Integrating Open Source Statistical Packages with ArcGIS

Mark V. Janikas
Liang-Huan Chin
Outline

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

• Direct
  - SciPy (Scientific Python)
  - PANDAS (Python Data Analysis Library)
  - PySAL (Python Spatial Analysis Library)

• Indirect
  - R (matlab, SPSS, SAS)
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
  - +60 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
Directly Extendable Via Python

- **IPython**
  - [http://ipython.org/](http://ipython.org/)
  - Notebook (HTML Option)

- **SciPy**

- **PANDAS**

- **PySAL**

*Esri UC2013: Technical Workshop: Integrating Open Source Statistical Packages with ArcGIS*
Direct Python – ArcGIS Interaction Model

Input Data

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

SSDataObject

Output Data

NumPy
- Spatial Weights
- Open-Source Analytical Function

NumPy
Directly Extendible

Using the IPython Notebook to Demonstrate How ArcGIS Can Leverage Python Modules

Using the ArcGIS Script Tool Interface to Wrap Advanced Spatial Data Analysis Functions
Conclusions

- SciPy, PANDAS, PySAL
  - Advanced spatial analytic techniques
  - Combined with SSDataObject and Utilities
    - NumPy - Directly compatible
    - Python Harness Implementation
    - BSD
- R
  - Needs a collaborative effort to grow
  - Must adjust for short release cycles
Software Links

- PySAL
  - https://geodacenter.asu.edu/pysal
  - http://code.google.com/p/pysal/

- NumPy and SciPy
  - http://www.numpy.org/

- PANDAS
  - http://pandas.pydata.org/

- R
  - http://www.r-project.org/index.html
Additional Resources

• This Presentation (Slides, Data, IPython Notebook)
  - Public GitHub Repository:

• ArcGIS – PySAL Toolbox
  - http://geodacenter.asu.edu/software

- mjanikas@esri.com
Additional Resources (cont.)

• Using R in ArcGIS (Versions 10 and 10.1)
  - https://github.com/Esri/R-toolbox-py

• Spatial Statistics Resource Blog
Thank you…

Please fill out the session evaluation

First Offering ID: 1284

Online – www.esri.com/ucsessionsurveys
Paper – pick up and put in drop box