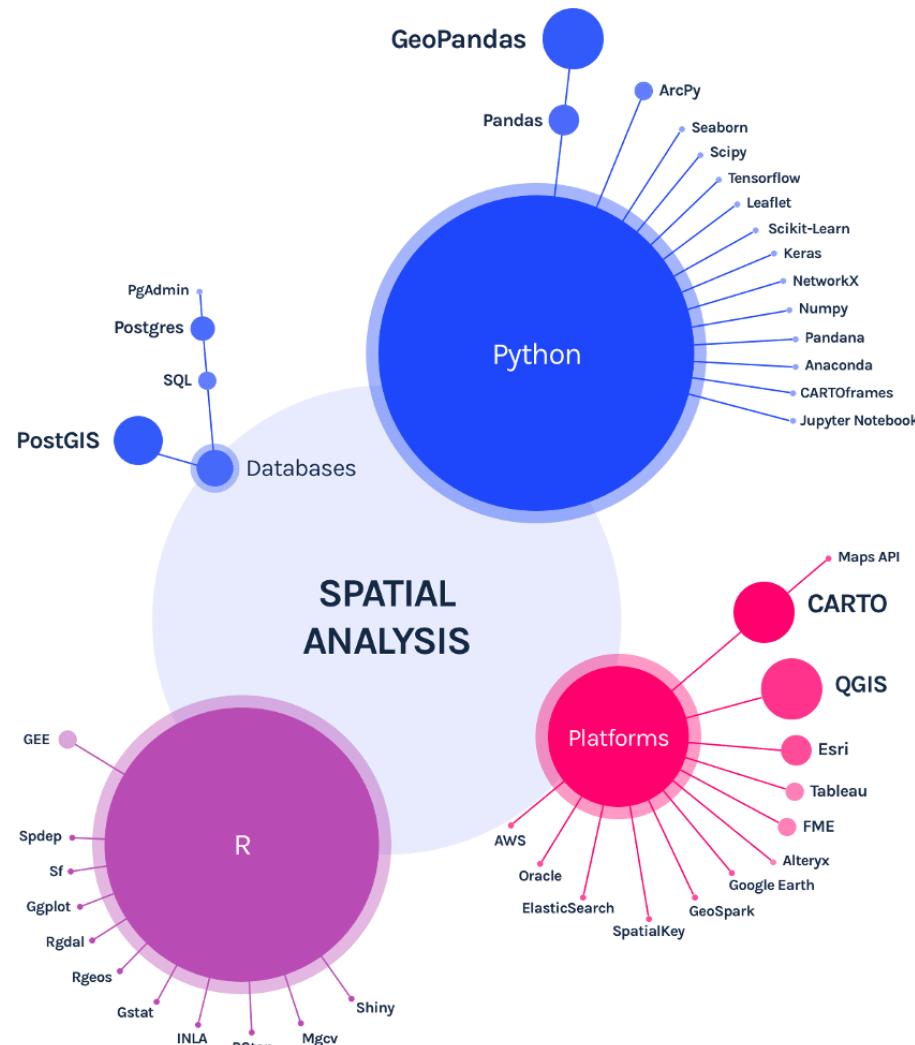
 Pietrostefani, Elisabetta 17:22 Edited  
Welcome to ENVS363/563! Let me introduce the team. Cabrera-Arnau, Carmen and will be co-teaching this module. Please feel free to introduce yourself in this thread. Looking forward to meeting you all next week! 

 Cabrera-Arnau, Carmen 18:22  
Welcome to ENVS353/563! I am also looking forward to the first week of teaching 

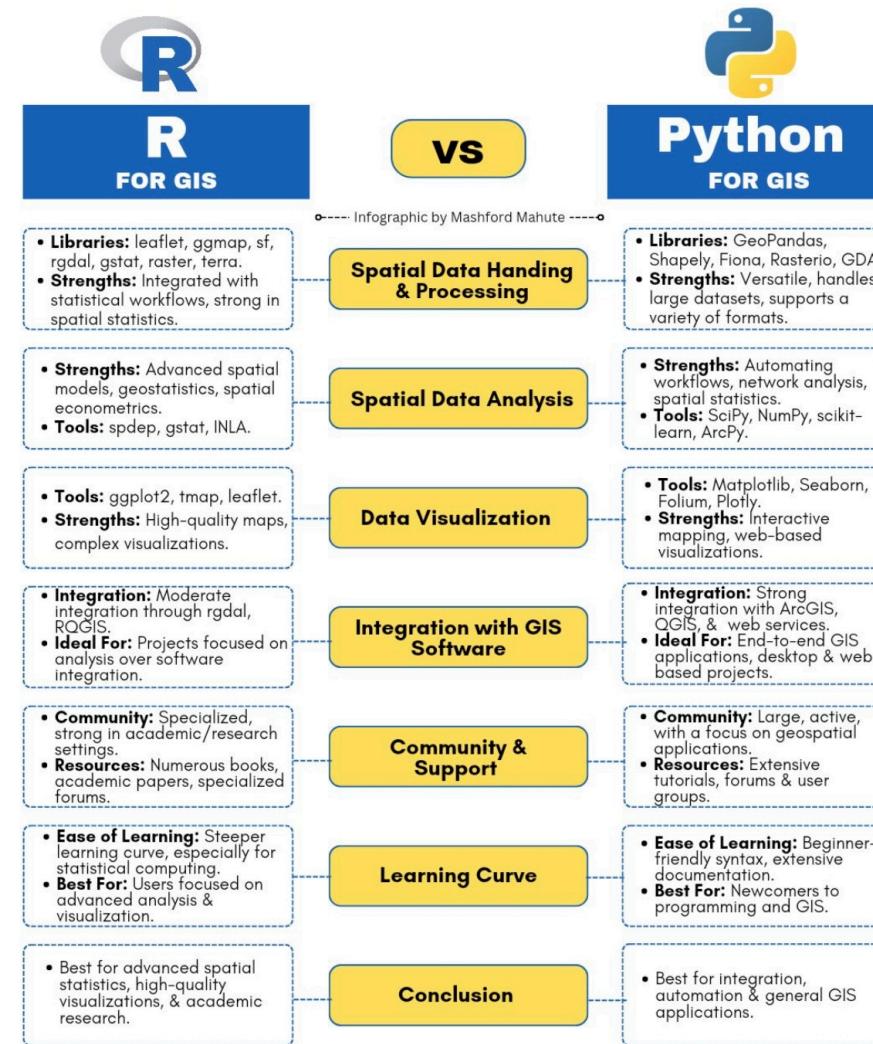
Reply



# Code



# Code



# Code



Illustrating having funR: Artwork by '@allison\_horst'

# Code



# Code



# Website

- Syllabus
- Assessments
- Environment

Let's have a look

# 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

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

## 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



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