ENVS3/563 - Module Handbook

Dani Arribas-Bel

Contents

ENVS363/563

Geographic Data Science

Welcome to Geographic Data Science, a course taught by Dr. Dani Arribas-Bel in the Autumn of 2019 at the University of Liverpool.

The timetable for the course is:

- **Lectures**: *Monday* 12:00pm-1:00pm, `ERB-ERT`
- **Computer Labs**: *Thursdays* 1:00pm-3:00pm, `CTL-6-PCTC`

Locations

- `ERB-ERT`: Eleanor Rathbone Building - Eleanor Rathbone Theatre [URL]
- `CTL-6-PCTC`: Central Teaching Labs, PC Teaching Centre (Red, Green, Blue) [URL]

Contact

Dani Arribas-Bel - D.Arribas-Bel [at] liverpool.ac.uk \ Lecturer in Geographic Data Science \ Office 508, Roxby Building, \ University of Liverpool - 74 Bedford St S, \ Liverpool, L69 7ZT, \ United Kingdom.

Overview

Aims

The module provides students with little or no prior knowledge core competences in Geographic Data Science (GDS). This includes the following:

- Advancing their statistical and numerical literacy.
- Introducing basic principles of programming and state-of-the-art computational tools for GDS.
- Presenting 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.
- Focusing on real world applications of these techniques in a geographical and applied context.

Learning outcomes

By the end of the course, students will be able to:

- Demonstrate advanced GIS/GDS concepts and be able to use the tools programmatically to import, manipulate and analyse spatial data in different formats.
- Understand the motivation and inner workings of the main methodological approcahes of GDS, both analytical and visual.
- Critically evaluate the suitability of a specific technique, what it can offer and how it can help answer questions of interest.
- Apply a number of spatial analysis techniques and explain how to interpret the results, in a process of turning data into information.
- When faced with a new data-set, work independently using GIS/GDS tools programmatically to extract valuable insight.

Feedback strategy

The student will receive feedback through the following channels:

- Formal assessment of three summative assignments: two tests and a computational essay. This will be on the form of reasoning of the mark assigned as well as comments specifying how the mark could be improved. This will be provided no later than three working weeks after the deadline of the assignment submission.

- Direct interaction with Module Leader and demonstrators in the computer labs. This will take place in each of the scheduled lab sessions of the course.
- Online forum maintained by the Module Leader where students can contribute by asking and answering questions related to the module.

Key texts and learning resources

Access to materials, including lecture slides and lab notebooks, is centralized through the use of a course website available in the following url: `http://darribas.org/gds19`

Specific readings, videos, and/or podcasts, as well as academic references will be provided for each lecture and lab, and can be accessed through the course website.

Syllabus

Week 1: Introduction

- Lecture: Geographic Data Science.
- Tutorial: Tools + Manipulating data in Python - Tidy Data.

Week 2: Spatial Data

- Lecture: Geo-Data.
- Tutorial: Manipulating data in Python - Advanced Tricks.

Week 3: (Geo)Visualization

- Lecture: (Geo)Visualization.
- Tutorial: Manipulating geospatial data in Python.

Week 4: Choropleth mapping

- Lecture: Choropleth mapping.
- Tutorial: Mapping deprivation.

Week 5: Spatial Weights

- Lecture: Spatial Weights.
- Tutorial:
  - **TEST 1** (1h): Thursday Oct. 24th
  - Spatial Weights with PySAL.

Week 6: ESDA

- Lecture: Exploratory Spatial Data Analysis (ESDA).
- Tutorial: ESDA in Python.

Week 7: Clustering

- Lecture: Clustering.
- Tutorial: Geodemographic analysis.

Week 8: Point Data

- Lecture: Point Data.
- Tutorial: Exploring Twitter patterns.

Week 9

- Lecture: Assignment preparation.
- Tutorial:
  - **TEST 2** (1h): Thursday Nov. 21st
  - Assignment Clinic

Week 10: (Spatial) causal inference

- Lecture: Spatial causal inference.
- Tutorial: Assignment Clinic.

Week 11: Geographic Data Science in Action

- Lecture: Geographic Data Science *in the wild*.
- Tutorial: Assignment Clinic.

ASSIGNMENT due on Thursday, December 5th-2019.

Assessment

The final mark for the course is composed of the following three components:

- Contribution to module's discussion forum (5%)
- Assignment 1 (20%)
- Assignment 2 (25%)
- Assignment 3 (50%)

Assignments 1 and 2 are described below. Students should keep in mind the following information regarding the submission of assignments:

- **Submission is electronic only** and will be managed through Turnitin.
- Assignments will be prepared in the Jupyter Notebook file format and then converted into a self-contained HTML file that will then be submitted on Turnitin.

Marking Criteria

This course follows the standard marking criteria (the general ones and those relating to GIS assignments in particular) set by the School of Environmental Sciences. In addition to these generic criteria, the following specific criteria relating to the code provided will be used:

- **0-15**: the code does not run and there is no documentation to follow it.
- **16-39**: the code does not run, or runs but it does not produce the expected outcome. There is some documentation explaining its logic.
- **40-49**: the code runs and produces the expected output. There is some documentation explaining its logic.
- **50-59**: the code runs and produces the expected output. There is extensive documentation explaining its logic.
- **60-69**: the code runs and produces the expected output. There is extensive documentation, properly formatted, explaining its logic.
- **70-79**: all as above, plus the code design includes clear evidence of skills presented in advanced sections of the course (e.g. custom methods, list comprehensions, etc.).
- **80-100**: all as above, plus the code contains novel contributions that extend/improve the functionality the student was provided with (e.g. algorithm optimizations, novel methods to perform the task, etc.).

Assignment 1 - In-lab computer test I

- Type: `Coursework`
- 1h.
- Date: **Thursday, October 24th-2019** (Week 5).
- 20% of the final mark
- Chance to be reassessed
- In-lab electronic submission.

This assessment will consist of an in-lab computer test with multi-option and short answer questions covering topics introduced in lectures and labs 1-4. More details will be provided during class in advance.

Assignment 2 - In-lab computer test II

- Type: `Coursework`
- 1h.
- Date: **Friday, November 21st-2019** (Week 9).
- 25% of the final mark
- Chance to be reassessed
- In-lab electronic submission.

This assessment will consist of an in-lab computer test with multi-option and short answer

questions covering topics introduced in lectures and labs 5-9. More details will be provided during class in advance.

Assignment 3 - *Targetting areas*

- Type: `Coursework`
- [Equivalent to 2,500 words] Three maps/tables, code and 750 words.
- 50% of the final mark
- Chance to be reassessed
- Final Assessment
- Due on **Thursday, December 5th-2019** (Week 11).
- Electronic submission only. Static HTML with NO interactive cells.

In this assignment, you will take the role of a real-world data scientist tasked to identify areas to direct investments. You are consulting for the City of Liverpool on a program to target investments towards particularly disadvantaged areas that are nevertheless popular or have the potential to become so. The Economic Development division knows that only five local super output areas (LSOAs) will be funded but would like to know which ones.

Develop a data strategy, deploy it, and present the results in a rigorous but intuitive fashion, together with the code.

Minimum requirements (complete *all*)

- Combine *at least two* datasets, potentially among those used in the course.
- Employ *at least two* techniques from the set of analytics covered in the course.
- Justify why you have chosen the methods you use and how they help you answer the question at hand. Critically discuss their limitations too.
- Provide a list of the top five areas that you recommend be funded for improvement.
- Explain clearly how you have arrived at the list and how the decision has been informed by the data analysis and the methodologies employed.
- Include documented code that demonstrate programming understanding of the student.

Suggestive paths (optional)

- Combine a LISA analysis of deprivation with kernel density maps of Twitter activity to identify areas of high values at both.
- Combine several relevant variables into a geodemographic analysis to obtain candidate areas and display the results in an aesthetically pleasant choropleth.

Data

This assignment can use any of the datasets employed in the course, and/or any other datasets you consider useful. If you are thinking of including additional datasets, or have ideas in this respect, please get in touch with the module lead (Dani Arribas-Bel).

IMPORTANT - MSc Students: MSc students are *required* to use alternative datasets.