Home

GDS4AE - *Geographic Data Science for Applied Economists*

- Dani Arribas-Bel [@darribas]
- Diego Puga [@ProfDiegoPuga]

Contact

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

Diego Puga - diego.puga [at] cemfi.es
Professor
CEMFI,
Casado del Alisal 5,
28014 Madrid,
Spain.

Citation

If you use materials from this resource in your own work, we recommend the following citation:

```
@article{darribas_gds_course,
  author = {Dani Arribas-Bel and Diego Puga},
  title = {Geographic Data Science for Applied Economists},
  year = 2021,
  annote = {\href{https://darribas.org/gds4ae}}
}
```

Overview

This resource provides an introduction to Geographic Data Science for applied economists using Python. It has been designed to be delivered within 15 hours of teaching, split into ten sessions of 1.5h each.

How to follow along

GDS4AE is best followed if you can interactively tinker with its content. To do that, you will need two things:

1. A computer set up with the Jupyter Lab environment and all the required libraries (please see the Software stack part in the Infrastructure section for instructions)

2. A local copy of the materials that you can run on your own computer (see the repository section in the Infrastructure section for instructions)

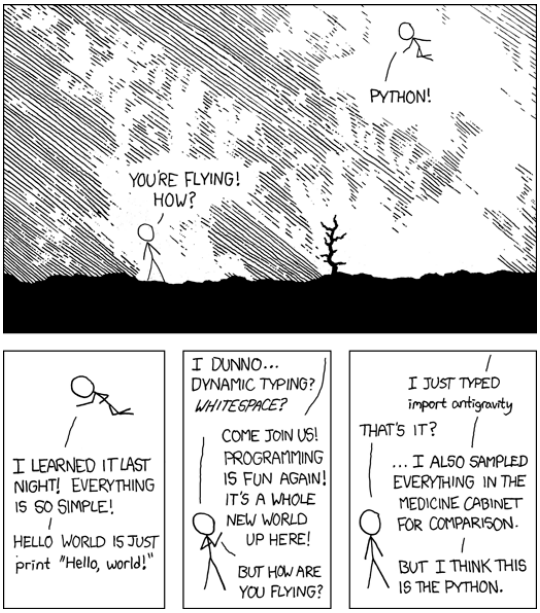
Content

The structure of content is divided in nine blocks:

- *Introduction*: get familiar with the computational envirionment of modern data science
- *Spatial Data*: what do spatial data look like in Python?
- *Geovisualisation*: make (good) data maps
- *Spatial Feature Engineering* (Part I and Part II)
- *Spatial Networks* (Part I and Part II)
- *Transport Costs*
- *Visual challenges*

Each block has its own section and is designed to be delivered in 1.5 hours approximately. The content of some of these blocks relies on external resources, all of them freely available. When that is the case, enough detail is provided in the to understand how additional material fits in.

Why Python?



Source: XKCD

Data

All the datasets used in this resource is freely available. Some of them have been developed in the context of the resource, others are borrowed from other resources. A full list of the datasets used, together with links to the original source, or to reproducible code to generate the data used is available in the Datasets page.

License

Infrastructure

This page covers a few technical aspects on how the course is built, kept up to date, and how you can create a computational environment to run all the code it includes.

Software stack

This course is best followed if you can not only read its content but also interact with its code and even branch out to write your own code and play on your own. For that, you will need to have installed on your computer a series of interconnected software packages; this is what we call a *stack*.

Instructions on how to install a software stack that allows you to run the materials of this course depend on the operating system you are using. Detailed guides are available for the main systems on the following resource, provided by the Geographic Data Science Lab:


@gdsl-ul/soft_install

Github repository

All the materials for this course and this website are available on the following Github repository:


@darribas/gds4ae

If you are interested, you can download a compressed `.zip` file with the most up-to-date version of all the materials, including the HTML for this website at:

Icon made by Freepik from www.flaticon.com


@darribas/gds4ae_zip

Containerised backend

The course is developed, built and tested using the `gds_env`, a containerised platform for Geographic Data Science. You can read more about the `gds_env` project at:

Binder

Binder is service that allows you to run scientific projects in the cloud for free. Binder can spin up *"ephemeral"* instances that allow you to run code on the browser without any local setup. It is possible to run the course on Binder by clicking on the button below:

> **Warning**
>
> It is important to note Binder instances are *ephemeral* in the sense that the data and content created in a session is **NOT** saved anywhere and is deleted as soon as the browser tab is closed.

Binder is also the backend this website relies on when you click on the rocket icon (🚀) on a page with code. Remember, you can play with the code interactively but, once you close the tab, all the changes are lost.

Introduction

Geographic Data Science

> **Note**
>
> This section is adapted from Block A of the GDS Course [AB19].

Before we learn *how* to do Geographic Data Science or even *why* you would want to do it, let's start with *what* it is. We will rely on two resources:

- First, in this video, Dani Arribas-Bel covers the building blocks at the First Spatial Data Science Conference, organised by CARTO.


```
- Second, *Geographic Data Science*, by Alex Singleton and Dani Arribas-Bel
{cite}`singleton2019geographic`
```

The computational stack

Spatial Data

Geovisualisation

Spatial Feature Engineering (I)

Spatial Feature Engineering (II)

Spatial Networks (I)

Spatial Networks (II)

Transport costs

Visual challenges and opportunities

Student presentations

In this session, students will present their projects to the group.

Datasets

Further Resources

If this course is successful, it will leave you wanting to learn more about using Python for (Geographic) Data Science. See below a few resources that are good "next steps".

Courses

- The "Automating GIS processes", by Vuokko Heikinheimo and Henrikki Tenkanen is a great overview of GIS with a modern Python stack:

> https://autogis-site.readthedocs.io/

- The "GDS Course" by Dani Arribas-Bel [AB19] is an introductory level overview of Geographic Data Science, including notebooks, slides and video clips.

> https://darribas.org/gds_course

Books

- *"Python for Geographic Data Analysis"*, by Henrikki Tenkanen, Vuokko Heikinheimo and David Whipp:

> https://pythongis.org/

- *"Geographic Data Science in Python"*, by Sergio J. Rey, Dani Arribas-Bel and Levi J. Wolf:

> https://geographicdata.science

Bibliography

[AB19]
Dani Arribas-Bel. A course on geographic data science. *The Journal of Open Source Education*, 2019. doi:https://doi.org/10.21105/jose.00042.

By Dani Arribas-Bel & Diego Puga