



Data Science Made Easy in ArcGIS Using Python and R

Mark Janikas and Marjean Pobuda

mjanikas@esri.com; mpobuda@esri.com

**GIS
INSPIRING
WHAT'S
NEXT**

Data Science

From Core to Community

- 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

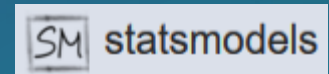
Data Science Community

Python

IP[y]: IPython
Interactive Computing

- Numeric/Scientific Python Modules
 - <http://wiki.python.org/moin/NumericAndScientific>
 - +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
 - <http://www.lfd.uci.edu/~gohlke/pythonlibs/>

 The Blaze Ecosystem



Data Science Community

R

- Well over 12,000 packages to enhance core
- Most widely used statistical software in the world
- Diverse and powerful
 - Universities, Government, Industry
 - Finance, Ecology, Statistics
 - Machine learning, predictive analytics



Battle of Bands

Which one is best?

- KD nuggets (2015)
 - <http://www.kdnuggets.com/2015/05/r-vs-python-data-science.html>
 - 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>
 - R Sample Toolbox
 - <https://github.com/R-ArcGIS/r-sample-tools>
 - Microsoft Data Science VM
 - <https://azuremarketplace.microsoft.com/en-us/marketplace/apps/microsoft-ads.standard-data-science-vm>



Integration

GUI Interface

Tool Properties: Automatic Model Search

| General | Name |
|------------|--|
| Parameters | autoModel |
| Validation | Label |
| | Automatic Model Search |
| | Script File |
| | C:\git\PySAL-ArcGIS-Toolbox\Scripts\AutoModel.py |

AutoModel.py

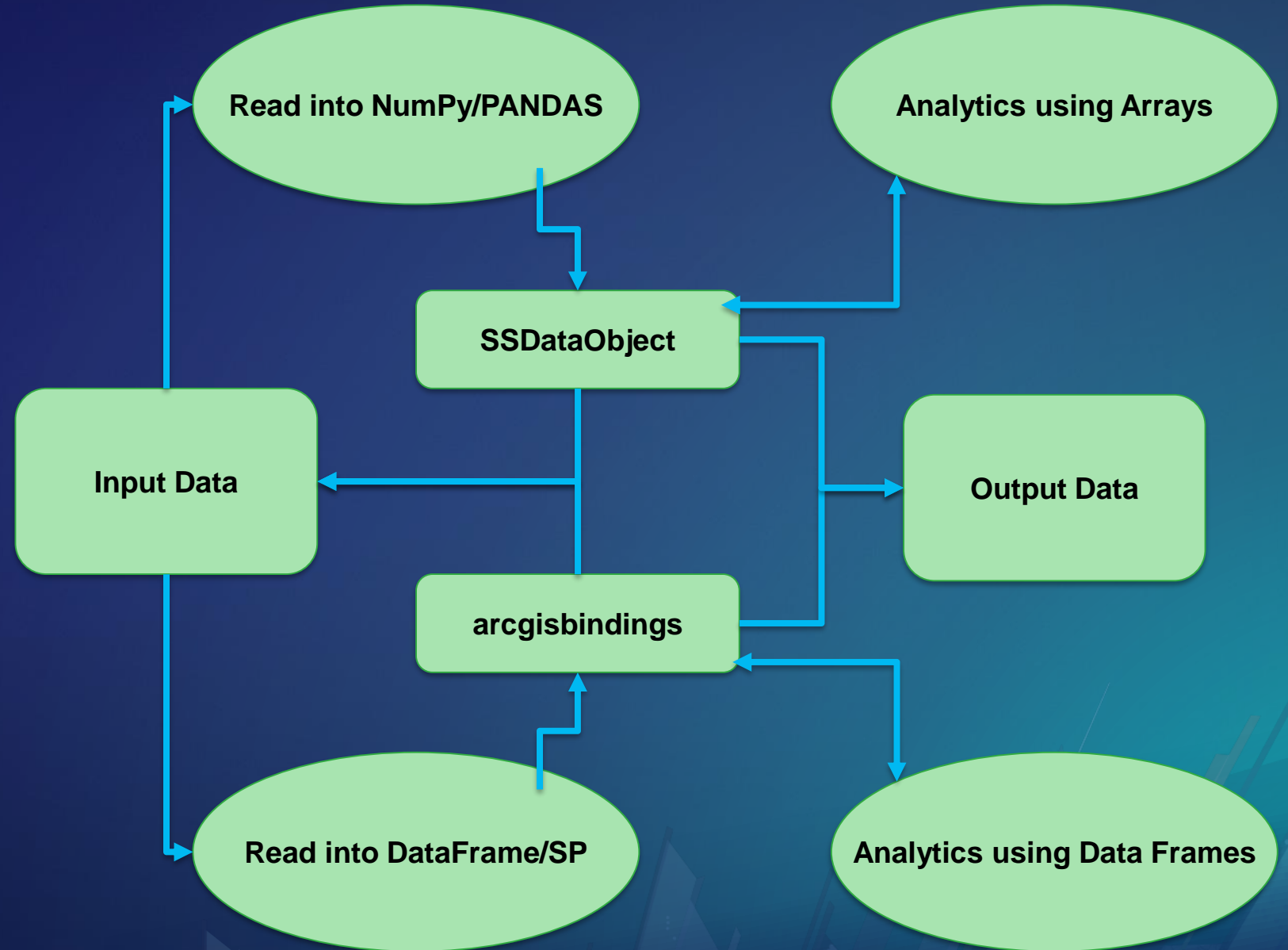
Tool Properties: Semiparametric Regression

| General | Name |
|------------|---|
| Parameters | SemiparametricRegression |
| Validation | Label |
| | Semiparametric Regression |
| | Script File |
| | C:\git\r-sample-tools\scripts\semi_par_regression.R |

semi_par_regression.R

Integration

Data IO Flow Chart



Integration

Simple API for reading data

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

```
In [ ]: ##### Loading dataset #####
        info <- arc.open(inputFC)
        ##### Create Data.Frame #####
        df <- arc.select(info, c('MYID', 'GROWTH', 'PCR1970', 'POPDEN70', 'PERCNOHS'))
        head(df)
```


Integration

Simple API for writing data

```
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

Panel Data IO, Jupyter Notebooks
and the PySAL Library

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
 - <https://github.com/Esri/gis-stat-analysis-py-tutor>
- **PySAL-ArcGIS-Toolbox**
 - Spatial Econometrics Made Easy
 - spreg module
 - <https://github.com/Esri/PySAL-ArcGIS-Toolbox>



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



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 (ArcGIS Pro 2.2), SSCube and SSPanel**
 - **Native Data Access in arcpy.da**

```
def getSpatialDataFrame(self):  
    """Creates an ArcGIS Python API Spatial Data Frame out of all the  
    fields in the DataObject. If the requireGeometry boolean option on  
    the obtainData call was False then True Centroids will be returned instead  
    of the actual geometries.  
    """  
    import arcgis as ARCGIS  
  
    df = self.getDataFrame()  
    shapes = self.getShapesAsArray()  
    return ARCGIS.features.SpatialDataFrame(df, geometry = shapes, sr = self.spatialRef)
```

The R-ArcGIS Bridge

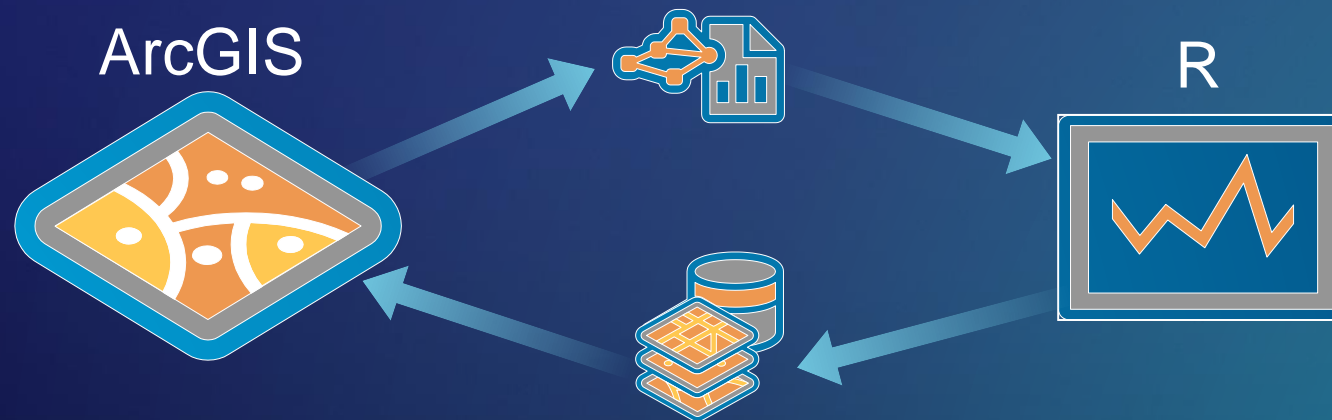
Some Background



Introducing the R-ArcGIS Bridge

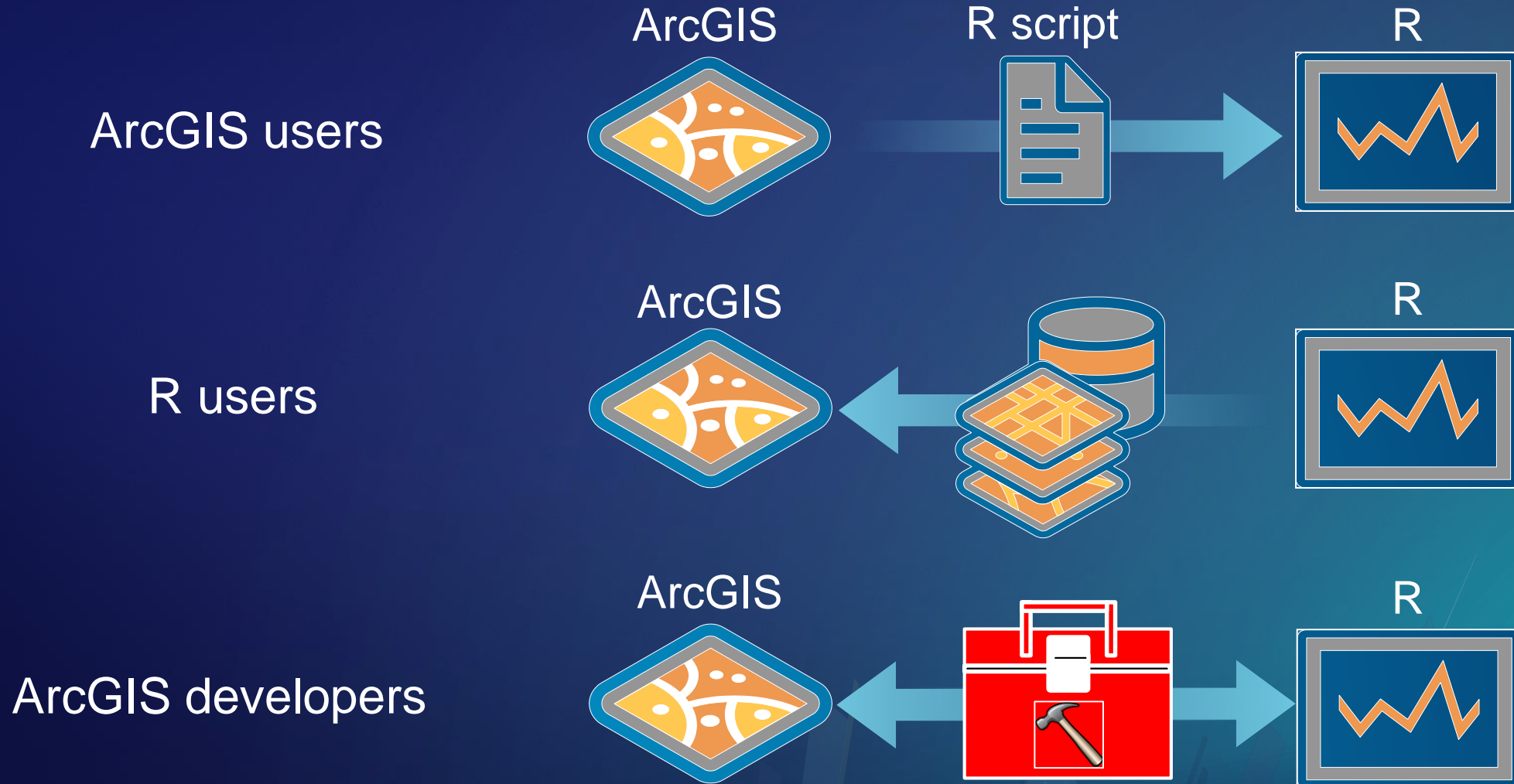
The R-ArcGIS Bridge

- 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 integrate R and ArcGIS functionality.



Who Can Use the R-ArcGIS Bridge?

The R-ArcGIS Bridge



Requirements for the R-ArcGIS Bridge

Installing The Bridge

ArcGIS Pro



1.1 (or later)

ArcMap



10.3.1 (or later)

R



3.2.2 (or later)

RStudio



Optional

Vector Support

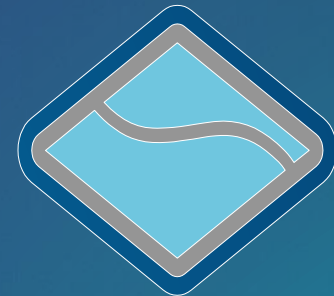
The R-ArcGIS Bridge

- Ability to read and write vector data
- Support for key R objects and spatial packages
 - R data frame object
 - Compatibility with **sp**
 - Compatibility with **sf**
- Customize data manipulations
 - Craft SQL queries to make selections
 - Subset by specific columns
 - Reproject data as needed
- Maintain spatial geometries when working with **dplyr**

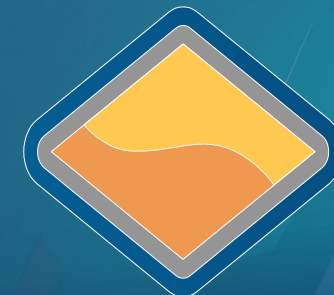
Points



Lines



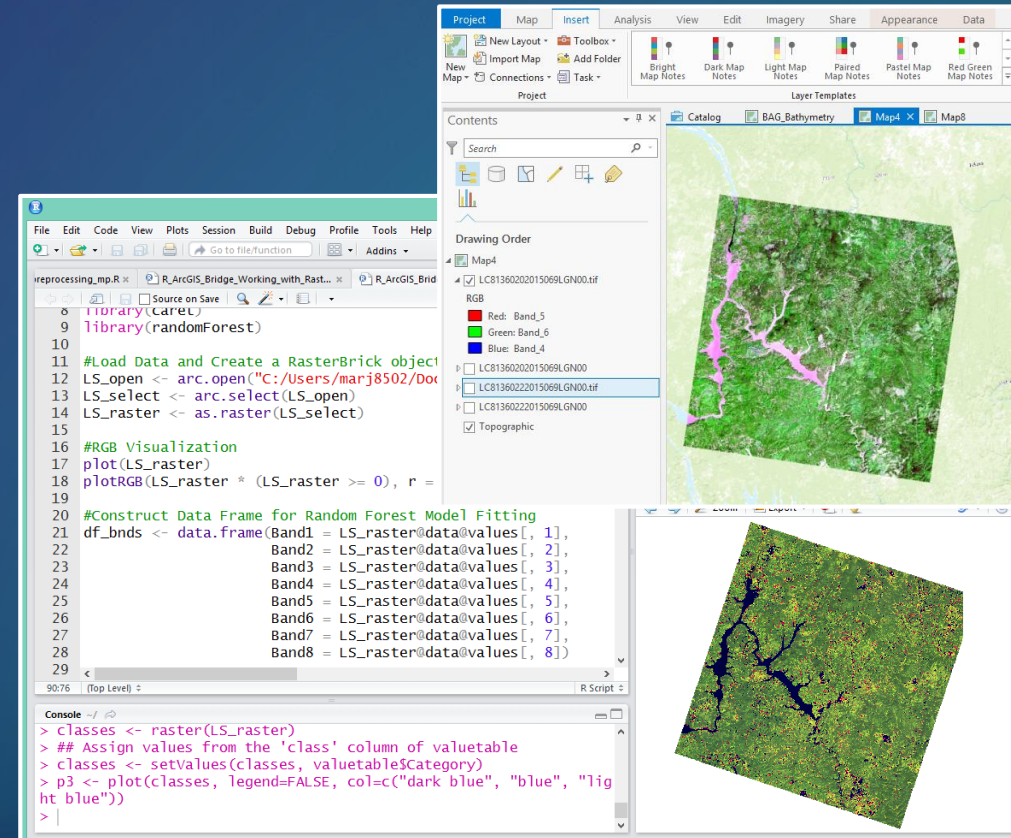
Polygons



Raster Support

The R-ArcGIS Bridge

- Ability to read and write raster data
 - Handle big data 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



Using the R-ArcGIS Bridge with Microsoft R

Microsoft R

- Microsoft Open R is a publicly available R-version for big data
- Contains almost all CRAN libraries
- Window-based operations and image operators speed up drastically
- Set-up and Usage from Pro is exactly the same as traditional R



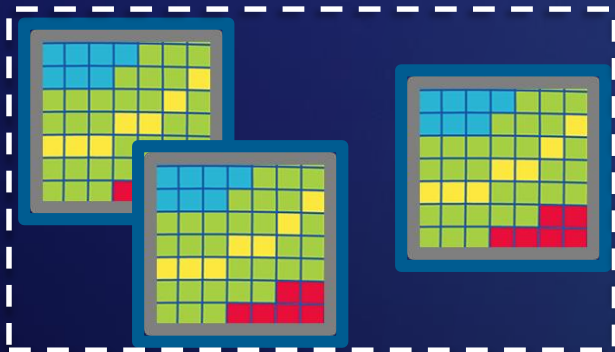
How is Microsoft R useful?

Microsoft R

Raster data can become a big data problem, quickly

Mosaics: Data structure to store/process rasters in space and time

Mosaics



Time-Series Rasters/Mosaic





R-ArcGIS Bridge Demo

Expanding Workflows and Creating
Script Tools

Resources

Learn More on Using the R-ArcGIS Bridge

Resources from UC 2018:

- <https://github.com/R-ArcGIS>

Getting Started:

- [Analyzing Crime Using Statistics and the R-ArcGIS Bridge Learn Lesson](#)
- [Using the R-ArcGIS Bridge Introductory Web Course](#)

Creating R Script Tools:

- [Integrating R Scripts into Geoprocessing Tools Web Course](#)
- [arcgisbinding Package Vignette](#)

Powerful, In-depth Workflows in ArcGIS and R

- [Identify an Ecological Niche for African Buffalo](#)

Upcoming Sessions

A Deeper Dive

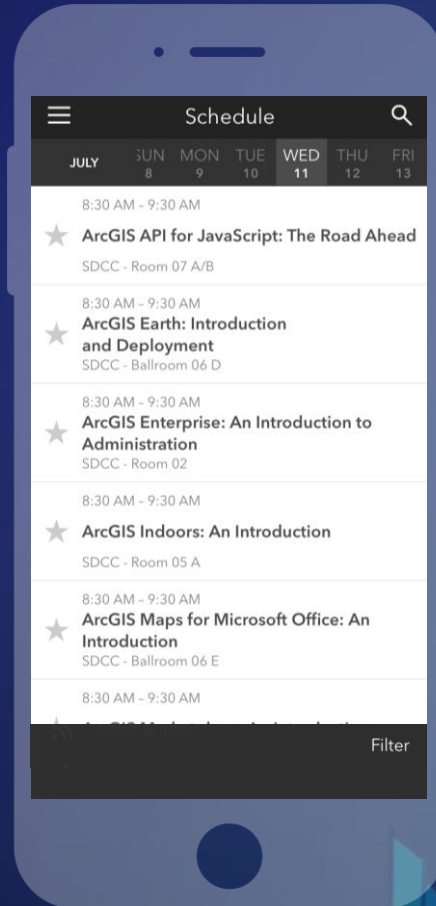
- **ArcGIS API for Python**
 - Tuesday, July 10th @ 1:00pm
- **GeoAI Deep Dive: Implementing Machine Learning Solutions with ArcGIS**
 - Tuesday, July 10th @ 1:00pm
- **GeoAI Use-Cases: Machine Learning Meets ArcGIS**
 - Tuesday, July 10th @ 2:30pm
- **Spatial Analysis: The Road Ahead**
 - Wednesday, July 11th @ 1:00pm
- **Integrating R and ArcGIS for Advanced Analysis**
 - Wednesday, July 11th @ 8:30am
 - Thursday, July 12th @ 1:00pm
- **Machine Learning in ArcGIS**
 - Thursday, July 12th @ 8:30am
 - Thursday, July 12th @ 2:30pm

Please Take Our Survey on the App

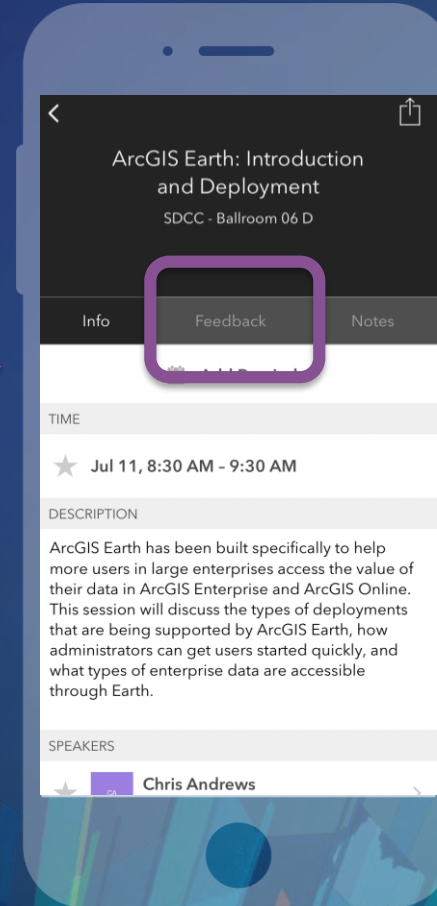
Download the Esri Events app and find your event



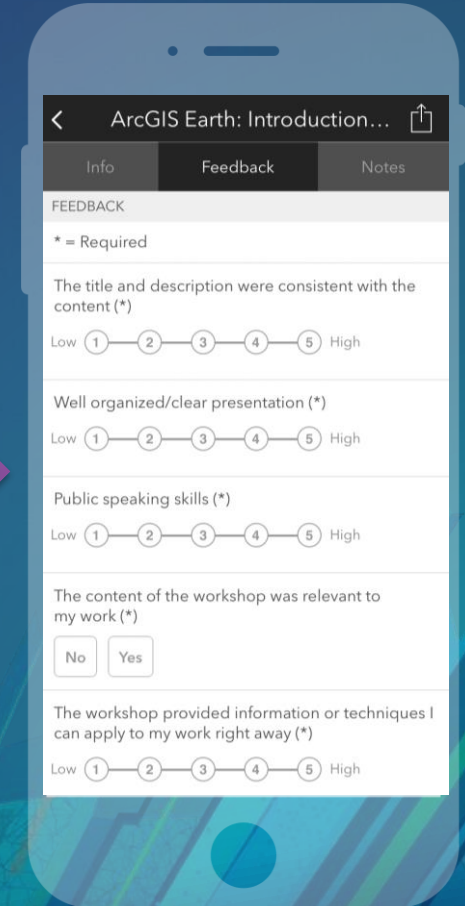
Select the session you attended



Scroll down to find the feedback section



Complete answers and select "Submit"



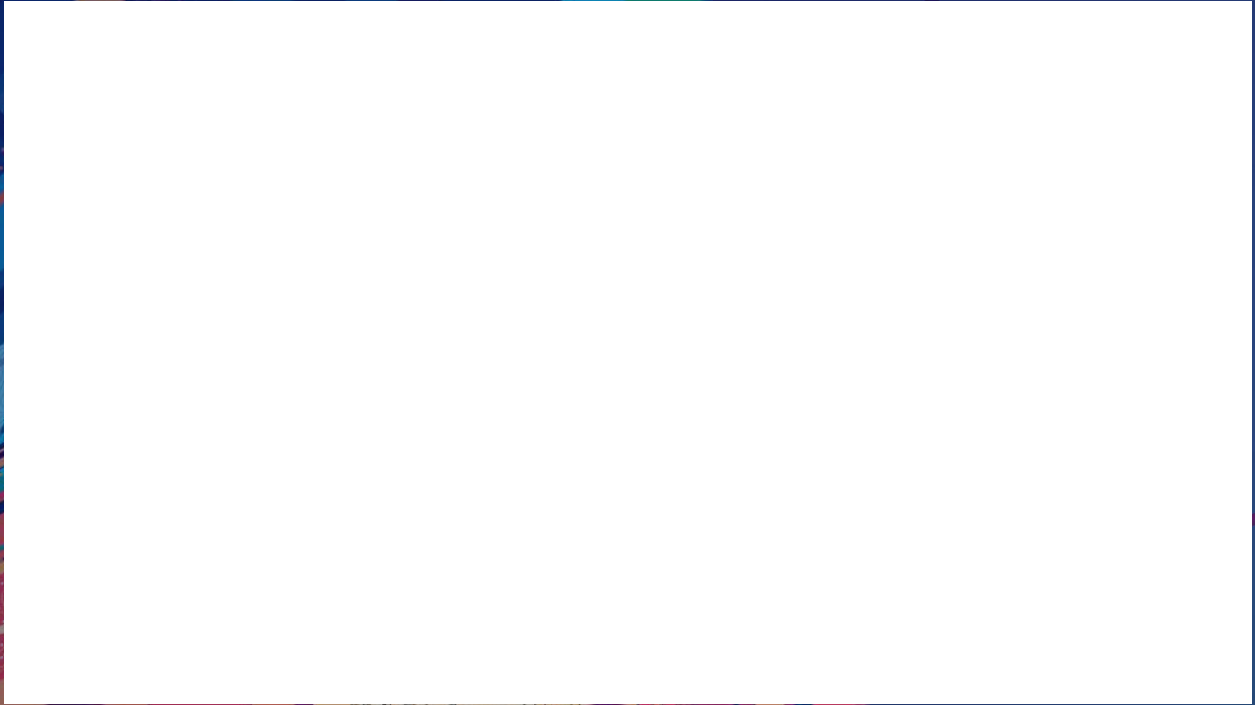


esri

**THE
SCIENCE
OF
WHERE**





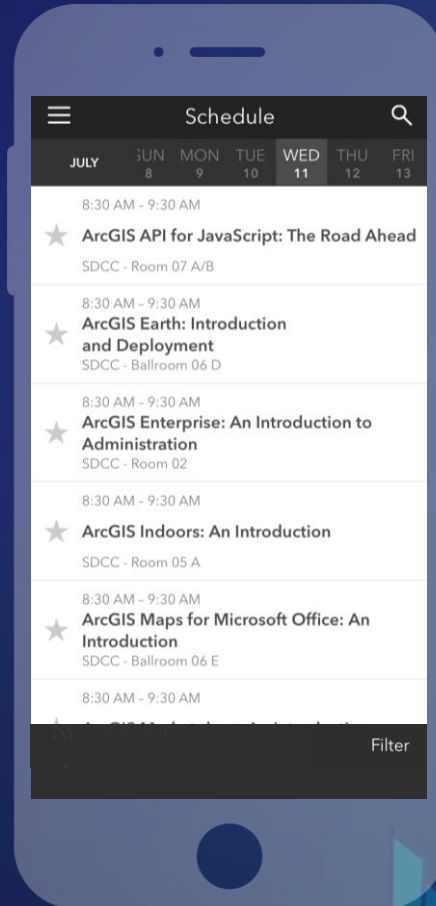


Please Take Our Survey on the App

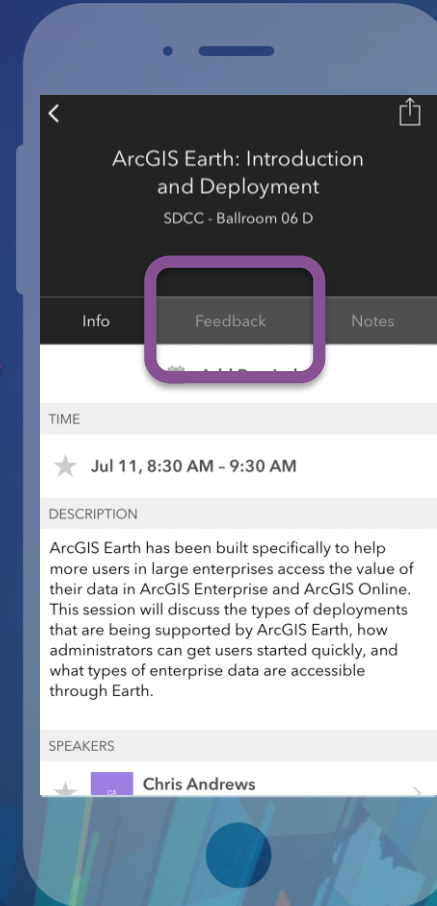
Download the Esri Events app and find your event



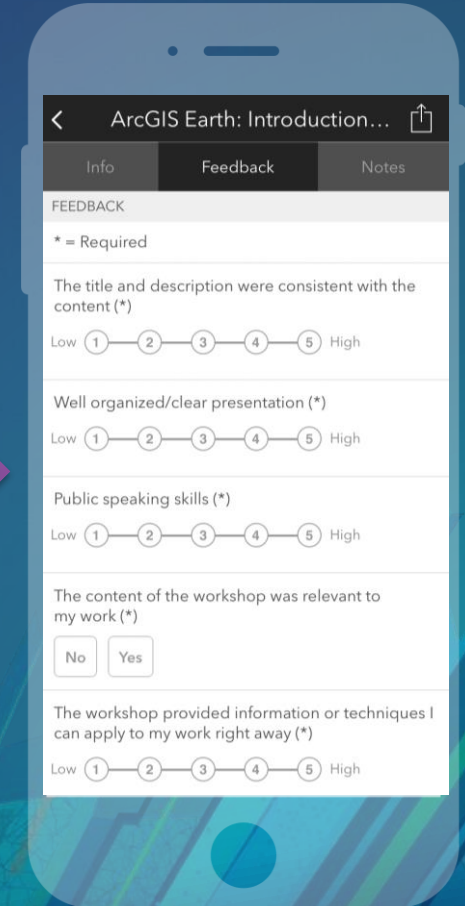
Select the session you attended



Scroll down to find the feedback section



Complete answers and select "Submit"



See Us Here

| WORKSHOP | LOCATION | TIME FRAME |
|---|--|--|
| <ul style="list-style-type: none"><li data-bbox="191 385 305 428">• Title | <ul style="list-style-type: none"><li data-bbox="930 385 1082 428">• Where | <ul style="list-style-type: none"><li data-bbox="1668 385 1936 428">• 12:00 – 1:00 |