> Geographic Data Science

with

PySAL

and the

pydata stack

Sergio J. Rey
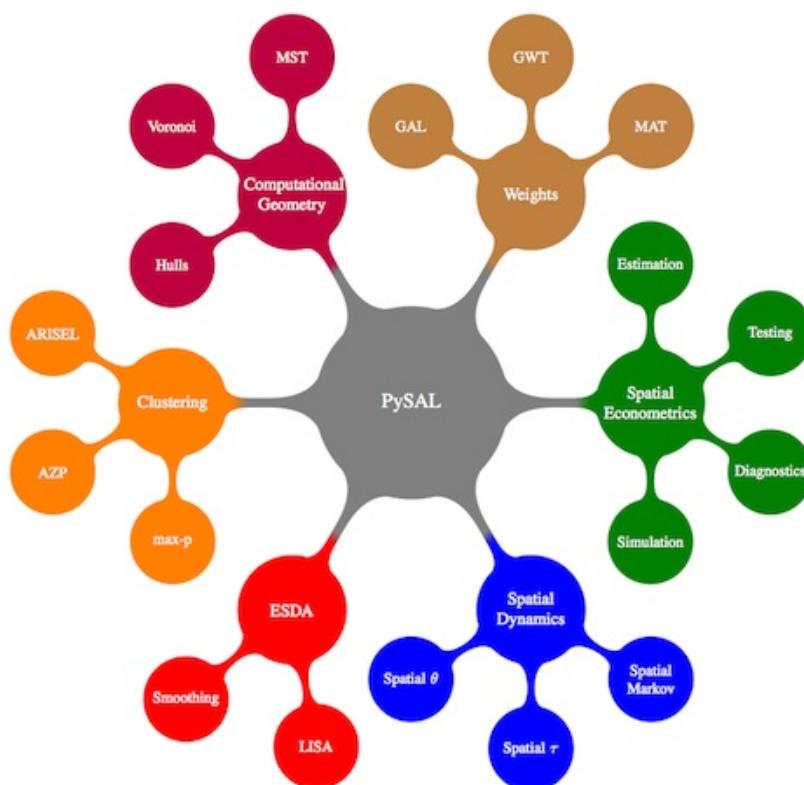
Dani Arribas-Bel

Table of Contents

Geographic Data Science with PySAL and the pydata stack

This two-part tutorial will first provide participants with a gentle introduction to Python for geospatial analysis, and an introduction to version `PySAL` 1.11 and the related eco-system of libraries to facilitate common tasks for Geographic Data Scientists. The first part will cover munging geo-data and exploring relations over space. This includes importing data in different formats (e.g. shapefile, GeoJSON), visualizing, combining and tidying them up for analysis, and will use libraries such as `pandas`, `geopandas`, `PySAL`, or `rasterio`. The second part will provide a gentle overview to demonstrate several techniques that allow to extract geospatial insight from the data. This includes spatial clustering and regression and point pattern analysis, and will use libraries such as `PySAL`, `scikit-learn`, or `clusterpy`. A particular emphasis will be set on presenting concepts through visualization, for which libraries such as `matplotlib`, `seaborn`, and `folium` will be used.

Distribution

[URL] [PDF] [EPUB] [MOBI] [IPYNB]

License

Geographic Data Science with PySAL and the pydata stack by Sergio J. Rey and Dani Arribas-Bel is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

About the authors

Sergio Rey is professor of geographical sciences and core faculty member of the GeoDa Center for Geospatial Analysis and Computation at the Arizona State University. His research interests include open science, spatial and spatio-temporal data analysis, spatial econometrics, visualization, high performance geocomputation, spatial inequality dynamics, integrated multiregional modeling, and regional science. He co-founded the Python Spatial Analysis Library (PySAL) in 2007 and continues to direct the PySAL project. Rey is a fellow of the spatial econometrics association and editor of the journal Geographical Analysis.

Dani Arribas-Bel is Lecturer in Geographic Data Science and member of the Geographic Data Science Lab at the University of Liverpool (UK). Dani is interested in undestanding cities as well as in the quantitative and computational methods required to leverage the power of the large amount of urban data increasingly becoming available. He is also part of the team of core developers of PySAL, the open-source library written in Python for spatial analysis. Dani regularly teaches Geographic Data Science and Python courses at the University of Liverpool and has designed and developed several workshops at different levels on spatial analysis and econometrics, Python and open source scientific computing.

Outline

Part I

1. Software and Tools Installation (10 min)

2. Spatial data processing with PySAL (45 min)

   a. Input-output

   b. Visualization and Mapping

   c. Spatial weights

3. Exercise (10 min.)

4. ESDA with PySAL (45 min)

   a. Global Autocorrelation

   b. Local Autocorrelation

   c. Space-Time exploratory analysis

5. Exercise (10 min)

Part II

1. Spatial clustering a (30 min)

   a. Geodemographic analysis

   b. Regionalization

2. Exercise (10 min)

3. Spatial Regression (30 min)

   a. Overview

   c. Basic spatial regression: spatial lag and error model

4. Exercise (30 min)

5. Point Patterns (30 min)

   a. Kernel Density Estimation visualization

   b. Centrography and distance based statistics

6. Exercise (10 min)

Data

Description of datasets here.

Part I

Software and Tools Installation

Dependencies

Participants should have installed the following dependencies:

- [Anaconda](#) or [MiniConda](#) Python distributions for Python 2.7. See installation instructions on the links.
- `git`
- A `conda` environment loaded with all the dependencies can be installed by running the `pydata.sh` script available as part of the `envs` repository ([Github link](#)). To install it, follow these instructions:

    - Clone the repository on your machine:

        ```
        > git clone https://github.com/darribas/envs.git
        ```

    - Navigate into the folder:

        ```
        > cd envs
        ```

    - Run the script:

        ```
        > bash pydata.sh
        ```

Once installed, you need to activate the environment to run the notebooks. In Windows, open up [PowerShell](#) and type:

```
> activate pydata
```

And if you are on GNU/Linux or OSX:

```
> source activate pydata
```

Get started

Instructions to fire up a notebook here.

Spatial Data Processing

Notebook here.

```
%matplotlib inline

import pysal as ps
```

ESDA with `PySAL`

```
%matplotlib inline

import pysal as ps
```

```
/home/dani/anaconda/envs/pydata/lib/python2.7/site-packages/matplotlib/font_manager.py:273: UserWarning: Matplotlib i
s building the font cache using fc-list. This may take a moment.
  warnings.warn('Matplotlib is building the font cache using fc-list. This may take a moment.')
```

Part II

Spatial Clustering

Spatial Regression

Point Patterns