Geospatial Analysis and Representation for Data Science

course for the master in Data Science University of Trento
2023/2024

Document release date: 16 Nov 2023

Summary

This document presents the course program organized into lessons and provides indications about the exam.
For the course's exam, students are required to perform a data analysis project and submit a report.
Recommended readings.

Lecturers

Diego Giuliani <diego.giuliani@unitn.it>
Maurizio Napolitano <napolitano@fbk.eu>

Summary **1**
Lecturers 1
1. Lectures schedule **2**
1.1 15/09/23 - Lecture 1 2
1.2 22/09/23 - Lecture 2 2
1.3 29/09/23 - Lecture 3 2
1.4 06/10/23 - Lecture 4 2
1.5 13/10/23 - Lecture 5 3
1.6 20/10/23 - Lecture 6 3
1.7 27/10/23 - Lecture 7 3
1.8 19/11/23 - Lecture 8 3
1.9 17/11/23 - Lecture 9 3
1.10 Notes: 4
2. Tools used during the course **4**
2.2 Software 4
3. Final project for the exam **4**
3.1 Specification of a research question (abstract) 4
3.2 Description of data 5
3.3 Data analysis oriented by the research question 5
3.4 Conclusions 5
3.5 References 5
3.6 Codes, scripts, softwares 5

3.7 Final notes 5
4.0 Readings **6**
4.1 Books 6
4.2 Blogs 6
4.3 Courses 6
4.4 Workshops 6
4.5 Tutorials 6

1. Lectures schedule

1.1 15/09/23 - Lecture 1

Lecturer: Maurizio Napolitano
Geospatial vector data

- Course introduction
- The basics of geographic data
- Vector data and projections

1.2 22/09/23 - Lecture 2

Lecturer: Maurizio Napolitano
Geospatial vector data

- Solutions to Exercises
- Spatial relations and operations
- Spatial SQL

1.3 29/09/23 - Lecture 3

Lecturer: Maurizio Napolitano
Geospatial vector data

- Solutions to Exercises
- Sources of vector data
- Use of OpenStreetmap data

=1.4 06/10/23 - Lecture 4

Lecturer: Diego Giuliani
Spatial statistics

- Spatial dependence and spatial autocorrelation
- Spatial weight matrix
- Spatial regression

1.5 13/10/23 - Lecture 5

Lecturer: Diego Giuliani
Spatial statistics

- Spatially continuous variables
- Spatial interpolation
- Geostatistics

1.6 20/10/23 - Lecture 6

Lecturer: Diego Giuliani
Spatial statistics

- Spatial point pattern data
- Point processes
- Point pattern analysis

1.7 27/10/23 - Lecture 7

Lecturer: Maurizio Napolitano
Analysis of flow data

- Solutions to Exercises
- Network analysis
- Analysis of GPX data

1.8 19/11/23 - Lecture 8

Lecturer: Maurizio Napolitano
Raster data

- Solutions to Exercises
- Basic operations on raster data
- Raster data and vector data

1.9 17/11/23 - Lecture 9

Lecturers: Maurizio Napolitano and Diego Giuliani
Representation of geospatial data

- Solutions to Exercises
- Important caveats about the visualization of geospatial data

Preparation of the final project for the exam

- Presentation of an example of final project

1.10 Notes:

Lectures topics may be subject to change

2. Tools used during the course

The lecture approach follows the "hands-on-learning" approach.
Students are advised to bring a laptop.
Most of the lessons take place through presentation of code in Python or R by using Jupyter Notebook.
Participants are asked to take care to install the software that will be proposed on their machines and to have the environment to work.
We recommend using Visual Studio Code.
All softwares used during the course are open source.

2.2 Software

- Python
  https://python.org
  and libraries that will be presented during the course
- R
  https://www.r-project.org/
  and libraries that will be presented during the course
- Quantum GIS
  https://www.qgis.org
- DuckDB
  https://duckdb.org
- Spatialite
  https://www.gaia-gis.it/fossil/libspatialite/index
- GDAL
  https://gdal.org
- Visual Studio Code
  https://code.visualstudio.com/

3. Final project for the exam

For the course's exam, students are required to perform a data analysis project and submit a report, which should respect the following structure:

3.1 Specification of a research question (abstract)

The data analysis has to be performed according to a clear research question supported by some scientific literature and should concern a well-defined geographical area.

The research question does not necessarily need to be original, it is possible to replicate an existing study for a different geographical context.

3.2 Description of data

Metadata and sources of the datasets used for the analysis should be provided, as well as any possible restriction of use.
For any dataset, an exploratory preliminary analysis showing why it is useful for the project should be performed. It is also important to clarify which data cleaning and wrangling operations have been used.

3.3 Data analysis oriented by the research question

This section should illustrate all the analyses that have been made to empirically validate the research question. All the results, tables, maps and graphs should be reported here.

3.4 Conclusions

Interpretation of the main findings and implication and discussion on how they relate to the research question.

3.5 References

Links to all the employed sources (data, software software tools and bibliographic references). They should be presented in the form of a list or table.

3.6 Codes, scripts, softwares

All the codes written to conduct the project have to be submitted with the report. It can be integrated in the report's text (for example through jupyter notebook or R markdown) or not.
The work should be done using one or more of the tools shown during the lectures (Python, R, QGIS, gdal, duckdb, spatialite …).
For R and Python, it is important to list the employed libraries, together with their version number and, in case, installation instructions.
In the case of softwares with a user interface, (e.g. QGIS), the basic sequence of commands needs to be listed.
It is essential to put the lecturers in the position to reproduce the steps proposed by the delivered project work.

3.7 Final notes

The report can be prepared as a PDF file or a Notebook and should be submitted by email or providing a link to a repository (such as github).
Extra work, such as the application of methodologies that are not treated during the course or the creation of websites, is welcome but it will not have a great effect on the final mark.
Group works are not allowed.

At the end of the course, the lecturers, or some companies interested in offering internship programs, will propose some research questions.

4 Readings

4.1 Books

Geographic Data Science with Python
By Sergio Rey, Dani Arribas-Bel, Levi John Wolf
https://geographicdata.science/

Spatial Data Science: With Applications in R (1st ed.)
Pebesma, E.; Bivand, R. (2023). Chapman and Hall/CRC.
https://doi.org/10.1201/9780429459016
https://r-spatial.org/book/

4.2 Blogs

Geohackweek
https://geohackweek.github.io/
Open Source Spatial Analysis Tools for Python: A Quick Guide
https://makepath.com/open-source-spatial-analysis-tools-a-quick-guide/

4.3 Courses

Automating GIS-processes
https://automating-gis-processes.github.io/site/
Spatial data science for sustainable development
https://sustainability-gis.readthedocs.io/en/latest/
Geographic Data Science for Applied Economists
http://darribas.org/gds4ae/content/pages/home.html

4.4 Workshops

GIS Open Source Python
https://www.earthdatascience.org/workshops/gis-open-source-python/

4.5 Tutorials

Earth Data Science
https://www.earthdatascience.org/

5. Traces for Exam

5.1 Analyzing Key Influences on Airbnb Pricing

Assessing the determinant factors of Airbnb's prices using spatial regression models.
The analysis can focus on a specific city or region.
Airbnb data: http://insideairbnb.com/get-the-data.
Some inspiring examples:
https://link.springer.com/article/10.1007/s00168-021-01064-z
https://www.sciencedirect.com/science/article/pii/S0261517719301797

5.2 Assessing Local Benefits of Urban Greenery

Measuring the benefit of greenness using the Normalised Difference Vegetation Index, or similar indices, at local level.
The analysis can focus on the statistical association between a vegetation index and other environmental, social or economic variables.
Some inspiring examples:
https://www.sciencedirect.com/science/article/pii/S2212095522001316
https://www.tandfonline.com/doi/full/10.1080/09640560802423509

5.3 Spatial analysis of crime

The analysis can focus on the determinants of crime or on the impact of crime on other dimensions.
Some inspiring examples:
https://ica-proc.copernicus.org/articles/4/5/2021/
https://www.emerald.com/insight/content/doi/10.1108/JES-07-2014-0121
Istat crime data at province level:
http://dati.istat.it/Index.aspx?DataSetCode=dccv_delittips