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.

Overview

Delivery modes

The course is delivered following a flipped, hybrid model by which some of the content is available to students *before* class time, and some is delivered in group sessions together with the instructor. The former is called *asynchronous* delivery, while the latter is referred to as *synchronous*.

Throughout the course content, each learning object (typically subsections in the materials) is tagged with one of the following two labels for reference:

- `[Async]` To be completed asynchronously, on your own, and *before* class
- `[Sync]` To be completed synchronously, with course instructor, remotely through video link

Similarly, the overview below contains them for a quick, general sense at what is expected from you before we meet in synchronous sessions.

Day 1 - Introduction

- `[Async]` Introduction to the course: *data, data, data*
- `[Async]` Computational building blocks

Day 2 - Tabular data

- `[Sync]` Manipulation
- `[Sync]` Visualisation
- `[Sync]` Advanced manipulation

Contents

Day 3 - Unsupervised learning

- Unsupervised learning
  - `[Async]` Overview
  - `[Sync]` Hands-on

Day 4 - Supervised learning

- Introduction
  - `[Async]` Overview
  - `[Sync]` Hands-on
- `[Sync]` Inference
- `[Sync]` Overfitting & cross-validation

Day 5

- Data Studio
  - `[Async]` Assignment brief
  - `[Sync]` Studio time

Data

All the data required for this course is contained in the zip compressed bundled download provided via Github.

Data preparation code is available, if you are so inclined, for you to peruse over here.

Further materials

A list of further materials where you can continue learning is available at:

```
http://darribas.org/gds19/further_resources.html
```

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/data_science_studio

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:

@darribas/data_science_studio_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

Geovisualisation

Spatial Feature Engineering (pt.I)

Spatial Feature Engineering (pt.II)

Spatial Networks (pt.I)

Spatial Networks (pt.II)

Transport costs

Visual challenges and opportunities

Student presentations

Datasets

Bibliography