

“Open by Default”

Developing reproducible, computational research

Dani Arribas-Bel

@darribas



The
Alan Turing
Institute

Martin Fleischmann

@martinfleis



Geographic
Data Science
Lab



Open research

- Open! Open? “Open”; *Open*; Open
 - 1. Transparent
 - 2. Accessible
 - 3. Transferrable
- Open data is *not* necessary, *nor* sufficient

“Open by default”

- Like happiness, it's about the journey not (only) the destination
- Treat research artifacts as OSS from Day 1
- Use (and nurture) the *existing* OSS infrastructure

For example



The Alan Turing Institute



Menu

[Home](#) + [Research](#) + [Research projects](#)

Urban Grammar

Learning an urban grammar from satellite data through AI

Learn more ↓

Project status
Ongoing

Related programmes
[Urban analytics](#)

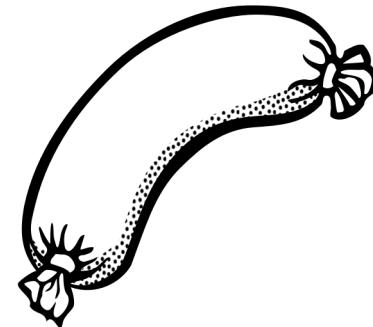
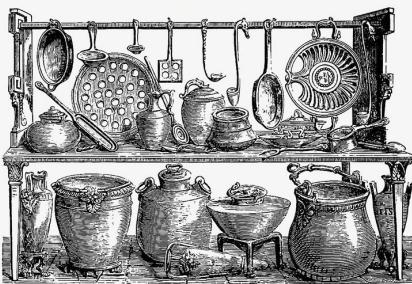


Urban Grammar

*A characterisation of space based on form and function
designed to understand urban environments*

+ Satellite imagery & AI

Needs



- Data, data, data
- (New) methods,
(new) code
- Infrastructure
- (Academic)
dissemination
- Open Data Products
- Journaling/reporting

The kitchen

Data

- Standard (formats) better than niche
- Open better than proprietary
- Static better than DB

E.g., parquet, COGs

New methods



Contribution to tobler: Speeding up X +

Contribution to tobler: Speeding up areal interpolation

Have you ever needed to link two sources of data, each attached to a different geometry? In our work in the [WP2](#), we do. We have to transfer data from various sources, linked to output areas, urban blocks or other spatial units to our own bespoke set of geographies. Therefore, we often need to do areal interpolation to correctly map data from one layer to another. Luckily, the open-source Python ecosystem can help.

Tobler, a part of [PySAL family](#), is a library for areal interpolation and dasymetric mapping which already offered what we needed. However, our data tends to be large, up to 15 million rows on which we need to interpolate several hundreds of thousands of rows of input data. That can take a while, so each performance improvement can help a lot.

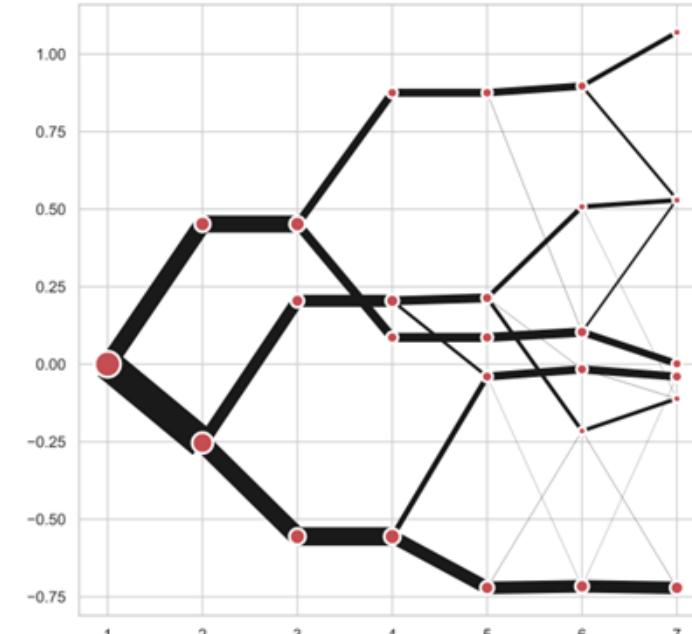
Urban Grammar
AI
research project

GitHub - martinfleis/clustergram X +

https://github.com/martinfleis/clustergram

README.md

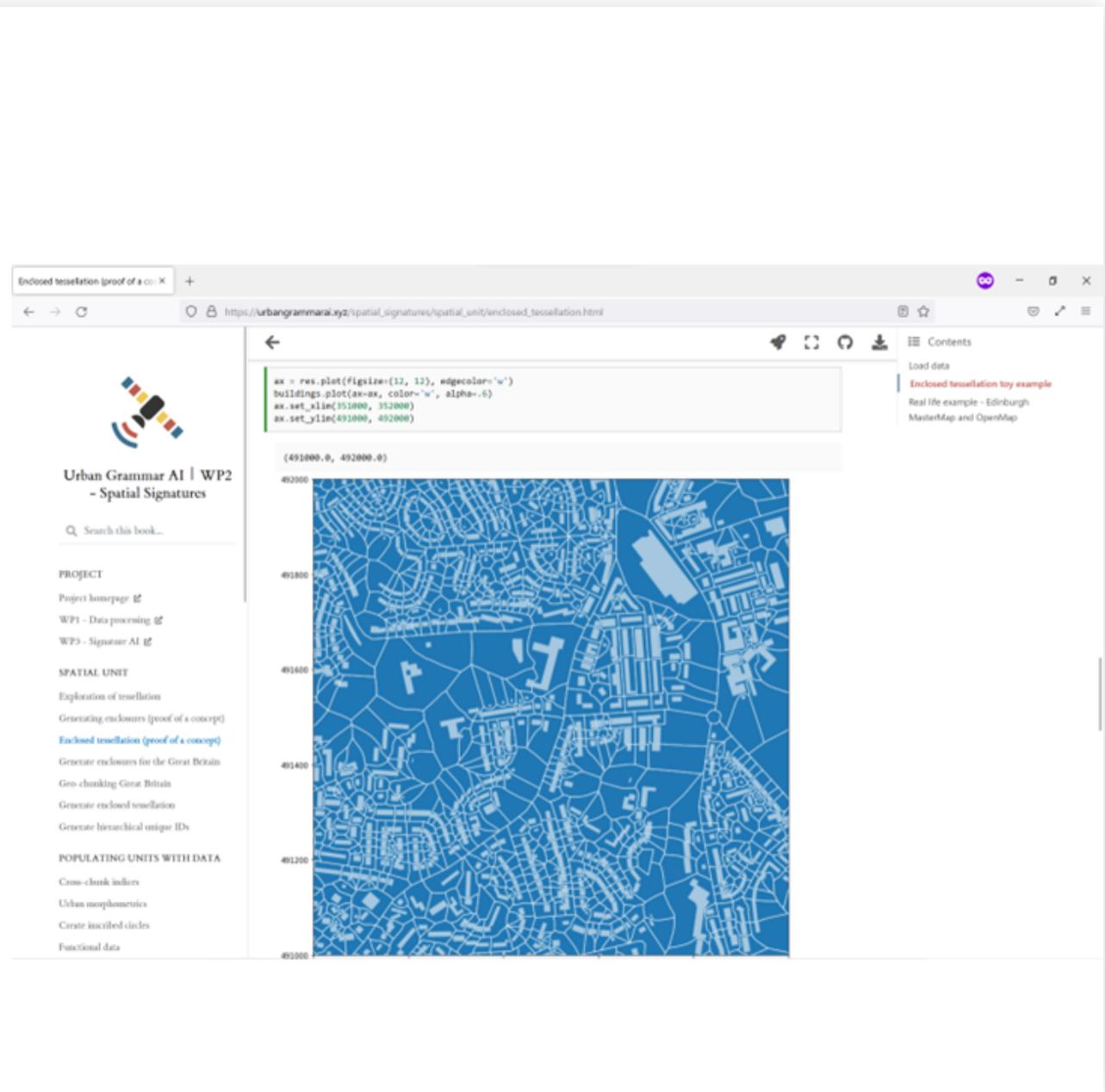
Clustergram



DOI 10.5281/zenodo.4750483

Visualization and diagnostics for cluster analysis

New code



Infrastructure

Home | gds_env

A containerised platform for Geographic Data Science: `gds_env`

[launch binder](#)

• Dani Arribas-Bel [[@darribas](#)]

The `gds_env` (short for “GDS environment”) provides a modern platform for Geographic Data Science. The project is a [Jupyter](#)-based stack that includes state-of-the-art **geospatial** libraries for **Python** and **R**. The `gds_env` is based on **container** technology to make it a transferrable platform for reproducibility. The source code is released under an [open source license](#) and the build process is transparent.

The `gds_env` extends the official [Jupyter Docker Stack](#) to include geospatial functionality in both Python and R. To offer more flexibility, this extension is provided in three different flavours, or stacks (to): `gds_py`, `gds` and `gds_dev`. Each of them builds on each other and adds further functionality. Please check the [Stacks section](#) for more information.

The goal of the `gds_env` is to make using Python and R for geospatial easy to set up in a large variety of contexts. The `gds_env` can support research and teaching activities, but is also suitable for data scientists using Python and R “in the field”. The stacks can be used in a range of environments, including: Windows/Mac/Linux laptops and desktops, servers, compute clusters, supercomputers or in the cloud (e.g. you can deploy them on [Binder](#)). For more information on how to build or install any of the stacks, check the [Guides section](#).

Building blocks

The `gds_env` stands on the shoulders of giants. Here are the core open technologies it is built with:



This site uses [Just the Docs](#), a documentation theme for Jekyll.

... and the “sausage”

Dissemination (I): Papers

The screenshot shows a GitHub repository page for the project "spatial_signatures_concept". The repository is private and has 245 commits across 5 branches. The "Code" tab is selected. The repository has 0 stars, 1 watcher, and 0 forks. It includes sections for "About", "Releases", "Packages", "Contributors", and "Languages".

About
Conceptual paper on Spatial Signatures
0 stars
1 watching
0 forks

Releases
No releases published
Create a new release

Packages
No packages published
Publish your first package

Contributors 2

- martinfleis** Martin Fleischmann
- darribas** Dani Arribas-Bel

Languages

https://github.com/urbangrammarai/spatial_signatures_concept/projects?type=beta

Dissemination (II): Talks

talks | Archive of public talks

Archive of public talks for the [Urban Grammar](#) project.

- March'21 - UBDC [\[PDF\]](#) [YouTube](#)
- March'21 - SAD seminars [\[PDF\]](#) [YouTube](#)
- April'21 - GDSL internal seminars [\[PDF\]](#)
- June'21 - ISUF conference [\[PDF\]](#) [YouTube](#)
- June'21 - Turing Urban Analytics [\[PDF\]](#) [YouTube](#)
- November'21 - Turing Townhall [\[PDF\]](#)
- November'21 - CASA Seminars [\[PDF\]](#) [YouTube](#)
- November'21 - Turing 2.0 (Leeds) [\[PDF\]](#)
- December'21 - Turing/Met Office [\[PDF\]](#)
- April'22 - GISRUK [\[PDF\]](#)

This site is open source. [Improve this page.](#)

urbangrammarai/talks: Archive

Archive of public talks

urbangrammarai.github.io/talks/

BSD-3-Clause License

3 stars 2 forks

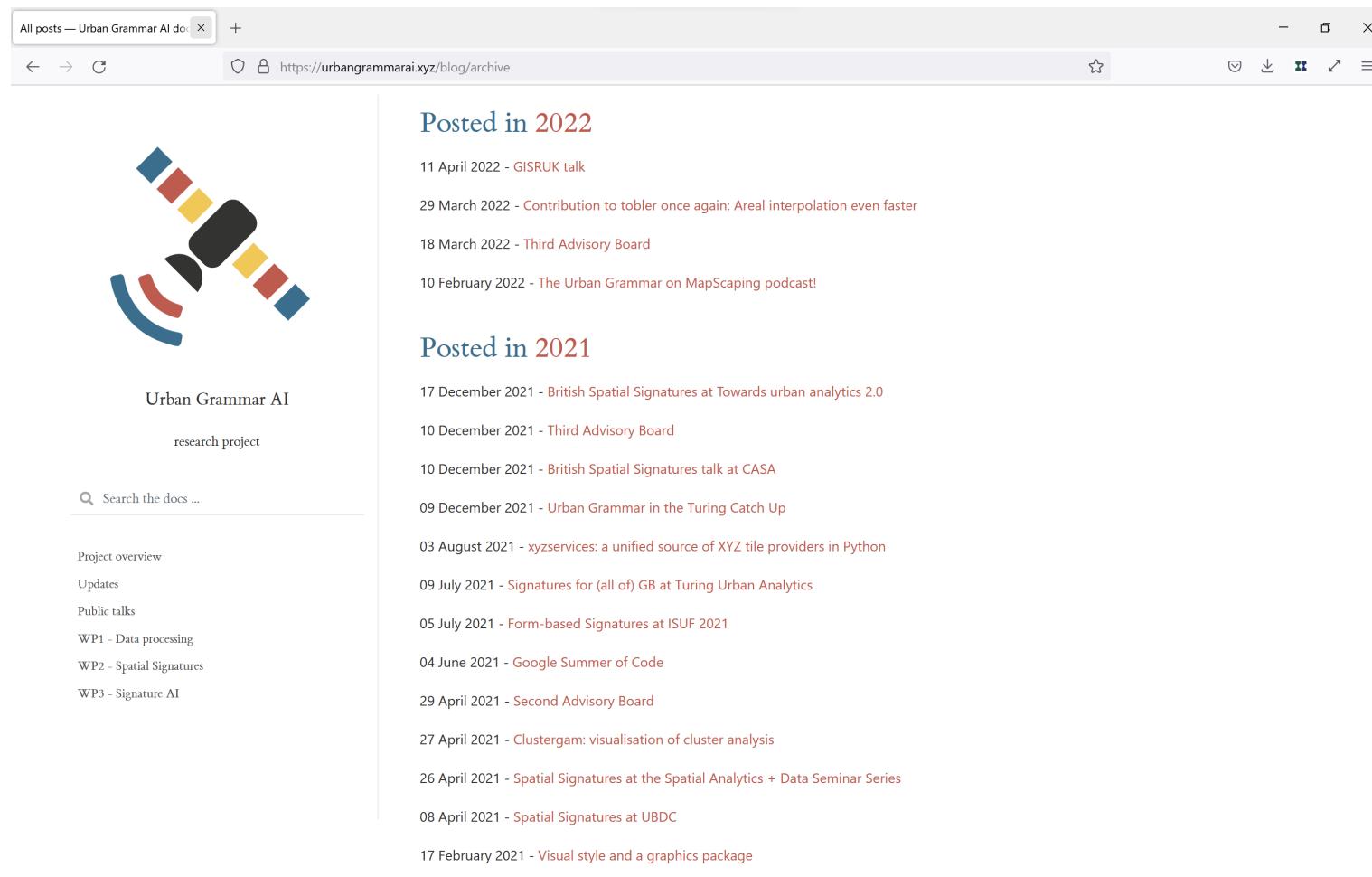
Star Unwatch

Code Issues 1 Pull requests Actions Projects Wiki ...

main

File	Last Commit
martinfleis.pdf	28 days ago 52
202103_sad	13 months ago
202103_ubdc	13 months ago
202104_gdsl	13 months ago
202106_ati	10 months ago
202106_isuf	10 months ago
202111_ati	6 months ago
202111_ati_leeds	5 months ago
202111_casa	5 months ago
202112_mo	5 months ago

Dissemination (III): Other



All posts — Urban Grammar AI do +

https://urbangrammarai.xyz/blog/archive

Posted in 2022

- 11 April 2022 - [GISRUK talk](#)
- 29 March 2022 - [Contribution to tobler once again: Areal interpolation even faster](#)
- 18 March 2022 - [Third Advisory Board](#)
- 10 February 2022 - [The Urban Grammar on MapScaping podcast!](#)

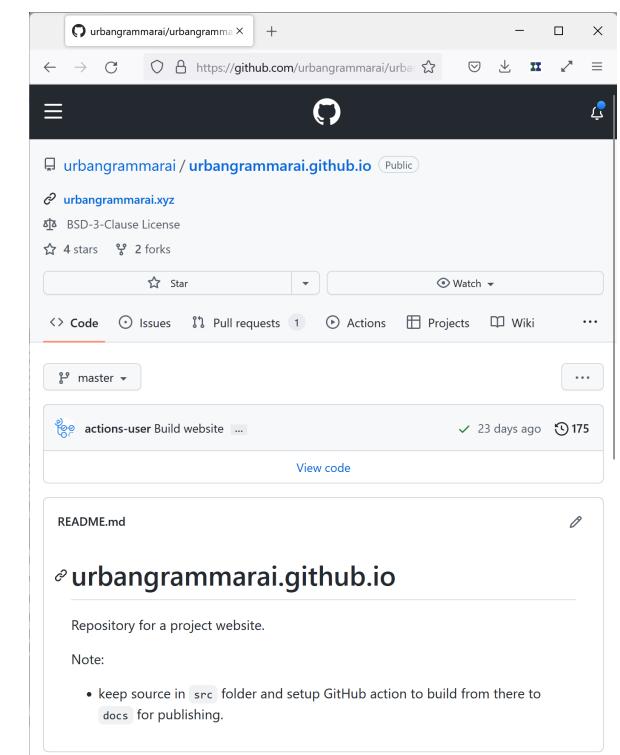
Posted in 2021

- 17 December 2021 - [British Spatial Signatures at Towards urban analytics 2.0](#)
- 10 December 2021 - [Third Advisory Board](#)
- 10 December 2021 - [British Spatial Signatures talk at CASA](#)
- 09 December 2021 - [Urban Grammar in the Turing Catch Up](#)
- 03 August 2021 - [xyzservices: a unified source of XYZ tile providers in Python](#)
- 09 July 2021 - [Signatures for \(all of\) GB at Turing Urban Analytics](#)
- 05 July 2021 - [Form-based Signatures at ISUF 2021](#)
- 04 June 2021 - [Google Summer of Code](#)
- 29 April 2021 - [Second Advisory Board](#)
- 27 April 2021 - [Clustergam: visualisation of cluster analysis](#)
- 26 April 2021 - [Spatial Signatures at the Spatial Analytics + Data Seminar Series](#)
- 08 April 2021 - [Spatial Signatures at UBDC](#)
- 17 February 2021 - [Visual style and a graphics package](#)

Urban Grammar AI
research project

Search the docs ...

Project overview
Updates
Public talks
WP1 – Data processing
WP2 – Spatial Signatures
WP3 – Signature AI



urbangrammarai/urbangrammarai Public

urbangrammaraixyz
BSD-3-Clause License
4 stars 2 forks

Star Watch

Code Issues Pull requests 1 Actions Projects Wiki

master

actions-user Build website ... 23 days ago 175 View code

README.md

urbangrammarai.github.io

Repository for a project website.

Note:

- keep source in `src` folder and setup GitHub action to build from there to `docs` for publishing.

Open Data Products (I): Data

Spatial Signatures of Great Britain

Population & Mobility Transport & Movement

Geographical Characterisation of British Urban Form and Function using the Spatial Signatures Framework

Spatial signatures characterise space based on form and function in a way designed to understand urban environments. This data product, part of the Urban Grammar project, contains a typology of spatial signatures in Great Britain. Each type has a distinct character capturing what the place looks like (form) and how it is used (function).

The data product contains bespoke Signature geometry with signature type, summary of input variables per each geometry and per each type, interpolation of signature types to OA and LSOA geometry and short pen portraits for the typology, shorthand descriptions of the characteristics of each signature type.

The interactive map showing the typology is available at <https://urbangrammarai.xyz/great-britain/>. More details about the project can be found at the project website <https://urbangrammarai.xyz>.

Related publications:

Fleischmann, Martin; Arribas-Bel, Daniel (2021): Geographical Characterisation of British Urban Form and Function using the Spatial Signatures Framework. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.16691575.v1>

Data Extent

Leaflet | Map data © OpenStreetMap

License

UK Open Government Licence (OGL)

Open Data Products (II): Exploratory tools

The image displays two side-by-side screenshots of open data products related to spatial signatures in Great Britain.

Left Screenshot: A web browser window showing a map of Great Britain with a color-coded spatial signature typology. A modal window titled "Spatial Signatures in Great Britain" provides a detailed description of the project, mentioning the Urban Grammar project, the Consumer Data Research Centre, and the project website. Logos for the Geographic Data Science Lab, the University of Liverpool, and The Alan Turing Institute are displayed. The map shows various regions with distinct colors representing different signature types. A legend at the bottom right of the map provides a key for these colors.

Right Screenshot: A GitHub repository page for "urbangrammarai/great-britain". The repository is public and has 2 pull requests, 0 forks, and 0 stars. It includes sections for Code (with master branch), Issues (1 issue), Pull requests, Actions, Projects, Wiki, Security, and About. The About section contains information about the project, links to the GitHub page, and statistics like 0 stars, 2 watching, and 0 forks. It also features a "Help people interested in this repository understand your project by adding a README." button and a "Add a README" button.

The takeaway

- “Release early, release often”
- Make open a feature, not a bug
- *Share your knowledge, it is a way to achieve immortality* (Dalai Lama)

A bit more to read

Rey, S., Arribas-Bel, D., & Wolf, L. (*in press*). Computational Tools for Geographic Data Science. Chapter 2 in *Geographic Data Science with Python*. Available at:

<https://geographicdata.science/book>

Boeing, G., & Arribas-Bel, D. (2021). GIS and Computational Notebooks. *Geographic Information Science & Technology Body of Knowledge*, 2021(Q1).

Arribas-Bel, D., Green, M., Rowe, F., & Singleton, A. (2021). Open data products—A framework for creating valuable analysis ready data. *Journal of Geographical Systems*, 23(4), 497–514.

<https://urbangrammarai.xyz>

“Open by Default”

Developing reproducible, computational research

Dani Arribas-Bel

@darribas

Martin Fleischmann

@martinfleis



The
Alan Turing
Institute



Geographic
Data Science
Lab