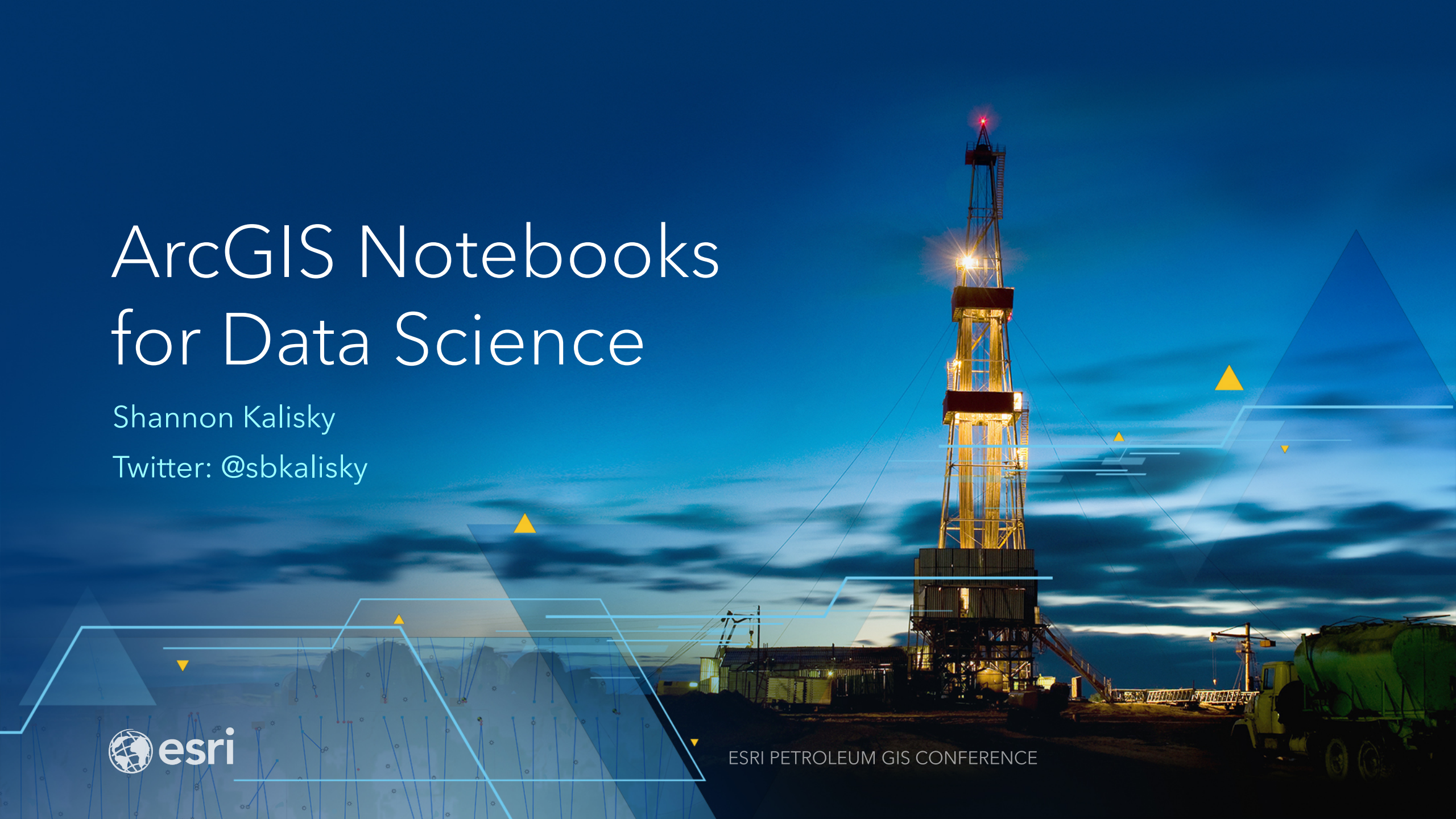
ArcGIS Notebooks for Data Science

Shannon Kalisky

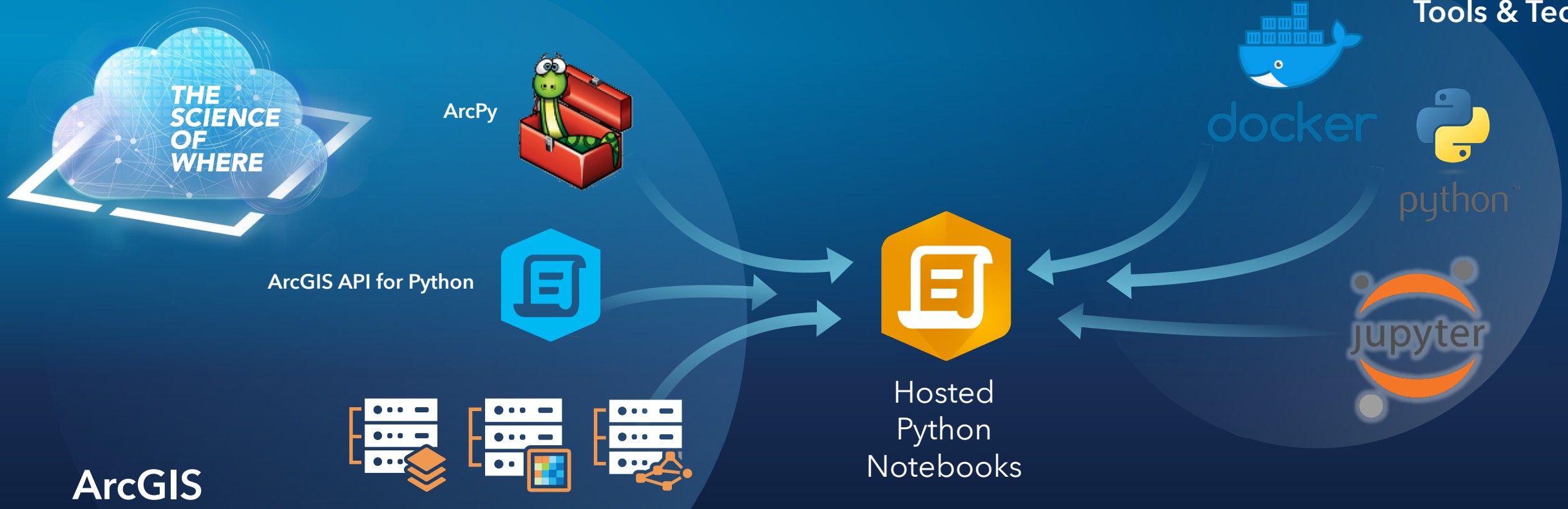
Twitter: @sbkalisky



esri

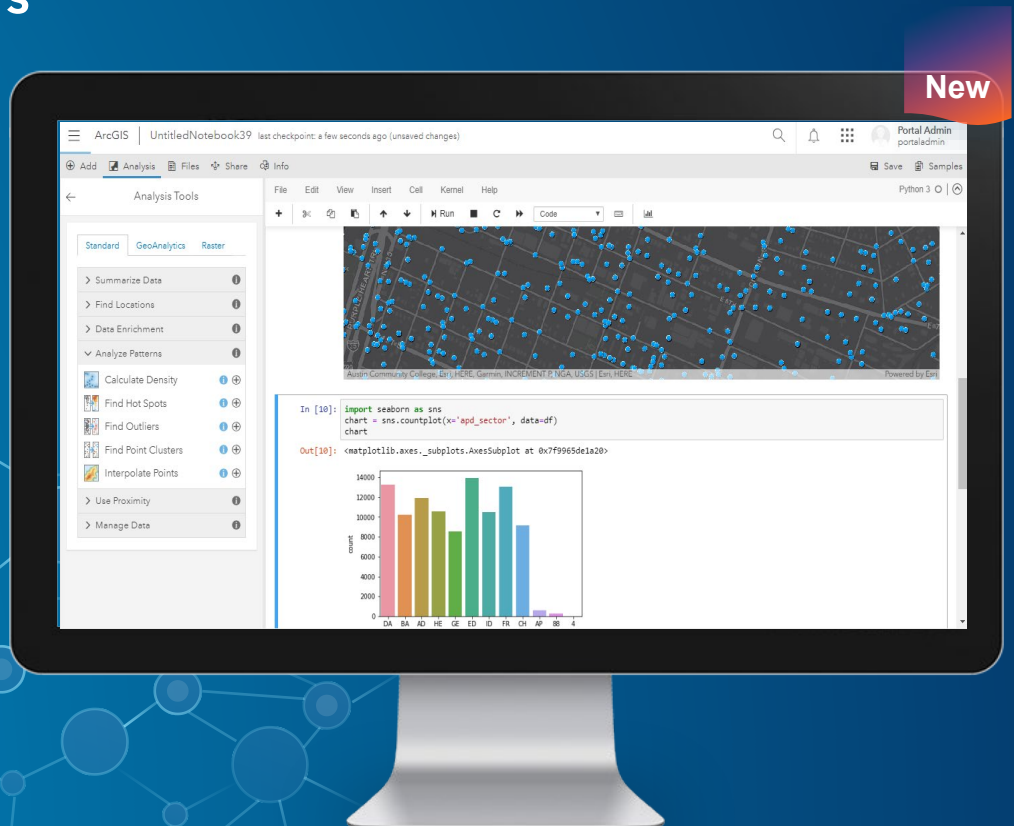
ESRI PETROLEUM GIS CONFERENCE

What is ArcGIS Notebooks?



ArcGIS Notebooks Top 5 things to remember

1. New at ArcGIS Enterprise 10.7
2. Integrated with the portal – notebooks run within ArcGIS Enterprise
3. Uses Python 3.x libraries/kernels
4. “Notebook runtimes” sync the libraries across users
5. Includes a gallery of sample notebooks



Deployment and Setup

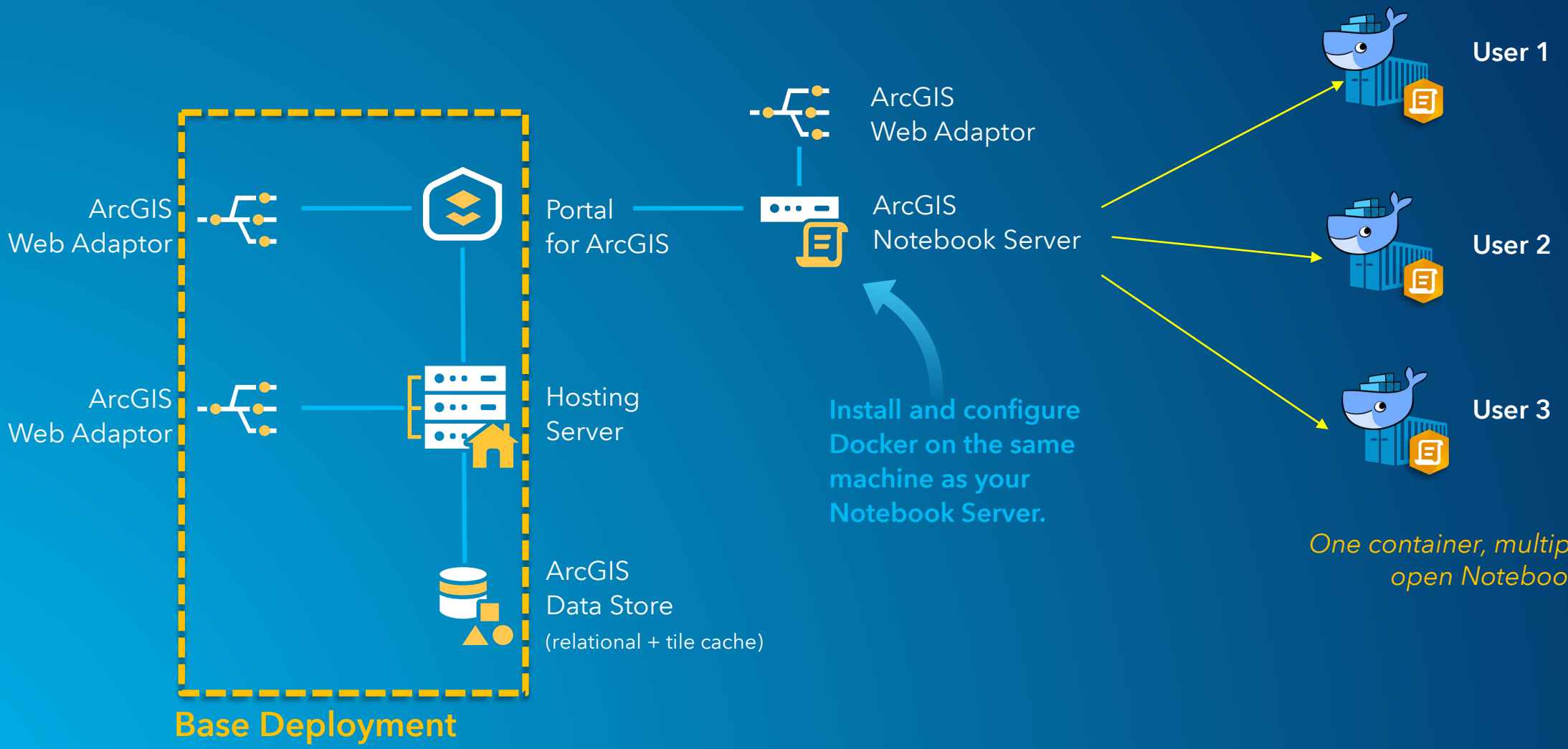
Quick Overview

	Windows	Linux

Prerequisites	• HyperV
	• Docker Desktop
	  - Docker must use Linux containers
		• Docker
		  - The specific version of Docker will depend on the OS selected

Hardware Requirements	• 8 GB RAM • 50 GB disk space	• 8 GB RAM • 50 GB on /var directory
Operating Systems	• Window Server 2016 (Standard or higher) • Windows Server 2019	• Red Hat Enterprise Linux Server 7 • Ubuntu LTS Server 16.04 or 18.04
Browsers	• Chrome • Firefox • Edge	• Chrome • Firefox • Edge

Conceptual Architecture



Number of machines in a site

- At 10.7 Notebook Server supports single machine sites only
  - If you want to scale, make the machine bigger
- At 10.7.1 Notebook Server we are planning to support multi-machine sites

But what about high availability?

- You can setup and prepare for a fail-over situation
- Notebooks are not like typical map services
  - Notebooks are stateful, transaction based

Sizing the site: How to plan for capacity?

Minimum recommendations:

- **For users who will use the ArcGIS Notebook Server Standard runtime:**
  - **1 core per user**
  - **4 GB RAM per user**
- **For users who will use the ArcGIS Notebook Server Advanced runtime:**
  - **2 cores per user**
  - **6 GB RAM per user**

Licensing: ArcGIS Notebook Server

- 2 Levels
  - Standard
  - Advanced
- No additional cost for Notebook Server Standard and there is no limit to the number of cores you can deploy
- ArcGIS Notebook Server Advanced is priced as a typical server licensing role – talk to your account manager for details
- The difference between the two is that Advanced includes ArcPy

The basics of ArcGIS Notebooks

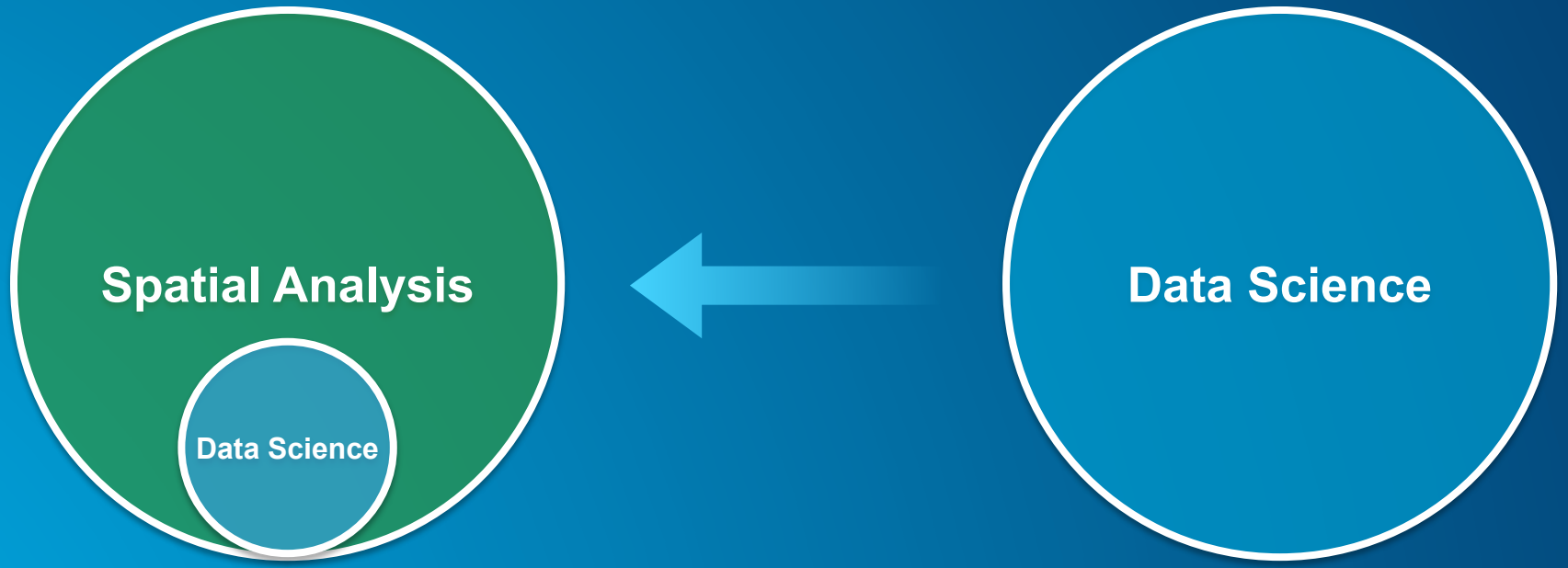
UI and Admin API walk-through

What is spatial data science?

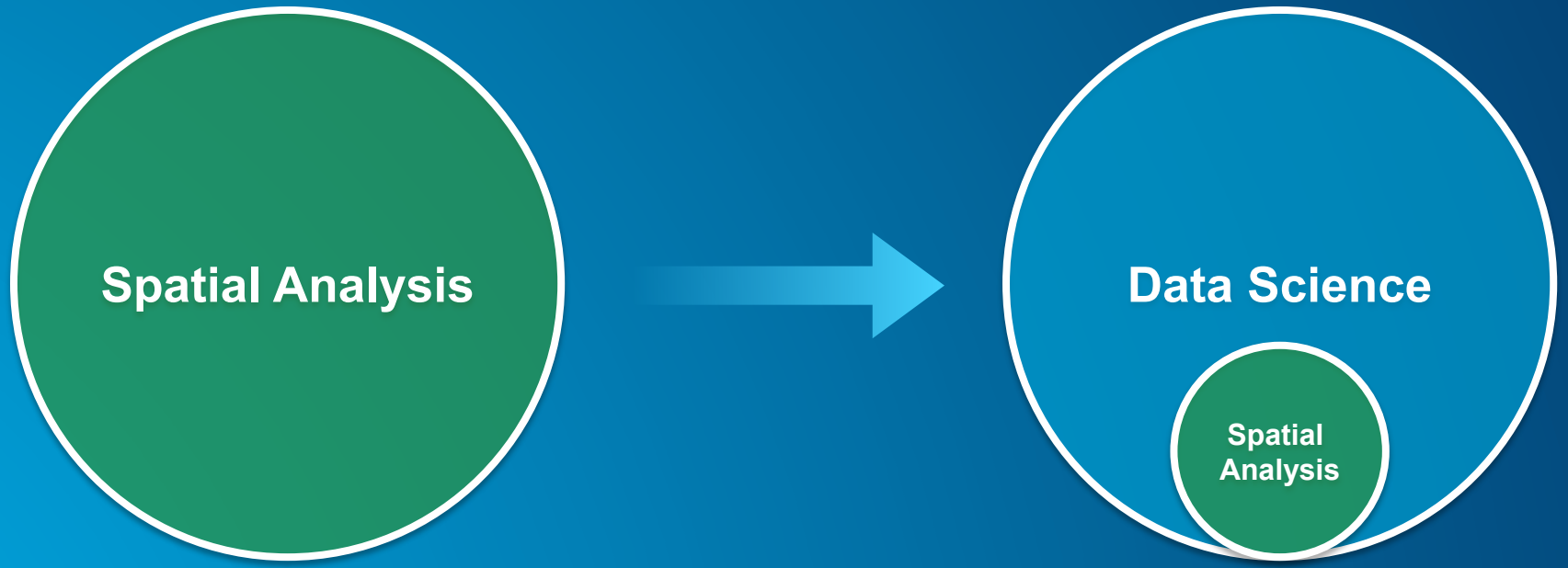
Spatial Analysis

Data Science

How some people describe it

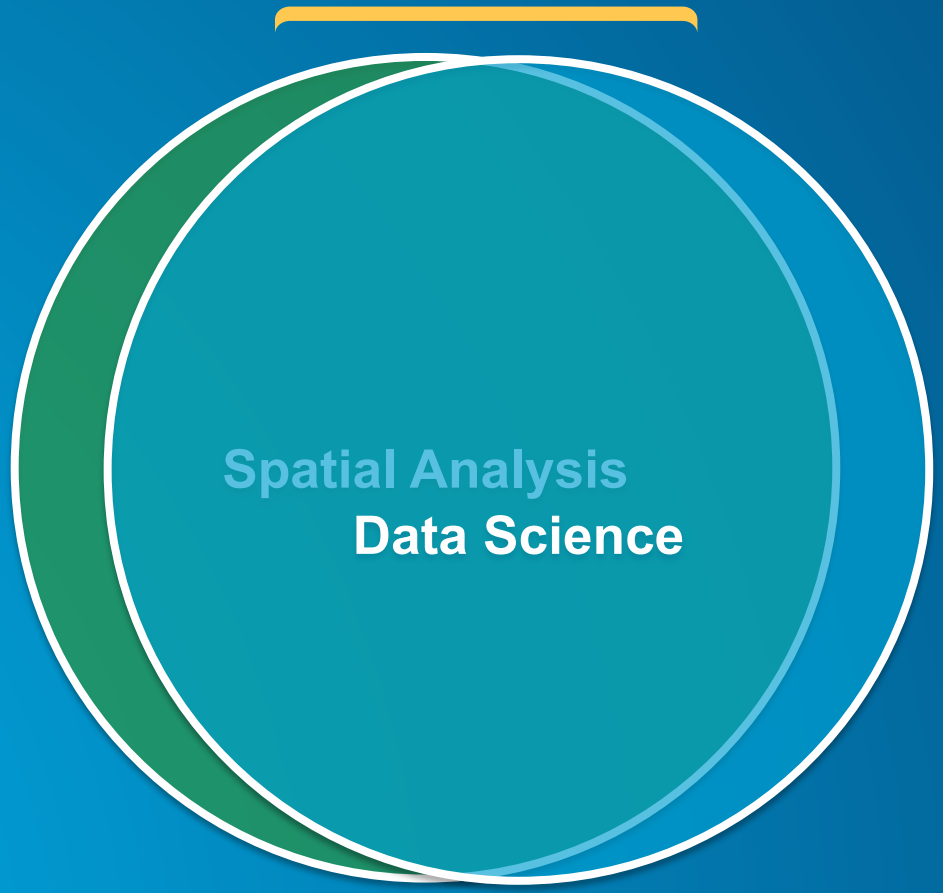


How other people describe it



What it is

The overlap between spatial analysis and data science



- Method > Interface
- Involves spatial data or location
- A map is not always the answer, but it usually helps
- It is about the problem you are solving, not about job titles

The relationship between spatial analysis and data science is not binary or mutually exclusive.

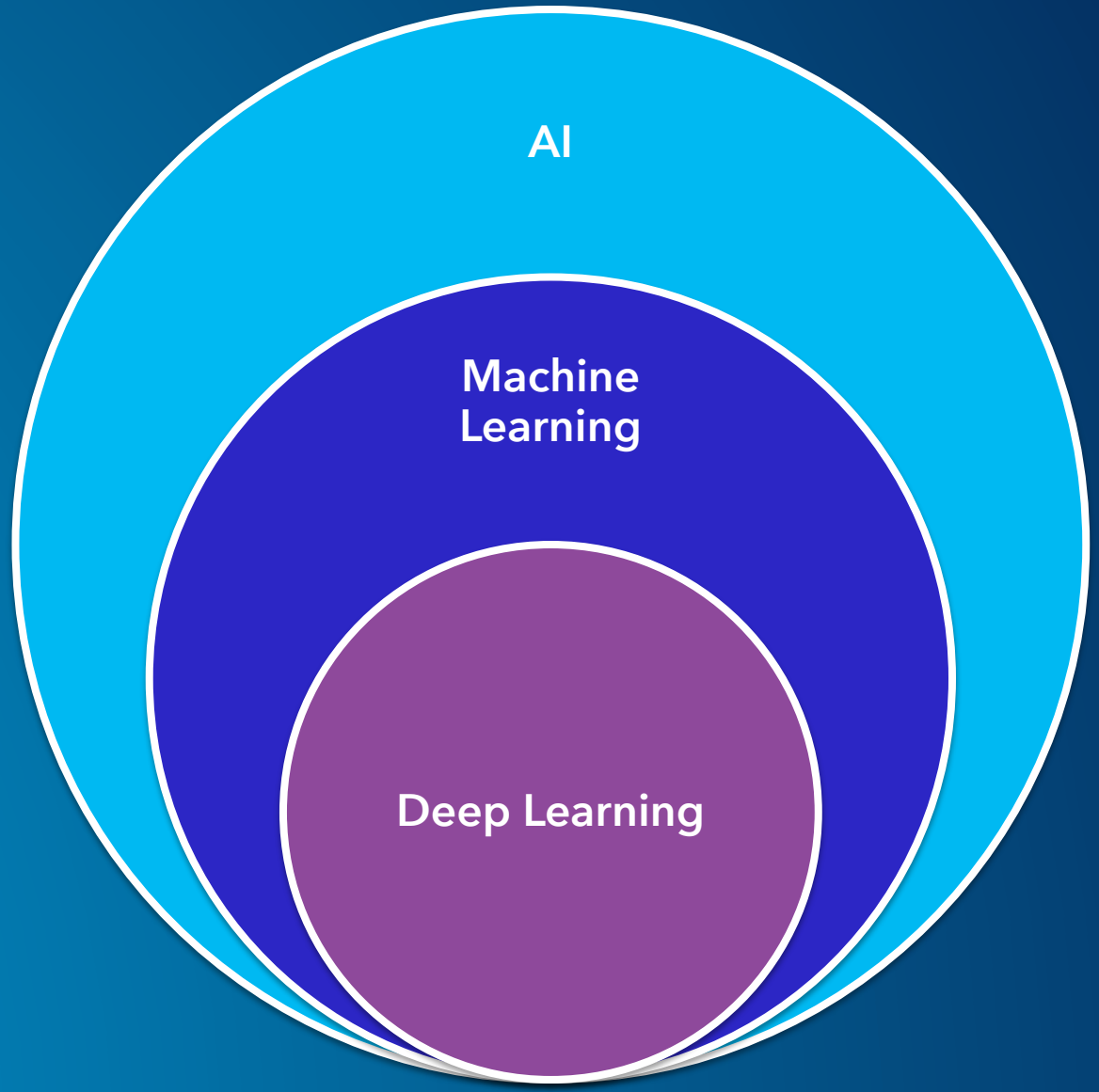
Machine Learning through ArcGIS

Click-button interface:

- **ArcGIS Pro, geoprocessing tools**
- **GeoAnalytics, web tools**

Scripting interface:

- **ArcGIS API for Python**
  - **arcgis.learn module (deep learning)**
- **ArcGIS Notebooks**
- **Open source Python libraries**
  - **scikit-learn, TensorFlow, CNTK, Keras, …**



Spatial Machine Learning with ArcGIS Notebooks

Find Restaurant Clusters

The ethics of AI and the role of spatial

Explainable AI > An AI “black box”

- AI itself is a black box
  - We understand the design of the model, but can’t say how exactly it made a prediction
- Privacy and business liability in the age of algorithms
  - Still being debated around the world, the next few years will be transformational
  - AI algorithms are hungry creatures – GIS is a gatekeeper of organizational data
- We use AI to drive action, our actions impact the world
  - GIS is the software that is designed to quantify the impact that human actions have on the human and natural worlds
    - **GIS is more important than ever.**

Be sure to

- **Use markdown to explain the steps of your analysis and describe your data**
- **Don’t enter credentials in plain text**
- **Remember the Zen of Python**

```
Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than *right* now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!
```

Use GIS to…

- Label data for machine learning and deep learning model
  - Labelling in data science = classifying data, it is not cartographic labels
- Evaluate the accuracy of a model
  - Score the model
- Perform spatiotemporal analysis to see the impact of the model over time
  - Is it working
  - Is it not
- Build your model
  - ModelBuilder
  - ArcGIS API for Python
  - ArcGIS Notebooks
- Tell a story that resonates and is relatable

What's next?

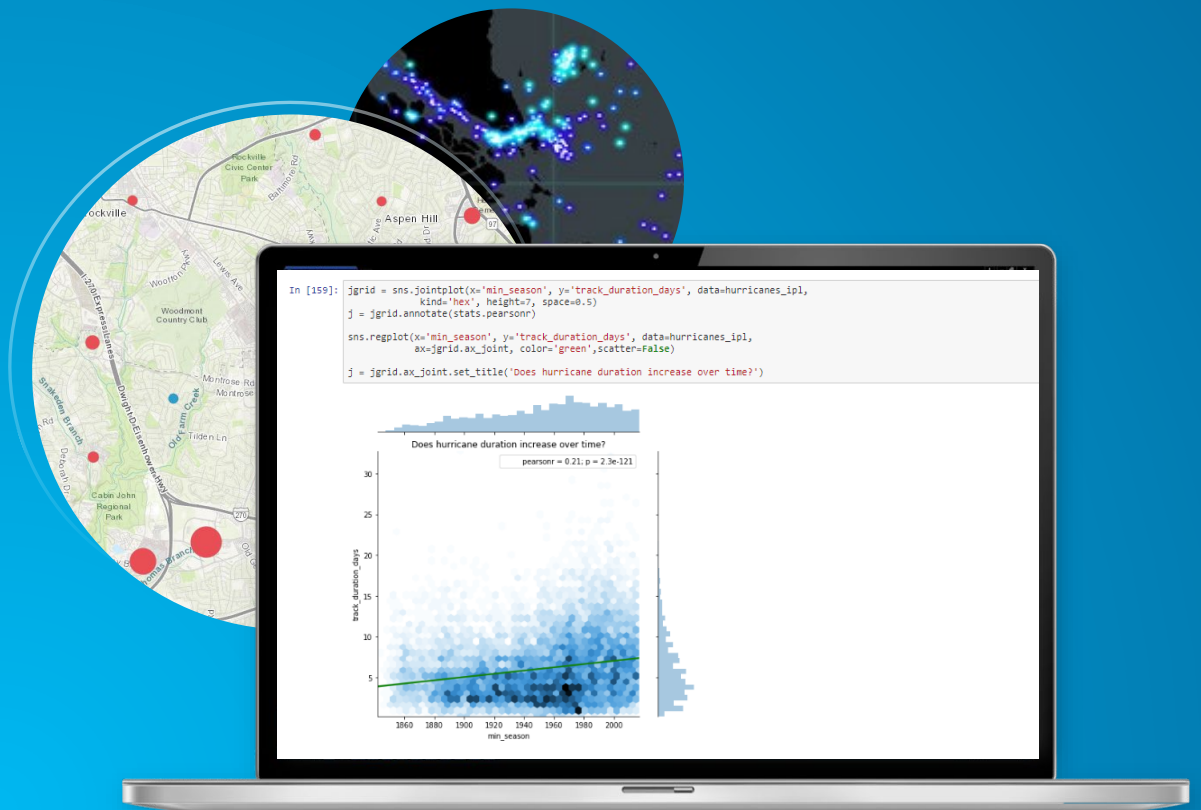
The road ahead for ArcGIS Notebooks

Near term

- More samples
- More tools
- Multi-machine sites
- ArcGIS Notebooks in Online

Mid to long term

- Scheduling
- Versioning
- Notebooks in Pro
- Notebook workbench
- Notebook Server Manager
- Backup and restore with WebGISDR
- …




esri
THE
SCIENCE
OF
WHERE