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)
  - R (via IPython and RPy2 or Python Win Extensions)

• 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, NetCDF4-Python
  - **Effort to Support Scientific Community**
    - SciPy, PANDAS, PySAL
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
    - pip, github, easy_install, svn
    - Unofficial Windows Binaries for Python Extensions – Christoph Gohlke, UC Irvine
      - http://www.lfd.uci.edu/~gohlke/pythonlibs/

• Indirect
  - Alternative Languages
  - No Python Hooks or Module
  - Python Serves as Active Script and OS
  - Out of Process
  - Using R in ArcGIS (Version Independent)
    - https://github.com/Esri/R-toolbox-py
Directly Extendable Via Python

- **IPython**
  - [http://ipython.org/](http://ipython.org/)
  - Notebook (HTML Option)
- **SciPy**
- **PANDAS**
- **PySAL**
- **R (Rpy Revisited)**
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

Output Data

NumPy

Spatial Weights

Open-Source Analytical Function

NumPy

Esri UC 2014 | Technical Workshop | Integrating Open Source Statistical Packages with ArcGIS
SSDataObject NumPy Arrays to PANDAS DataFrame

```python
In [8]: ssdo = SSDO.SSDataObject(inputFC)
years = NUM.arange(1975, 2015, 5)
fieldNames = ['PCR' + str(i) for i in years]
fieldNamesAll = fieldNames + ['NEW_NAME', 'SOCAL']
ssdo.getData("MYID", fieldNamesAll)
ids = [ssdo.order2Master[i] for i in xrange(ssdo.numObs)]
convertDictDF = {}
for fieldName, fieldObject in ssdo.fields.iteritems():
    convertDictDF[fieldName] = fieldObject.data
df = PANDA.DataFrame(convertDictDF, index = ids)
print df[0:5]
```

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>158</td>
<td>Alameda</td>
<td>1.169255</td>
<td>1.195712</td>
<td>1.200988</td>
<td>1.165406</td>
<td>1.158115</td>
<td>1.307115</td>
<td>1.248997</td>
</tr>
<tr>
<td>159</td>
<td>Alpine</td>
<td>0.844546</td>
<td>0.906803</td>
<td>0.855655</td>
<td>0.924508</td>
<td>0.820581</td>
<td>0.949886</td>
<td>0.930033</td>
</tr>
<tr>
<td>160</td>
<td>Amador</td>
<td>0.991467</td>
<td>0.963228</td>
<td>0.921839</td>
<td>0.823639</td>
<td>0.815521</td>
<td>0.814954</td>
<td>0.864324</td>
</tr>
<tr>
<td>161</td>
<td>Butte</td>
<td>0.910668</td>
<td>0.898385</td>
<td>0.817796</td>
<td>0.794387</td>
<td>0.773955</td>
<td>0.763665</td>
<td>0.790418</td>
</tr>
<tr>
<td>162</td>
<td>Calaveras</td>
<td>0.941372</td>
<td>0.875469</td>
<td>0.891595</td>
<td>0.870938</td>
<td>0.806776</td>
<td>0.867385</td>
<td>0.880388</td>
</tr>
</tbody>
</table>
Analysis Using PANDAS, SSDataObject Makes Output Easy

Example: Calculating the Trend of Rolling Means

```python
In [11]:
    pcr = df.ix[:, 1:9]
    rollMeans = NUM.apply_along_axis(PANDA.rolling_mean, 1, pcr, 4)
    timeInts = NUM.arange(0, 5)
    outArray = NUM.empty((ssdo.numObs, 5), float)
    for i in xrange(ssdo.numObs):
        outArray[i] = SCIPY.stats.linregress(timeInts, rollMeans[i, 3:]))
```

Write to Output (Same as Always...)

```python
In [12]:
    outputFC = OS.path.abspath(r'../data/testMyRollingMeanInfo.shp')
    outFields = ["SLOPE", "INTERCEPT", "R_SQRAURED", "P_VALUE", "STD_ERR"]
    outDict = {}
    for fieldInd, fieldName in enumerate(outFields):
        outDict[fieldName] = SSDO.CandidateField(fieldName, "DOUBLE", outArray[:, fieldInd])
    ssdo.output2NewFC(outputFC, outDict, fieldOrder = outFields)
    del ssdo
```
Advanced Example: Spatially Constrained Clustering Using PySAL

```python
ssdo = SSDO.SSDataObject(inputFC)
ssdo.obtainData(ssdo.ioName, ['GROWTH', 'POP1970', 'PERCNOHS'])
w = PYBAL.weights.knnW(ssdo.xyCoords, k=5)
X = NUM.empty((ssdo.numObs,2), float)
X[:,0] = ssdo.fields['GROWTH'].data
X[:,1] = ssdo.fields['PERCNOHS'].data
floorVal = 1000000.0
floorVar = ssdo.fields['POP1970'].returnDouble()
maxp = PYBAL.region.Maxp(w, X, floorVal, floor_variable = floorVar)
outArray = NUM.empty((ssdo.numObs,), int)
for regionID, orderIDs in enumerate(maxp.regions):
    outArray[orderIDs] = regionID
    print regionID, orderIDs
```
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
    - New Tools on GitHub
  - Revisit In Proc Methodology
    - Installation Process is still a roadblock
Additional Resources

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

• ArcGIS – PySAL Toolbox
  - http://geodacenter.asu.edu/software
  - Keep checking for release version… Coming soon on GitHub!

• Mark Janikas, Ph. D.
  - mjanikas@esri.com

• Shaun Walbridge
  - swalbridge@esri.com
### Additional Resources (Cont.)

- **Using R in ArcGIS (Version Independent – Out of Proc)**
  - [https://github.com/Esri/R-toolbox-py](https://github.com/Esri/R-toolbox-py)

- **Spatial Statistics Resource Blog**

<table>
<thead>
<tr>
<th>Book Title</th>
<th>Formats</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIS Tutorial for Python Scripting</td>
<td>Paperback and e-book</td>
<td>Just released! Offers several hands-on tutorial exercises.</td>
</tr>
<tr>
<td>Esri Press, 2014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Python Scripting for ArcGIS</td>
<td>Paperback and e-book</td>
<td>Good reference text</td>
</tr>
<tr>
<td>Esri Press, 2013</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Software Links

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

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

- IPython
  - http://ipython.org/

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

- R
  - http://www.r-project.org/index.html
Thank you...

• Please fill out the session survey:

First Offering ID: 193 / 1736

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