Integrating Open Source Statistical Packages with ArcGIS

Mark V. Janikas
Carlos Osorio-Murillo
Introduction

- Traditional Spatial Analysis
- Spatial Analyst
- Geostatistics
- Spatial Statistics
- Most Useful Tools
- Best Implementation
Spatial Analytics in ArcGIS: Moving Forward

- Python
  - Spatial Analyst
    - Raster
    - SciPy
  - Spatial Statistics and Geostatistics
    - Data Access Module
    - Vector
      - NumPy
    - Spatial Statistics Data Object and Utilities
      - SciPy, PANDAS, Matplotlib, NetCDF4-Python
  - Effort to Support Scientific Community
    - PySAL, scikit-learn, statsmodels, bokeh
The Great and Extendable Python

- Direct
  - Numeric/Scientific Python Modules
  - http://wiki.python.org/moin/NumericAndScientific
    - +60 Modules Listed
  - Check Compatibility... Then Plug and Play
    - ArcGIS Desktop, Server, ArcGIS Pro < 1.3
    - pip, GitHub, easy_install, svn
    - Unofficial Windows Binaries for Python Extensions – Christoph Gohlke, UC Irvine
      - http://www.lfd.uci.edu/~gohlke/pythonlibs/
  - Conda Effort
    - ArcGIS Pro 1.3+
The Jupyter Notebook

- Open-Source Web Application
  - One-Stop-Shop for Programming, Images, Equations and Text
    - Education, Business, Software Development, Science etc...

- Installation
  - ArcGIS Pro 1.3+: Conda
  - pip or Windows Binaries

- gis-stat-analysis-py-tutor
  - Series of Tutorials for Extending ArcGIS Directly in Python and R
SSDataObject

1. Environment Settings (Except Extent)
2. Bad Records
3. Error/Warning Messages
4. Localization
5. Feature Accounting

- Cursors and DataAccess are not assured to read attributes in order.
- Keeps track of the shapes and their attributes so that one can create output features w/o post-joins.
- Unique ID works with Spatial Weights Formats in ArcGIS, PySAL, R, Matlab, GeoDa etc.
import SSDataObject as SSDO
ssdo = SSDO.SSDataObject(inputFC)
ssdo.loadData(ssdo.oIdName, ['PCR2000', 'POP2000'])
print(ssdo.fields['PCR2000'].data)

[1.30711491, 0.94988574, 0.81495358, 0.76366472, 0.867]
[1.50388847, 0.61251602, 1.23401339, 0.75898862, 0.666]
[0.62599316, 0.83824116, 0.70966432, 0.55805988, 0.795]
[0.98591008, 0.62640028, 2.30772254, 0.76350391, 0.856]
[0.69603809, 0.90614532, 1.04151493, 1.28209386, 1.064]
Data Frames: Preferred Data Structure for Analysis

Using PANDAS to get that R Feel

```python
import pandas as PANDAS
df = ssdo.getDataFrame()
print(df)
```

<table>
<thead>
<tr>
<th></th>
<th>PC2000</th>
<th>POP2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.307115</td>
<td>1449840</td>
</tr>
<tr>
<td>1</td>
<td>0.949886</td>
<td>1209</td>
</tr>
<tr>
<td>2</td>
<td>0.814954</td>
<td>35153</td>
</tr>
<tr>
<td>3</td>
<td>0.763665</td>
<td>203807</td>
</tr>
<tr>
<td>4</td>
<td>0.867385</td>
<td>40645</td>
</tr>
<tr>
<td>5</td>
<td>0.813593</td>
<td>18817</td>
</tr>
<tr>
<td>6</td>
<td>1.503888</td>
<td>952810</td>
</tr>
</tbody>
</table>
Advanced Analysis Using Your Library of Choice

```python
import numpy as NUM
import scipy.cluster.vq as CLUST
import arcgis.scripting as ARC

X = NUM.empty((ssdo.numObs,2), dtype = float)
X[:,0] = df['PCR2000']
X[:,1] = df['POP2000']
whiteData = CLUST.whiten(X)
centers, distortion = CLUST.kmeans(whiteData, 5)
groups = ARC.ss.closest_centroid(whiteData, centers)
print(groups)
```
Output Resulting Feature Class with Ease

```python
ARCPY.env.overwriteOutput = True
outputFC = r'C:\Data\Output\kmeans_ca.shp'
outField = SSDO.CandidateField('SS_GROUP', 'LONG', groups)
outFields = {'SS_GROUP': outField}
ssdo.output2NewFC(outputFC, outFields,
appendFields = ['NEW_NAME', 'PERCNOHS'])
```

https://github.com/Esri/gis-stat-analysis-py-tutor
PySAL – ArcGIS Toolbox
Advanced Spatial Econometrics made easy

- **Spatial Econometrics**, ESDA, Inequality, Network Analytics and more…
- **Integration Strategy** – `spreg` module
  - GUI Interface provides easy-to-use framework for regression in the presence of spatial dependence.
  - Object-oriented design coupled with the SSDataObject and supporting utilities to facilitate scripting.
  - GMM Estimators and Spatial Weights Conversion Routines included.
  - Example of the possible symbiosis between Open-Source and For-Profit Software
- **Git Repository**
  - [https://github.com/Esri/PySAL-ArcGIS-Toolbox](https://github.com/Esri/PySAL-ArcGIS-Toolbox)
Demo

pysal
Python Spatial Analysis Library

ESRI
R Integration

- Highly Active Community
  - Over 6000 Libraries
- Old Method (Indirect)
  - Out of Proc
  - Python as the Glue
- New Method
  - In Proc
  - Native Data Access
  - Honors Selection Sets and Projections
  - Vector Data
  - Charts and Graphs
  - GUI Interface
R-ArcGIS Links

- [https://r-arcgis.github.io](https://r-arcgis.github.io)
- [https://github.com/R-ArcGIS](https://github.com/R-ArcGIS)

Welcome to the R – ArcGIS Community

*Combine the power of ArcGIS and R to solve your spatial problems*

The R – ArcGIS Community is a community driven collection of free, open source projects making it easier and faster for R users to work with ArcGIS data, and ArcGIS users to leverage the analysis capabilities of R.

Need the R Statistical Software? Download it now.
Installation

Via Python Toolbox
R Package Manager
Sample Tools

- Model Based Clustering
- Semi-Parametric Regression
- Spatial Statistical Data Analysis for GIS Users
  - Konstantin Krivoruchko, Esri Press
  - Tool Help/Documentation
  - Two More Examples to Come
- Community
  - Please Join Us!
Documentation for package 'arCGISbinding' version 1.0.0.111

- DESCRIPTION file
- User guides, package vignettes and other documentation

arcgisbinding-package
arc
arc.check_product
arc.data2sp
arc.dataset-class
arc.env
arc.feature-class
arc.fromPROJ4

- Bindings for ArcGIS
- ArcGIS product and license information
- Convert an arc.dataframe object to an sp SpatialDataFrame object
- Class "arc.dataset"
- Get geoprocessing environment settings
- Class "arc.dataset"
- Convert PROJ4 Coordinate Reference System string to Well Known Text
Learn Lessons

R-bridge in Jupyter Notebook

https://github.com/R-ArcGIS/r-bridge-install.git

R kernel for Jupyter Notebook

https://github.com/IRkernel/IRkernel.git
https://irkernel.github.io/

https://jupyter.org/
https://ipython.org/

ESRI

IP[y]:
Installation of R kernel for Jupyter

```python
##### Python arcgispro-py3 environment #####
##### install Jupyter using pip #####
pip install jupyter

##### install Jupyter using conda #####
conda install jupyter

##### Install IRKernel in R #####
install.packages('devtools')
devtools::install_github('IRkernel/IRkernel')
# or devtools::install_local('IRkernel-master.tar.gz')
IRKernel::installspec()  # to register the kernel in the current R installation
```
Moving Forward

- Listen to our Users, Listen to the Experts
  - Build the Best Tools Around
  - Support their Endeavors
- Python
  - Conda
    - Data Science Python Environments
    - Tighter Integration w/ PANDAS
    - Panel Data Model – Space-Time Pattern Mining
- R
  - The Community Must Lead the Way
  - Raster Support
- Microsoft Azure Data Science VM