Open Source Spatial Analysis
Lessons for research and education from PySAL

Sergio Rey

GeoDa Center for Geospatial Analysis and Computation
School of Geographical Sciences and Urban Planning
Arizona State University

Open Source Geospatial Research & Education Symposium
Yverdon-les-Bains, Switzerland
Oct. 26, 2012
Outline

1. Open Source
2. PySAL
3. Lessons for Education
4. Lessons for Research
5. Conclusions
Outline

1. Open Source
2. PySAL
3. Lessons for Education
4. Lessons for Research
5. Conclusions
How it all started (for me)
How it all started (for me)

Jim LeSage
How it all started (for me)

Jim LeSage
How it all started (for me)

Jim LeSage

\LaTeX
How it all started (for me)

Jim LeSage

\LaTeX

hamm (2.0)
Open Source Freedom: As in Beer

- Kid in a candy shop
- No longer constrained to proprietary tools
- Fundamental change in my research productivity

apt-get install *

Yes. “The first hit of heroin is always free.” (Scott McNealy)
Open Source Freedom: As in Beer

apt-get install *

- Kid in a candy shop
- No longer constrained to proprietary tools
- Fundamental change in my research productivity

This open source stuff is really free?
apt-get install *

- Kid in a candy shop
- No longer constrained to proprietary tools
- Fundamental change in my research productivity

This open source stuff is really free?

Yes. “The first hit of heroin is always free.” (Scott McNealy)
Open Source

“The twentieth century’s only true innovative concept in business, representing all that is truly new in the new economy.”
(Sandred 2001, p. xlii)
What about Science and Open Source?

Analogies often made

- Peer review
- Standing on the shoulders of giants
- Community and culture
What about Science and Open Source?

Analogies often made

- Peer review
- Standing on the shoulders of giants
- Community and culture

But are these accurate?
Open Source and Academia: Separated at Birth?
Open Source and Academia: Separated at Birth?

Saint IGNUcius Stallman left MIT
Saint IGNUcius Stallman left MIT
Open Source and Academia: Separated at Birth?

Saint IGNUcius Stallman left MIT

Torvalds left U. of Helsinki
Ancient History: Spatial Analysis and Open Source

Center for Spatially Integrated Social Science

Specialist Meeting on Spatial Data Analysis Software Tools

This meeting will bring together software developers from both the public/academic sector as well as the private sector who deal with tools to visualize spatial data (geovisualization), carry out exploratory spatial data analysis (ESDA) and facilitate spatial modeling (spatial regression modeling, spatial econometrics, geostatistics), with a special focus on the potential for social science applications.

Dr. Sergio Rey, Organizer

Dr. Luc Anselin, Organizer

Specialist Meeting on Spatial Data Analysis Software Tools
Upham Hotel, Santa Barbara, CA
May 10-11, 2002
Early Doubts: Spatial Analysis and Open Source

“...doubts held by some developers of spatial analysis software about the value of open source code and the underlying development model. Levine (2001) argued that once the code for a complex spatial analysis program was no longer the province of a single producer there would be the possibility of security breaches and quality control problems.”
Outline

1. Open Source
2. PySAL
3. Lessons for Education
4. Lessons for Research
5. Conclusions
PySAL Objectives

Leverage Existing Tools Development
- GeoDa/PySpace
- STARS

Develop Core Library
- spatial data *analytical* functions
- enhanced specialization, modularity
- fill void in geospatial Python libraries

Flexible Delivery Mechanisms
- interactive shell
- GUI
- Toolkits
- webservices
Acknowledgments

- **NSF** CyberGIS Software Integration for Sustained Geospatial Innovation
- **NIJ** Flexible Geospatial Visual Analytics and Simulation Technologies to Enhance Criminal Justice Decision Support Systems
- **NIH** Geospatial Factors and Impacts: Measurement and Use
- **NSF** Spatial Analytical Framework for Examining Sex Offender Residency Issues Over Space and Time
- **NSF** An Exploratory Space-Time Data Analysis Toolkit for Spatial Social Science Research
- **NSF** Hedonic Models of Location Decisions with Applications to Geospatial Microdata
Team

Serge Rey          Luc Anselin
Charles Schmidt   Dave Folch
Myunghwa Hwang    Dani Arribas
Phil Stephens     Julia Koschinsky
Pedro Amaral      Nick Malizia
Xing Kang         Xun Li
Xinyue Ye         Andrew Winslow
Mark McCann        Ran Wei
Nancy Lozano      Jing Yao

and contributions from many others!
Welcome

PySAL is an open source cross-platform library of spatial analysis functions written in Python. It is intended to support the development of high level applications for spatial analysis.

Documentation

- PySAL 1.4, released 2012 07 31

All Versions

- Unreleased development version
- PySAL 1.3, released 2012 01 31
- PySAL 1.2, released 2011 07 31
- PySAL 1.1, released 2011 01 31
- PySAL 1.0, released 2010 08 01

News

2012-07-31 PySAL 1.4 Stable (Downloads)
2012-01-31 PySAL 1.3 released
2012-01-19 PySAL 1.3 release code sprint on January 23
2011-08-04 PySAL now part of pythonanywhere
# Spatial Weights

## Types
- Contiguity
- Distance Based
- Kernel Weights
- Regimes
- Hybrid

## Operations
- Standardizations
- Higher orders
- Summary properties
- Set operations
- Conversions
Supported Weights Formats

In [7]: pysal.open.check()

PySAL File I/O understands the following file extensions:

Ext: '.shp', Modes: ['r', 'wb', 'w', 'rb']
Ext: '.mtx', Modes: ['r', 'w']
Ext: '.swm', Modes: ['r', 'w']
Ext: '.mat', Modes: ['r', 'w']
Ext: '.shx', Modes: ['r', 'wb', 'w', 'rb']
Ext: '.stata_text', Modes: ['r', 'w']
Ext: '.geoda_txt', Modes: ['r']
Ext: '.dbf', Modes: ['r', 'w']
Ext: '.dat', Modes: ['r', 'w']
Ext: '.gwt', Modes: ['r', 'w']
Ext: '.gal', Modes: ['r', 'w']
Ext: '.arcpyis_text', Modes: ['r', 'w']
Ext: '.wk1', Modes: ['r', 'w']
Ext: '.arcgis_dbf', Modes: ['r', 'w']
Ext: '.geobugs_text', Modes: ['r', 'w']
Ext: '.csv', Modes: ['r']
Ext: '.wkt', Modes: ['r']
W from Shapefiles

```python
>>> import pysal
>>> w = pysal.rook_from_shapefile('south.shp')
>>> w.n
1412
>>> w.s0
7700.0
>>> w.n**2
1993744
>>> w.pct_nonzero
0.0038620805880795125
>>> wq = pysal.queen_from_shapefile('south.shp')
>>> wq.s0
8096.0
>>> w.histogram
[(1, 16), (2, 32), (3, 65), (4, 187), (5, 378), (6, 435), (7, 230),
(8, 56), (9, 11), (10, 2)]
>>> wq.histogram
[(1, 16), (2, 29), (3, 54), (4, 141), (5, 306), (6, 437), (7, 319),
(8, 86), (9, 18), (10, 5), (11, 1)]
```
Space-Time EDA

Spatial pattern of dynamics
- Temporal co-movements
- Spatial clustering of dyads (and beyond)
- Optimal currency areas
- Convergence clubs, spatial poverty traps

Temporal stability of spatial patterns
- Choropleth map and raster comparisons
- Emergent properties (ABM)
- Hot and cold spot detection
Spatial Markov

```python
>>> import pysal
>>> f = pysal.open(pysal.examples.get_path("usjoin.csv"))
>>> pci = np.array([[f.by_col[str(y)] for y in range(1929,2010)]])
>>> pci = pci.transpose()
>>> rpci = pci/(pci.mean(axis=0))
>>> w = pysal.open(pysal.examples.get_path("states48.gal"))
>>> w.transform = 'r'
>>> sm = Spatial_Markov(rpci, w, fixed=True, k=5)
>>> for p in sm.P:
...     print p
...
[[ 0.96341463  0.0304878  0.00609756  0.    0.    ]
 [ 0.06040268  0.83221477  0.10738255  0.    0.    ]
 [ 0.    0.14     0.74     0.12    0.    ]
 [ 0.    0.03571429  0.32142857  0.57142857 0.07142857]
 [ 0.    0.        0.        0.16666667 0.83333333]]
[[ 0.79831933  0.16806723  0.03361345  0.    0.    ]
 [ 0.0754717  0.88207547  0.04245283  0.    0.    ]
 [ 0.00537634  0.06989247  0.8655914  0.05913978 0.    ]
 [ 0.    0.06372549  0.90196078  0.03431373 0.    ]
 [ 0.    0.        0.        0.19444444 0.80555556]]
```

Sergio Rey (ASU)
PySAL
OGRS 2012
Space-Time Rank Concordance

```python
>>> import pysal
>>> f=pysal.open(pysal.examples.get_path("mexico.csv"))
>>> vnames=["pcgdp%d"%dec for dec in range(1940,2010,10)]
>>> y=np.transpose(np.array([f.by_col[v] for v in vnames]))
>>> regime=np.array(f.by_col["esquivel99"])  
>>> w=pysal.weights.regime_weights(regime)
>>> np.random.seed(12345)
>>> res=[pysal.SpatialTau(y[:,i],y[:,i+1],w,99) for i in range(6)]
>>> for r in res:
...    ev = r.taus.mean()
...    print("%8.3f %8.3f %8.3f\n"%r.tau_spatial, ev, r.tau_spatial_psim)
...
'  0.281  0.466  0.010'
'  0.348  0.499  0.010'
'  0.460  0.546  0.020'
'  0.505  0.532  0.210'
'  0.483  0.499  0.270'
'  0.572  0.579  0.280'
```
# pvs.py
import pysal
import numpy as np
from scipy.stats import kendalltau
import time
ns = [25000, 50000, 100000, 200000]
for n in ns:
    x = np.arange(n); y = np.arange(n)
y[40:50] = 20
y = np.random.permutation(y)
t1 = time.time(); res_p = pysal.Tau(x,y); t2 = time.time()
res_s = kendalltau(x,y); t3 = time.time()
print 'n=%d scipy/pysal: %8.3f'%(n, (t3-t2)/(t2-t1))

>>> run pvs.py
n=25000 scipy/pysal:  3.691
n=50000 scipy/pysal:  4.030
n=100000 scipy/pysal:  3.921
n=200000 scipy/pysal:  3.891
Global and Local Autocorrelation

Global Autocorrelation

\[ l_t = \left( \frac{n}{S_0} \right) \frac{\sum_i \sum_j z_{i,t} w_{i,j} z_{j,t}}{\sum_i z_{i,t}^2} \]  

Local Autocorrelation

\[ l_{i,t} = \left( \frac{z_{i,t}}{m_2} \right) \sum_j w_{i,j} z_{j,t} \]

\[ m_2 = \sum_i z_{i,t}^2 / n \]
LISA Markov

![Graphs showing spatial lag of relative income](image-url)
LISA Markov

```python
>>> import numpy as np
>>> f = pysal.open(pysal.examples.get_path("usjoin.csv"))
>>> pci = np.array([f.by_col[str(y)] for y in range(1929,2010)]).transpose()
>>> w = pysal.open(pysal.examples.get_path("states48.gal")).read()
>>> lm = LISA_Markov(pci,w)
>>> lm.classes
array([1, 2, 3, 4])
>>> lm.steady_state
matrix([[ 0.28561505],
        [ 0.14190226],
        [ 0.40493672],
        [ 0.16754598]])
>>> lm.p
matrix([[ 0.92985458, 0.03763901, 0.00342173, 0.02908469],
        [ 0.07481752, 0.85766423, 0.06569343, 0.00182482],
        [ 0.00333333, 0.02266667, 0.948 , 0.026          ],
        [ 0.04815409, 0.00160514, 0.06420546, 0.88603531]])
```
LISA Markov - Spillover and Components

```python
>>> lm_random = pysal.LISA_Markov(pci, w, permutations=99)
>>> r = lm_random.spillover()
>>> r['components'][:,12]
array([ 0., 1., 0., 1., 0., 2., 2., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 2., 2., 0., 0., 0., 0., 0., 0., 1.,
       2., 2., 0., 2., 0., 0., 0., 1., 2., 2., 0., 0.,
       0., 0., 0., 2., 0., 0., 0., 0., 0., 0.])
>>> r['components'][:,13]
array([ 0., 2., 0., 2., 0., 1., 1., 0., 0., 2., 0., 0., 0., 0.,
       0., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0., 0.,
       0., 1., 0., 1., 0., 0., 0., 0., 2., 1., 1., 0., 0.,
       0., 0., 2., 1., 0., 2., 0., 0., 0.])
>>> r['spill_over'][:,12]
array([ 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0.,
       0., 0., 1., 0., 0., 0., 0., 0.])
```
CAST: Crime Analytics for Space-Time

NIJ Project
- Under development
- Stand-alone program based on PySAL’s spatial dynamics module
- WxPython front-end
Linked Maps
Linked Views
LISA Markov Plots

![LISA Markov Plot](image)
Spatial Econometrics

- Still Rare in Commercial Econometrics Packages
  - exception: Stata (ado files)
- Many Specialized Scripts
  - SAS, SPSS, Xlispstat, RATS, etc
- Several Open Toolboxes
  - spdep R, LeSage-Pace-Elhorst Matlab
Goals

- Handle large problems
- Efficient spatial weights
- Modularity and reusability
- Model complexity
Spatial Lag Model

**Specification**

\[ y = \rho Wy + X\beta + u, \]  

with \( u \sim N(0, \Sigma_\theta) \).

**Log Likelihood**

\[
L = -(n/2)(\ln 2\pi) - (1/2) \ln |\Sigma_\theta| + \ln |I - \rho W| \\
- (1/2)(y - \rho Wy - X\beta)' \Sigma_\theta^{-1}(y - \rho Wy - X\beta). 
\]

**Prediction**

\[
\hat{y} = (I - \hat{\rho} W)^{-1} X\hat{\beta}. 
\]
spreg

Functionality

- OLS, Two Stage Least Squares
- Spatial Two Stage Least Squares
- GM Error (Kelejian and Prucha 98-99)
- GM Error Homoskedasticity (Drukker et al. 2010)
- GM Error Heteroskedasticity (Arraiz et al. 2010)
- Spatial HAC variance-covariance estimation
- Anselin-Kelejian tests for residual spatial autocorrelation from IV regression
- LM diagnostics spatial error and spatial lag
- Robust-LM diagnostics spatial error and spatial lag
- Probit
- Probit with spatial diagnostics
OLS with Diagnostics
Model Setup: ModelType = Standard, Endogenous = N
Data: south.dbf N: 1412 df: 1406
Dependent Variable: HR90
Spatial Weights: south.gal Type: binary
R2: 0.30915681 Adjusted R2: 0.30670005

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>8.96249956</td>
<td>1.78133335</td>
<td>5.03134326</td>
<td>0.000001</td>
</tr>
<tr>
<td>RD90</td>
<td>4.58777939</td>
<td>0.21456950</td>
<td>21.38132067</td>
<td>0.000000</td>
</tr>
<tr>
<td>PS90</td>
<td>1.95589350</td>
<td>0.20540066</td>
<td>9.52233312</td>
<td>0.000000</td>
</tr>
<tr>
<td>UE90</td>
<td>-0.52440205</td>
<td>0.07002751</td>
<td>-7.48851453</td>
<td>0.000000</td>
</tr>
<tr>
<td>DV90</td>
<td>0.46159381</td>
<td>0.11517235</td>
<td>4.00785263</td>
<td>0.000064</td>
</tr>
<tr>
<td>MA90</td>
<td>-0.04948175</td>
<td>0.04890142</td>
<td>-1.01186726</td>
<td>0.311776</td>
</tr>
</tbody>
</table>

LM Tests for Spatial Autocorrelation
LM-err: 53.65 p: 0.000000
LM-lag: 50.58 p: 0.000000
RLM-err: 5.52 p: 0.018832
RLM-lag: 2.45 p: 0.117848
Estimation: Spatial Lag Model

GeoDaSpace Regression: untitled*

Data File
- south.dbf

Specification
- Y
  - HR90
  - YE

- X
  - RD90
  - PS90
  - DV90
  - MA90
  - UE90

Model Weights
- southqueen.gal

Kernel Weights
- southk20.gwt

Model Estimation

Model Type
- Standard
- Spatial Lag
- Spatial Error
- Spatial Lag+Error

Spatial Tests
- LM
- White
- HAC

Standard Errors

PySAL

OGRS 2012
Estimation: Spatial Lag Model

Model Setup: ModelType = Spatial Lag, Endogenous = N
Data: south.dbf N: 1412 df: 1405
Dependent Variable: HR90
Instruments:
W_RD90 W_PS90 W_DV90 W_MA90 W_UE90
Spatial Weights: southqueen.gal Type: binary
Kernel Weights: southk20.gwt Type: epanechn
R2 (var): 0.31428562       R2 (corr): 0.31031186

2SLS Results, HAC Variance with kernel epanechn

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>6.04710382</td>
<td>2.15452446</td>
<td>2.80670001</td>
<td>0.005074</td>
</tr>
<tr>
<td>RD90</td>
<td>4.16857898</td>
<td>0.47894879</td>
<td>8.70360057</td>
<td>0.000000</td>
</tr>
<tr>
<td>PS90</td>
<td>1.83298568</td>
<td>0.41516097</td>
<td>4.41512043</td>
<td>0.000011</td>
</tr>
<tr>
<td>DV90</td>
<td>0.46930008</td>
<td>0.11689093</td>
<td>4.01485457</td>
<td>0.000063</td>
</tr>
<tr>
<td>MA90</td>
<td>-0.01977798</td>
<td>0.04740126</td>
<td>-0.41724587</td>
<td>0.676562</td>
</tr>
<tr>
<td>UE90</td>
<td>-0.45986481</td>
<td>0.10429535</td>
<td>-4.40925509</td>
<td>0.000011</td>
</tr>
<tr>
<td>W_HR90</td>
<td>0.17004958</td>
<td>0.09196025</td>
<td>1.84916405</td>
<td>0.064644</td>
</tr>
</tbody>
</table>

Anselin-Kelejian Test for Residual Spatial Autocorrelation
Moran's I: 0.0395 LM: 1.35 p: 0.245306
```python
>>> from pysal.spreg import GM_Error_Het
>>> db = pysal.open('south.dbf', 'r')
>>> y = np.array(db.by_col('HR90'))
>>> y.shape = (len(y),1)
>>> X = []
>>> rhs = "RD90", "PS90", "DV90", "MA90", "UE90"
>>> tmp = [X.append(db.by_col(col)) for col in rhs]
>>> X = np.array(X).T
>>> w = pysal.queen_from_shapefile('south.shp')
>>> w.transform = 'r'
>>> reg = GM_Error_Het(y, X, w)
>>> reg.betas
array([[ 6.74524073],
       [ 4.41126522],
       [ 1.78024214],
       [ 0.48884888],
       [-0.01499858],
       [-0.39338759],
       [ 0.3132412 ]])
```
ArcGIS Toolkit

ArcToolbox
- 3D Analyst Tools
- Analysis Tools
- Cartography Tools
- Conversion Tools
- Data Interoperability Tools
- Data Management Tools
- Editing Tools
- Geocoding Tools
- Geostatistical Analyst Tools
- Linear Referencing Tools
- Multidimension Tools
- Network Analyst Tools
- Parcel Fabric Tools
- Python Spatial Analysis Library (PySAL)
- Spatial Regression Tools
  - Ordinary Least Squares (OLS)
  - Spatial Error Model
  - Spatial Lag Model
- Weight Tools
  - Conversion
  - Spatial Weight File Conversion
  - Creation
    - Create Adaptive Kernel Spatial Weights
    - Create Contiguity-Based Spatial Weights
    - Create Distance-Based Spatial Weights
  - Transformation
    - Application of Transformation
- Schematics Tools
- Server Tools
- Spatial Analyst Tools
- Spatial Statistics Tools
- Tracking Analyst Tools

Spatial Lag Model
- Input Data File
- Input Spatial Weights
- Output File Report (optional)
- Dependent Variable (Y)
- Independent Variables (X)
QGIS Toolkit

Figure 1: QGIS Plugin Design

Figure 2: PySAL Tree Diagram
6 Month Release Cycle

<table>
<thead>
<tr>
<th>Month</th>
<th>Yr</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul</td>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td>Jan</td>
<td>11</td>
<td>1.1</td>
</tr>
<tr>
<td>Jul</td>
<td>11</td>
<td>1.2</td>
</tr>
<tr>
<td>Jan</td>
<td>12</td>
<td>1.3</td>
</tr>
<tr>
<td>Jul</td>
<td>12</td>
<td>1.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEP and Vote</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Development/Sprints</td>
<td>16 weeks</td>
</tr>
<tr>
<td>Freeze and Testing</td>
<td>4 weeks</td>
</tr>
<tr>
<td>Release</td>
<td>2 weeks</td>
</tr>
</tbody>
</table>
PySAL Enhancement Proposals (PEP)

- PEP 0001 Spatial Dynamics Module
- PEP 0002 Residential Segregation Module
- PEP 0003 Spatial Smoothing Module
- PEP 0004 Geographically Nested Inequality based on the Geary Statistic
- PEP 0005 Space Time Event Clustering Module
- PEP 0006 Kernel Density Estimation
- PEP 0007 Spatial Econometrics
- PEP 0008 Spatial Database Module
- PEP 0009 Add Python 3.x Support
- PEP 0010 Add pure Python rtree
# PEP 0007 Spatial Econometrics

**Author**
Luc Anselin <luc.anselin@asu.edu>, Sergio Rey <sjsrev@gmail.com>, David Folch <dfolch@asu.edu>, Daniel Arribas-Bel <daniel.arribas.bel@gmail.com>, Pedro Amaral <pymda2@cam.ac.uk>, Nicholas Malizia <nmalizia@gmail.com>, Ran Wei <rwei5@asu.edu>, Jing Yao <jyao13@asu.edu>, Elizabeth Mack <Elizabeth.A.Mack@asu.edu>

**Status**
Approved 1.1

**Created**
12-Oct-2010

**Updated**
12-Oct-2010

## Abstract

The spatial econometrics module will provide a uniform interface to the spatial econometric functionality contained in the former PySpace and current GeoDaSpace efforts. This module would centralize all specification, estimation, diagnostic testing and prediction/simulation for spatial econometric models.

## Motivation

Spatial econometric methodology is at the core of GeoDa and GeoDaSpace. This module would allow access to state of the art methods at the source code level.

## Reference Implementation

We suggest adding the module `pysal.sreg`. As development progresses, there may be a need for submodules dealing with pure cross sectional regression, spatial panel models and spatial probit.

Core methods to be implemented include:

- OLS estimation with diagnostics for spatial effects
- 2SLS estimation with diagnostics for spatial effects
Discipline

You can do anything, but you can’t do everything.

David Allen
Discipline

Before Release Cycles

- Brainstorming: 70.0%
- Development: 20.0%
- Testing: 5.0%
- Release: 5.0%

After Release Cycles

- Development: 75.0%
- Testing: 10.0%
- Release: 10.0%
- Brainstorming: 5.0%
Documentation is like sex... 

when it is good, it is very, very good; and when it is bad, it is better than nothing.
- Dick Brandon
A map is like sex... when it is good, it is very, very good; and when it is bad, it is better than nothing.
- Me
Welcome

PySAL is an open source cross-platform library of spatial analysis functions written in Python. It is intended to support the development of high level applications for spatial analysis.

Getting Involved

Prospective developers can learn more in the developer guide.

You can track PySAL development, submit bug reports, and make feature requests here.

Downloads are here.

Documentation

- User Guide
- Developer Guide
- API Reference
Outline

1 Open Source

2 PySAL

3 Lessons for Education

4 Lessons for Research

5 Conclusions
Challenge: Conservative Institutions (Take 1)
Challenge: Conservative Institutions (Take 1)

- SDSU CTO: “Has this software under this grant been licensed?”
SDSU CTO: “Has this software under this grant been licensed?”
Me: “Yes, it has been under the GPL since before the proposal.”
Challenge: Conservative Institutions (Take 1)

- SDSU CTO: “Has this software under this grant been licensed?”
- Me: “Yes, it has been under the GPL since before the proposal.”
- SDSU CTO: “What is the GPL?”
Challenge: Conservative Institutions (Take 2)

Linear thinking in an era of exponential rates of technological change

- Institutional inertia
- Resistance to change
- Short run vs. long run
Challenge: Digital Natives?

Reality
- Irony of tools versus wills
- Job training versus learning
- Entertain

“I took this course as I heard we would be taught ArcGIS. I don’t care about the science and the algorithms underneath the software, I want a job”.

Possibilities
- New types of students
- Change agents
- Interdisciplinary catalysts
- Coaching and collaborating
Course: Open Source Geocomputation

Introduction

This course will introduce students to theory and practice of geocomputation. Broadly defined, geocomputation can be viewed as both the practice of using computational techniques to solve spatial problems, and as a new way of doing science in a geographical context.

Contemporary research in analytical geography has placed an increasing demand on the computational skills of its practitioners. The advances in spatial data analysis and geographical modeling have also largely out-paced the capabilities of standard statistical software. At the same time, the multidisciplinary nature of the spatial sciences often translates into the need to deal with disparate data sources, formats and programming languages. As such, students undertaking research are often confronted with a daunting set of tasks that are seldom covered in an integrated fashion in course work. This course is designed to address this situation by assisting students in acquiring a core set of geocomputational capabilities.

Objectives

1. Acquire basic skill sets used in quantitative spatial analysis in computational environments
2. Introduce students to fundamental algorithms in spatial analysis
3. Introduce students to contemporary research in geocomputation and dynamic visualization
4. Familiarize students with the fundamental tools used in collaborative programming and research projects in an open source and cross-platform
Flattening the Learning Curve

Classical learning curves for some common editors

Notepad

Visual Studio

Pico

vi

emacs
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Readings</th>
<th>Out</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>08/23/12</td>
<td>Course Introduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>08/28/12</td>
<td><strong>Software Installation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>08/30/12</td>
<td>Vim</td>
<td><strong>Vim Tutorial</strong></td>
<td>Exercise 1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>09/04/12</td>
<td>Python Introduction</td>
<td>D 1,2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>09/06/12</td>
<td>IPython</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>09/11/12</td>
<td>Operators and Sequences</td>
<td>D 8,10,11,12</td>
<td>Exercise 2</td>
<td>Exercise 1</td>
</tr>
<tr>
<td></td>
<td>09/13/12</td>
<td>Conditional Execution</td>
<td>D 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>09/18/12</td>
<td>Files, Functions and Modules</td>
<td>D 14</td>
<td>Exercise 3</td>
<td>Exercise 2</td>
</tr>
<tr>
<td></td>
<td>09/20/12</td>
<td>Object Orientation</td>
<td>D 15,16,17,18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>09/25/12</td>
<td>Object Orientation</td>
<td></td>
<td>Exercise 4</td>
<td>Exercise 3</td>
</tr>
<tr>
<td></td>
<td>09/27/12</td>
<td>Review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>10/02/12</td>
<td>Numerical Programming</td>
<td></td>
<td></td>
<td>Exercise 4</td>
</tr>
<tr>
<td></td>
<td>10/04/12</td>
<td>Code Repositories</td>
<td></td>
<td>Exercise 5</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10/09/12</td>
<td>PySAL and STARS</td>
<td></td>
<td></td>
<td>Exercise 5</td>
</tr>
</tbody>
</table>
Geocomputation Content

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>10/09/12</td>
<td>PySAL and STARS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/11/12</td>
<td>Bounding Regions</td>
<td>Exercise 5</td>
</tr>
<tr>
<td>9</td>
<td>10/16/12</td>
<td>Spatial Indexing</td>
<td>MR 4</td>
</tr>
<tr>
<td></td>
<td>10/18/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10/23/12</td>
<td>Tessellations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10/25/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10/30/12</td>
<td>Searching and Nearest Neighbor Problems</td>
<td>MR 6</td>
</tr>
<tr>
<td></td>
<td>11/01/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11/06/12</td>
<td>Graphs and Networks</td>
<td>MR 7</td>
</tr>
<tr>
<td></td>
<td>11/08/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11/13/12</td>
<td>Spatial Partitioning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11/15/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>11/20/12</td>
<td>Classification</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11/22/12</td>
<td>Thanksgiving Holiday</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>11/27/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11/29/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>12/04/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12/06/12</td>
<td>Studio</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>12/11/12</td>
<td>Future Directions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12/15/12</td>
<td>Presentations</td>
<td></td>
</tr>
</tbody>
</table>

Sergio Rey (ASU)
PySAL
OGRS 2012
Code as Text

**PySAL**
- Pure Python
- Pedagogical Goal
- Trade-off

**Students**
- Free as in beer
- Free as in speech
- No black boxes

**Challenge**
- Currently methods are looked at as something to learn/use
- Need to change this to see methods as areas to research
- Code=Text=Paper
Outline

1 Open Source

2 PySAL

3 Lessons for Education

4 Lessons for Research

5 Conclusions
New Ways of Doing Better Science

Challenges: Gap between OS and Academia
- Academic Reward System
- Peer Review
- Consumers versus Producers

Opportunities
- Provenance and Replication Mechanisms
- Real Peer Review
- Scientist as Developer
- Network Effects
Academic Reward System

Code isn’t viewed as text

- Tools are viewed as inputs to, not output of, the research process
- A journal article counts, documentation does not
- A book counts, but a book about software becomes dated quickly

This is why:

- Funding for tools development is difficult to obtain
- Documentation is “scarce”
- So few professors are on open source projects

The fundamental problem is the social returns (to the community) are at odds with the returns to the individual.
Academic Reward System

Code isn’t viewed as text
- Tools are viewed as inputs to, not output of, the research process
- A journal article counts, documentation does not
- A book counts, but a book about software becomes dated quickly

This is why:
- Funding for tools development is difficult to obtain
- Documentation is “scarce”
- So few professors are on open source projects
Academic Reward System

Code isn’t viewed as text
- Tools are viewed as inputs to, not output of, the research process
- A journal article counts, documentation does not
- A book counts, but a book about software becomes dated quickly

This is why:
- Funding for tools development is difficult to obtain
- Documentation is “scarce”
- So few professors are on open source projects

The fundamental problem is the social returns (to the community) are at odds with the returns to the individual.
Peer Review

Scientific Journal Review Process

- Double-blind (mostly)
- Reviews done out of professional obligation
- Miracle it works

Publishing

- We subsidize journals (refereeing, editorial boards)
- We pay to obtain copyright to our work (“Open Access”)
- Cracks in the foundation
Should Economists Use Open Source Software for Doing Research?

A. Talha Yalta · A. Yasemin Yalta

Accepted: 17 February 2010 / Published online: 5 March 2010
© Springer Science+Business Media, LLC. 2010

Abstract   We survey the literature on the accuracy of econometric software. We also assess the advantages of open source software from the point of view of reliability and discuss its potential in applied economics, which has now become fully dependent on computers. As a case study, we apply various accuracy tests on GNU Regression, Econometrics and Time-series Library (gretl) and demonstrate that the open source nature of the program made it possible to see the cause, facilitated a rapid fix, and enabled verifying the correction of a number of flaws that we uncovered. We also run the same tests on four widely-used proprietary econometric packages and observe the known accuracy errors that remained uncorrected for more than 5 years.

Keywords   Open source · Econometric software · Gretl · Accuracy · Software reliability
Scientist as Developer

From Consumer to Producer

“Increasingly researchers are relying on numerical simulation for results as closed-form solutions are not available for emerging research questions” (Atkins et al. 2003, p. 11).

Blurring distinction between scientist and developer

- the days of solving challenging research problems with off-the-self applications are gone
- rolling your own increasingly necessary
- downstream effects of giving back
- early days but some evidence this is happening
Outline

1. Open Source
2. PySAL
3. Lessons for Education
4. Lessons for Research
5. Conclusions
The Future is Bright

"It's tough to make predictions, especially about the future." (Yogi Berra)

"The best way to predict the future is to create it." (Abraham Lincoln)

"I see this as the essence of open source projects. The energy and creativity of many people with diverse goals together can work miracles!" Guido van Rossum

Sergio Rey (ASU)
The Future is Bright

“It’s tough to make predictions, especially about the future.” (Yogi Berra)
The Future is Bright

“It’s tough to make predictions, especially about the future.” (Yogi Berra)
“The best way to predict the future is to create it.” (Abraham Lincoln)
The Future is Bright

“It’s tough to make predictions, especially about the future.” (Yogi Berra)
“The best way to predict the future is to create it.” (Abraham Lincoln)

“I see this as the essence of open source projects. The energy and creativity of many people with diverse goals together can work miracles!”
Guido van Rossum
For More Information

http://pysal.org
http://geodacenter.asu.edu
http://sjrey.org
email: srey@asu.edu