Data Science Made Easy in ArcGIS Using Python and R

Mark Janikas and Marjean Pobuda
Techniques and methodologies continue to develop
  - Across disciplines
  - Subject to an ever-increasing amount of data

Core analytics in ArcGIS
  - Maximize performance and utility
  - E.g. Spatial Statistics, Geostatistics, Spatial Analyst
  - E.g. GeoAnalytics, Insights, ArcGIS Python SDK

Community is vast and evolving
  - Broad and specific
  - Techniques can come to market quickly
  - ArcGIS extends directly via scripting APIs
    - E.g. Python, R, Java
Data Science Community

Python

- Numeric/Scientific Python Modules
  - +60 Modules Listed
  - E.g. Life sciences, visualization, mathematics, GIS

- Python as a glue language
  - E.g. C++, Java, R, Hadoop/Spark, NetCDF/HDF-5

- Conda
  - Pro 1.3+

- 10x (pip)
  - Unofficial Windows Binaries for Python Extensions – Christoph Gohlke, UC Irvine
Data Science Community

- Well over 6000 packages to enhance core
- Most widely used statistical software in the world
- Diverse
  - Universities, Government, Industry
  - Finance, Ecology, Statistics
Battle of Bands
Which one is best?

• KD nuggets (2015)
  - Pros and Cons
  - R has a broader set of modules specific to a variety of methodologies
  - Python is a more fully functional programming language

• A ton to consider

• ArcGIS has you covered
  - PySAL – ArcGIS Toolbox
    - [https://github.com/Esri/PySAL-ArcGIS-Toolbox](https://github.com/Esri/PySAL-ArcGIS-Toolbox)
  - R Sample Toolbox
    - [https://github.com/R-ArcGIS/r-sample-tools](https://github.com/R-ArcGIS/r-sample-tools)
  - Microsoft Data Science VM
### Integration

#### GUI Interface

#### Tool Properties: Automatic Model Search

<table>
<thead>
<tr>
<th>General</th>
<th>Name</th>
<th>autoModel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>Label</td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td>Automatic Model Search</td>
<td></td>
</tr>
<tr>
<td>Script File</td>
<td>C:\git\PySAL-ArcGIS-Toolbox\Scripts\AutoModel.py</td>
<td></td>
</tr>
</tbody>
</table>

#### Tool Properties: Semiparametric Regression

<table>
<thead>
<tr>
<th>General</th>
<th>Name</th>
<th>SemiParRgression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>Label</td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td>Semiparametric Regression</td>
<td></td>
</tr>
<tr>
<td>Script File</td>
<td>C:\git\r-sample-tools\scripts\semi_par_regression.R</td>
<td></td>
</tr>
</tbody>
</table>
Integration
Data IO Flow Chart

Input Data

SSDataObject

arcgisbindings

Read into NumPy/PANDAS

Analytics using Arrays

Output Data

Read into DataFrame/SP

Analytics using Data Frames
Integration
Simple API for reading data

```python
In [ ]:  ###### Loading dataset ######
ssdo = SSDO.SSDataObject(inputFC)
##### Create DataFrame #####
ssdo.obtainData('MYID', ['GROWTH', 'PCR1970', 'POPDEN70', 'PERCNOHS'])
df = ssdo.getDataFrame()
print(df.head())
```

```python
In [ ]:  ###### Loading dataset ######
info <- arc.open(inputFC)
##### Create DataFrame #####
df <- arc.select(info, c('MYID', 'GROWTH', 'PCR1970', 'POPDEN70', 'PERCNOHS'))
head(df)
```
Integration
Simple API for writing data

```python
In [ ]: outDict = {}
outField = SSDO.CandidateField('STDNORM', 'DOUBLE', outArray, alias = 'Standard Normal')
outDict[outField.name] = outField
ssdo.output2NewFC(outputFC, outDict, appendFields = ['GROWTH', 'PERCNOHS', 'NEW_NAME'])

In [ ]: df['STDNORM'] = randnorm
arc.write(outputFC, df)
```
Python Demo

Data IO, Neighborhood Searching and Advanced Analysis using Jupyter Notebooks
Links to Related Python Projects

Python

- **gis-stat-analysis-py-tutor**
  - Jupyter Notebooks
  - Integrating Open Source Projects Using Python
  - Neighborhood Searching
  - Past Conferences
  - PySAL/ArcGIS API

- **PySAL-ArcGIS-Toolbox**
  - Spatial Econometrics Made Easy
  - spreg module
  - [https://github.com/Esri/PySAL-ArcGIS-Toolbox](https://github.com/Esri/PySAL-ArcGIS-Toolbox)
Future Directions

Python

• Tighter integration between ArcGIS Python SDK and ArcPy
  - SDK to be included in Pro Core
  - Spatial Data Frames
    - Geometry Operators: Intersection, Touches, Within etc..
    - Add time
    - Integration with SSDataObject, SSCube and SSPanel

• Possible GeoAnalytics Python API?
ArcGIS and R
Introducing the R-ArcGIS Bridge

The R-ArcGIS bridge allows you to connect ArcGIS to R and enables the seamless transfer of data back and forth, along with the ability to create Geoprocessing tools based on R scripts.

Reasons you might need the bridge:
- A particular tool or action is easier to perform in ArcGIS or R.
- A particular tool or analysis only exists in ArcGIS or R.
- You want to share R functionality with others who are not familiar with R.
- You need easy access to spatial data contained in shapefiles, geodatabases or stored online.

Requirements for using the bridge:
- ArcGIS (Pro 1.1+ or ArcGIS 10.3.1+)
- R (3.1.0+)
R-ArcGIS Bridge Demo
Expanding Workflows and Creating Script Tools
Future Directions
R-ArcGIS Bridge

• Raster Support (officially coming in Pro 2.1)
  - Ability to read and write raster data
    - Handle big raster data with the ability to read in chunks by bands.
    - Compatibility with CRF format and Mosaic Datasets
  - Customize selections and subsets
    - Create subsets by bands or pixel rows and columns
    - Resample options available
    - Select desired pixel format for specific analyses
Resources
Learn More on Using the R-ArcGIS Bridge

Getting Started:
- Analyzing Crime Using Statistics and the R-ArcGIS Bridge Learn Lesson
- Using the R-ArcGIS Bridge Introductory Web Course
  (https://www.esri.com/training/catalog/58b5e417b89b7e000d8bfe45/using-the-r-arcgis-bridge/)

Creating R Script Tools:
- Integrating R Scripts into Geoprocessing Tools Web Course
  (https://www.esri.com/training/catalog/58b5e578b89b7e000d8bfff4/integrating-r-scripts-into-arcgis-geoprocessing-tools/)
- arcgisbinding Package Vignette
  (https://r-arcgis.github.io/assets/arcgisbinding-vignette.html)

Upcoming Live Seminar Training
- Go Deeping with Data Analytics Using ArcGIS Pro and R – Thursday, August 31st
Upcoming Sessions
A Deeper Dive

AI for Earth: Microsoft AI and the R Bridge to ArcGIS
- Tuesday, July 11th, 2:30-3:30pm, SDCC – Esri Showcase: Sustainable Worlds Theaterette
  (https://userconference2017.schedule.esri.com/schedule/1834935949/)

Bridging the Gap: Integrating R and ArcGIS for Advanced Analysis
- Wednesday, July 12th, 10:00-10:30am, SDCC – Tech Theater 17 Exhibit Hall A
  (https://userconference2017.schedule.esri.com/schedule/1087523793/)

Statistics Special Interest Group (SIG) Meeting
- Wednesday, July 12th, 12:00-1:00pm, SDCC – Room 26 B
  (https://userconference2017.schedule.esri.com/schedule/1125815194/)
Please Take Our Survey on the **Esri Events App**!

**Download the Esri Events app and find your event**

**Select the session you attended**

**Scroll down to find the survey**

**Complete Answers and Select “Submit”**