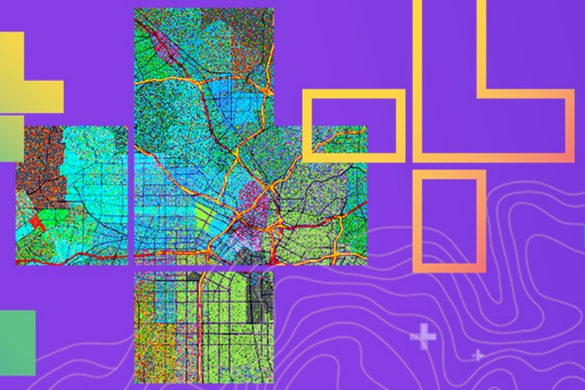
Spatial Data Science in ArcGIS: The Ecosystem

2020 ESRI DEVELOPER SUMMIT | Palm Springs, CA

Shaun Walbridge

Kevin Butler

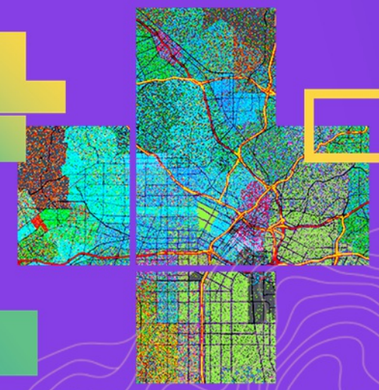


esri | THE SCIENCE OF WHERE™

https://github.com/scw/ds-scipy-devsummit-2020-talk

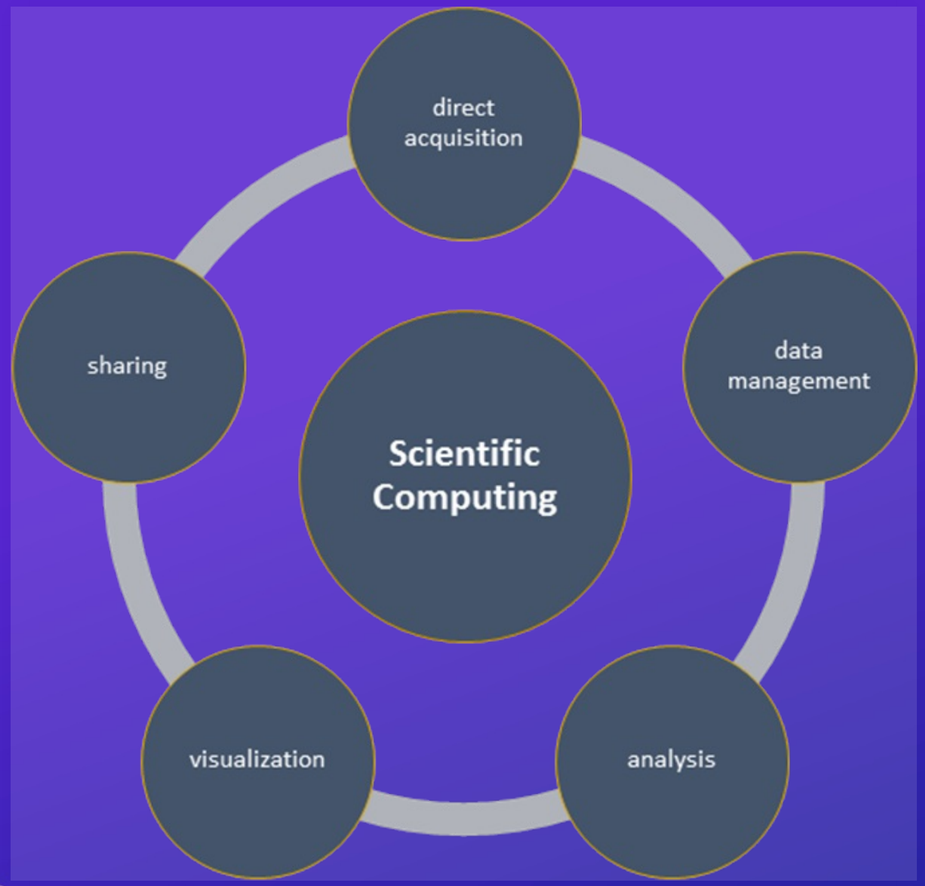
High Quality PDF (5MB)

Resources Section



Data Science

Data Science



The application of computational methods to all aspects of the process of scientific investigation – data acquisition, data management, analysis, visualization, and sharing of methods and results.

ArcGIS for *spatial* data science

- ArcGIS is a *system of record*. Combine data and analysis from many fields and into a common environment.
- Why extend? Can’t do it all, we support over 1600 GP tools — enabling *integration* with other environments to extend the platform.
- ArcGIS is an ecosystem that lends itself very nicely to the way that spatial data scientists already work.

What’s in the Ecosystem

Python in ArcGIS

- Python API for driving ArcGIS Desktop and Server
- A fully integrated module: `import arcpy`
- Interactive Window, Python Addins, Python Tooboxes
- ArcGIS API for Python
- *Hosted Notebooks*
- Notebooks in ArcGIS Pro

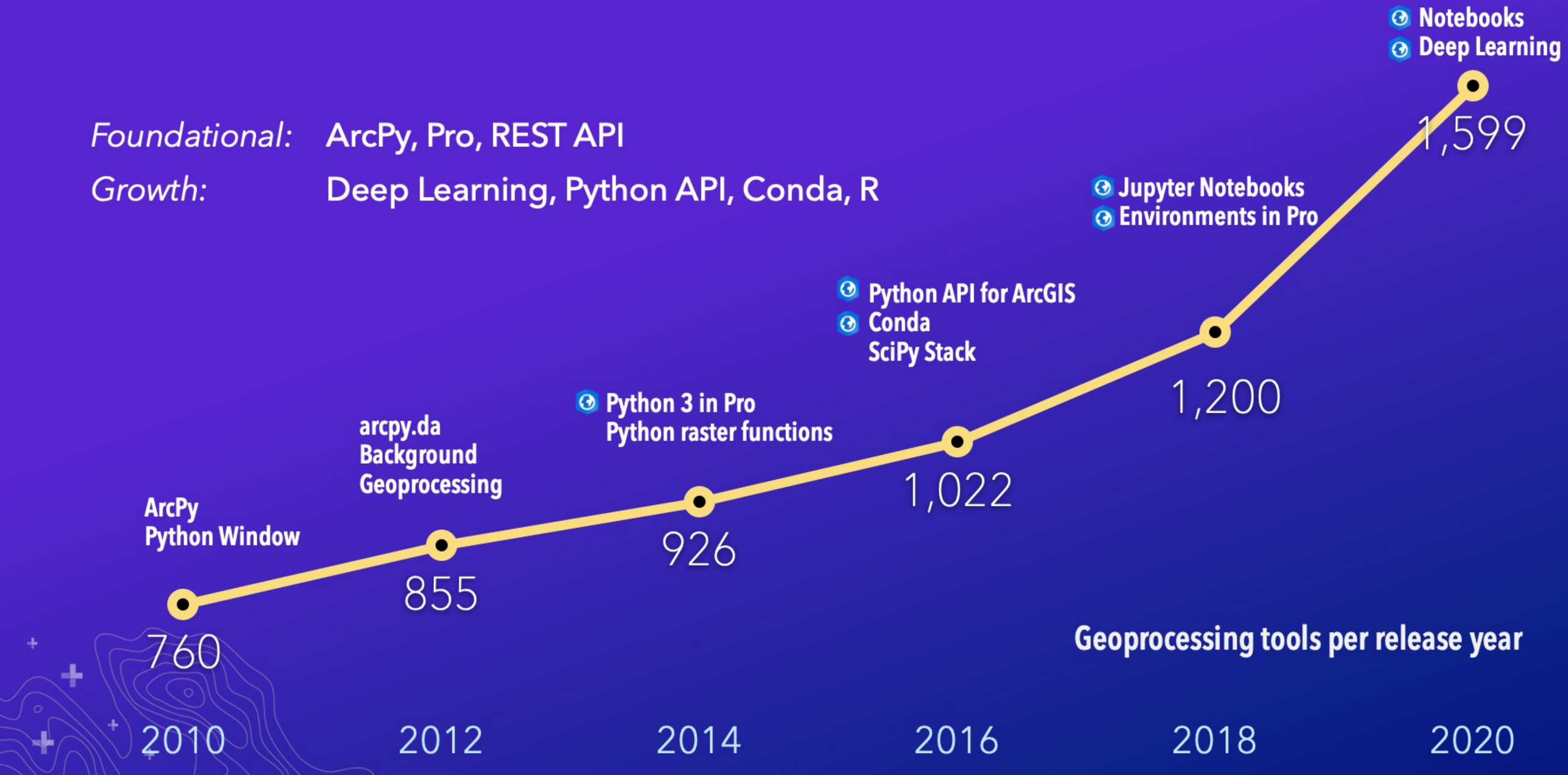
Python Everywhere

ANACONDA®

The last decade of Python with ArcGIS

Foundational: ArcPy, Pro, REST API

Growth: Deep Learning, Python API, Conda, R



Demo: Notebooks in Pro

Core Python Libraries

Why SciPy?

- Most languages don’t support things useful for science, e.g.:
  - Vector primitives
  - Complex numbers
  - Statistics
- Object oriented programming isn’t always the right paradigm for analysis applications, but is the only way to go in many modern languages
- SciPy brings the pieces that matter for scientific problems to Python.

Included SciPy

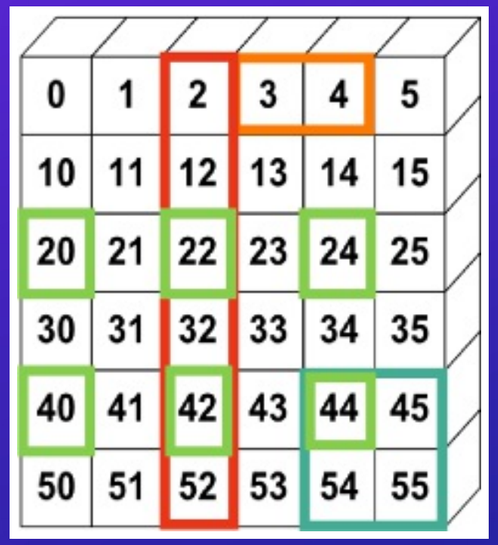
Package	KLOC	Contributors	Stars
dask	52	229	4293
IPython	36	587	13408
JupyterLab	85	214	7396
NumPy	236	738	9868
Pandas	183	1433	18431
SciPy	387	699	5522
SymPy	243	730	5617

- Plotting library and API for NumPy data
- Matplotlib Gallery
- Pro *also* includes `arcpy.chart` for plotting via Pro charts
- UC 2020: Embedded Pro charts in notebooks

ArcGIS with NumPy

NumPy

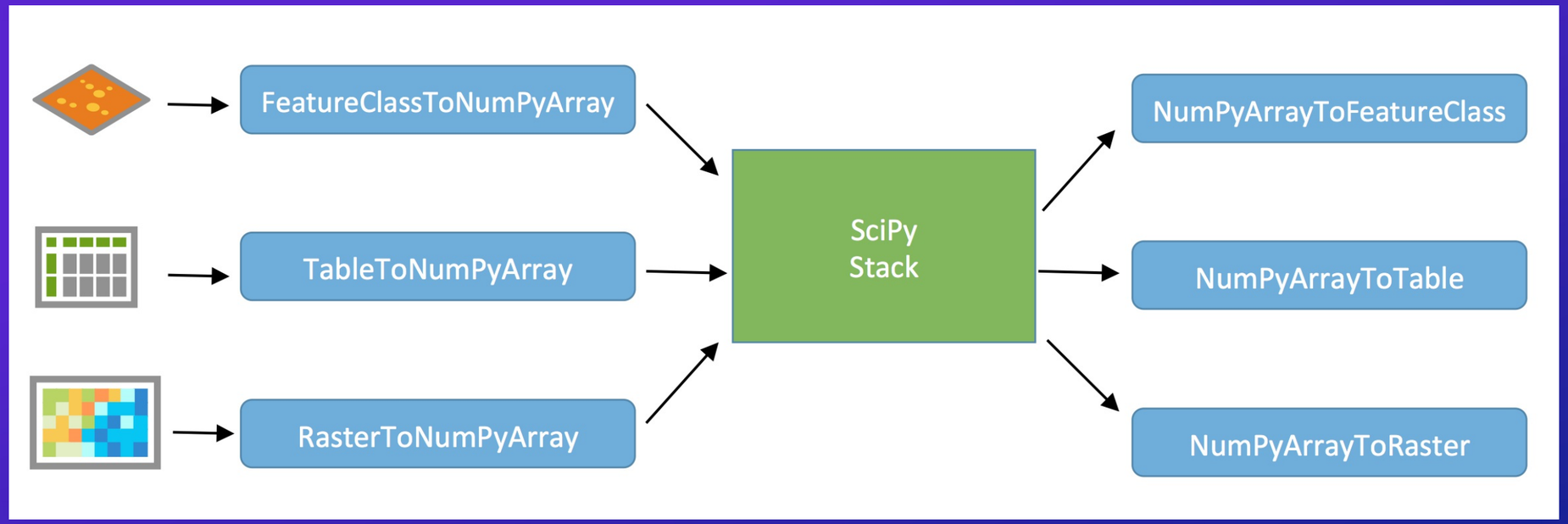
1. An array object of arbitrary homogeneous items
2. Fast mathematical operations over arrays



SciPy Lectures, CC-BY

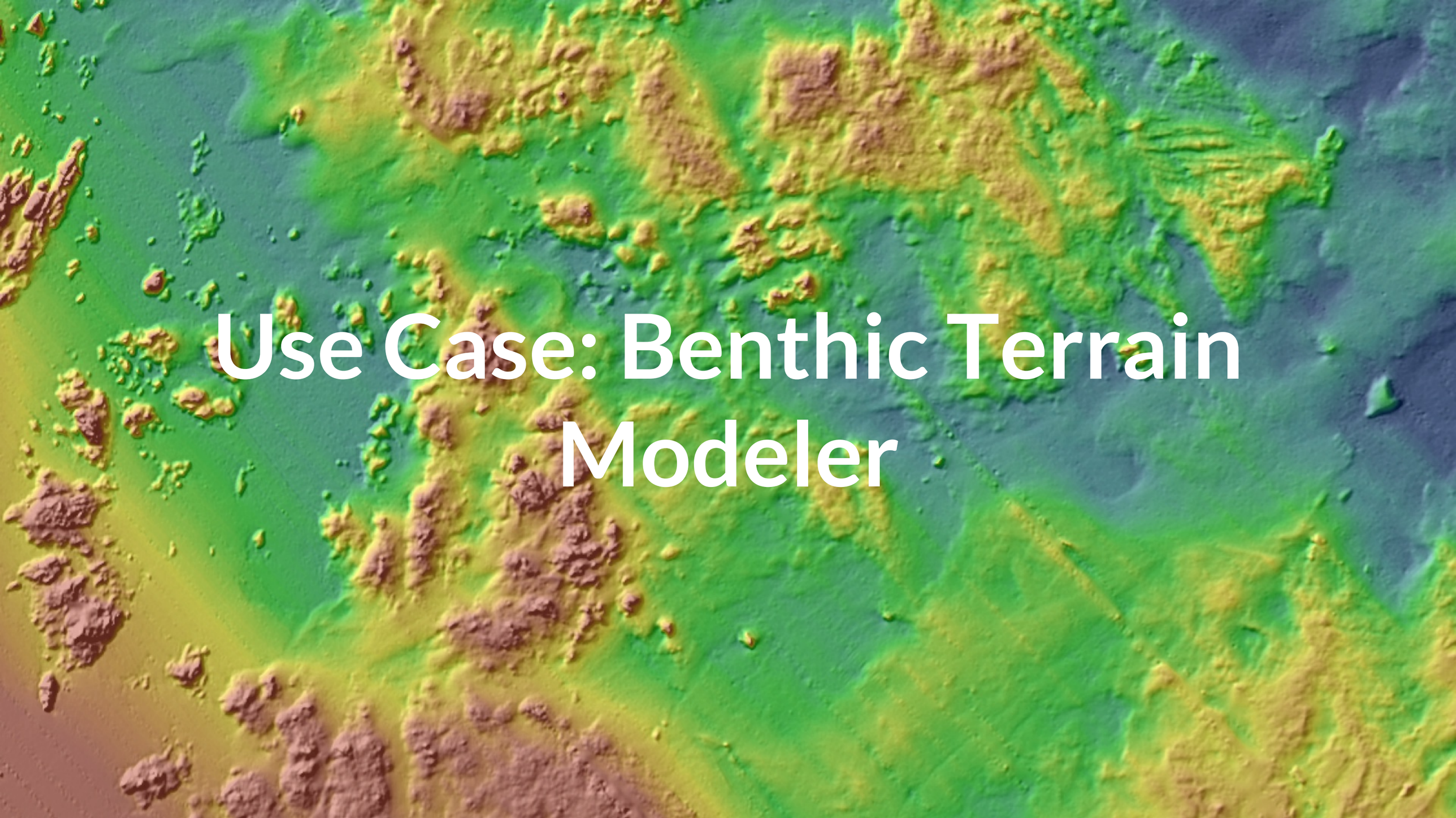
- ArcGIS and NumPy can interoperate on raster, table, and feature data.
- See Working with NumPy in ArcGIS
- In-memory data model. Example script to process by blocks if working with larger data.
- Use `arcgis`' SeDF if you need a high-level interface for feature data

ArcGIS with NumPy



Computational methods for:

- Integration (scipy.integrate)
- Optimization (scipy.optimize)
- Interpolation (scipy.interpolate)
- Fourier Transforms (scipy.fft)
- Signal Processing (scipy.signal)
- Linear Algebra (scipy.linalg)
- Spatial (scipy.spatial)
- **Statistics** (scipy.stats)
- **Multidimensional image processing** (scipy.ndimage)



Use Case: Benthic Terrain Modeler

Lightweight SciPy Integration

- Using `scipy.ndimage` to perform basic multiscale analysis
- Using `scipy.stats` to compute circular statistics

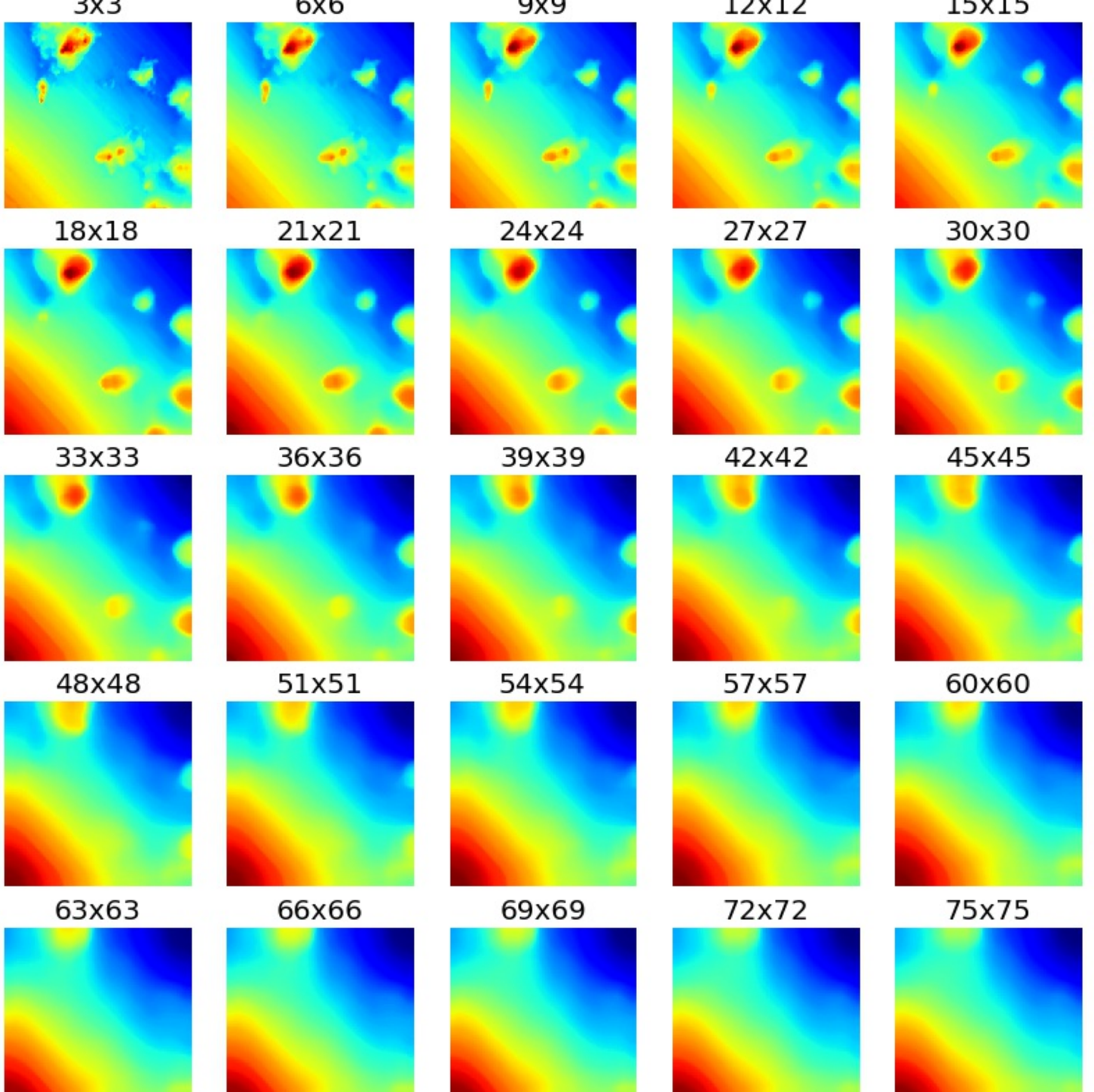
Lightweight SciPy Integration

Example source

```
import arcpy
import scipy.ndimage as nd
from matplotlib import pyplot as plt

ras = "data/input_raster.tif"
r = arcpy.RasterToNumPyArray(ras, "", 200, 200, 0)

fig = plt.figure(figsize=(10, 10))
```

Lightweight SciPy Integration

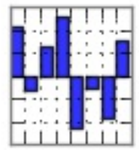
```
for i in xrange(25):
    size = (i+1) * 3
    print "running {}".format(size)
    med = nd.median_filter(r, size)

    a = fig.add_subplot(5, 5,i+1)
    plt.imshow(med, interpolation='nearest')
    a.set_title('{}x{}'.format(size, size))
    plt.axis('off')
```

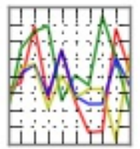

3x3
6x6
9x9
12x12
15x15
18x18
21x21
24x24
27x27
30x30
33x33
36x36
39x39
42x42
45x45
48x48
51x51
54x54
57x57
60x60
63x63
66x66
69x69
72x72
75x75

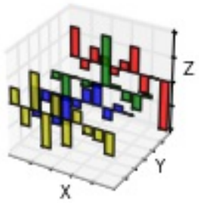
Pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$







- **Pan**el **Da**ta — like R “data frames”
- Bring a robust data *analysis* workflow to Python
- Data frames are fundamental — treat tabular (and multi-dimensional) data as a labeled, indexed series of observations.

Spatial Data Frames

- Same data frame model + geometries
- ArcPy + ArcGIS API for Python
- Continues to expand and improve performance

New in ArcPy

ArcPy Improvements

- `arcpy.metadata` for transforming your metadata
- `arcpy.nax` for rich network analysis
- Raster cell iterators for custom per-cell raster analysis without needing to copy data using NumPy `#DOCELLRISES`
- `arcpy.SetParameterSymbology` for rich analytical results like Charts and popups

ArcPy Improvements

- Rich representations for data like `arcpy` geometries, rasters
- More coming UC 2020

Integration

Integration

- OK, so we've covered core libraries that exist within the Pro Python distribution. What about going beyond this?

Integration

- What kind of code is being run?

Bring your own Existing libraries	Your components and ecosystem tools
Domain specific tools Tools built and supported by Esri	The frameworks + tools that bind to them

- The Principle of stack minimization

Demo: MetPy

- Massive data parallelism through Python
- Computes graphs of the computational structure

Demo: Dask & Tying It Together

Integration

Leverage the broad data science ecosystems of R and Python

ANACONDA

Drive Integrated Code

Pro Python Distribution
Environment Management
Docker Runtimes (Hosted)
https://anaconda.org/esri

ArcPy and ArcGIS API

Integration includes:
- NumPy
- Pandas
- PyTorch
- Jupyter Notebooks

R-ArcGIS Bridge

RStudio
Geoprocessing Tools
Web Tools
Jupyter Notebooks

R

- R Statistical Programming Language
- Powerful core data structures for analysis
- Unparalleled breath of statistical routines

R-ArcGIS Bridge

- Access to local and remote data
- Transform to native R spatial types (`sf`, `sp`, `raster`)
- Call ArcPy through `reticulate`
- Use in RStudio
- Make GP tools which call R
- Jupyter Notebooks with R: `conda install r-arcgis-essentials`

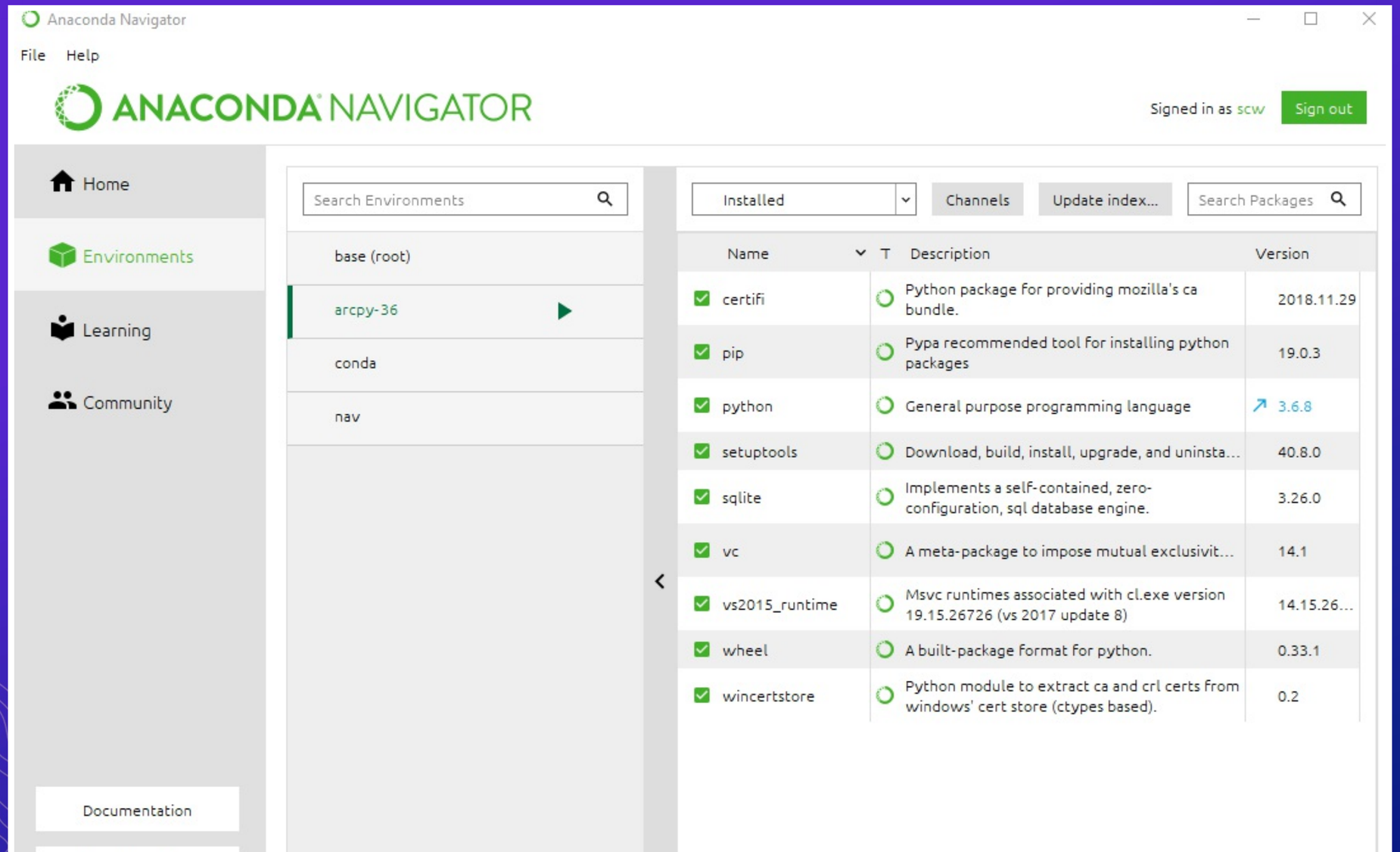
Demo: R

from future import *

Road Ahead

- Continued improvements in Deep Learning in Pro — make this experience as seamless and as simple as possible
- Rich representations (`__repr__`) for many objects in ArcPy and Pro
- ArcPy in External Conda environments (detects Pro)

Pro External Environments



Resources

New to Python

- Courses:
  - Programming for Everybody
  - Codecademy: Python Track
- Books:
  - Learn Python the Hard Way
  - How to Think Like a Computer Scientist

GIS Focused

- Python Scripting for ArcGIS
- ArcPy and ArcGIS - Geospatial Analysis with Python
- Python Developers GeoNet Community
- GIS Stackexchange

Scientific

Courses:

- Python Scientific Lecture Notes
- High Performance Scientific Computing
- Coding the Matrix: Linear Algebra through Computer Science Applications
- The Data Scientist’s Toolbox

Scientific

Books:

- Free:
  - Probabilistic Programming & Bayesian Methods for Hackers
    - very compelling book on Bayesian methods in Python, uses SciPy + PyMC.
  - Kalman and Bayesian Filters in Python

Scientific

- Paid:
  - Coding the Matrix
    - How to use linear algebra and Python to solve amazing problems.
  - Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython
    - The cannonical book on Pandas and analysis.

Packages

Only require SciPy Stack:

- Scikit-learn:
  - Lecture material
  - Includes SVMs, can use those for image processing among other things…
- FilterPy, Kalman filtering and optimal estimation:
  - FilterPy on GitHub
- An extensive list of machine learning packages

Code

- ArcPy + SciPy on Github
- raster-functions
  - An open source collection of function chains to show how to do complex things using NumPy + scipy on the fly for visualization purposes
- statistics library with a handful of descriptive statistics included in Python 3.4+.
- *TIP*: Want a codebase that runs in Python 2 and 3? Check out future, which helps maintain a single codebase that supports both. Includes the `futurize` script to initially a project written for one version.

Scientific ArcGIS Extensions

- PySAL ArcGIS Toolbox
- Movement Ecology Tools for ArcGIS (ArcMET)
- Marine Geospatial Ecology Tools (MGET)
  - Combines Python, R, and MATLAB to solve a wide variety of problems
- SDMToolbox
  - species distribution & maximum entropy models
- Benthic Terrain Modeler
- Geospatial Modeling Environment

Conferences

- PyCon
  - The largest gathering of Pythonistas in the world
- SciPy
  - A meeting of Scientific Python users from all walks
- GeoPython
  - The Python event for Python and Geo enthusiasts
- PyVideo
  - Talks from Python conferences around the world available freely online.
  - PyVideo GIS talks

Closing

Thanks

- Geoprocessing Team
- ArcGIS API for Python Team
- The many amazing contributors to the projects demonstrated here.
  - Get involved! All are on GitHub and happily accept contributions.

esri®

THE SCIENCE OF WHERE™