Integrating Open-Source Statistical Packages with ArcGIS

Mark V. Janikas, Ph. D.
Xing Kang
Outline

• Introduction to Spatial Data Analysis in ArcGIS
  - Spatial Statistics, Geostatistics and Spatial Analyst
  - Python: Directly and Indirectly Extendable
  - Collaborative Motivation

• PySAL (Python Spatial Analysis Library)
  - Direct

• R
  - Indirect

• Conclusions and Future Directions
Spatial Analytics in ArcGIS: Past and Present

- **Traditional Spatial Analysis**
  - Core tools continue to evolve

- **Spatial Analyst**
  - Raster
  - Map Algebra

- **Geostatistics**
  - Raster and Vector
  - Continuous Data

- **Spatial Statistics**
  - Vector
  - Exhaustive Data
  - Python
Spatial Analytics in ArcGIS: Moving Forward

- **Python**
  - Spatial Analyst
    - Raster ↔ NumPy
    - SciPy
  - Spatial Statistics and Geostatistics
    - Data Access Module
    - Vector ↔ NumPy
    - Spatial Statistics Data Object and Utilities
    - matplotlib
The Great and Extendable Python

• Direct
  - Numeric/Scientific Python Modules
  - +50 Modules Listed
  - Check Compatibility… Then Plug and Play

• Indirect
  - Alternative Languages
  - No Python Hooks or Module
  - Compatibility Fails
  - Python Serves as Active Script and OS
  - Out of Process
Collaborative Motivation

- ArcGIS Users
  - Substantive Interest
  - Methodological Choice
  - Quantitative State of Mind
  - Enhancement Requests
    - User Conference
    - Resource Center and Forums
    - Email... You know who you are...

- What we build and why
  - What is left?
  - PySAL and R
Collaborative Motivation: 2nd Side of the Coin

- **Value Added in the Other Direction**
  - Direct or indirect linkage should focus on the strengths of each
  - **A Simple (Humorous?) Example**
    - The R module “Midas”
    - **Method**
      - Inputs: X, Y, Z Coordinates, Image Reading at Location
      - Output: Probability that Gold Exists at each location
    - I’m rich!... But what if “they” get there first?
  - **ArcGIS**
    - Network Analyst: How do I get there? Best route?
    - Spatial Statistics: Hot-Spots to identify an area to focus on
    - Geostatistics: Kriging to predict and unknown locations
Introduction to PySAL

- Open Source Python Library for Spatial Analytical Functions
  - ASU GeoDa Center for Geospatial Analysis and Computation
  - Luc Anselin
    - PySpace
  - Sergio Rey
    - STARS
  - BSD License
Anatomy of PySAL
Collaborative Advantages: PySAL and ArcGIS

- **PySAL**
  - Advance Spatial Analysis code base with novel functions
  - E.g. Regionalization, Spatial Econometrics
  - Do not have to “reinvent the wheel”

- **ArcGIS**
  - GIS User Interface
  - ~800 GP Tools
  - Easy-to-use Script Tool Framework
  - Enriched functionalities from ArcGIS Python scripts
    - arcpy, SSDataObject, SSUtilities, SSReport etc.
    - Multiple input/output data format
    - Error messages
    - Pyharness framework for robust testing
Basic PySAL – ArcGIS Interaction Model

Input Data -> SSDataObject

SSDataObject

Environment Settings
- Projections
- Field Qualification
- Z/M Values
- Bad Records
- Error/Warning Messages
- Localization
- Feature Accounting

SSUtilities

Output Data -> NumPy

NumPy

PySAL Analytical Function

Spatial Weights
PySAL – ArcGIS Toolbox Demonstration: Regional Income Distributions
What is R? Why should I use it?

• R *(The R Project for Statistical Computing)*
  - Over 60 CRAN sites across 30+ countries
  - Its Free
    - GNU GENERAL PUBLIC LICENSE
  - Base is powerful
    - Statistics, Linear Algebra, Visualization, etc...
  - Its extendible
    - 1800+ Contributed Extensions
    - splancs, spatstat, spdep, rgdal, maptools, shapefiles
Indirect Integration Strategy

• **Python and R: “Decoupled”**
  - Used as the core script tool
  - Hooks into the Operating System to call R
  - Post-Processor
  - “Out of Process”

• **RPy/RPy2**
  - Compatibility

• **win32com**
  - Windows only
  - Works for other programs as well
Basic R – ArcGIS Interaction Model

ArcGIS

Python

Retrieves Parameters
Organizes into R command
Executes R command
Post-Processing
  Apply Symbology
  Apply Projections
  Report

R

Enhanced Output Data

Input Params

Output Data
R – ArcGIS Toolbox Demonstration: Regional Income Distributions
Conclusions

- **PySAL**
  - Advanced spatial analytic techniques
  - Combined with SSDataObject and Utilities
    - Directly compatible
  - Python Harness Implementation
  - Spatial Econometrics and Spatial Weights Conversion
    - ESDA, Clustering, Spatial Dynamics etc.
  - BSD
Conclusions

• R
  - Contains “cutting edge” data analysis techniques from a wide body of academic and applied fields
  - Extendible
  - Indirectly compatible
    - Direct via RPy/RPy2 and win32com
  - GNU
  - Revolution
  - esri continues to focus on improving the interaction in the future
Software Links

- **PySAL**
  - [https://geodacenter.asu.edu/pysal](https://geodacenter.asu.edu/pysal)

- **NumPy and SciPy**

- **R**
  - [http://www.r-project.org/index.html](http://www.r-project.org/index.html)
- Using R in ArcGIS (Version 10)
- R Point Clustering Tools for ArcGIS (Version 9.3)
- Using R in ArcGIS (Version 11) – Coming Soon!
- Using PySAL in ArcGIS (Version 11)
  - Either here or at the GeoDa Center
Related Sessions

- **Search the Online Agenda**
  - [http://events.esri.com/uc/2012/infoWeb/OnlineAgenda/](http://events.esri.com/uc/2012/infoWeb/OnlineAgenda/)
- **Keywords**
  - Python
  - Spatial Statistics
  - Geostatistics
  - Spatial Analyst
  - Business Analyst