Teaching Spatial Data Science

Dr. Song Gao
Geospatial Data Science Lab
Department of Geography & Data Science Institute
University of Wisconsin – Madison

https://geods.geography.wisc.edu/

Data Science Major as the top-1 fast growing major

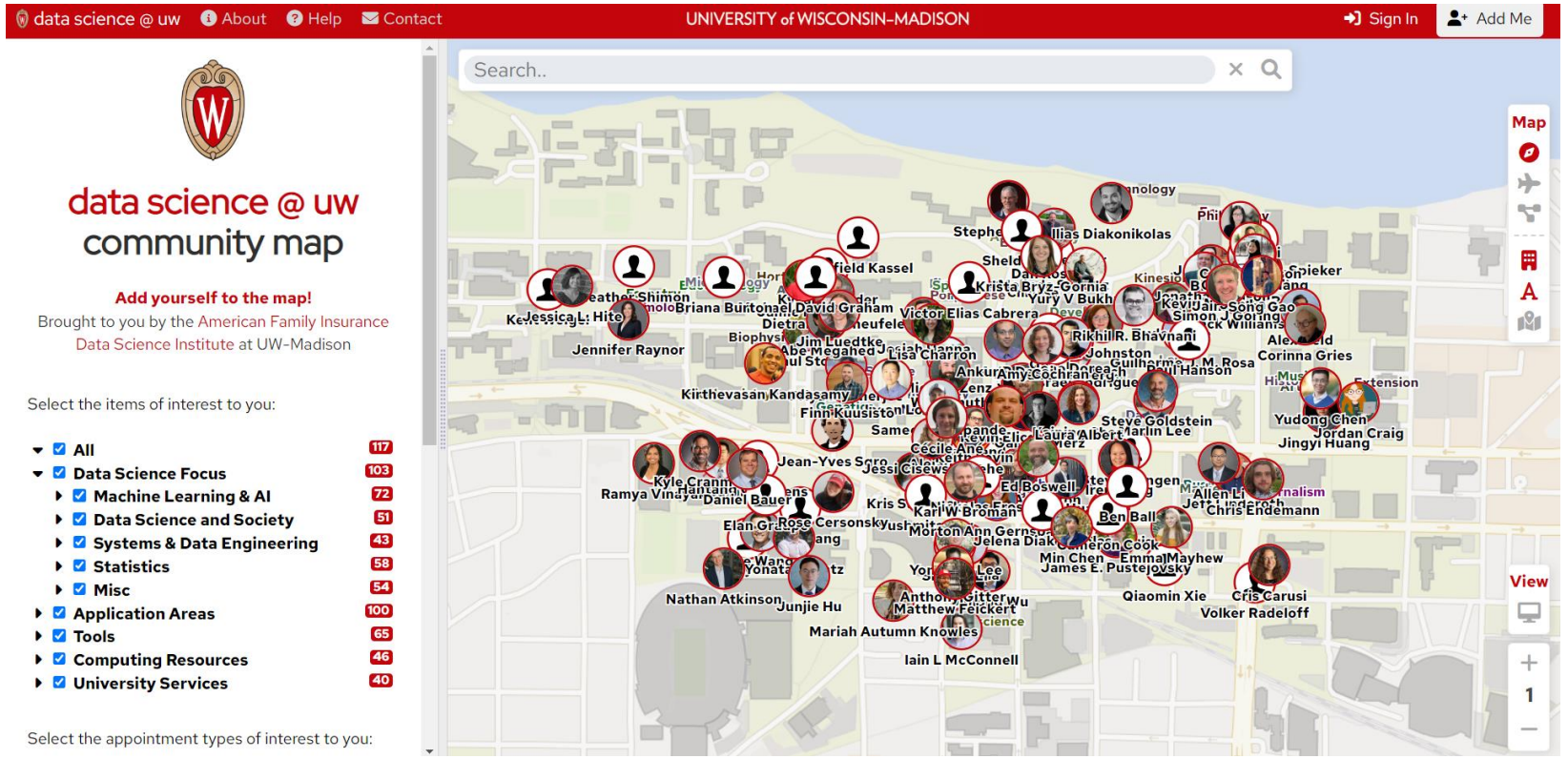
No. 1: Data Science (+900 since inception in 2019)

https://news.wisc.edu/uws-5-hottest-majors-student-demand-workforce-trends-drive-enrollment-gains

New Data Science Institute (DSI)

Mission

- Perform innovative research in the fundamentals and applications of data science
- Advance scientific discovery in collaboration with researchers across campus and beyond



https://datascience.wisc.edu/

Data Science Major and Spatial Data Science Curriculum

https://datascience.wisc.edu/

Data Science Major and Spatial Data Science Curriculum

Foundational Data Science Courses

STAT 240	🎓 Introduction to Data Modeling I
STAT 340	Introduction to Data Modeling II
COMP SCI 220	🎓 Data Science Programming I
or COMP SCI 300	🎓 Programming II
COMP SCI 320	Data Science Programming II
L I S 461	Data and Algorithms: Ethics and Policy

GEOG 573	Advanced Geocomputing and Geospatial Big Data Analytics
GEOG 574	Geospatial Database Design and Development

Statistical Modeling

Machine Learning

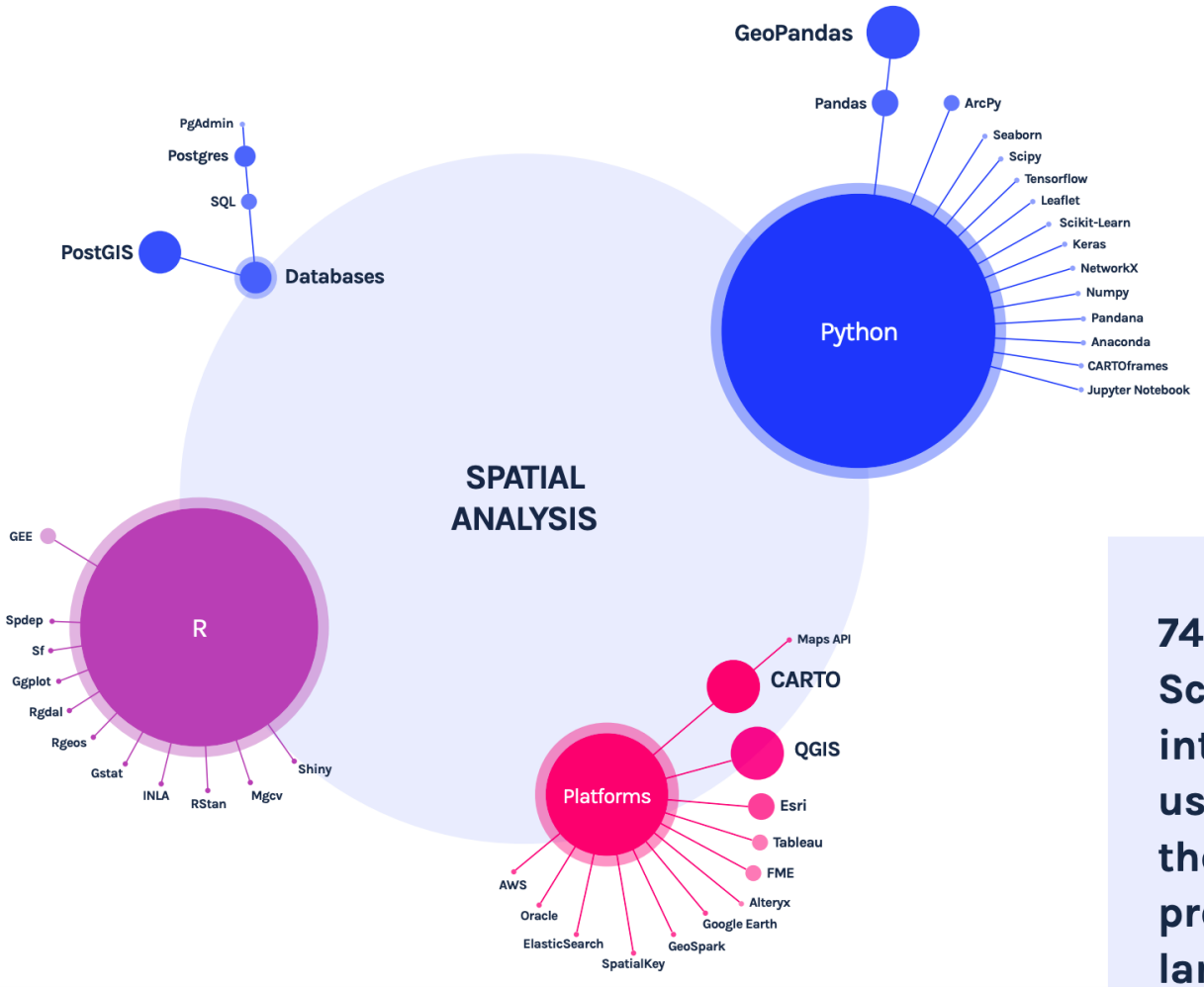
Complete one of the following:

COMP SCI/ E C E/M E 532	Matrix Methods in Machine Learning [1]
COMP SCI/ E C E/M E 539	Introduction to Artificial Neural Networks
COMP SCI 540	Introduction to Artificial Intelligence
MATH 535	Mathematical Methods in Data Science
STAT 451	Introduction to Machine Learning and Statistical Pattern Classification
STAT 453	Introduction to Deep Learning and Generative Models

https://datascience.wisc.edu/

Become a Spatial Data Scientist

Skills contributed to the collection, management, modeling, representation, retrieval, integration, visualization, and analysis of spatial data.



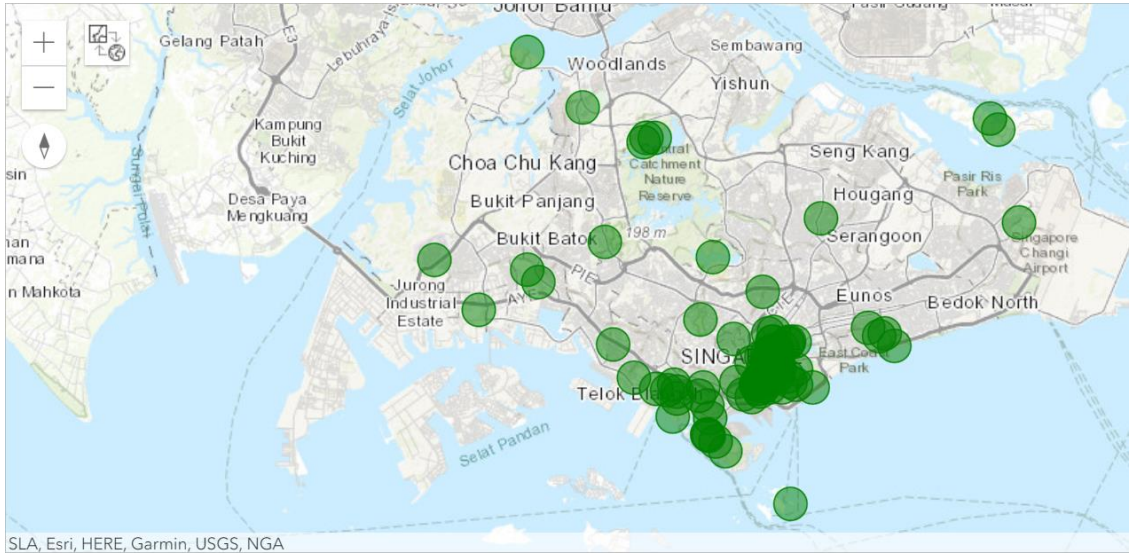
74% of Data Science teams interviewed use Python as their principal programming language

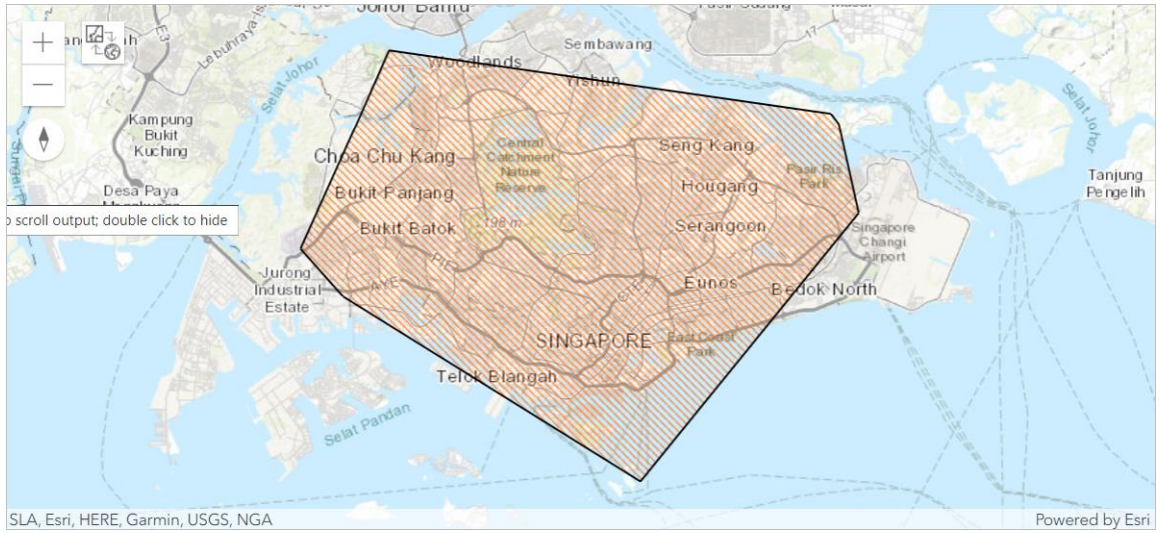
Course Topics

- **Geospatial Big Data Storage** (e.g., compression)
- **Vector Data Algorithms** (e.g., geometry computation, spatial data clustering)
- **Raster Data Algorithms** (e.g., classification, change detection);
- **Network Data Algorithms** (e.g., shortest path, centrality)
- **Geospatial Big Data Visualization**
- **Spatial Statistics** (e.g. spatial autocorrelation, regression, spatiotemporal analytics)
- **Special Topics**: GPS Trajectory data analytics and Remote Sensing imagery analytics, etc.

Use ArcGIS Notebooks

- **Minimum Bounding Geometry Computation with ArcPy**

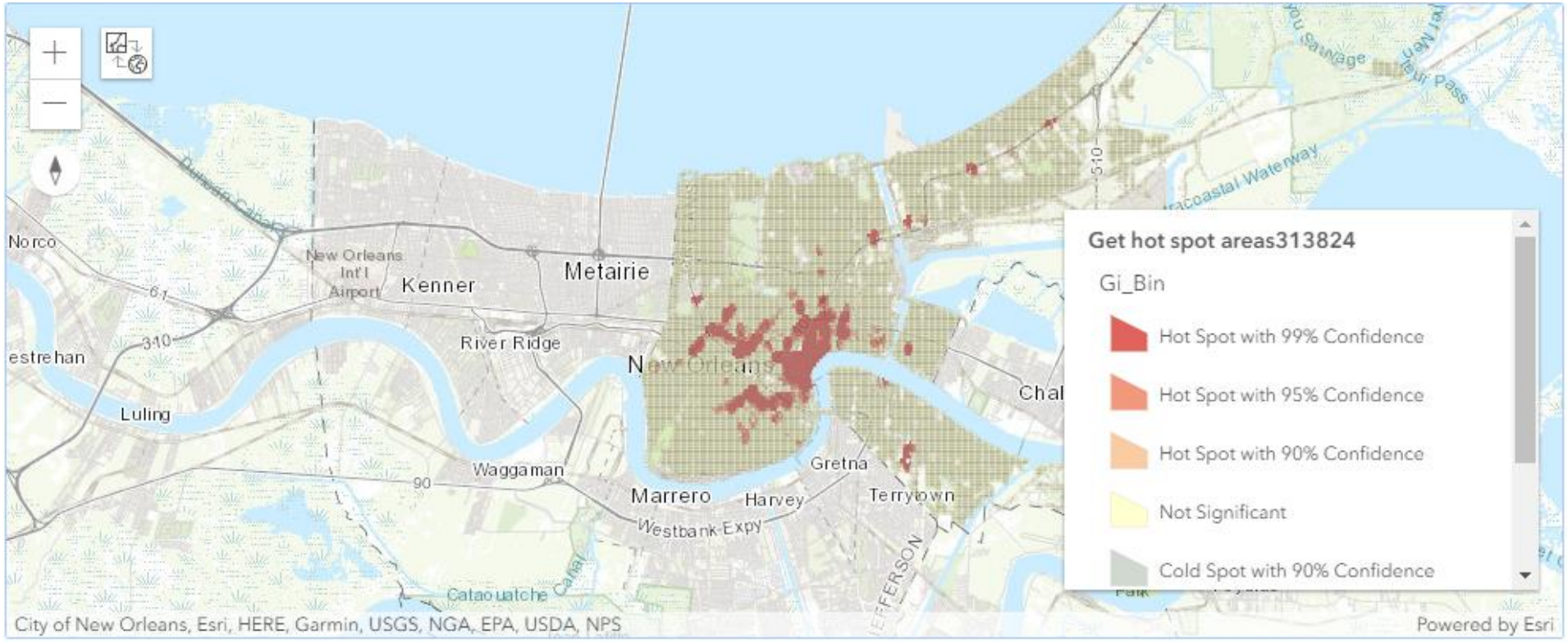





```
arcpy.management.MinimumBoundingGeometry("Tourist_Attractions_Join",
                                          "Tourist_Attractions_Boundary_YOURINITIALS",
                                          "CONVEX_HULL")
```

Use ArcGIS Notebooks

- **Find Statistically Significant Hotspots/Coldspots**



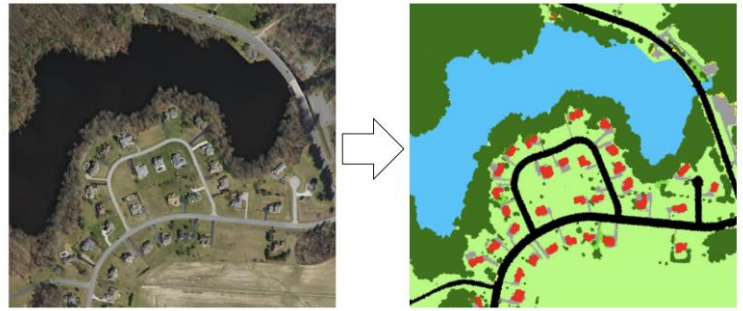
arcgis.geoanalytics.analyze_patterns.find_hot_spots()

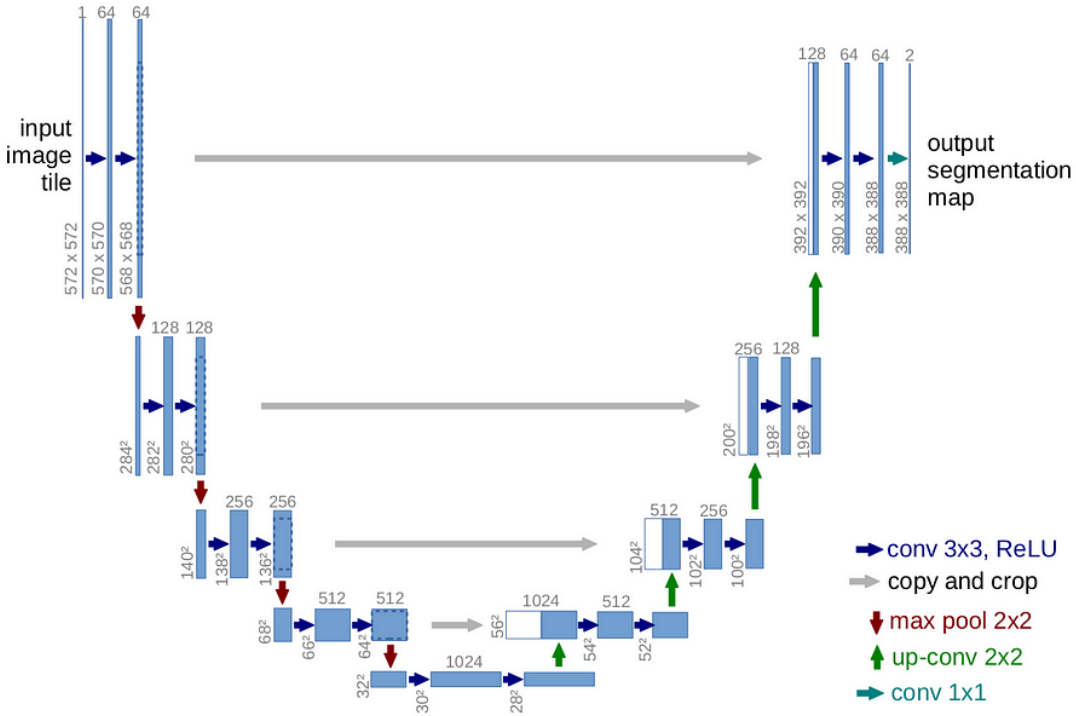
Use ArcGIS Notebooks

- **Deep Learning for semantic segmentation**

Keras implementation

```
def build_model(input_layer, start_neurons):
    conv1 = Conv2D(start_neurons * 1, (3, 3), activation="relu", padding="same")(input_layer)
    conv1 = Conv2D(start_neurons * 1, (3, 3), activation="relu", padding="same")(conv1)
    pool1 = MaxPooling2D((2, 2))(conv1)
    pool1 = Dropout(0.25)(pool1)

    conv2 = Conv2D(start_neurons * 2, (3, 3), activation="relu", padding="same")(pool1)
    conv2 = Conv2D(start_neurons * 2, (3, 3), activation="relu", padding="same")(conv2)
    pool2 = MaxPooling2D((2, 2))(conv2)
    pool2 = Dropout(0.5)(pool2)

    conv3 = Conv2D(start_neurons * 4, (3, 3), activation="relu", padding="same")(pool2)
    conv3 = Conv2D(start_neurons * 4, (3, 3), activation="relu", padding="same")(conv3)
    pool3 = MaxPooling2D((2, 2))(conv3)
    pool3 = Dropout(0.5)(pool3)

    conv4 = Conv2D(start_neurons * 8, (3, 3), activation="relu", padding="same")(pool3)
    conv4 = Conv2D(start_neurons * 8, (3, 3), activation="relu", padding="same")(conv4)
    pool4 = MaxPooling2D((2, 2))(conv4)
    pool4 = Dropout(0.5)(pool4)

    # Middle
    convm = Conv2D(start_neurons * 16, (3, 3), activation="relu", padding="same")(pool4)
    convm = Conv2D(start_neurons * 16, (3, 3), activation="relu", padding="same")(convm)

    deconv4 = Conv2DTranspose(start_neurons * 8, (3, 3), strides=(2, 2), padding="same")(convm)
    uconv4 = concatenate([deconv4, conv4])
    uconv4 = Dropout(0.5)(uconv4)
    uconv4 = Conv2D(start_neurons * 8, (3, 3), activation="relu", padding="same")(uconv4)
    uconv4 = Conv2D(start_neurons * 8, (3, 3), activation="relu", padding="same")(uconv4)

    deconv3 = Conv2DTranspose(start_neurons * 4, (3, 3), strides=(2, 2), padding="same")(uconv4)
    uconv3 = concatenate([deconv3, conv3])
    uconv3 = Dropout(0.5)(uconv3)
    uconv3 = Conv2D(start_neurons * 4, (3, 3), activation="relu", padding="same")(uconv3)
    uconv3 = Conv2D(start_neurons * 4, (3, 3), activation="relu", padding="same")(uconv3)

    deconv2 = Conv2DTranspose(start_neurons * 2, (3, 3), strides=(2, 2), padding="same")(uconv3)
    uconv2 = concatenate([deconv2, conv2])
    uconv2 = Dropout(0.5)(uconv2)
    uconv2 = Conv2D(start_neurons * 2, (3, 3), activation="relu", padding="same")(uconv2)
    uconv2 = Conv2D(start_neurons * 2, (3, 3), activation="relu", padding="same")(uconv2)

    deconv1 = Conv2DTranspose(start_neurons * 1, (3, 3), strides=(2, 2), padding="same")(uconv2)
    uconv1 = concatenate([deconv1, conv1])
    uconv1 = Dropout(0.5)(uconv
    uconv1 = Conv2D(start_neurons * 1, (3, 3), activation="relu", padding="same")(uconv1)
```





unet = **UnetClassifier**(data, backbone='resnet34)

Credit: towardsdatascience

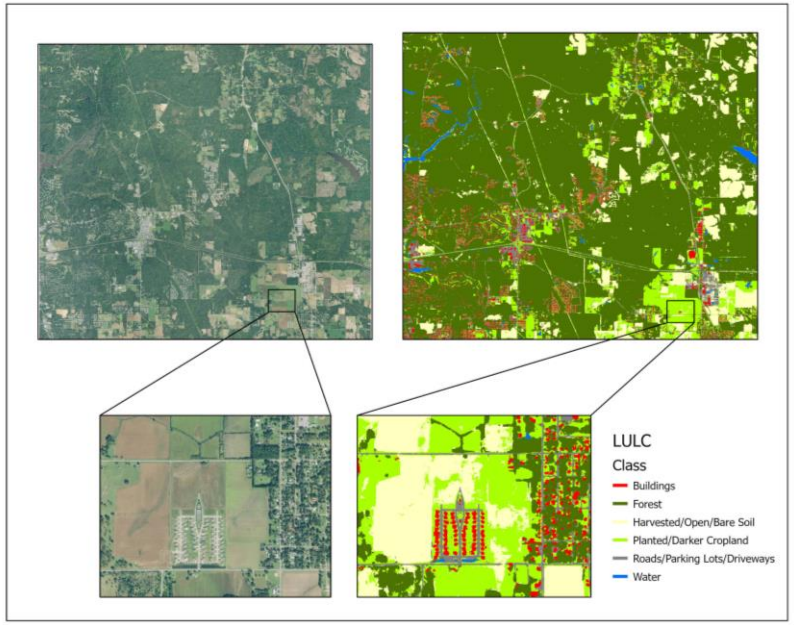
Use ArcGIS Notebooks

- **Land Use and Land Cover Classification with Deep Learning**

Training data can be exported as **a GIS data layer** by using ArcGIS tools

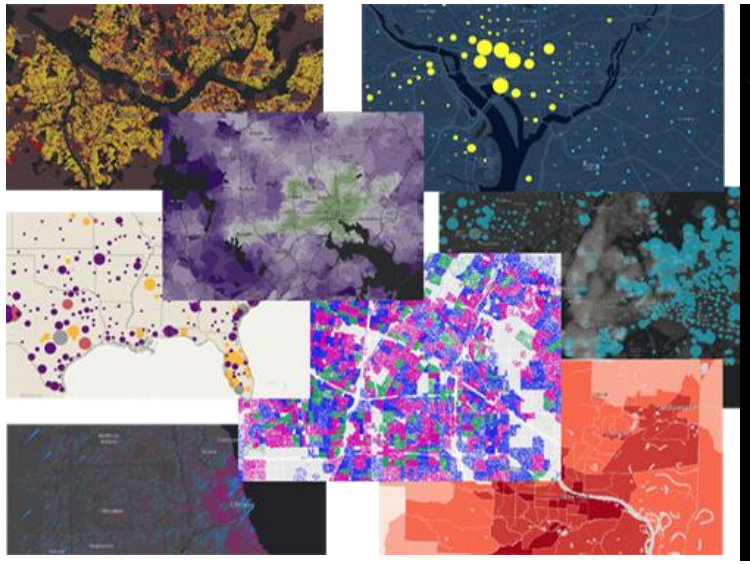
#Call the **Built-in** U-Net **Deep Learning Model**

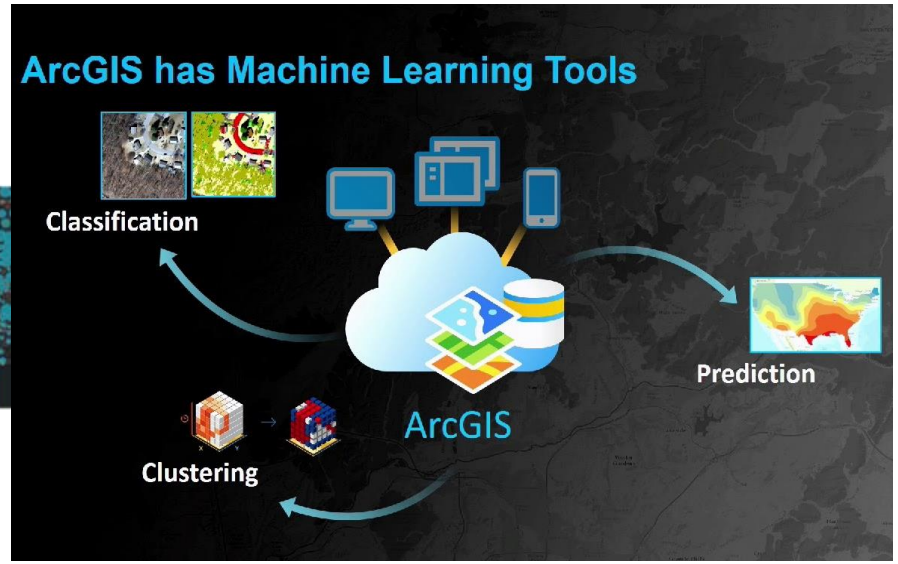
*unet = **UnetClassifier**(data, backbone='resnet34', ignore_classes=[0])*
unet.fit ()

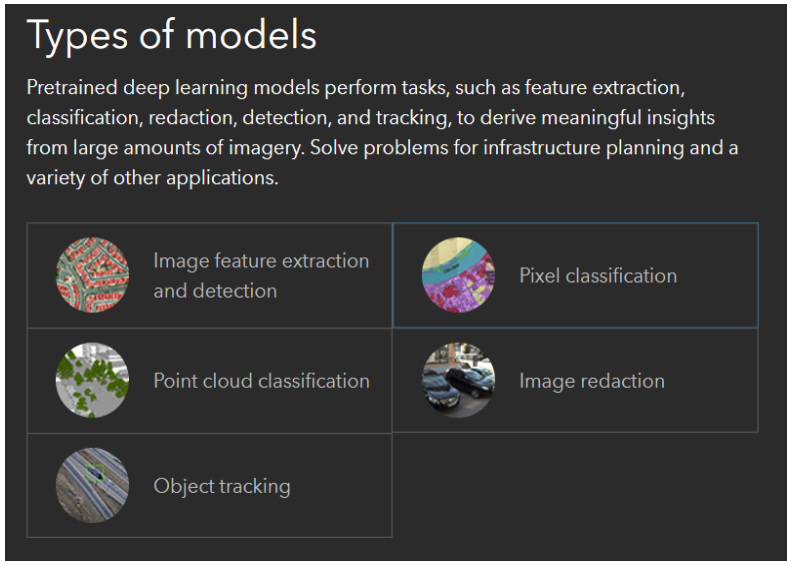


Use ArcGIS Notebooks

- Access to large-scale GIS data resources hosted on ArcGIS Online
- Integrate powerful spatial analytics tools and workflow in ArcGIS
- Utilize pretrained GeoAI models on GPU-enabled Online Notebook







Two different groups of mindsets

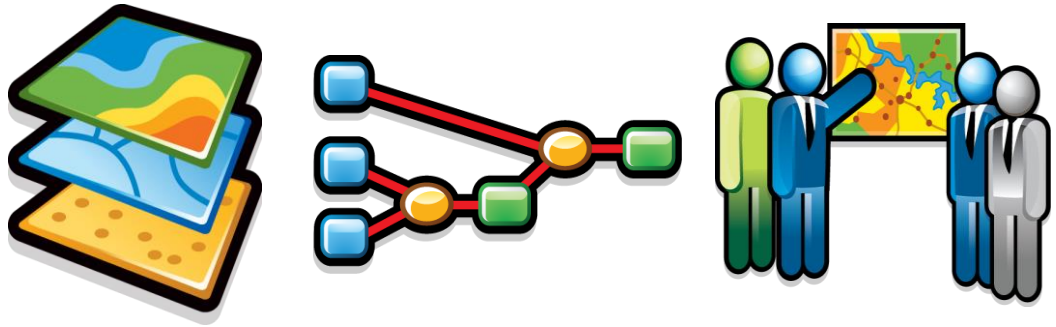
Geography elected

Data Science elected

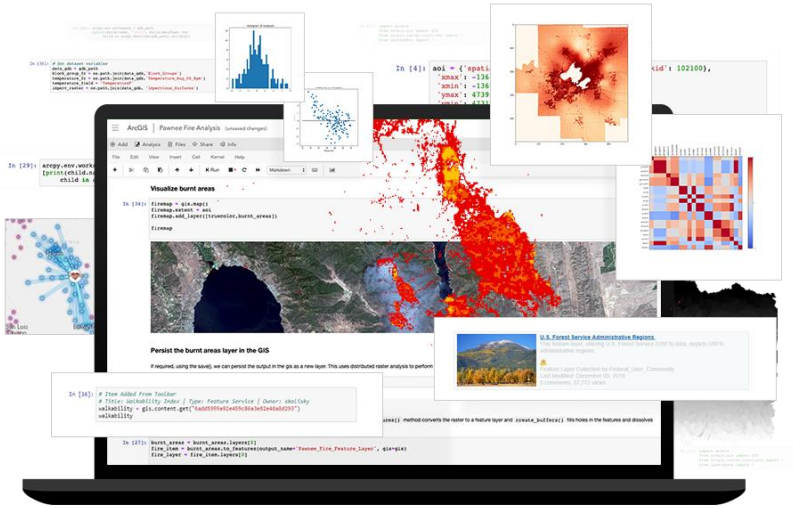
Image Source: https://www.ptcnews.tv/why-should-you-choose-data-science-as-your-career-en

Learned from experiences

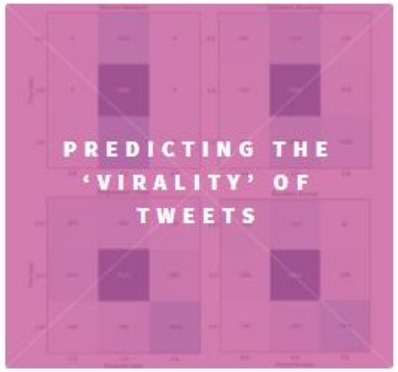
- Acknowledge differences in thinking, knowledge, and skills
- Geoprocessing is awesome

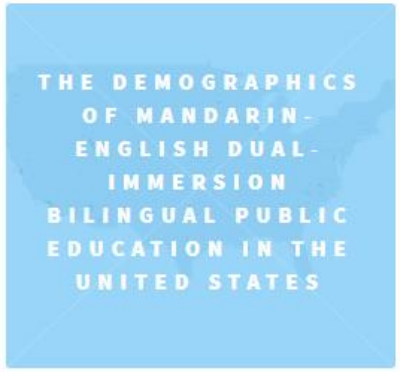


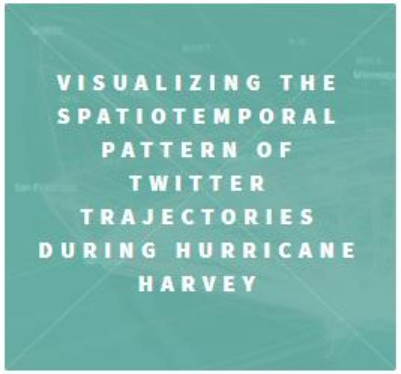
- Engage with real-world problems/applications
- Students love notebook code examples



Final Projects







Higher Ed Guide to Esri E-Learning for Spatial Data Science

Overview

Spatial data science allows analysts to extract deeper insight from data using a comprehensive set of analytical methods and spatial algorithms, including machine learning and deep learning techniques. Spatial data science topics may be included in a range of courses, including data science, business, and GIS.

This guide is an aid for instructors who want to use authoritative Esri web-based learning resources as part of college or university courses. This list is not a comprehensive curriculum, nor is it a course outline. It is intended to help instructors quickly identify and select those resources that best support their goals and students. Listed items are available as of September 2021 and are expected to be available through at least December 2021. New listings are in orange.

Full descriptions can be found at the links provided. All items listed are web courses unless otherwise noted. The complete Esri catalog can be found at esri.com/training/catalog. The information provided in this guide is subject to change without notice. Please email GIStraining@esri.com or call (800) 447-9778, ext. 5757 with questions about courses.

You and your students may be eligible for unlimited access to the entire collection of self-paced e-Learning (web courses, training seminars, and more) if your institution has a qualifying product with a current maintenance subscription. To determine if this applies to you, contact your Esri software license administrator, check online, or email educationinfo@esri.com.

This guide is organized into three main sections:

- **Learning Plans:** Esri-curated sets of e-Learning offerings with a suggested order.
- **Technology:** Individual e-Learning offerings that provide foundational concepts and skills to support spatial data science workflows.
- **Capabilities:** Individual offerings about specific spatial data science analysis techniques.

Thank you!