Integrating Open Source Statistics Packages with ArcGIS

Brett Rose
Our site's users

Subscribers to Martha Stewart Living

Consumers of furry pornography

The business implications are clear.

Pet Peeve #208:
Geographic profile maps which are basically just population maps
Why are we here?

Why open source statistics?
You’ve asked us

Substantive Interest

Quantitative minds

Mythological choice
Spatial analytics can mean a lot of thing
With spatial analysis we will map to see possible patterns, describe to improve understanding, and measure to minimize subjectivity.
Tools in ArcGIS
Spatial Analyst

# This is map algebra
outRas = Raster("inraster1") + Raster("inraster2")
Geostatistics
Spatial Statistics
Continuing with spatial analytics
Spatial Analyst
Spatial & Geo Stats

Data Access Module

Spatial Statistics Data Object and Utilities

python NumPy

matplotlib
Two kinds of “integration”
 Numeric/Scientific python modules
  Here: https://wiki.python.org/moin/NumericAndScientific

50+ Modules

Check compatibility
Not direct

• Alternative languages
• No python hooks
• Incompatible
• Python servers as active script and OS
• Out of process
got it

Now what
pysal

- Open Source Python Library for Spatial Analytical Functions
- ASU GeoDa Center for Geospatial Analysis and Computation
- Luc Anselin
  - PySpace (<GeoDaSpace>)
- Sergio Rey
  - STARS

BSD License
Collaborative Advantages:

**PySAL** and **ArcGIS**

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

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

Input Data

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

Output Data

Spatial Weights
PySal Analytical Functions
NumPy

NumPy
ssdo = SSDO.SSDaDataObject(inputFile, templateFC=outputFC)

masterField = UTILS.setUniqueIDField(ssdo, weightsFile = weightsFile)
ssdo.obtainData(masterField, fields=allVarList)

depArray = ssdo.fields[dependVar].data
PySAL – ArcGIS Toolbox Demonstration: Regional Income Distributions
R

• 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 Model

• 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
Python

Retrieves Parameters
Organizes into R command
Executes R command Post-Processing
Apply Symbology
Apply Projections Report
```python
inputFC = '***' + ARCPY.GetParameterAsText(0) + '***
outputFC = '***' + ARCPY.GetParameterAsText(1) + '***
numClusters = ARCPY.GetParameterAsText(2)
clusterMethod = ARCPY.GetParameterAsText(3)
### Create R Command ###
pyScript = SYS.argv[0]
toolDir = OS.path.dirname(pyScript)
rScript = OS.path.join(toolDir, "PointClusters.r")
rScript = '***' + rScript + '***
ARCPY.SetProgressor("default", "Executing R Script...")
args = "".join([inputFC, outputFC,
                 numClusters, clusterMethodStr,
                 varNames, useLocation])
RCMD = "R --slave --vanilla --args "
cmd = RCMD + args + " < " + rScript
```
R et al. → Output Data → Python → Enhanced Output Data

```python
### Execute Command ###
OS.system(cmd)

### Project the Data ###
DM.DefineProjection(outputFC.strip('""'), inputFC.strip('""'))

### Render the Results ###
params = ARCPY.gp.GetParameterInfo()
renderFile = OS.path.join(toolDir, "RenderClusters.lyr")
params[1].Symbology = renderFile
```
R – ArcGIS Toolbox Demonstration: Regional Income Distributions
• 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
• **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
links

• pysal
  • https://geodacenter.asu.edu/pysal
  • https://github.com/pysal

• SciPy and NumPy
  • http://www.scipy.org/

• R
  • http://www.r-project.org
Try for yourself

https://github.com/Esri/gis-stat-analysis-py-tutor

http://esri.github.io/
“Essentially, all models are wrong, but some are useful.”

-George E. P. Box