Workshop Proposal: PySAL: Python Spatial Analysis Library

Sergio Rey
sergio.rey@ucr.edu
University of California-Riverside
Riverside, California

Elijah Knaap
knaap@ucr.edu
University of California-Riverside
Riverside, California

ABSTRACT

PySAL is an open source library for cross platform geospatial data science. PySAL contains a family of 15 packages for different specialized areas of spatial data analysis, from geoprocessing, geovisualization, exploratory spatial data analysis, spatial machine learning, and confirmatory spatial analysis. The workshop focuses on a subset of PySAL dealing with geoprocessing, geovisualization, exploratory spatial data analysis, and regionalization.

KEYWORDS

spatial analysis, data science, urban analytics

ACM Reference Format:

Sergio Rey and Elijah Knaap. 2020. Workshop Proposal: PySAL: Python Spatial Analysis Library. In *2nd ACM SIGSPATIAL International Workshop on Geospatial Data Access and Processing APIs (SpatialAPI'20)*, November 3–6, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3423452.3430696>

This work is licensed under a Creative Commons Attribution International 4.0 License.

SpatialAPI'20, November 3–6, 2020, Seattle, WA, USA
ACM ISBN 978-1-4503-8164-2/20/11
<https://doi.org/10.1145/3423452.3430696>

1 INTRODUCTION

PySAL is an open source library for cross platform geospatial data science. It is comprised of 16 python packages for different specialized areas of spatial data analysis, from geoprocessing, geovisualization, exploratory spatial data analysis, spatial machine learning, and confirmatory spatial analysis. Since its initial public release in 2010 (Rey 2019), PySAL has demonstrated the benefits of an open source geographic science library and seen widespread adoption across a diverse set of applications. As a software library, PySAL is relied upon by a number of other analysis packages to develop specialized tools for spatial analysis, prominent examples include geopandas, geoplots, momepy, and geosnap. This workshop provides attendees with an introduction to the library for applied spatial data science, focusing on a subset of PySAL dealing with geoprocessing, geovisualization, exploratory spatial data analysis, and regionalization.

1.1 Background

PySAL is a family of packages divided into four major components: lib, model, explore, and viz. The lib layer provides tools to solve a wide variety of computational geometry problems including graph construction from polygonal lattices, lines, and points, construction and interactive editing of spatial weights matrices and graphs - computation of alpha shapes, spatial indices, and spatial-topological relationships, and reading and writing of sparse graph data, as well as pure python readers of spatial vector data. Unlike other PySAL layers, these functions are exposed together as a single package. The explore layer includes packages for exploratory analysis of spatial and spatio-temporal data. The model layer focuses on confirmatory analysis. In particular, its packages focus on the estimation of spatial

relationships in data with a variety of linear, generalized-linear, generalized-additive, nonlinear, multi-level, and local regression models. The viz layer provides functionality to support the creation of geovisualisations and visual representations of outputs from a variety of spatial analyses.

During this tutorial, attendees will be introduced to modules from each of the four sections of the PySAL library, and the instructors will describe how the modules can be used in concert to execute spatial analyses ranging from simple to highly complex.

2 TUTORIAL PLAN

2.1 Part I

Geoprocessing & Fundamentals of Spatial Analysis

- (1) • Spatial data input-output: Reading and writing GIS file formats, spatial data wrangling, changing coordinate transformation systems.
- Spatial weights & feature engineering: Introduction to the spatial weights matrix for formally encoding geographic relationships.

Geovisualization

- (1) • Choropleth mapping: Introduction to choropleth map classification using mapclassify.
- Interactive geovisualization: Basic visualization with GeoPandas, and matplotlib as well as interactive visualization via folium, leaflet and geoviews/hvplot,

2.2 Part II

- (1) Hot and cold spot detection
 - Global spatial autocorrelation: Exploratory spatial data analysis and overview of measures of spatial autocorrelation such as Moran's I and the join-count statistic.
 - Local indicators of spatial autocorrelation: Local Indicators of Spatial Association (LISA)
- (2) Regionalization
 - Agglomerative spatial clustering: Introduction to classic and spatially-constrained geodemographics (regionalization). An overview of integrating scikit-learn and pysal to develop socio-demographic cluster models that optionally include a spatial constraint.
 - Max-p: An optimization model for maximizing the number of neighborhood groups

3 MATERIALS/PREREQUISITES

The infrastructure for the tutorial will be similar to that used in previous workshops on PySAL. Instructions for installation of all software and tutorial data sets will be made available prior to the workshop. Additionally, a JupyterHub instance that runs all the software will be available for participants to interact live without having to do local installations.