



# Integrating Open Source Statistical Packages with ArcGIS

Mark V. Janikas

Liang-Huan (Leo) Chin

# Introduction

- **Traditional Spatial Analysis**
- **Spatial Analyst**
- **Geostatistics**
- **Spatial Statistics**
- **Most Useful Tools**
- **Best Implementation**

# 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

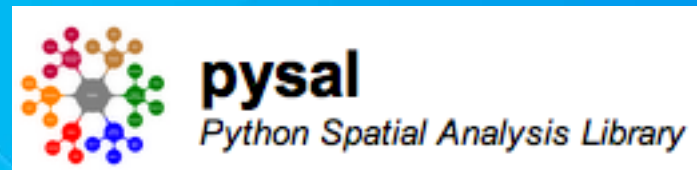
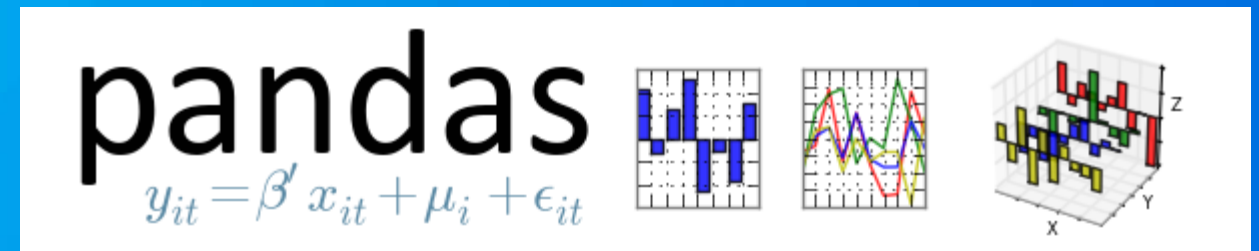
- pip, GitHub, easy\_install, svn

- Unofficial Windows Binaries for Python Extensions – Christoph Gohlke, UC Irvine

- <http://www.lfd.uci.edu/~gohlke/pythonlibs/>

- Conda Effort

IP[y]: IPython  
Interactive Computing



# Enhanced Data Access in the Spatial Statistics Codebase

## **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..

## Attribute Data as NumPy Arrays with Value Added

```
import SSDataObject as SSDO
ssdo = SSDO.SSDataObject(inputFC)
ssdo.obtainData(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

```
import pandas as PANDAS
ids = [ssdo.order2Master[i] for i in range(ssdo.numObs)]
convertDictDF = {}
for fieldName, fieldObject in ssdo.fields.items():
    convertDictDF[fieldName] = fieldObject.data
df = PANDAS.DataFrame(convertDictDF, index = ids)
print(df)
```

	PCR2000	POP2000
0	1.307115	1449840
1	0.949886	1209

## Advanced Analysis Using Your Library of Choice

### Advanced Analysis [SciPy Example]

```
import numpy as NUM
import scipy.cluster.vq as CLUST
import arcgisscripting 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)
```

```
[3 4 4 4 4 4 3 0 3 4 0 4 0 4 0 0 4 0 1 0 2 4 4 0 0 4
```



## Output Resulting Feature Class with Ease

```
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"])
```



Esri / [gis-stat-analysis-py-tutor](https://github.com/Esri/gis-stat-analysis-py-tutor)

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

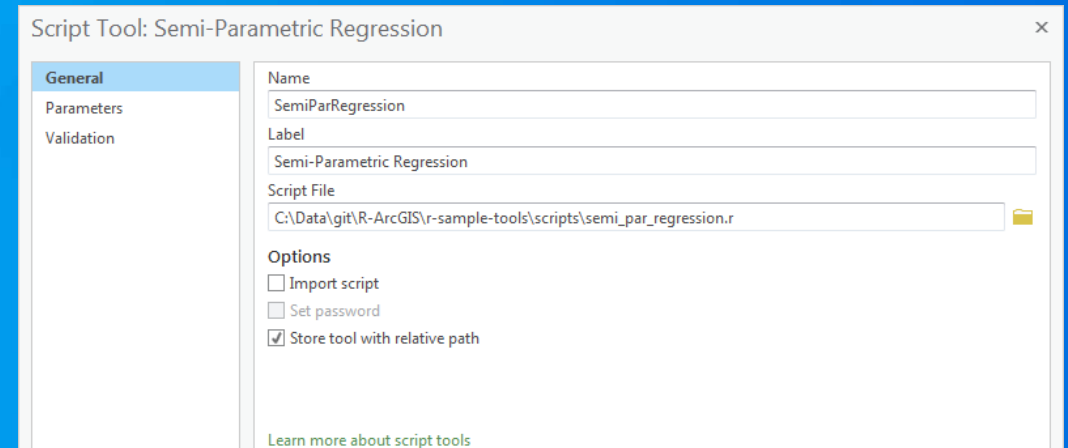
# R Integration

- **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

Available CRAN Packages By Name  
[A](#)[B](#)[C](#)[D](#)[E](#)[F](#)[G](#)[H](#)[I](#)[J](#)[K](#)[L](#)[M](#)[N](#)[O](#)[P](#)[Q](#)[R](#)[S](#)[T](#)[U](#)[V](#)[W](#)[X](#)[Y](#)[Z](#)



## Script File

C:\Data\git\R-ArcGIS\r-sample-tools\scripts\semi\_par\_regression.r

## R-ArcGIS Links

- <https://r-arcgis.github.io>
- <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.

**R** ↔ **ArcGIS**

Need the R Statistical Software? Download it now.

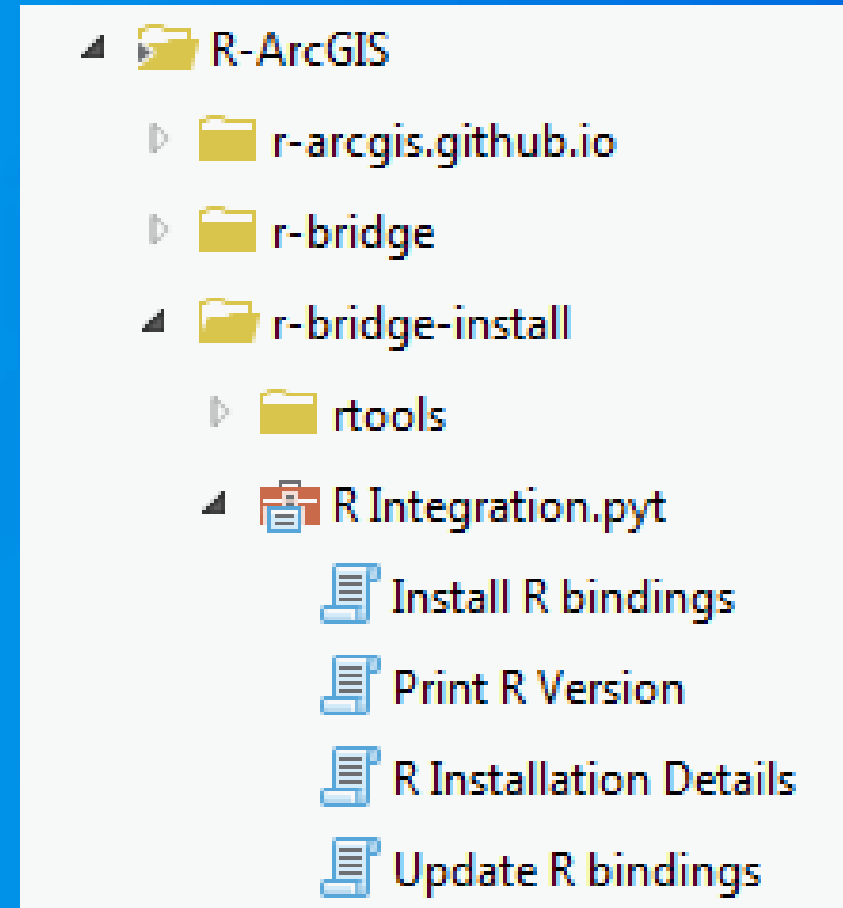
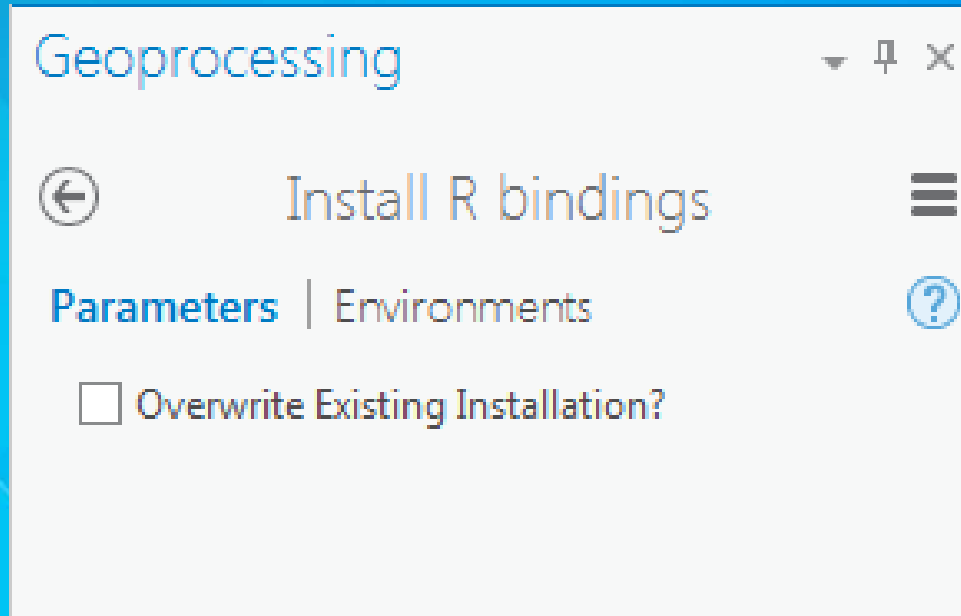
The banner features a topographic map background. At the bottom, there is a diagram showing two dark grey rounded rectangles labeled 'R' and 'ArcGIS'. Two thick orange arrows connect them, one pointing from R to ArcGIS and one from ArcGIS to R. Above the arrows is a white icon of a network graph, and below is a white icon of a database cylinder.

# Installation



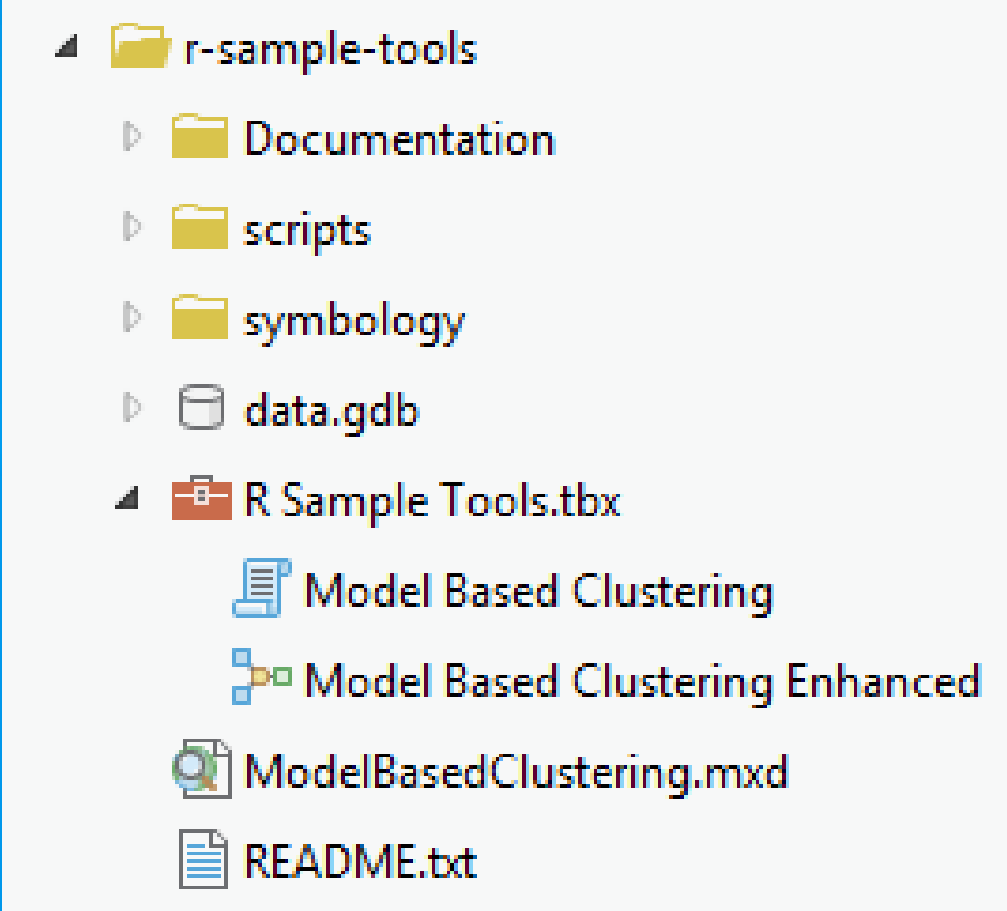










R-ArcGIS 

Via Python Toolbox  
R Package Manager



# Sample Tools

- **Model Based Clustering**
- **Spatial Statistical Data Analysis for GIS Users**
  - Konstantin Krivoruchko, Esri Press
  - Tool Help/Documentation
  - Two More Examples to Come
- **Community**

- 
- ▲  r-sample-tools
    - ▶  Documentation
    - ▶  scripts
    - ▶  symbology
    - ▶  data.gdb
    - ▲  R Sample Tools.tbx
      -  Model Based Clustering
      -  Model Based Clustering Enhanced
    -  ModelBasedClustering.mxd
    -  README.txt

# Standard R Documentation

Bindings for ArcGIS



Documentation for package 'arcbinding' version 1.0.0.111

- [DESCRIPTION file.](#)
- [User guides, package vignettes and other documentation.](#)

## Help Pages

[arcbinding-package](#)

Bindings for ArcGIS

[arc](#)

Bindings for ArcGIS

[arc.check\\_product](#)

ArcGIS product and license information

[arc.data2sp](#)

Convert an arc.dataframe object to an sp SpatialDataFrame object

[arc.dataset-class](#)

Class "arc.dataset"

[arc.env](#)

Get geoprocessing environment settings

[arc.feature-class](#)

Class "arc.dataset"

[arc.fromP4ToWkt](#)

Convert PROJ4 Coordinate Reference System string to Well-known Text

# Demo



# Moving Forward

- **Listen to our Users**
  - **Build the Best Tools Around**
  - **Support their Endeavors**
- **Python**
  - **Aim to be Agnostic to Python Versioning**
  - **Conda**
- **R**
  - **The Community Must Lead the Way**





Understanding our world.