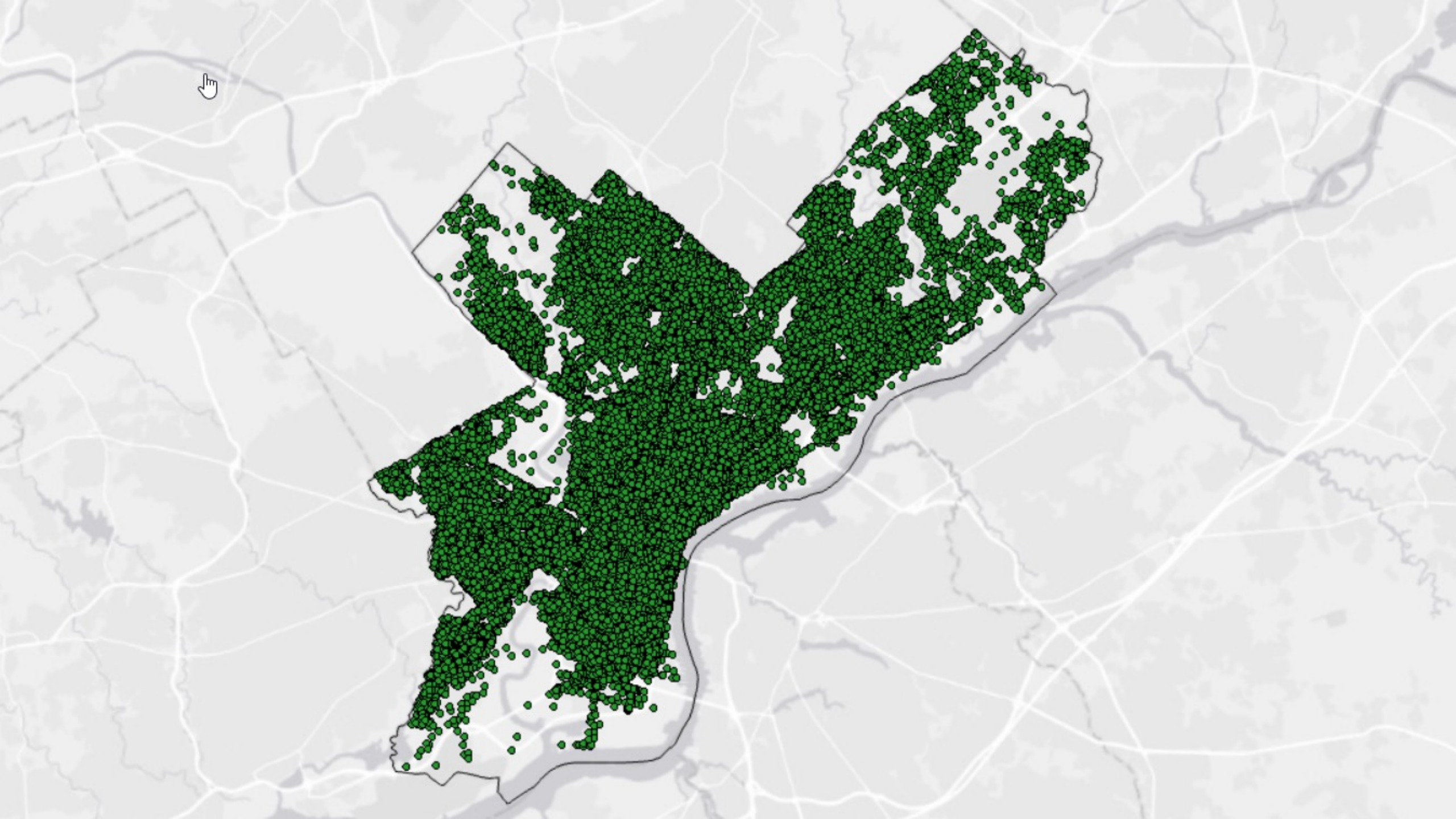
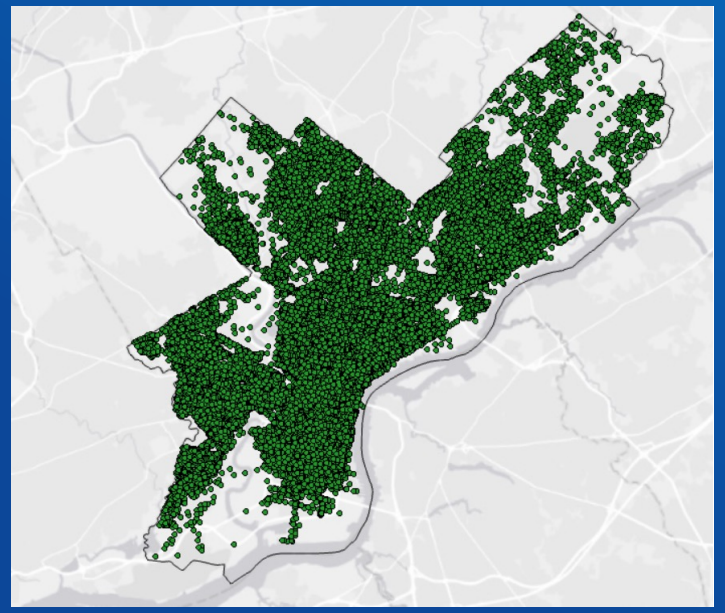
Spatial Data Science Using R and Python

Josiah Parry, Nick Giner

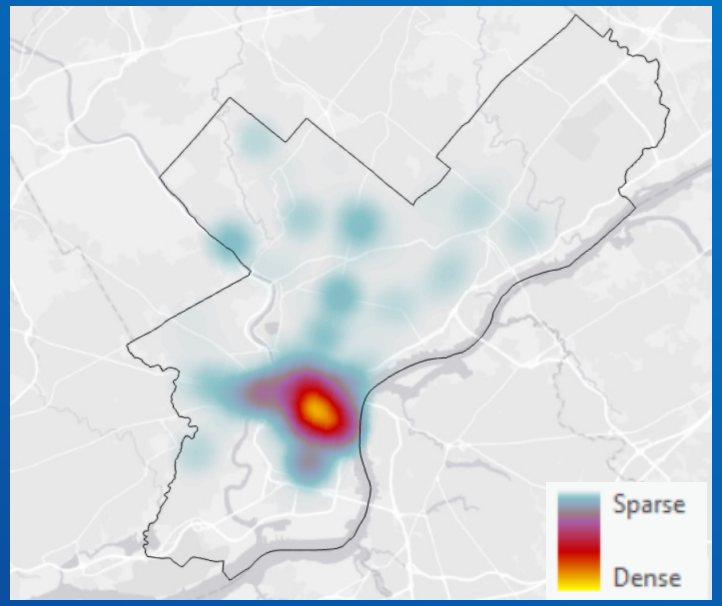
October 21, 2024



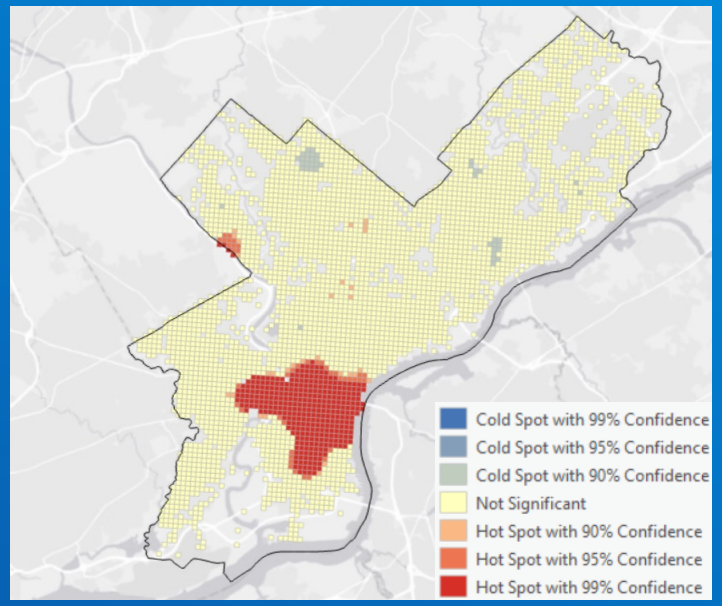
What is data science?



Raw data



Initial understanding



Actionable information

What is *spatial* data science?

area

coincidence

adjacency

distance

orientation

proximity

length

direction

Building blocks of spatial data science

Data Integration & Engineering

Visualization & Exploration

Spatial Analysis

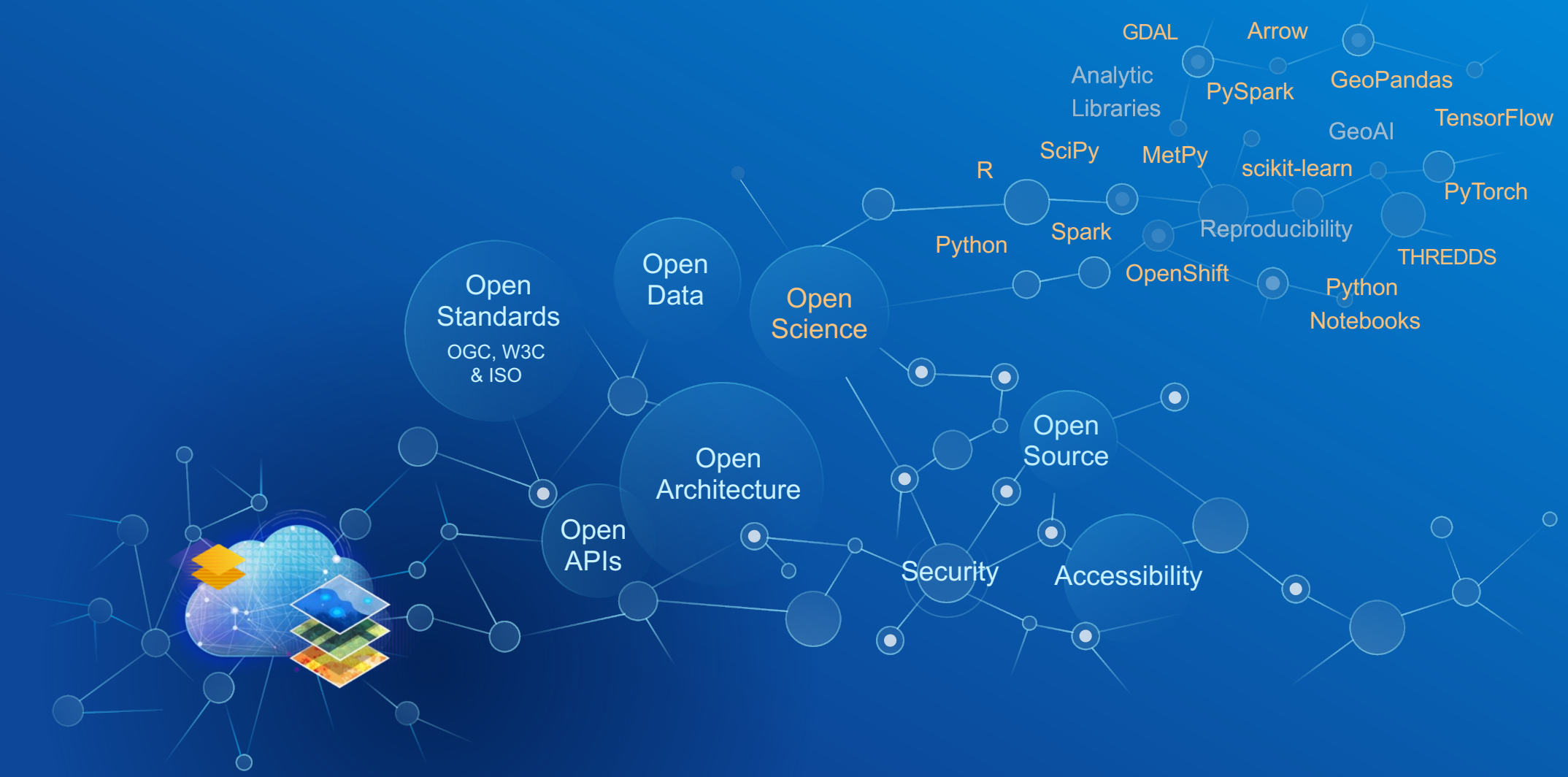
GeoAI

Big Data Analytics

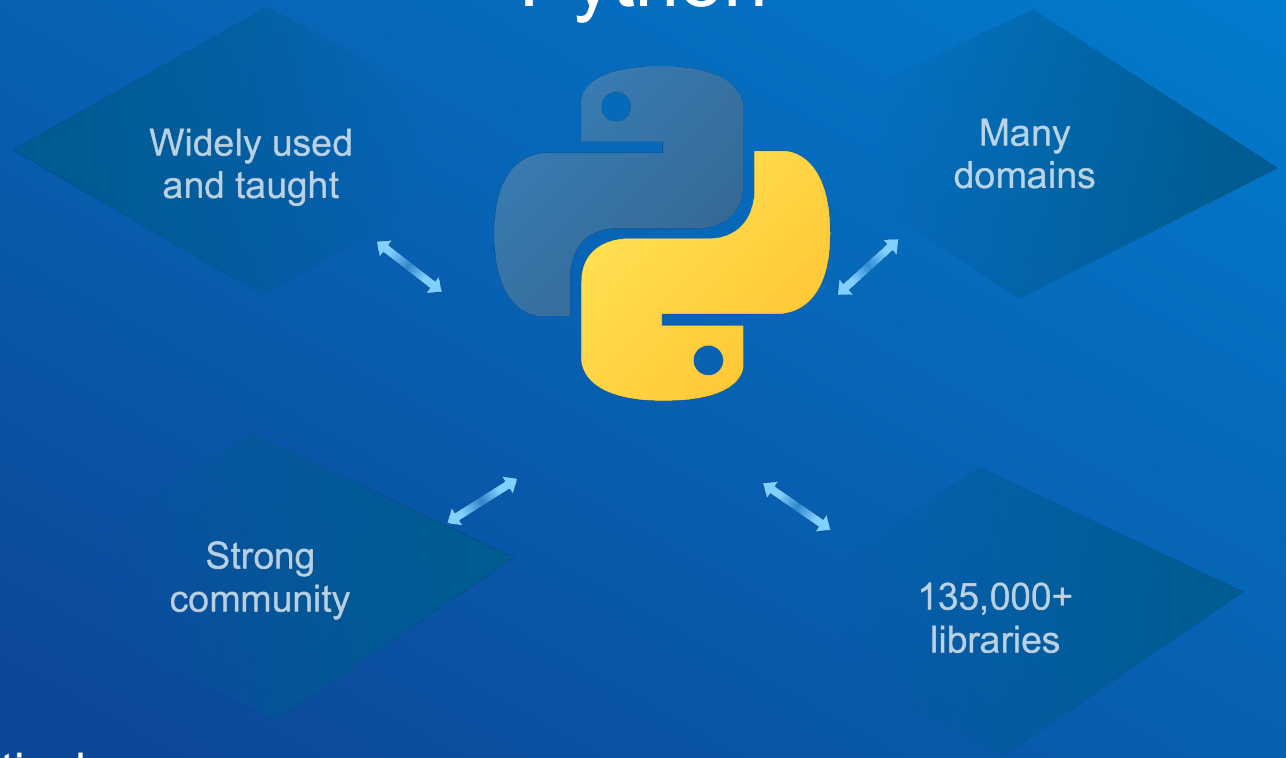
Modeling & Scripting

Sharing & Collaboration

ArcGIS supports open data science integration



Python



Software & web development

Task automation

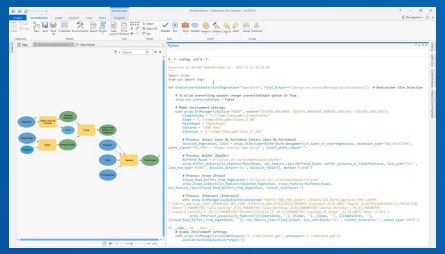
Scientific, mathematical, AI/ML, & statistical analysis

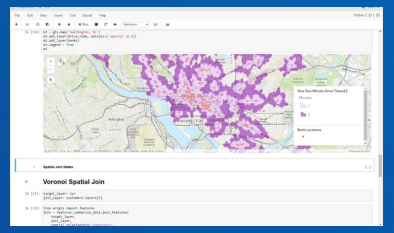
Geospatial data analysis

ArcGIS Python Libraries

ArcPy

- Scripting geoprocessing
- 2000+ functions
- Data management, spatial analysis, mapping





ArcGIS API for Python

- Web GIS, services
- Administration, manage content, data science & GeoAI
- Deploy anywhere
- Spatially-Enabled DataFrame

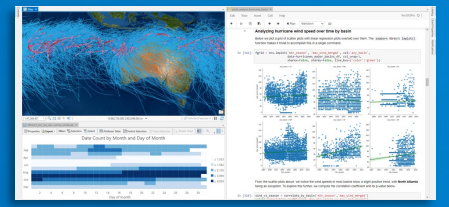
Open-source integration

pandas, jupyter, SciPy, Keras, scikit learn, fast.ai, matplotlib, CONDA, dmlc XGBoost, TensorFlow, PYTORCH, OpenCV

Experiences

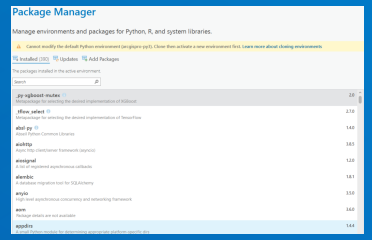
ArcGIS Notebooks

- Enterprise, Online, Pro
- Python environments
  - 500+ libraries



ArcGIS Pro

- Notebooks & Python window
- Python environments
  - 200+ libraries + AI

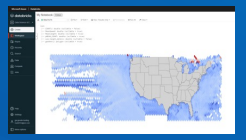


IDEs

- VS code, PyCharm, Spyder, etc.

ArcGIS GeoAnalytics Engine

- PySpark geospatial functions



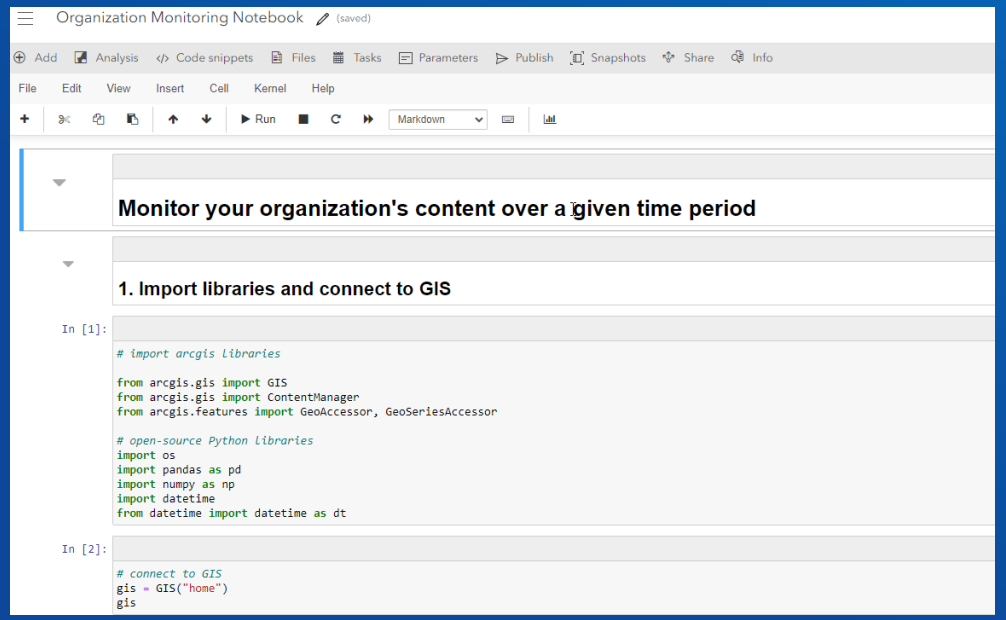
Apps

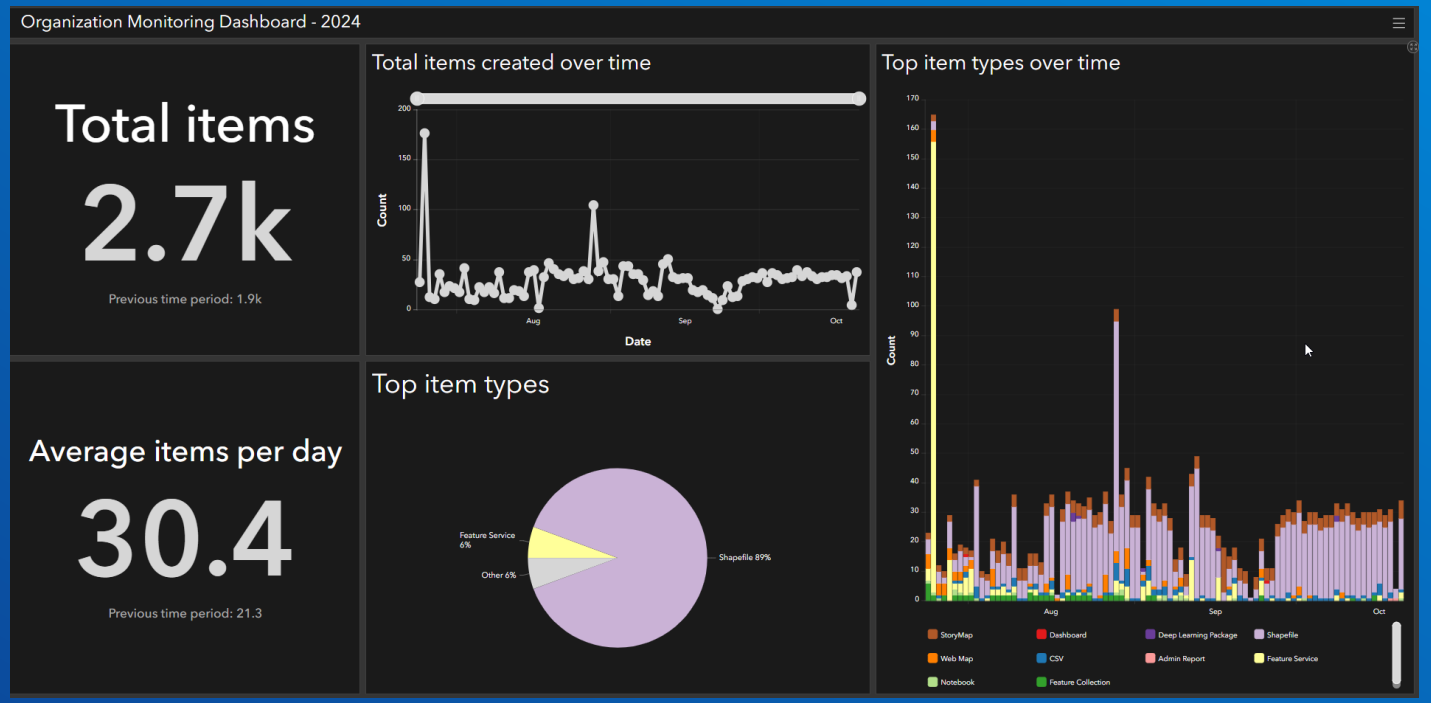
- Web tools
- GP Services

The scripting language of ArcGIS…

Python Demo

Automatically update ArcGIS Dashboard using ArcGIS API for Python and ArcGIS Notebooks





Python Hands-on Exercise (20 min)

esri Developer Documentation Features SDKs and APIs Products Support Sign in

ArcGIS API for Python / Samples Home Samples API Reference

Search topics

- Power users / Developers
- Org Administrators
- GIS analysts and data scientists
  - Chennai Floods 2015 - A Geographic Analysis
  - Analyzing violent crime
  - Finding hospitals closest to an incident
  - Which college district has the fewest low-income families?
  - Designate Bike Routes for Commuting Professionals
  - Locating a new retirement community
  - Finding grazing allotments
  - Analyzing growth factors of rental properties in New York City
  - How green is New Delhi?
  - Safe Streets to Schools
  - Which areas are good cougar habitats?
  - Data Visualization - Construction permits near Washington DC, part 1/2
  - Data Summarization - Construction permits near Washington DC, part 2/2

Which college district has the fewest low-income families?

A pilot program was run by a local cable operator in the county to provide low-cost computers and Internet access to low-income families with kids in high school. This showed a marked improvement in school performance for these kids, and the program has brought the company a fair amount of positive publicity and goodwill in the community.

Company officials now want to set up a similar program for community college students. The company provides Internet access to the five community college districts in the county, and officials are aware that the colleges are under a lot of pressure - they are facing funding cuts at the same time as increased demand for enrollment. To try to improve the situation the colleges are turning more and more to distance learning, primarily via the Internet. By providing computers and Internet access, the cable company can enable more low-income students to take advantage of online classes.

This case study uses ArcGIS API for Python to find districts that have the fewest low income families in order to empower these students.

We will use `summarize_within` tool to get the number of low-income families within each community district. We will also visualize this using the map widget.

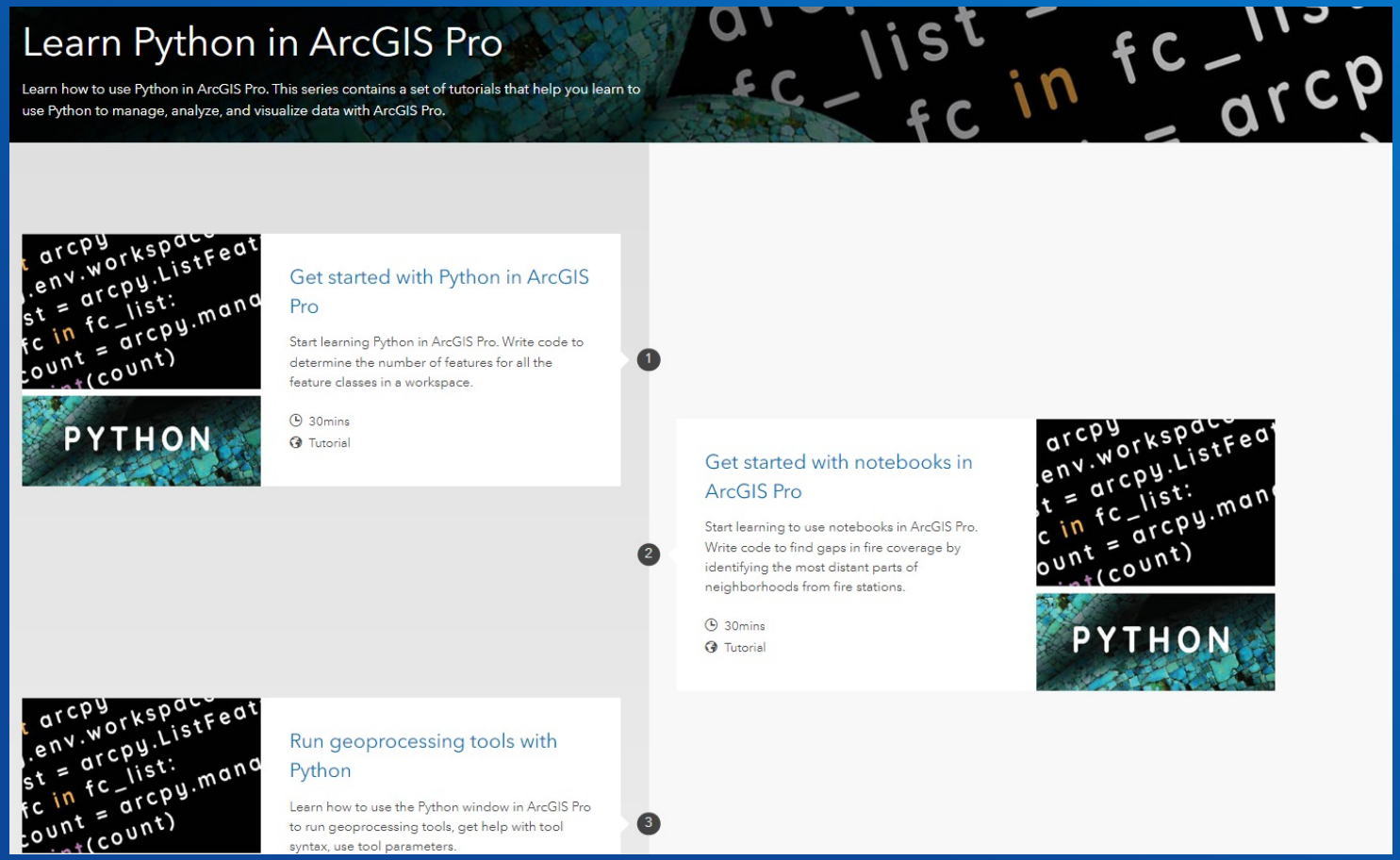
On this page

- Get data for analysis
  - Find the community college district with the fewest low income families
- Get the number of low-income households in each district
  - Visualization to show district with fewest households
- Conclusion

Try it live

Edit on GitHub

Python Resources



ArcGIS API for Python

ArcGIS Maps SDK (JS)

SDK for R??

R-ArcGIS Bridge

+

ArcGIS

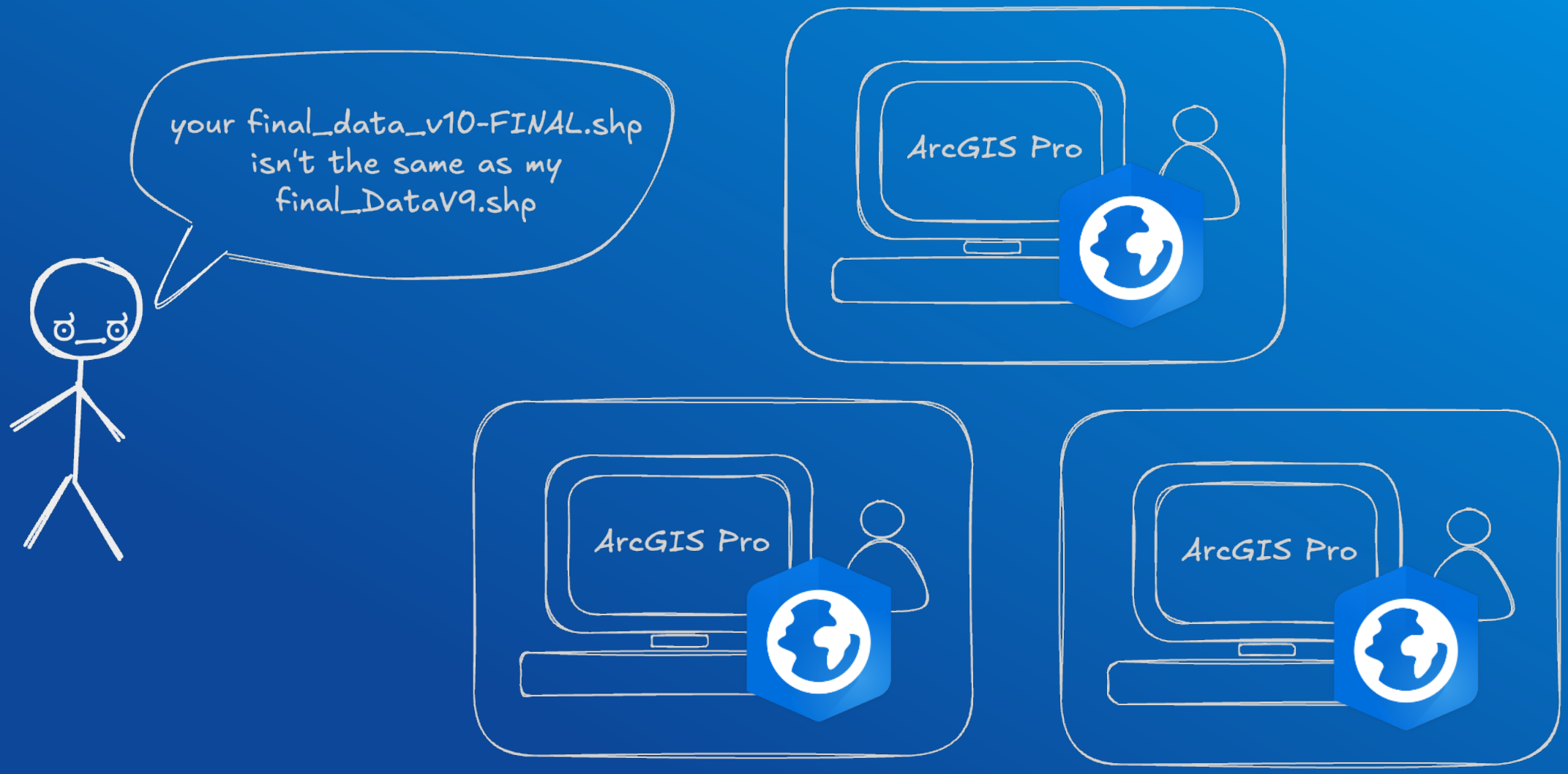
How I think about ArcGIS

- ArcGIS is a **System**
- **NOT** just ArcGIS Pro
- Two buckets:
  - “on-prem”
  - Cloud

“On-prem” GIS

- On premises (you can touch it)
- ArcGIS Pro
- Shapefile
- File Geodatabase
- *Sharing is tough*

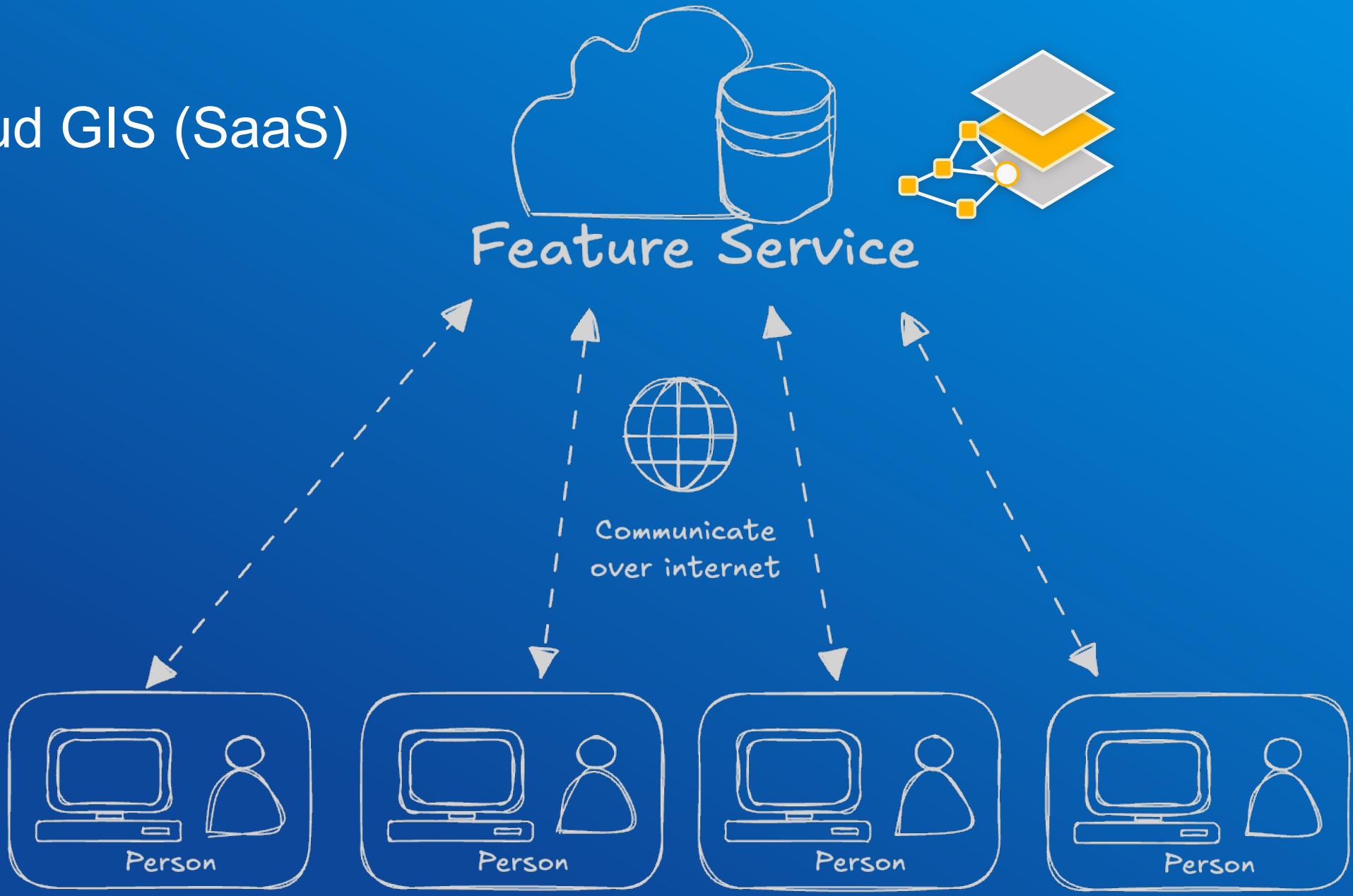
“On-prem” GIS


your final_data_v10-FINAL.shp isn't the same as my final_DataV9.shp
ArcGIS Pro
ArcGIS Pro
ArcGIS Pro

Cloud GIS (SaaS)

- Someone else owns the servers
- Accessed over the internet
- ArcGIS Online
- Survey123
- ArcGIS Dashboards
- Data shared as a **Feature Service**

Cloud GIS (SaaS)



Feature services are
everywhere

Feature Services power everything

Introduction

R-ArcGIS Bridge

{arcgisbinding}

- Connects ArcGIS Pro and R
- Data transfer between Pro and R
- R-driven Geoprocessing tools

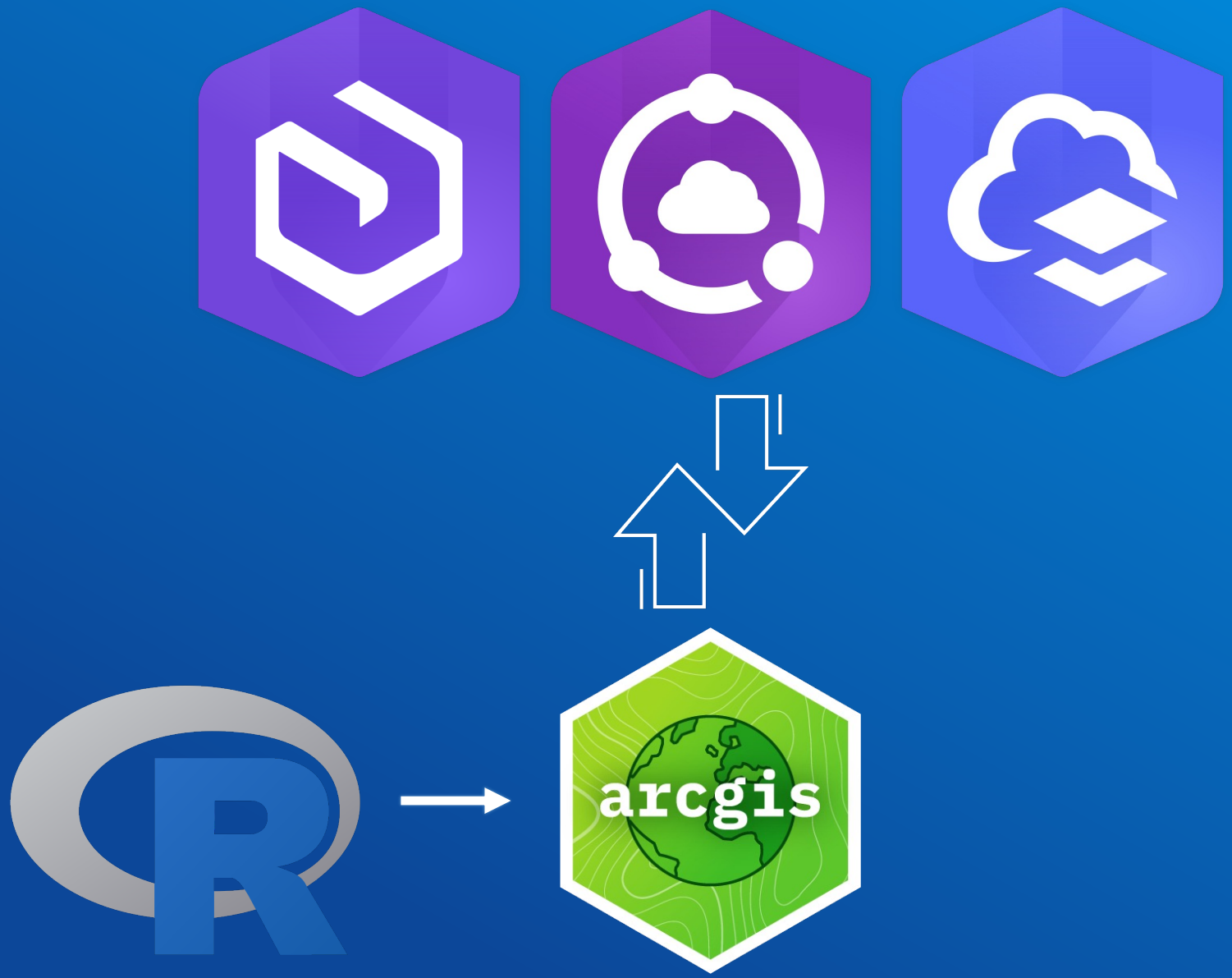
Brings R to GIS analysts in Pro

NEW!

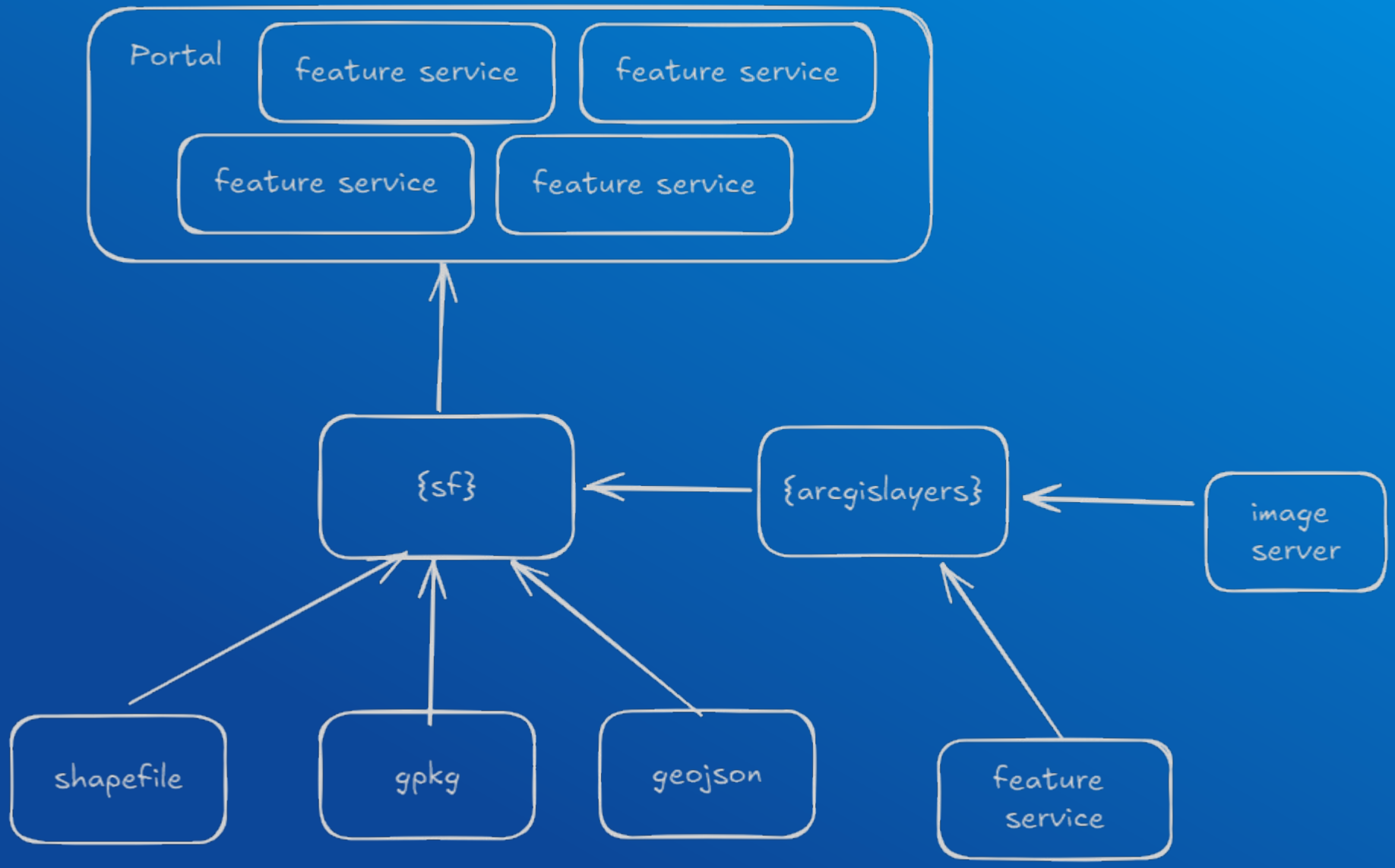
{arcgis}

- Access to ArcGIS Location Services in R
  (Hosted data I/O, Geocoding, Routing, Enrichment, etc.)
- R data scientists with ArcGIS Online, Enterprise, or Platform accounts (e.g. existing subscriptions)
- Free and open-source

Brings ArcGIS to R data scientists where they work


arcgis

The “big” picture



R & Feature Services

R & Survey123

Your turn! (10 min)

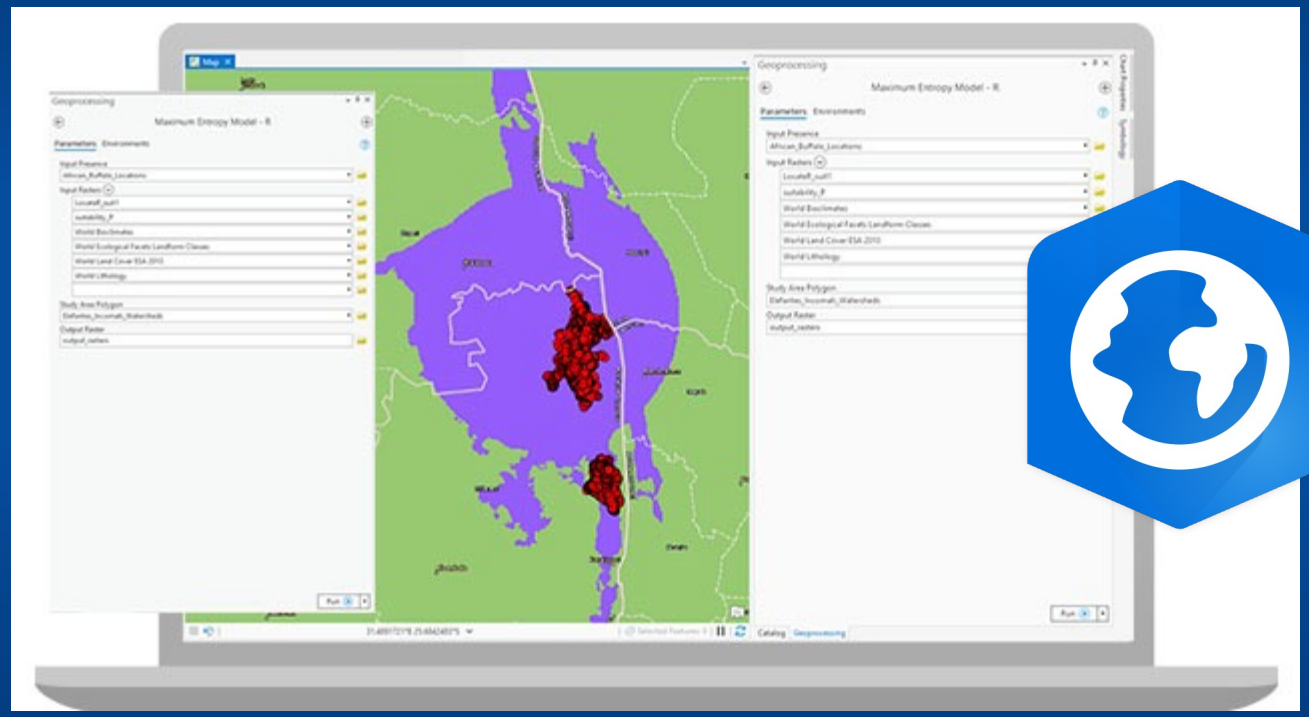
(Very fast) Geocoding

Performance Comparison

Method	Duration
`R-ArcGIS` (Batched addresses)	*3.5s*
`R-ArcGIS` (Single Address)	1m
`tidygeocoder` (Single Address)	17m
`arcgeocoder` (Single Address)	21m

n = 2,664 addresses in Boston, MA

Your turn (7-10min)



{arcgisbinding}

Workflow

Automate as R script

As Geoprocessing Tool in ArcGIS Pro

Deploy as Geoprocessing Service in ArcGIS Enterprise

Integrate in WebApp