Scientific Toolkit

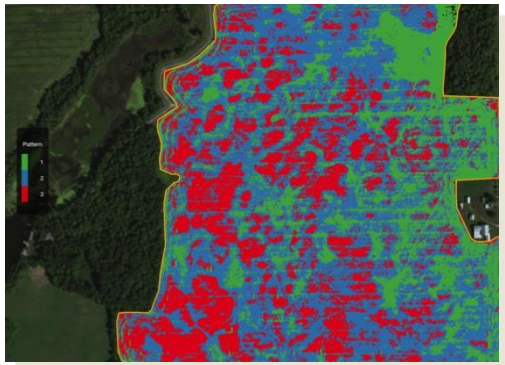
Brett Rose, PhD

brose@esri.com

Science begins with observations

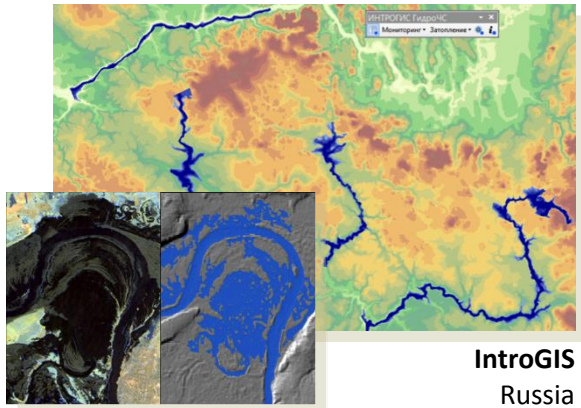
We use science everyday

Crop Health



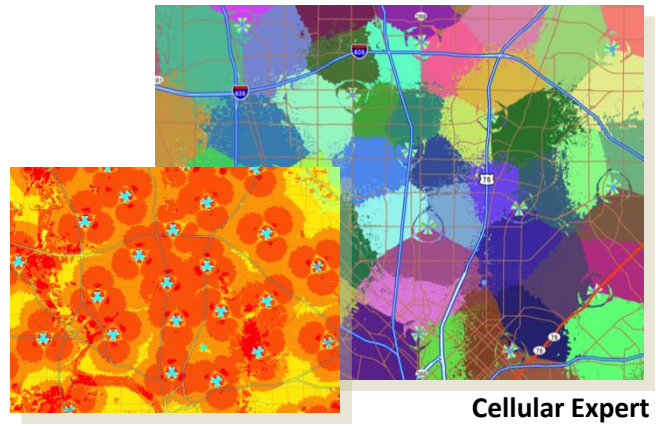
GeoSilos
Indiana

Hydro Analysis



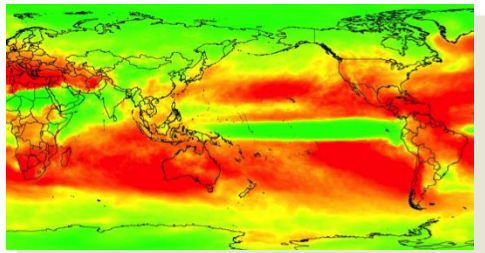
IntroGIS
Russia

Cellular Coverage Analysis



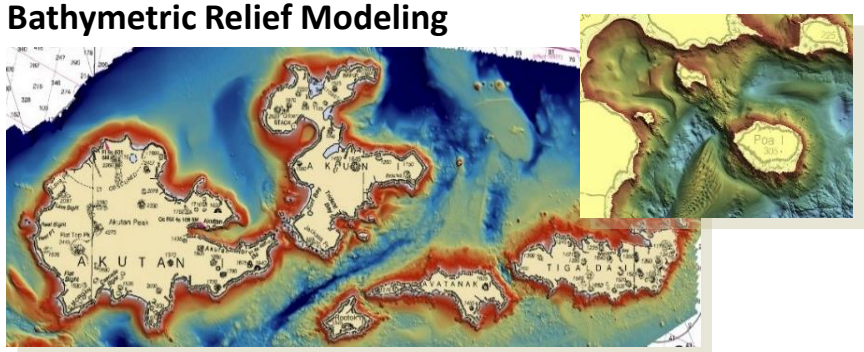
Cellular Expert
Texas

Global Rainfall



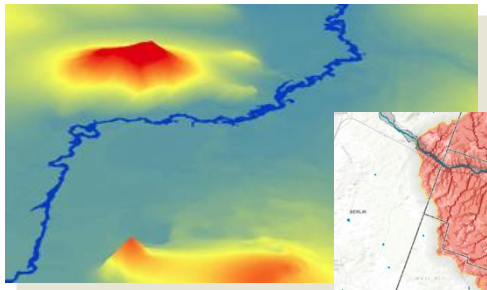
CLIMsystems
Global

Bathymetric Relief Modeling



Fugro Palegos
Alaska

Hydrographic Modeling

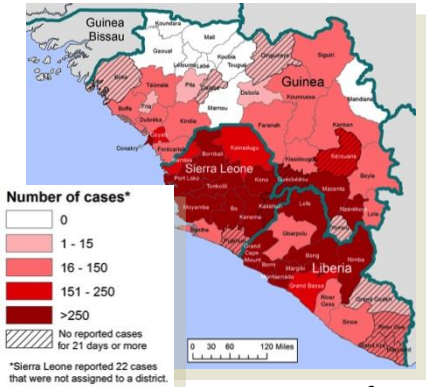


IntroGIS
Russia

Water Runoff Modeling

Stone Environmental
Vermont

Ebola Disease



West Africa

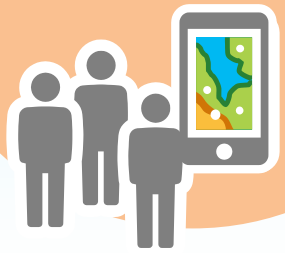
Scientific Process

Engaging EOS

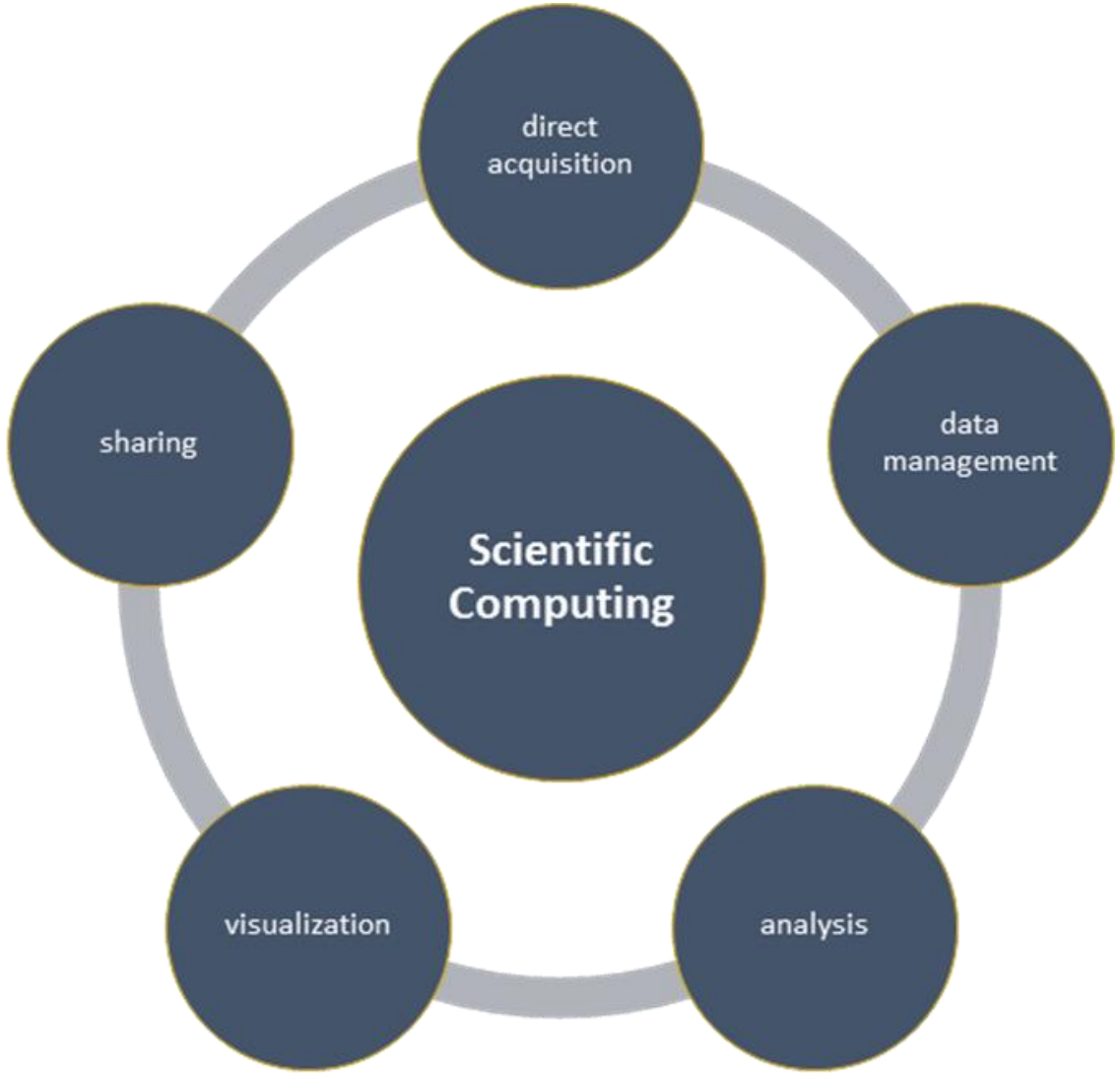
Science Toolkit

Communicating Science

Optimizing Field Collection



Scientific Computing



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.

To do science we need to

Understand where things are

Measure size, shape and distribution

Determine how places are related

Find best location and paths

Detect and quantify patterns

Make predictions

We can use Spatial Analysis

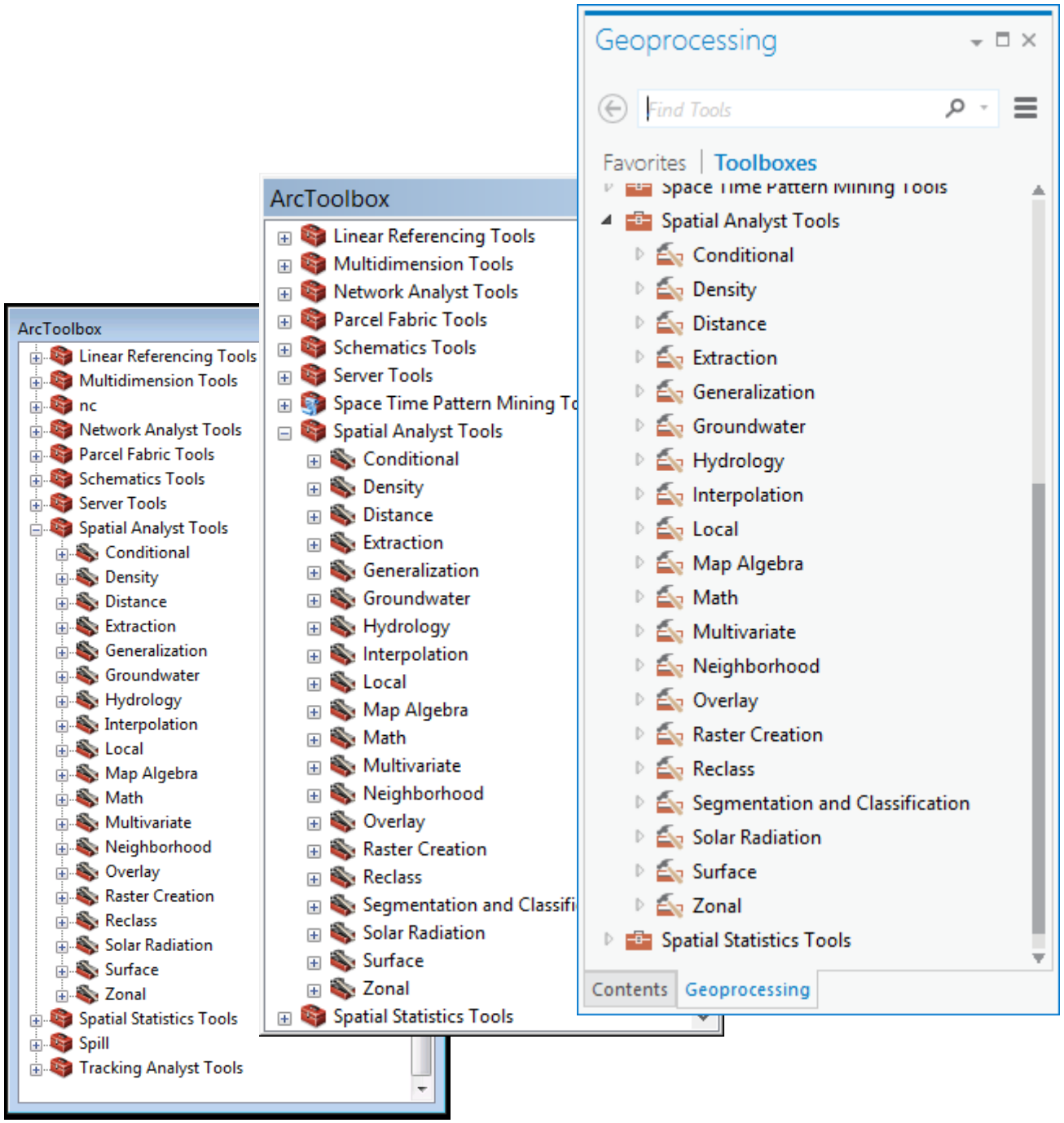
With spatial analytics we

map → to see possible patterns

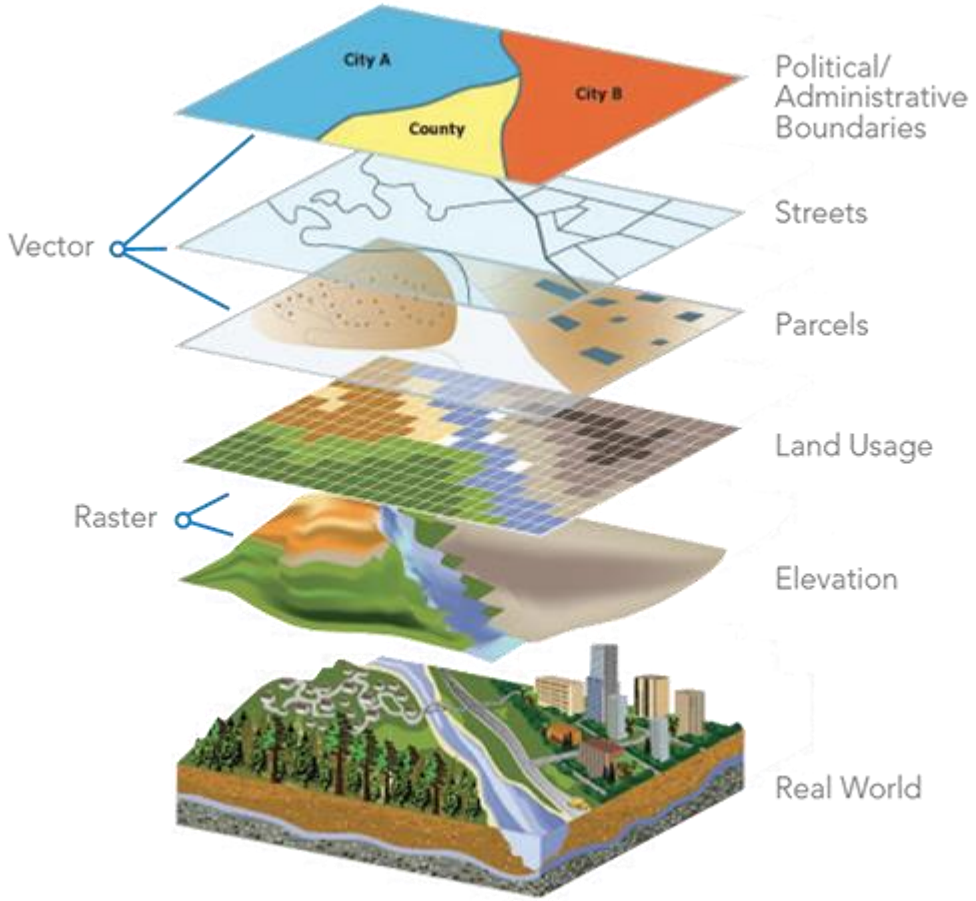
describe → to improve understanding

measure → to minimize subjectivity

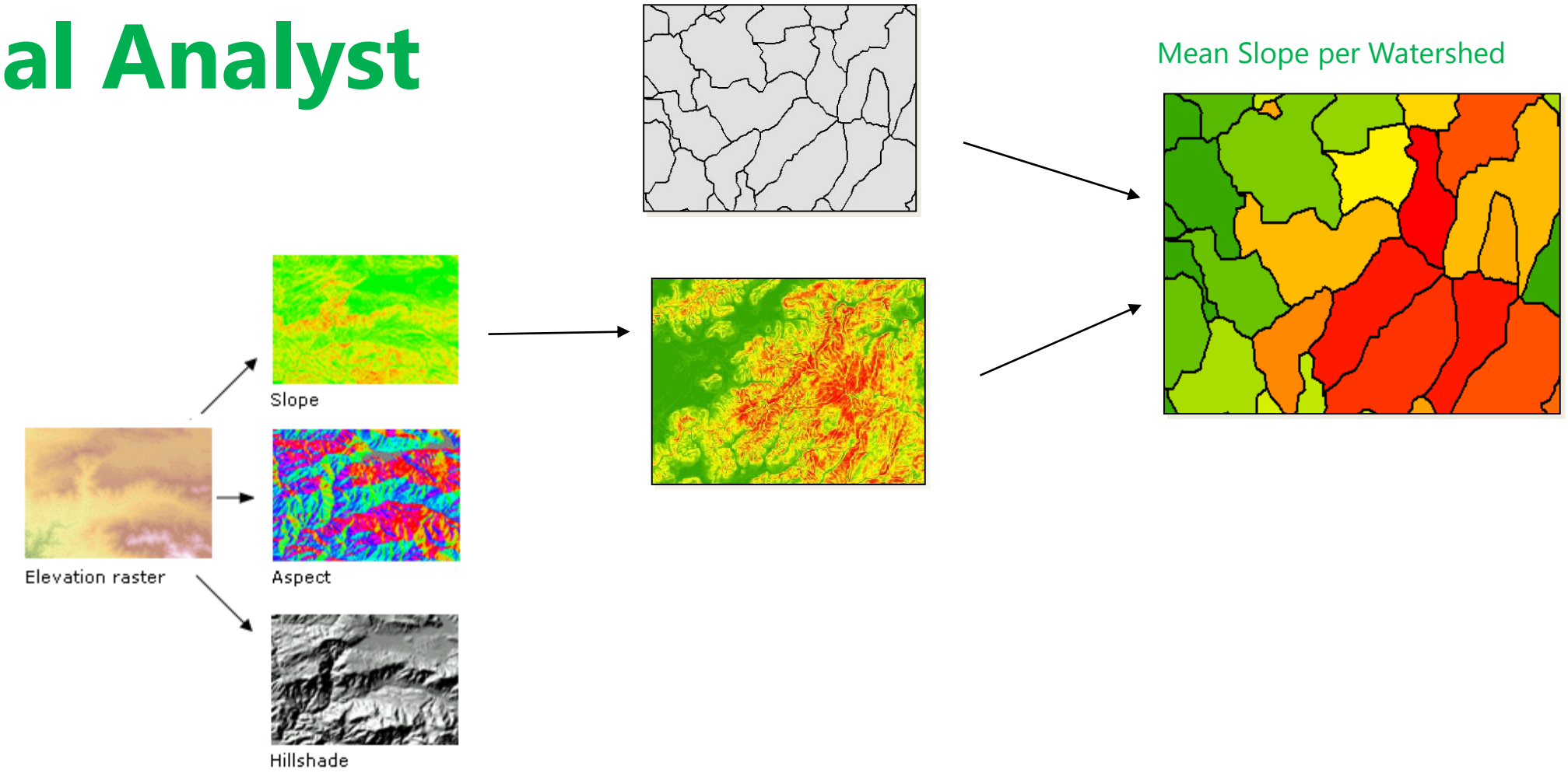
Tools in ArcGIS



Spatial data and spatial analysis



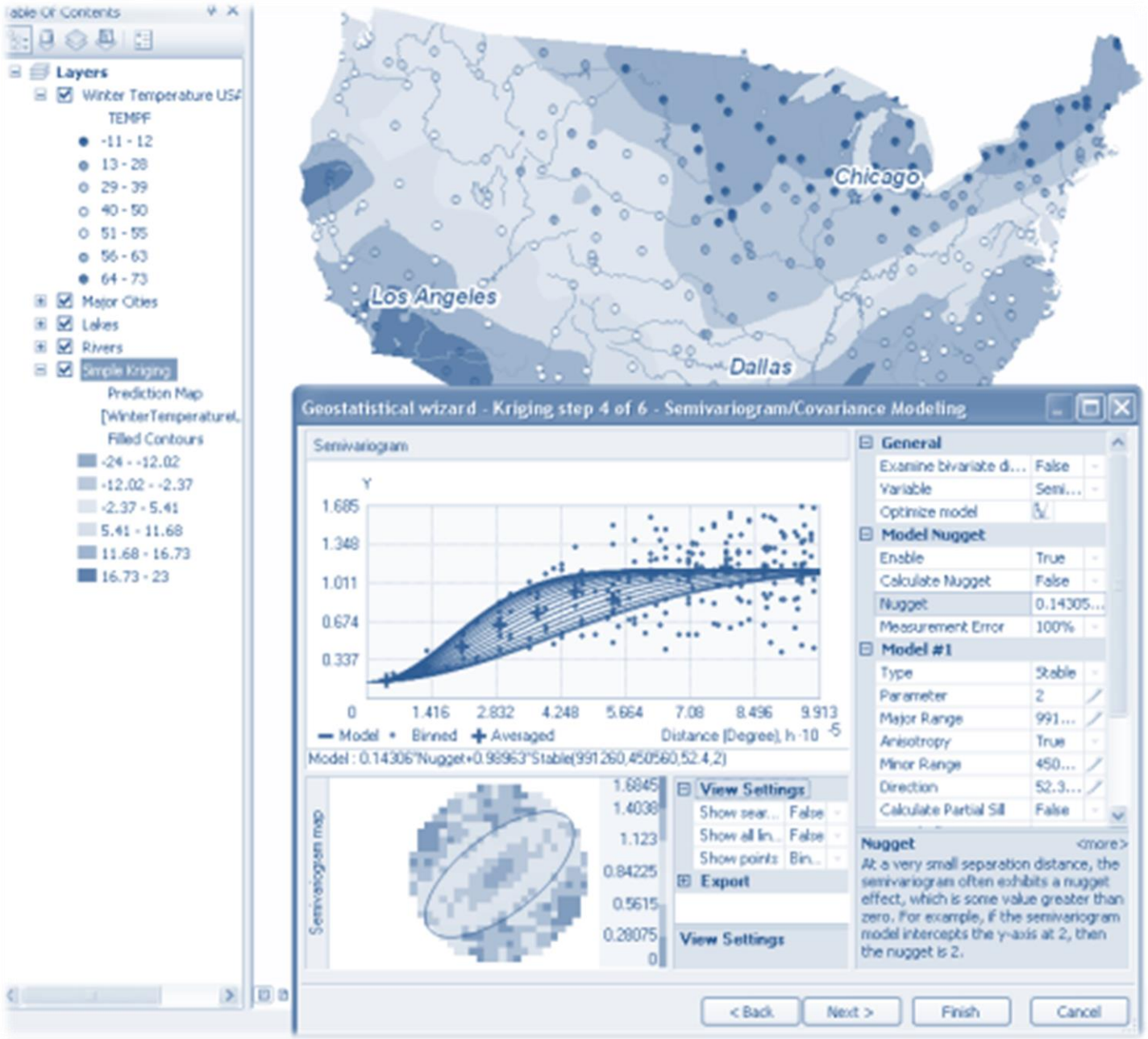
Spatial Analyst

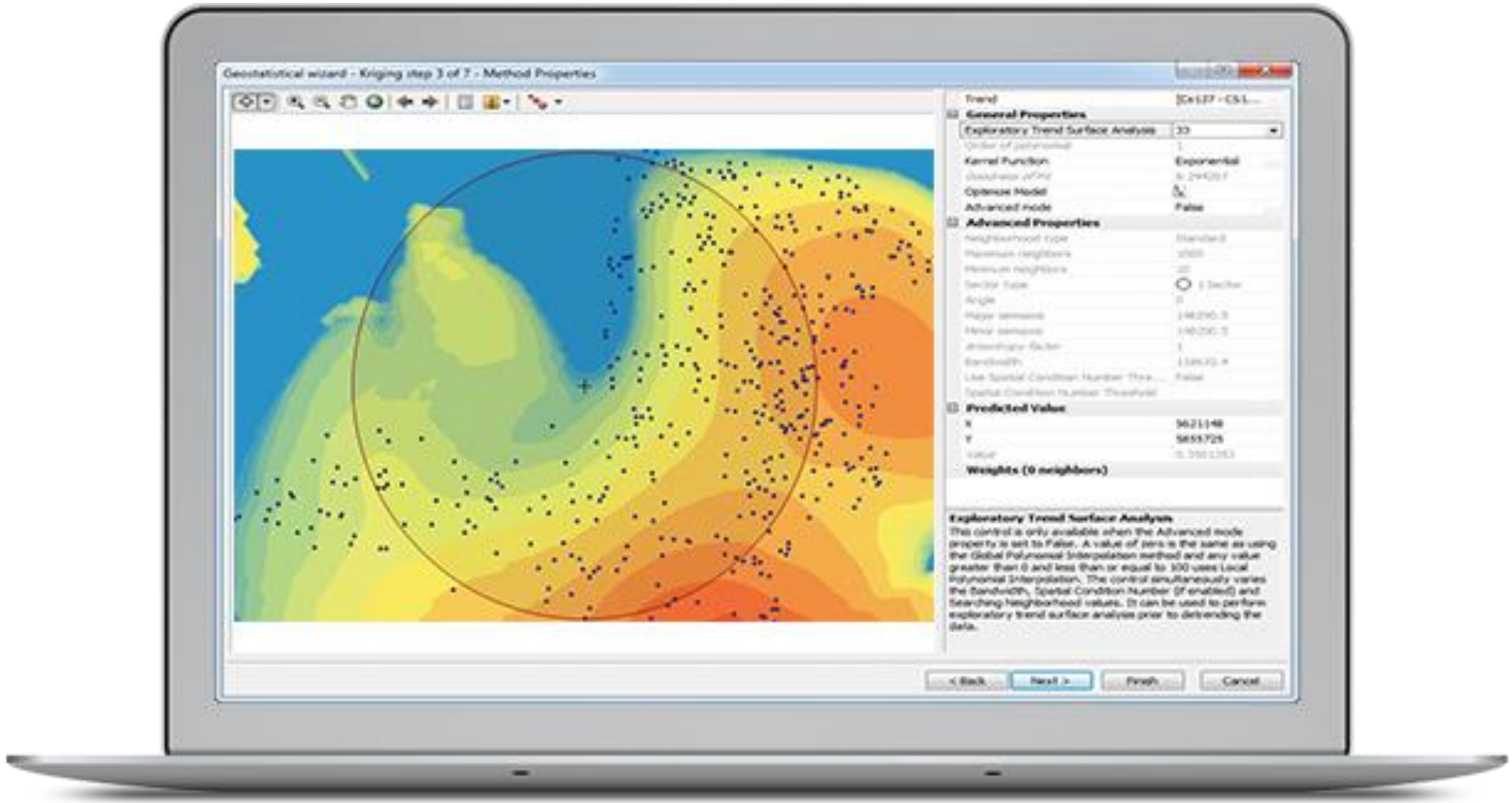


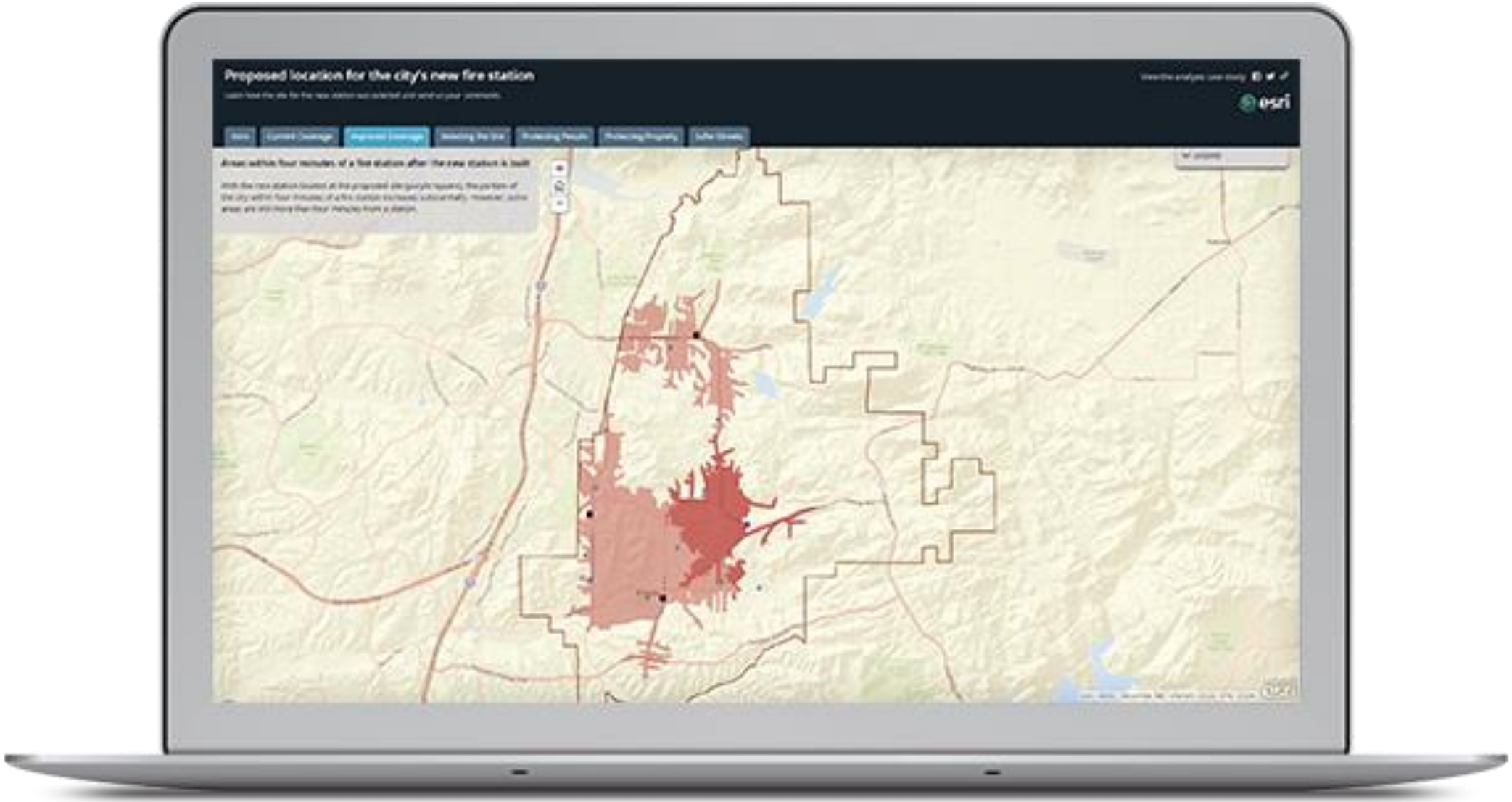

```
# This is map algebra
outRas = Raster("inraster1") + Raster("inraster2")
```


An example from Learn.arcgis.com
The Cougar Corridors of Southern California
The proposed wildlife corridors
Based on all the combined and weighted statistical data from the four layers, the resulting wildlife corridors are seen here. While not an exact trail per se, these paths represent the most likely routes of mating mountain lions attempting to travel between the core areas. As such, they represent the areas where attempts at cohesive, contiguous land preservation should be focused.
As we can see from this map, these corridors pass through some pretty densely developed areas. The reality of making these corridors permanently protected is going to be a huge challenge, and the survival of mountain lions in Southern California is by no means guaranteed, even with reliably protected corridors. But the alternative to not trying would be a Southern California with no more wild mountain lions.
The analysis method and the model outlined in this map journal, while simplified for clarity, is an excellent starting point for those seeking to carry out similar analyses in their own regions of the world. For a detailed step-by-step tutorial in which this analysis is conducted by you with the geoprocessing tools in ArcGIS for Desktop, visit the project page at Learn.ArcGIS.com.

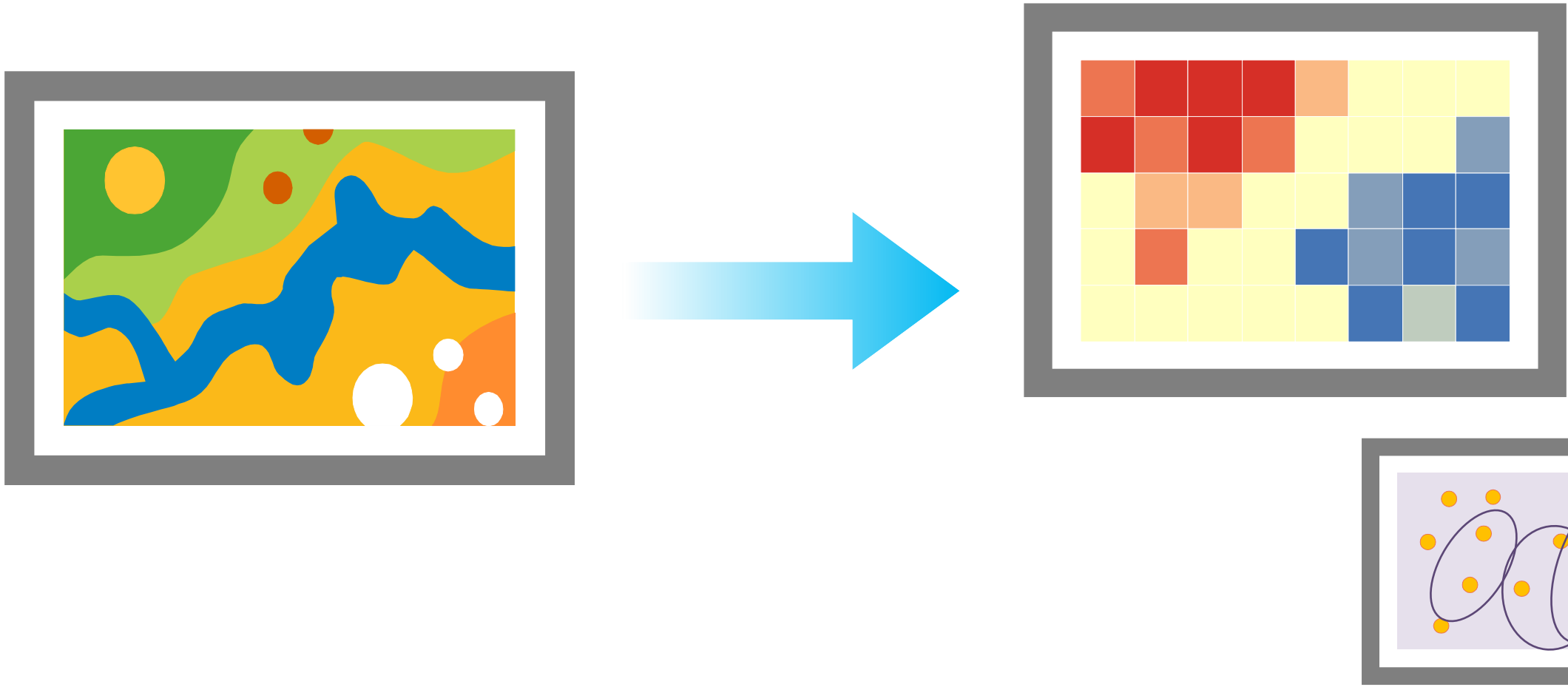
Geostatistics




Geostatistical wizard - Kriging step 3 of 7 - Method Properties
Trend
General Properties
Exploratory Trend Surface Analysis
33
Kernel Function
Exponential
Optimize Model
Advanced mode
False
Advanced Properties
Predicted Value
X
9621148
Y
5855725
Weights (0 neighbors)
Exploratory Trend Surface Analysis
This control is only available when the Advanced mode property is set to False. A value of zero is the same as using the Global Polynomial Interpolation method and any value greater than 0 and less than or equal to 100 uses Local Polynomial Interpolation. The control simultaneously varies the Bandwidth, Spatial Condition Number (if enabled) and Searching Neighborhood values. It can be used to perform exploratory trend surface analysis prior to detrending the data.
< Back
Next >
Finish
Cancel


Proposed location for the city's new fire station
esri

Spatial Statistics



Data Science

and Spatial Data Science

“A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician.”

— Josh Wills

Data Science

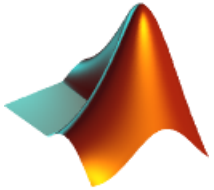
- A much-hyped phrase, but effectively is about the application of statistics and machine learning to real-world data, and developing formalized tools instead of one-off analyses. Combines diverse fields to solve problems.
- Us geographic folks also rely on knowledge from multiple domains.
- We know that spatial is more than just an x and y column in a table, and how to get value out of this data.

Data Science Languages

Languages commonly used in data science:


R


Python


Matlab

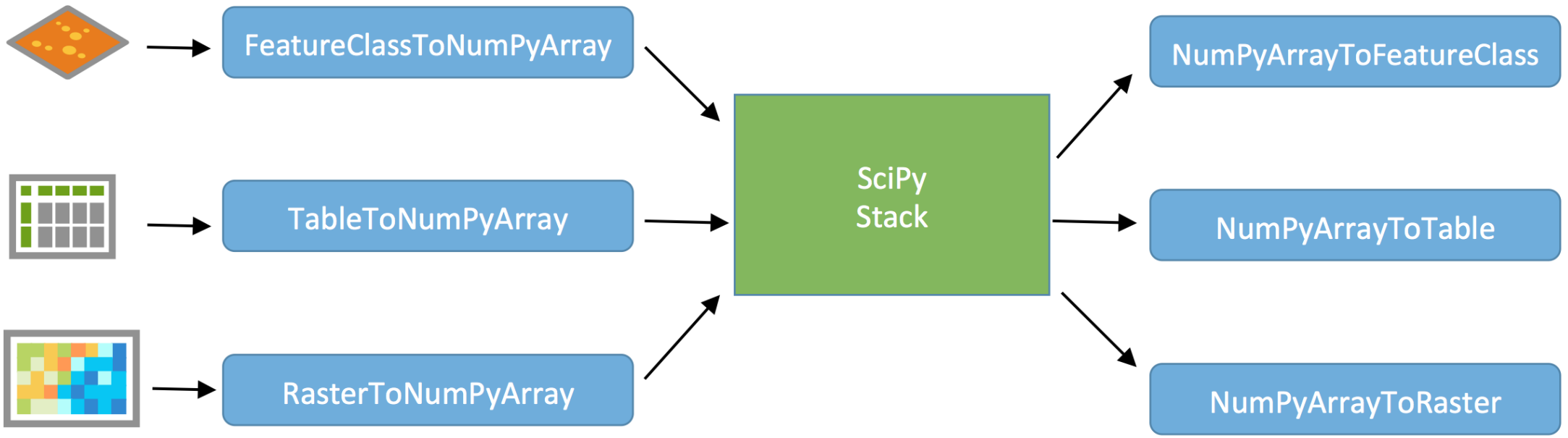

Julia

[R vs Python for Data Science]

python

- Spatial Statistics Tools
  - Analyzing Patterns
    - Average Nearest Neighbor
    - High/Low Clustering (Getis-Ord General G)
    - Incremental Spatial Autocorrelation
    - Multi-Distance Spatial Cluster Analysis (Ripleys K Function)
    - Spatial Autocorrelation (Morans I)
  - Mapping Clusters
    - Cluster and Outlier Analysis (Anselin Local Morans I)
    - Grouping Analysis
    - Hot Spot Analysis (Getis-Ord Gi*)
    - Optimized Hot Spot Analysis
    - Similarity Search
  - Measuring Geographic Distributions
    - Central Feature
    - Directional Distribution (Standard Deviational Ellipse)
    - Linear Directional Mean
    - Mean Center
    - Median Center
    - Standard Distance
  - Modeling Spatial Relationships
    - Exploratory Regression
    - Generate Network Spatial Weights
    - Generate Spatial Weights Matrix
    - Geographically Weighted Regression
    - Ordinary Least Squares

ArcGIS + SciPy



We use SciPy because

- 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

SciPy Stack





IP[y]:
IPython

nose

NumPy for numerical computation using arrays

SciPy a collection of numerical algorithms

Matplotlib for 2D and 3D plotting

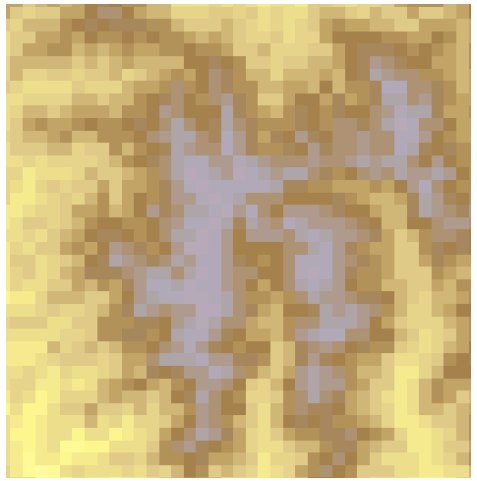
Pandas for high-performance data structures

SymPy for symbolic mathematics and computer algebra

IPython providing an interactive interface for quickly testing scripts and

nose which allows you to test your Python code.

Spatial Analyst

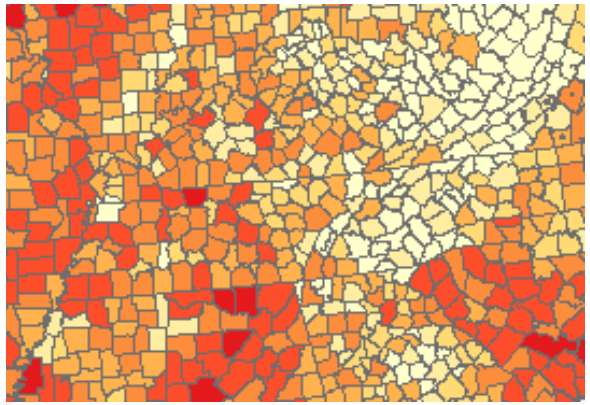


python™

NumPy

Spatial & Geo Stats

Data Access Module



python™

NumPy

Spatial Statistics Data Object and Utilities

matplotlib

NetCDF4

- Fast, HDF5 and NetCDF4 read+write support, OPeNDAP
- Heirarchical data structures
- Widely used in meterology, oceanography, climate communities
- Easier: Multidimensional Toolbox, but can be useful

```
import netCDF4
nc = netCDF4.Dataset('test.nc', 'r', format='NETCDF4')
print nc.file_format
# outputs: NETCDF4
nc.close()
```

Multi-Dimensional data

- Multidimensional formats: HDF, GRIB, NetCDF
- Access via OPeNDAP, vector renderer, Raster Function Chaining
- An example which combines mutli-D with time
- Multi-D supported as WMS, and in Mosaic datasets (10.2.1+)

Other integration

pysal

- Open Source Python Library for Spatial Analytical Functions
- ASU GeoDa Center for Geospatial Analysis and Computation
- Luc Anselin
  - PySpace (GeoDaSpace)
- Sergio Rey
  - STARS

BSD License

Why R?

R

- Powerful core data structures and operations
  - Data frames, functional programming
- Unparalleled breadth of statistical routines
  - The de facto language of Statisticians
- CRAN: 6400 packages for solving problems
- Versatile and powerful plotting

R

- Contains “cutting edge” data analysis techniques from a wide body of academic and applied fields
- Extendible
- Indirectly compatible
  - Direct via RPy/RPy2 and win32com
- GNU
- Revolution
- esri continues to focus on improving the interaction in the future


R — ArcGIS Bridge

links

pysal

https://geodacenter.asu.edu/pysal

https://github.com/pysal

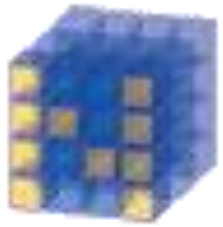
SciPy and **NumPy**

http://www.scipy.org/

R

http://www.r-project.org

IP[y]:
IPython



Get involved

http://esri.github.io/

GeoNet

Blogs

GeoDev Meetups

Ideas Portal

User Conference

Devsummit

Try for yourself

http://esri.github.io/

http://esriurl.com/scicomm

http://developers.arcgis.com

https://github.com/R-ArcGIS

Other

Demos