Data Science in ArcGIS Using Python and R

Kade Beem
Lu Zhang

2020 ESRI FEDERAL GIS CONFERENCE | WASHINGTON, D.C.
Agenda

• Introduction
• Performing Analysis using Python
  • ArcPy
  • ArcGIS API for Python
  • Demonstration
• Utilizing the R-ArcGIS Bridge
  • Installation / Use Cases
  • Sample Workflow
• Q&A
Data Science

Introduction

Drew Conway 2010
Data Science
From Core to Community

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

• The interoperability of the ArcGIS platform makes workflows more efficient
  - Techniques and methodologies continue to develop
  - Data availability continues to increase

• The data science community is vast and evolving
  - ArcGIS extends directly via scripting APIs
    - e.g. Python, R, Java
Spatial Analysis and Data Science Framework

Patterns of Use

- Data Engineering
- Visualization & Exploration
- Spatial Analysis
- Machine Learning & AI
- Big Data Analytics
- Modeling & Scripting
- Sharing & Collaboration
Spatial Analysis and Data Science Framework

Data Engineering

- Extract and prep
- Location-enable your data
- Enrich your data
- Access ready-to-use data and content
Spatial Analysis and Data Science Framework

Visualization & Exploration

- Mapping
- Charting
- Animation
Spatial Analysis and Data Science Framework

Spatial Analysis

- Pattern detection
- Space-time analytics
- Predictive analytics
- Suitability modeling and location optimization
- Route planning and logistics
- Terrain analysis
- Scale up your analysis
Spatial Analysis and Data Science Framework

Machine Learning & Artificial Intelligence

**Classification**
- Maximum Likelihood Classification
- Random Trees
- Support Vector Machine
- Find Dwell Locations

**Prediction**
- Empirical Bayesian Kriging
- Areal Interpolation
- EBK Regression Prediction
- Ordinary Least Squares Regression and Exploratory Regression
- Geographically Weighted Regression
- Forest-based Classification and Regression

**Clustering**
- Spatially Constrained Multivariate Clustering
- Multivariate Clustering
- Density-based Clustering
- Find Point Clusters
- Hot Spot Analysis
- Cluster and Outlier Analysis
- Space Time Pattern Mining
Spatial Analysis and Data Science Framework

Big Data Analytics

- Visualization
- Distributed computing
  - GeoAnalytics Server
  - GeoAnalytics Desktop
- Imagery and raster analytics
- Real-time analytics
  - GeoEvent Server
Spatial Analysis and Data Science Framework

Modeling & Scripting

- Train sophisticated models
- Save time by automating
- Extensibility and integration
- Transparency and reproducibility
Spatial Analysis and Data Science Framework
Sharing & Collaboration

- Work collaboratively and share result
- Integrate analytics across your organization
- Tell impactful stories to an engage audience
Data Science Community
Python

IP[y]: IPython Interactive Computing

BIG DATA

PYSAL: Python Spatial Analysis Library

ANACONDA

The Blaze Ecosystem

AI

arcsgis.learn

TensorFlow

NLP

Bokeh

statsmodels

scikit learn
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?

<table>
<thead>
<tr>
<th>General Programming Language</th>
<th>Functionality</th>
<th>Package Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python</td>
<td>Tailored towards statistics and data analysis</td>
<td>Yes (PyPI and Conda)</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td>Yes (CRAN and Conda)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scalability</th>
<th>Display of Data</th>
<th>Use Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual machines to distributed computing environments</td>
<td>Large number of libraries for graphical display of data</td>
<td>General Programming, Data Science, Web Development</td>
</tr>
<tr>
<td></td>
<td>Numerous libraries for making incredible graphics</td>
<td>Lingua franca of data science</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What are you most comfortable with?</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES!</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What is the best tool for the job?</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES!</td>
</tr>
</tbody>
</table>

Integrates with ArcGIS
Performing Analysis Using Python
Python
ArcPy vs ArcGIS API for Python

**ArcPy**
- Use, extend, automate, desktop GIS
- Access GP tools, extensions, functions, and classes
- Create standalone scripts to run with an IDE.
- Create script tools and added to a toolbox within ArcGIS Pro

**ArcGIS API for Python**
- Use, extend, automate, Web GIS
- Perform visualization, analysis, spatial data management, and GIS system administration
- Run interactively in a notebook
- Create script tools and added to a toolbox within ArcGIS Pro
Python

ArcPy

- **Tools**
- **Functions**
- **Classes**
- **Modules**
  - **Mapping – arcpy.mp**
    - Manipulate contents of ArcGIS Pro projects and layer files
    - Upgraded from arcpy.mapping
    - Leverages Python 3.x

```python
import arcpy
aprx = arcpy.mp.ArcGISProject(r"C:\Projects\Yosemite\Yosemite.aprx")
mp = aprx.listMaps('Yosemite National Park')
for lyr in mp.listLayers():
    if lyr.name == 'Roads':
        if lyr.maxThreshold == 10000:
            lyr.visible = True
        else:
            lyr.visible = False
aprx.save()
del aprx

arcpy.mp functions

ArcGisProject(aprx_path)

ColorRamp(color_ramp_name, (index))

ConvertLayoutFileToLayout(layout_file)

ConvertWebMapToArcGisProject(webmap_json, {template_page}, {mapframe_name}, {noti_gdb})

CreateWebLayerSDDraft(map_or_layers, out_sddraft, service_name, {server_type}, {service_type}, {folder_name}, {overwrite_existing_service}, {copy_data_to_server}, {enable_editing}, {allow_exporting}, {enable_sync}, {summary}, {tags}, {description}, {credits}, {use_limitations})

LayerFile(layer_file_path)

PDFDocumentCreate(pdf_path)

PDFDocumentOpen(pdf_path, {user_password}, {master_password})

Table(table_data_source)
```
Python
ArcGIS API for Python

- On your **Web GIS** perform:
  - GIS visualization
  - Analysis
  - Spatial data management

- Leverage SciPy ecosystem for
  - Automating workflows
  - Performing repetitive tasks

- Integrate with notebooks for **interactive scripting and visualization**
Data Science Process

- Ask a question
  - Download
  - Scrape
  - ETL
  - Enrich

- Get the data
  - Chart
  - Summary & Descriptive Statistics
  - Develop Hypotheses
  - Identify Patterns

- Explore the data
  - Regression
  - Machine Learning
  - "Big Data"

- Model the data
  - Notebooks
  - Written/Textual Reports
  - Web Apps
  - Story Maps
  - PowerPoint

- Communicate
  - Written/Textual Reports
  - Web Apps
  - Story Maps
  - PowerPoint
Get the Data

Non-GIS Data
- Large number of libraries to download, transform, condition, and prepare data
- Generic web libraries (Requests, urllib(2))
- API Specific libraries (Tweepy)
- Ability to scrape and parse existing web sites (Scrapy and BeautifulSoup)

GIS Data
- ArcPy allows native access to all Esri Data formats and the full functionality of ArcGIS Desktop
- ArcGIS API for Python allows for access and interaction with content within your WebGIS
Explore the Data

- **ArcGIS Notebooks** are your friend
- Ability to create **Spatially Enabled DataFrame (SEDF)**
- Core Pandas DataFrame, allows for rapid exploration of data
- Can create maps, charts and a variety of other objects quickly using a common syntax
- **Availability**
  - ArcGIS Pro Notebooks 2.5+
  - ArcGIS Notebooks for Enterprise 10.7+
  - We plan to open public beta for ArcGIS Online Notebooks in March with a full release later in the year if all goes well
Explore the Data
Spatially Enabled DataFrame

- **Spatial engine built directly into the DataFrame**
  - Use Pandas operations on attribute and spatial columns
- Read in shapefiles, Pandas DataFrames, feature classes, GeoJSON, Feature Layers
- New SEDF object requires ArcGIS API for Python version 1.5
  - Updated from old SDF object
  - Better memory management
  - Handle larger datasets
  - Pandas advocates as the path forward
Explore the Data

arcgis.learn Module

- Supports AI/ML workflows within the ArcGIS platform
  - Exporting training data
  - Data prep
  - Model training
  - Model deployment
  - Model inference
- Leverage tools across:
  - ArcGIS API for Python
  - ArcGIS Pro
  - ArcGIS Image Server
  - ArcGIS Notebooks
Modeling the Data

- Many native tools to enable modeling of data (Space-Time Pattern Mining, Density Based Clustering, etc…)

- Integration of popular third-party Machine Learning/Deep Learning libraries
  - Scikit-Learn
  - Tensorflow
  - PyTorch
  - NLTK
Python Demo

ArcGIS Notebooks
Data Engineering
Visualization & Exploration
Spatial Analysis
Resources

Documentation
- What is ArcPy?
- Overview of the ArcGIS API for Python
- Introducing ArcGIS Notebooks
- Spatial Analysis and Data Science

Community
- SciPy
- Python – Numeric and Scientific Libraries

Spatial Statistics GitHub
- Spatial Statistics Resources

GeoNet
- Spatial Statistics
- Python
- ArcGIS API for Python
Key Takeaways

- **Knowledge** – Use the best tool for the job - ArcPy and ArcGIS API for Python can help accomplish complex, data science workflows.

- **Integration** – ArcGIS is an open platform that supports end-to-end analytic workflows. Leverage third party libraries.

- **Communication** – Harness your portal to deliver information products to your organization/community.
Utilizing the R- ArcGIS Bridge
An Introduction
What is R?

• A widely used statistical programming language
  - Linear and non-linear modeling / Classification / Clustering …
  - Data Cleaning & Transformation
  - Data Visualization

• More than 12,000 Packages
  - General to Specific Use Cases

• Open-Source
  - Universities, Government, Industry
  - Finance, Ecology, Statistics
  - Machine learning, predictive analytics
What is the R-ArcGIS Bridge

• Provides the ability for users to integrate R and ArcGIS functionality

• Open source project, free to download and use
Who Can Use the R-ArcGIS Bridge?

- GIS Analyst
- Data Scientist
- Developers
Version Requirements for the R-ArcGIS Bridge

ArcGIS Pro
- 1.1 (or later)

ArcMap
- 10.3.1 (or later)

R
- 3.2.2 (or later)

RStudio / Notebook
- Recommended!
Installing the Bridge / Getting Started
ArcGIS Pro Project Tab
Vector Support

- Ability to read, write, and modify vector data

- Support / Conversion for key R objects and spatial packages
  - R data frame
  - Compatibility with sp package
  - Compatibility with sf package

- Customize data manipulations
  - SQL queries / Subsetting
  - Data Clean Up and Engineering

- Maintain spatial geometries when working with dplyr
**Raster Support**

- 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
Support for Mosaic Datasets

• Raster data can become a big data problem, quickly

**Mosaics:** Data structure to store/process rasters in space and time
Working with Big Data

- 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
R-ArcGIS Bridge Demo

Predict Home Vacancy Rates in Washington DC based on Vacant Home Locations in Baltimore (XY Coordinates)
Key Takeaways

• ArcGIS is an open platform that supports end to end analytic workflows

• Designed to integrate with your favorite modules in Python and R

• Communicate your results with the community through ArcGIS Online / Enterprise
Resources
Learn More on Using the R-ArcGIS Bridge

Resources from Spatial Statistics Team:
- 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
Relevant Upcoming Sessions: Tuesday February 11

• Data Science Using ArcGIS Notebooks
  - 150 A | 3:00 – 4:00 pm

• Machine Learning in ArcGIS
  - 146 C | 3:00 – 4:00 pm

• Using Deep Learning with Imagery in ArcGIS
  - 144 A | 3:00 – 4:00 pm

• From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations
  - 146 C | 4:15 – 5:15 pm

• Python: Beyond the Basics
  - 146 B | 4:15 – 5:15 pm
Relevant Upcoming Sessions: Wednesday February 12

- **Machine Learning in ArcGIS**
  - 146 C | 8:30 – 9:30 am

- **Using Deep Learning with Imagery in ArcGIS**
  - 144 A | 8:30 – 9:30 am

- **Data Visualization for Spatial Analysis**
  - 146 C | 11:00 am – 12:00 pm

- **From Means and Medians to Machine Learning: Spatial Statistics Basics and Innovations**
  - 146 C | 1:30 – 2:30 pm

- **Geospatial Deep Learning with arcgis.learn**
  - 101 | 1:30 – 2:30 pm
Relevant Upcoming Sessions: Wednesday February 12

- **Spatial Data Mining: Cluster Analysis and Space-Time Analysis**
  - 146 C | 2:45 – 3:45 pm

- **Python: Building Geoprocessing Tools**
  - 147 B | 4:00 – 5:00 pm

- **ArcGIS API for Python: Getting to Know Pandas and the Spatial Enabled DataFrame**
  - 146 C | 5:15 – 6:15 pm

- **Python: Map Automation in ArcGIS Pro**
  - 102 A | 5:15 – 6:15 pm

- **The Forest for the Trees: Making Prediction Using Forest-Based Classification and Regression**
  - 146 C | 5:15 – 6:15 pm
Print Your Certificate of Attendance

Print Stations Located in 150 Concourse Lobby

**Tuesday**
12:30 pm – 6:30 pm
Expo
Hall B

5:15 pm – 6:30 pm
Expo Social
Hall B

**Wednesday**
10:45 am – 5:15 pm
Expo
Hall B

6:30 pm – 9:30 pm
Networking Reception
Smithsonian National Museum of Natural History
Please Share Your Feedback in the App

Download the Esri Events app and find your event

Select the session you attended

Scroll down to “Survey”

Log in to access the survey

Complete the survey and select “Submit”