

# Spatial Analysis with ArcGIS® Pro

Official Esri Training Courseware



**esri**®

THE  
SCIENCE  
OF  
WHERE®



# Spatial Analysis with ArcGIS® Pro

STUDENT EDITION

Copyright © 2019 Esri  
All rights reserved.

Course version 3.0. Version release date February 2019.

Printed in the United States of America.

The information contained in this document is the exclusive property of Esri. This work is protected under United States copyright law and other international copyright treaties and conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by Esri. All requests should be sent to Attention: Director, Contracts and Legal, Esri, 380 New York Street, Redlands, CA 92373-8100, USA.

**Export Notice:** Use of these Materials is subject to U.S. export control laws and regulations including the U.S. Department of Commerce Export Administration Regulations (EAR). Diversion of these Materials contrary to U.S. law is prohibited.

The information contained in this document is subject to change without notice.

**Commercial Training Course Agreement Terms:** The Training Course and any software, documentation, course materials or data delivered with the Training Course is subject to the terms of the Master Agreement for Products and Services, which is available at <http://www.esri.com/~media/Files/Pdfs/legal/pdfs/ma-full/ma-full.pdf>. The license rights in the Master Agreement strictly govern Licensee's use, reproduction, or disclosure of the software, documentation, course materials and data. Training Course students may use the course materials for their personal use and may not copy or redistribute for any purpose. Contractor/Manufacturer is Esri, 380 New York Street, Redlands, CA 92373-8100, USA.

**Esri Trademarks:** Esri trademarks and product names mentioned herein are subject to the terms of use found at the following website: <http://www.esri.com/legal/copyright-trademarks.html>.

Other companies and products or services mentioned herein may be trademarks, service marks or registered marks of their respective mark owners.



# Table of Contents

Esri resources for your organization .....	ix
--	----

## Course introduction

Course introduction .....	1
Course goals .....	2
Installing the course data .....	2
Training Services account credentials .....	3
Icons used in this workbook .....	4
Understanding the ArcGIS platform .....	5

## 1 Building a foundation for spatial analysis

Lesson introduction .....	1-1
What is spatial analysis? .....	1-2
Benefits of spatial analysis .....	1-3
Common analysis problems .....	1-5
Spatial analysis tools .....	1-6
Spatial analysis workflow .....	1-8
Applying spatial analysis .....	1-10
Lesson review .....	1-11
Answers to Lesson 1 questions .....	1-12

## 2 Planning and preparing for spatial analysis

Lesson introduction .....	2-1
Data properties .....	2-2
Raster data considerations .....	2-3
Environment settings .....	2-5
Exercise 2: Prepare data for analysis .....	2-6
Set up an ArcGIS Pro project .....	2-7
Change the coordinate system for a feature class .....	2-8
Create a feature class from x,y coordinates .....	2-10
Enhance data using a table join .....	2-13
Import a map file for a different study area .....	2-13
Extract features using the Clip tool .....	2-14
Extract raster data using a mask .....	2-16
Lesson review .....	2-19
Answers to Lesson 2 questions .....	2-21

## 3 Proximity analysis

Lesson introduction .....	3-1
---------------------------	-----

Using proximity in everyday life .....	3-2
Choosing the best distance measure .....	3-3
Ways to measure distance .....	3-4
Outputs of proximity analysis .....	3-5
Buffering using different distance measures.....	3-7
Measuring cost .....	3-8
Exercise 3: Analyze proximity .....	3-9
Prepare the project .....	3-10
Select features based on distance .....	3-11
Create proximity zones .....	3-12
Determine the closest store to each customer .....	3-16
Add and calculate a field .....	3-17
Create desire lines.....	3-18
Create drive-time polygons .....	3-20
Create a distance surface.....	3-23
Lesson review.....	3-25
Answers to Lesson 3 questions.....	3-26

## 4 Overlay analysis

Lesson introduction .....	4-1
Introducing overlay .....	4-2
How overlay works.....	4-3
Overlay tools.....	4-5
Choosing the appropriate tool .....	4-7
Exercise 4: Perform overlay analysis .....	4-8
Make selections based on location.....	4-9
Overlay customers and driving times using the Intersect tool.....	4-11
Overlay customers and driving times using the Identity tool .....	4-14
Remove customers within 15 miles.....	4-16
Summarize stream length in a watershed .....	4-17
Calculate the amount of each land-use classification .....	4-19
Lesson review.....	4-21
Answers to Lesson 4 questions.....	4-22

## 5 Automating spatial analysis

Lesson introduction .....	5-1
Automating workflows .....	5-2
Automation methods in ArcGIS Pro.....	5-3
Batch geoprocessing .....	5-5
Exercise 5A: Build a model.....	5-6
Prepare ArcGIS Pro .....	5-7
Create a model .....	5-7

Add the XY Table To Point tool .....	5-8
Add the Near tool .....	5-9
Add the Make Feature Layer tool .....	5-10
Add the XY To Line tool .....	5-11
Run the model.....	5-12
Automating and sharing models .....	5-14
Exercise 5B: Use a model to process multiple inputs.....	5-16
Prepare ArcGIS Pro and make a copy of a model.....	5-17
Add an iterator to a model .....	5-18
Set model parameters.....	5-20
Change model element labels.....	5-24
Lesson review.....	5-28
Answers to Lesson 5 questions.....	5-29

## 6 Creating surfaces using interpolation

Lesson introduction .....	6-1
Tobler's First Law of Geography .....	6-2
What is interpolation?.....	6-3
Interpolation methods .....	6-5
Interpolation tools .....	6-7
Deterministic interpolation .....	6-8
Exercise 6: Interpolate surfaces .....	6-10
Examine data .....	6-11
Set geoprocessing environments.....	6-12
Interpolate using the Natural Neighbor tool .....	6-12
Interpolate using the Spline tool.....	6-13
Interpolate using inverse distance weighted interpolation.....	6-15
Examine interpolated values.....	6-17
Challenge: Challenge step.....	6-19
Lesson review.....	6-20
Answers to Lesson 6 questions.....	6-21
Exercise 6 challenge solution .....	6-22

## 7 Suitability modeling

Lesson introduction .....	7-1
What is suitability modeling?.....	7-2
Suitability modeling workflow.....	7-3
Evaluating analysis criteria .....	7-4
Choosing vector or raster overlay.....	7-5
Deriving surfaces from other sources .....	7-6
Raster functions and geoprocessing tools.....	7-7
Levels of measurement.....	7-8

Transforming values to a common scale.....	7-10
Exercise 7A: Build a model and classify data to a common scale .....	7-12
Prepare a project and set environments .....	7-13
Create a model .....	7-14
Add input layers and Euclidean Distance tools .....	7-14
Add the Slope tool and set parameters.....	7-17
Reclassify land-use values .....	7-17
Rescale the roads distance surface .....	7-19
Rescale the stream distance surface .....	7-20
Rescale the slope surface.....	7-21
Run the model.....	7-22
Types of raster overlay .....	7-24
The Raster Calculator.....	7-26
Locating and analyzing results.....	7-27
Exploring data sources .....	7-29
Exercise 7B: Perform suitability modeling .....	7-30
Overlay input rasters .....	7-31
Create regions.....	7-34
Lesson review.....	7-36
Answers to Lesson 7 questions.....	7-37

## 8 Spatial statistics

Lesson introduction .....	8-1
Spatial patterns.....	8-2
What are spatial statistics?.....	8-3
Types of spatial statistics.....	8-5
Interpreting inferential statistics.....	8-7
Descriptive versus inferential .....	8-9
Spatial statistics tools.....	8-12
Clusters and outliers .....	8-13
Clustering tools.....	8-15
Exercise 8A: Use spatial statistics to explore data .....	8-17
Prepare ArcGIS Pro .....	8-18
Locate directional trends in data.....	8-19
Run the Average Nearest Neighbor tool .....	8-20
Run the Spatial Autocorrelation tool.....	8-22
Run the Hot Spot Analysis tool .....	8-23
Create a density surface.....	8-25
Exercise 8B: Perform clustering and outlier analysis.....	8-27
Prepare the project .....	8-28
Perform density-based clustering .....	8-28
Perform optimized hot spot analysis.....	8-31

Perform optimized outlier analysis.....	8-34
Lesson review.....	8-37
Answers to Lesson 8 questions.....	8-38

## 9 Space-time analysis

Lesson introduction .....	9-1
Incorporating time into your analysis.....	9-2
Temporal analysis.....	9-3
Exercise 9A: Explore data .....	9-5
Use a chart to explore data .....	9-6
Space-time analysis .....	9-8
Emerging hot spot analysis.....	9-11
Space-time analysis workflow .....	9-13
Exercise 9B: Explore space-time pattern mining tools .....	9-15
Explore data using charts.....	9-16
Create a space-time cube .....	9-18
Run the Emerging Hot Spot Analysis tool.....	9-19
Visualize a space-time cube in 3D.....	9-21
Lesson review.....	9-24
Answers to Lesson 9 questions.....	9-25

## 10 Regression analysis

Lesson introduction .....	10-1
Explaining spatial patterns.....	10-2
Causes of spatial patterns.....	10-3
What is regression?.....	10-4
Regression equation .....	10-6
OLS regression.....	10-9
Checkpoint .....	10-11
Interpreting OLS diagnostics .....	10-12
Six OLS checks.....	10-14
OLS reports.....	10-17
Exploratory regression .....	10-19
Exercise 10: Find a properly specified regression model .....	10-21
Set up ArcGIS Pro .....	10-22
Perform exploratory data analysis.....	10-22
Use the Generalized Linear Regression tool to test for higher spending factors.....	10-24
Evaluate the spatial output from the GLR tool.....	10-25
Create a scatter plot matrix.....	10-27
Run the GLR tool on multiple dependent variables.....	10-30
Perform OLS checks.....	10-30
Lesson review.....	10-35

Enriching data for analysis .....	10-36
Answers to Lesson 10 questions.....	10-37

## 11 Geographically weighted regression

Lesson introduction .....	11-1
How relationships change over space .....	11-2
GWR characteristics .....	11-3
When to use GWR .....	11-4
GWR in action .....	11-6
Exercise 11: Perform GWR.....	11-9
Run GWR using a properly specified OLS model .....	11-10
Map coefficients to see variation over space .....	11-11
Predict using GWR .....	11-13
Lesson review.....	11-17
Answers to Lesson 11 questions.....	11-18

## 12 Geostatistical interpolation

Lesson introduction .....	12-1
Deterministic interpolation .....	12-2
Geostatistical interpolation.....	12-4
Kriging .....	12-5
Geostatistical workflow .....	12-6
Exercise 12: Use the Geostatistical Wizard to perform kriging.....	12-9
Set up the ArcGIS Pro project.....	12-10
Explore the data distribution .....	12-10
Perform kriging using the Geostatistical Wizard .....	12-12
Evaluate predicted value and error.....	12-15
Empirical Bayesian kriging (EBK) .....	12-17
Lesson review.....	12-20

## 13 3D analysis

Lesson introduction .....	13-1
When to use 3D analysis.....	13-2
3D analysis examples.....	13-3
Interactive 3D analysis .....	13-6
Exercise 13: Perform 3D analysis .....	13-8
Set up the project .....	13-9
Create sight lines .....	13-9
Perform line-of-sight analysis .....	13-11
Create a 3D buffer.....	13-13
Intersect 3D features.....	13-15

Lesson review..... 13-18  
Answers to Lesson 13 questions..... 13-19

## Appendices

Appendix A: Esri data license agreement ..... A-1  
Appendix B: Answers to lesson review questions ..... B-1  
Appendix C: Additional resources..... C-1





# Esri resources

Take advantage of these resources to develop ArcGIS software skills, discover applications of geospatial technology, and tap into the experience and knowledge of the ArcGIS community.

## Instructor-led and e-Learning resources

Esri instructor-led courses and e-Learning resources help you develop and apply ArcGIS skills, recommended workflows, and best practices. View all training options at [esri.com/training/catalog/search](https://esri.com/training/catalog/search).

## Planning for organizations

Esri training consultants partner with organizations to provide course recommendations for job roles, short-term training plans, and workforce development plans. Contact an Esri training consultant at [training@esri.com](mailto:training@esri.com).

## Esri technical certification

The Esri Technical Certification Program recognizes individuals who are proficient in best practices for using Esri software. Exams cover desktop, developer, and enterprise domains. Learn more at [esri.com/training/certification](https://esri.com/training/certification).

## Social media and publications

Twitter: [@EsriTraining](https://twitter.com/EsriTraining) and [@Esri](https://twitter.com/Esri)

Esri on LinkedIn: [linkedin.com/company/esri](https://www.linkedin.com/company/esri)

Esri training blog: [esri.com/trainingblog](https://esri.com/trainingblog)

Esri publications: Access online editions of ArcNews, ArcUser, and ArcWatch at [esri.com/esri-news/publications](https://esri.com/esri-news/publications)

Esri training newsletter: Subscribe at [go.esri.com/preferences](https://go.esri.com/preferences)

Other Esri newsletters: Subscribe to industry-specific newsletters at [go.esri.com/preferences](https://go.esri.com/preferences)

## Esri Press

Esri Press publishes books on the science and technology of GIS in numerous public and private sectors. [esripress.esri.com](https://esripress.esri.com)

# Esri resources (continued)

## GIS bibliography

A comprehensive index of journals, conference proceedings, books, and reports related to GIS, including references and full-text materials. [gis.library.esri.com](http://gis.library.esri.com)

## ArcGIS documentation and tutorials

In-depth information, tutorials, and documentation for ArcGIS products.

ArcGIS Online: [arcgis.com](http://arcgis.com)

ArcGIS Desktop: [desktop.arcgis.com](http://desktop.arcgis.com)

ArcGIS Enterprise: [enterprise.arcgis.com](http://enterprise.arcgis.com)

## GeoNet

Join the online community of GIS users and experts. [esri.com/geonet](http://esri.com/geonet)

## Esri events

Esri conferences and user group meetings offer a great way to network and learn how to achieve results with ArcGIS. [esri.com/events](http://esri.com/events)

## Esri Videos

View an extensive collection of videos by Esri leaders, event keynote speakers, and product experts. [youtube.com/user/esritv](http://youtube.com/user/esritv)

## ArcGIS for Personal Use

Improve your GIS skills at home and use ArcGIS to enhance your personal projects. The ArcGIS for Personal Use program includes a 12-month term license for ArcGIS Desktop, extension products, and an ArcGIS Online named user account with 100 service credits. [esri.com/personaluse](http://esri.com/personaluse)

## GIS Dictionary

This term browser defines and describes thousands of GIS terms. <http://support.esri.com/other-resources/gis-dictionary>

# Course introduction

Welcome to *Spatial Analysis with ArcGIS Pro*. In this course, you will learn essential concepts and a standard workflow that you can apply to any spatial analysis project. You will work with various ArcGIS tools to explore, analyze, and produce reliable information from data.

- There is a standard workflow that can be applied to any analysis.
- Every analysis should begin with a question.
- The analysis question and criteria drive the data and tools used in an analysis.
- There are four main types of analysis:

Proximity      Overlay      Statistical      Temporal

- Overlay combines features and attributes, and you can apportion numeric attributes for split features.
- Overlay can be performed on vector or raster data; each uses different tools.

This course will help you understand GIS analysis, which helps people answer questions about their data and the spatial relationships within the data. It teaches a standard GIS analysis workflow that can be applied to any analysis question.

After learning this workflow, you will follow it while performing the four types of analysis to answer real-world questions like the following:

- Where is the best place for bear habitats?
- What factors contribute to Medicare spending in an area?
- Where are all customers within a 15-minute drive time of each store?
- Where are hot spots of graffiti incidents?
- Is there a temporal pattern in graffiti incidents?

## Course goals

After completing this course, you will be able to perform the following tasks:

- Quantify spatial patterns using spatial statistics and analyze change over time to identify emerging hot spots.
- Use interpolation and regression analysis to explain why patterns occur and predict how patterns will change.
- Prepare data and choose appropriate tools and settings for an analysis.
- Examine features and distribution patterns within an area of interest and identify optimal locations using 2D and 3D analysis tools.

## Installing the course data

Some exercises in this workbook require data. Depending on the course format, the data is available on a DVD in the back of a printed workbook or as a data download. To install the data, place the DVD in your disc drive or double-click the data download and follow the instructions in the installation wizard. The data will automatically be installed in the C:\EsriTraining folder.



*DISCLAIMER:* Some courses use sample scripts or applications that are supplied either on the DVD or on the Internet. These samples are provided "AS IS," without warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or noninfringement. Esri shall not be liable for any damages under any theory of law related to the licensee's use of these samples, even if Esri is advised of the possibility of such damage.

# Training Services account credentials

Your instructor will provide a temporary account and group to use during class.

Record the information below:

User name: \_\_\_\_\_

Password: \_\_\_\_\_

Group name: \_\_\_\_\_

Organization URL: \_\_\_\_\_

After completing this course, you will need your own account to perform course exercises that require signing in to ArcGIS Online. The sign-in steps will vary based on your account type.

# Icons used in this workbook



**Estimated times** provide guidance on approximately how many minutes an exercise will take to complete.



**Notes** indicate additional information, exceptions, or special circumstances about specific course topics.



**Recommended practices** improve efficiency and save time.



**Esri Training resources** provide more in-depth training on related topics.



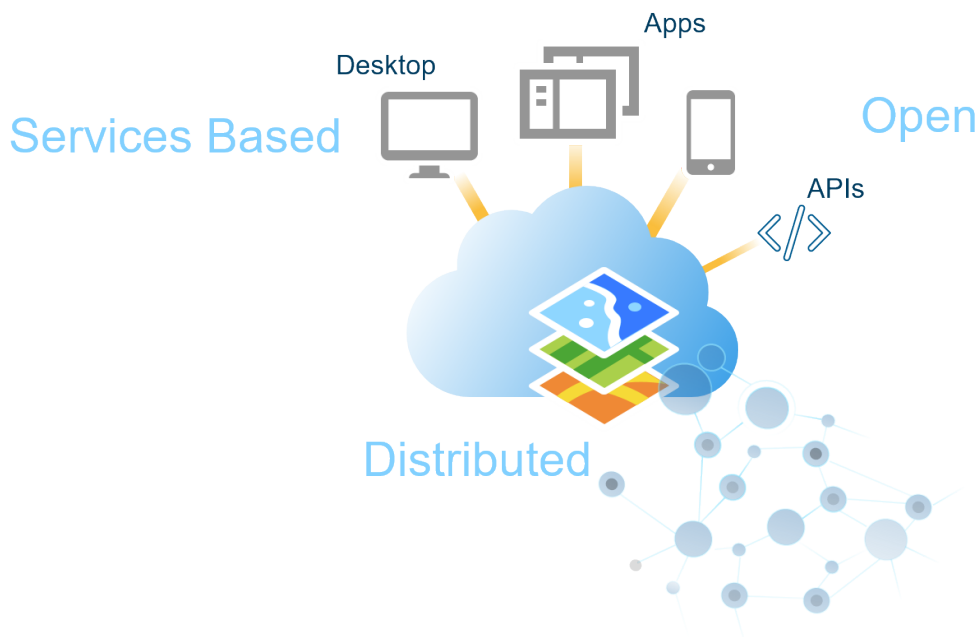
**Additional resources** provide additional information about related topics.



**Warnings** indicate potential problems or actions that should be avoided.

# Understanding the ArcGIS platform

ArcGIS is a Web GIS platform that you can use to deliver your authoritative maps, apps, geographic information layers, and analytics to wider audiences.



**Figure 1.** The ArcGIS platform.

- Individuals interact with ArcGIS through apps running on desktops, in web browsers, and on mobile devices.
- Organizations share their authoritative geospatial data, maps, and tools as web services to a central portal that supports self-service mapping, analytics, and collaboration. Organizations deploy portals in the cloud, in their own infrastructure, or in both.
- Individuals use ArcGIS apps and portals to find authoritative content, create web maps and web apps, perform analytics, and share results.
- Organizations leverage the information shared by individuals to make more informed decisions, communicate with partners and stakeholders, and engage the public.
- A portal is a collaborative space where users can create, analyze, organize, store, and share geospatial content. Within ArcGIS there are two ways to implement a portal: use ArcGIS Online or deploy ArcGIS Enterprise.





# 1

## Building a foundation for spatial analysis

Welcome to *Spatial Analysis Using ArcGIS Pro*, a course that will use spatial analysis to assist you in making important decisions in your work. This lesson introduces spatial analysis and presents a workflow that you can apply to any analytical project and data. You will also learn about the various types of spatial analysis, many of which you will use throughout the course.

### Topics covered

---

Defining and applying spatial analysis

Common analysis questions

ArcGIS Pro analysis tools

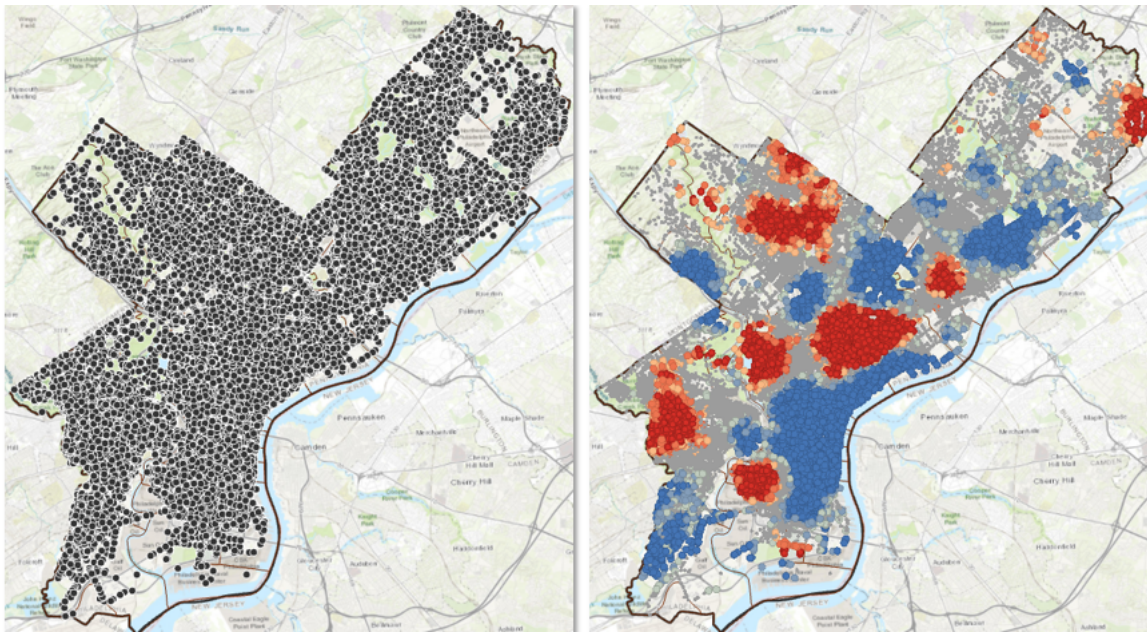
Standard spatial analysis workflow

## What is spatial analysis?

When you look at a map, you think about the features and relationships that you see. If the map illustrates wildlife habitats, you might conclude that some animals chose particular areas for forest cover or proximity to water. If the map illustrates fire risk areas, you might conclude that the fire risk comes from a certain vegetation type, the lack of rainfall, wind exposure, or a combination of them all. Based on what you see in a map, you draw conclusions that reflect your understanding of spatial data.

But sometimes the map's visual elements are not enough for you to understand what is occurring or why.

**Can you make any assumptions based on the following crime locations?**



**Figure 1.1.** Points on the left show where crimes have occurred. Results of spatial analysis on the right show hot and cold spots of crime incidents.

When you cannot rely solely on a map's visual elements to answer questions, you can perform spatial analysis. Spatial analysis is the process of examining the locations, attributes, and relationships of features in spatial data to help gain a better understanding or answer questions.

## Benefits of spatial analysis

---

In this short video, Lauren Bennett, a product engineer on the Esri spatial analysis team, discusses the benefits of spatial analysis.

**1. How do you plan to use spatial analysis in your work?**

---

---

---

## Benefits of spatial analysis (continued)

---

Spatial analysis can provide numerous benefits, such as reduction of costs and increases in efficiency, productivity, and revenue. Spatial analysis sets true GIS software, including ArcGIS Pro, apart from other map-viewing applications.

## Common analysis problems

---

You have probably performed some type of nonspatial analysis before. When you add a geographical element to your questions and subsequent decisions, you make analysis more complex by adding spatial properties like distance and direction. Spatial properties have a significant effect on the analytical methods that you use to solve a particular problem.

To help classify spatial properties, analytical problems are categorized into six groups. Each group reflects a set of related questions, described in the following table.

Analysis problem	Description
<b>Understand where</b>	At the most rudimentary level, you are lost if you do not know where you are or what is around you. Asking "Where?" is the first question in spatial analysis.
<b>Measure size, shape, and distribution</b>	You may want to describe an object in terms of its geometry, such as area, perimeter, or length. You may also want to describe the distribution of several objects.
<b>Determine how places are related</b>	You may need to describe and quantify the relationships among features to determine what is near, what is within, or how something overlaps in space and time.
<b>Find the best locations and paths</b>	You may need to find the best route to travel, or the best location to build a new storefront or station.
<b>Detect and quantify patterns</b>	You may need to find patterns in data, such as hot spots or outliers. You may also need to determine how those patterns change over time.
<b>Make predictions</b>	You may need to determine how things may appear in the future or how crime or fire danger will spread.

## Spatial analysis tools

You can use different types of spatial analysis to answer questions. Most real-world GIS spatial analyses may use several types at one time to solve spatial problems.

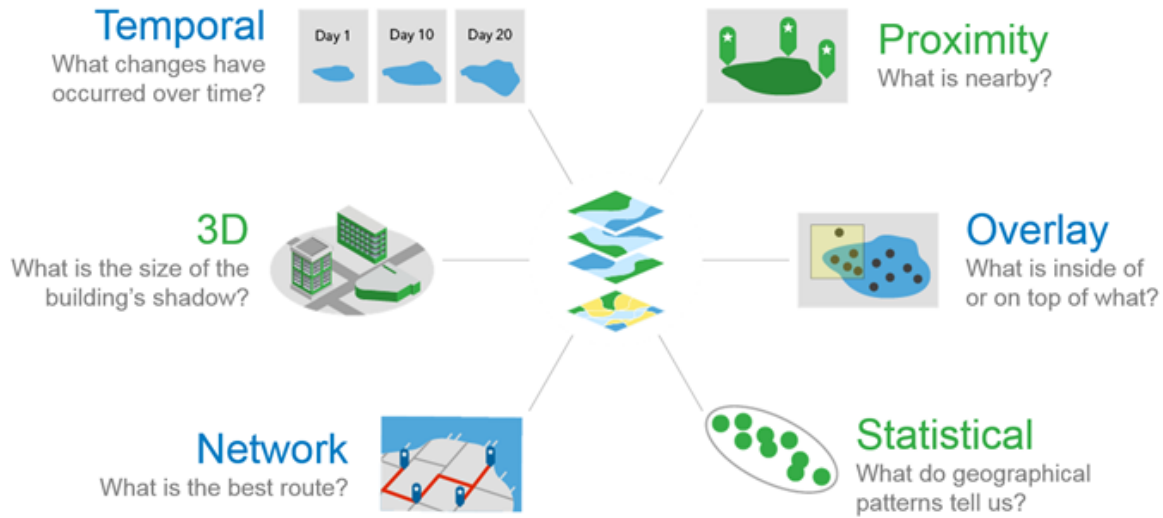


Figure 1.2. Six types of spatial analysis.

Analysis type	Description
<b>Temporal</b>	<ul style="list-style-type: none"> <li>• Clarifies patterns in specific types of data (such as incident data)</li> <li>• See how data or patterns change over time</li> </ul>
<b>Proximity</b>	<ul style="list-style-type: none"> <li>• Determines which features are close to other features, the exact distance between features, or which features are within a certain distance of other features</li> </ul>

## Spatial analysis tools (continued)

Analysis type	Description
<b>Overlay</b>	<ul style="list-style-type: none"><li>• Examines interactions among spatial phenomena</li><li>• GIS overlay tools combine features and attributes from multiple layers to create new information</li></ul>
<b>Statistical</b>	<ul style="list-style-type: none"><li>• Identifies and quantifies patterns or relationships in data to extract additional information that may not be obvious from maps</li><li>• Predicts data values at unknown locations or model relationships among data variables</li></ul>
<b>Network</b>	<ul style="list-style-type: none"><li>• Determines solutions for complex routing problems to help locate the best, most cost-effective path for delivering resources</li></ul>
<b>3D</b>	<ul style="list-style-type: none"><li>• Enables users to view and analyze data in 3D to solve more complex analysis questions, such as subsurface analysis</li></ul>

## Spatial analysis workflow

The analysis workflow provides a framework for you to plan, organize, execute, and share your spatial analysis project. The analysis process may not always be linear. Sometimes, after the initial examination of the analysis results, you may have more questions that require another smaller, more focused analysis before you can answer the initial question.



**Figure 1.3.** The spatial analysis workflow contains standard steps that you can apply to any analysis.

Workflow step	Description
<b>1. Ask questions</b>	<ul style="list-style-type: none"> <li>Determine the questions and the criteria, which determines the data</li> </ul>
<b>2. Explore and prepare data</b>	<ul style="list-style-type: none"> <li>Choose data based on criteria</li> <li>Review data for properties appropriate for the analysis (for example, vector or raster)</li> <li>Prepare data as necessary (for example, modify spatial reference, edit features)</li> </ul>



## Spatial analysis workflow (continued)

---

Workflow step	Description
<b>3. Analyze and model</b>	<ul style="list-style-type: none"><li>• Choose the appropriate methods and tools and run</li></ul>
<b>4. Interpret results</b>	<ul style="list-style-type: none"><li>• View and interpret results to identify flaws or errors in the process</li></ul>
<b>5. Repeat or modify</b>	<ul style="list-style-type: none"><li>• Refine analysis parameters and run tools again</li></ul>
<b>6. Present results</b>	<ul style="list-style-type: none"><li>• Show and discuss results with stakeholders</li><li>• Create maps, tools, and processes to share</li></ul>
<b>7. Make decisions</b>	<ul style="list-style-type: none"><li>• Use analysis results to answer the initial question and make decisions</li></ul>

## Applying spatial analysis

---

You have learned about various spatial analysis tools and a standard workflow. You will apply what you have learned to determine the possible ways to solve a spatial problem.

### Scenario 1: Siting a regional distribution center

An investment firm recently purchased a chain of neighborhood grocery stores that was almost bankrupt. As part of the firm's strategy for revitalizing the business, it is planning to expand into new markets in the next few years. The firm has identified one particular potential metropolitan region as a good candidate for expansion, and is looking for locations for a regional distribution center.

- 1. How can spatial analysis help identify the best location for the distribution center?**

---

---

---

---

- 2. Which types of analysis tools would you use to locate a suitable site for the distribution center?**

---

---

---

## Lesson review

---

**1. What are the six types of spatial analysis tools?**

---

---

---

**2. What helps you choose the appropriate datasets for your analysis?**

---

---

---

# Answers to Lesson 1 questions

## Benefits of spatial analysis (page 1-3)

1. How do you plan to use spatial analysis in your work?

**Answers will vary based on personal experience.**

## Applying spatial analysis (page 1-10)

### Scenario 1: Siting a regional distribution center

1. How can spatial analysis help identify the best location for the distribution center?

**Spatial analysis can provide the following information:**

- Sites that meet space requirements
- Sites near major highways or freeways for easy access and truck shipping
- Sites that are safe from flood zones
- Sites that are located near residential areas with prospective employees

2. Which types of analysis tools would you use to locate a suitable site for the distribution center?

- Proximity
- Overlay
- Network for routing and delivery

# 2

## Planning and preparing for spatial analysis

As you progress through the spatial analysis workflow, you will see that exploring your data can be a time-consuming process. It takes time to examine data quality and completeness, and to determine whether the analysis of your data will yield the results you want. In this lesson, you will explore data and learn how to modify your existing datasets efficiently to ensure optimal analysis results.

### Topics covered

---

■ Important data properties

■ Raster data considerations

■ Environment settings

## Data properties

---

Preparing to perform analysis involves criteria that are useful for answering your analysis questions. These criteria include metadata properties, which render your data easier to use and more effective. You will now discuss several useful metadata properties.

**What data properties are important to consider when performing analysis?**

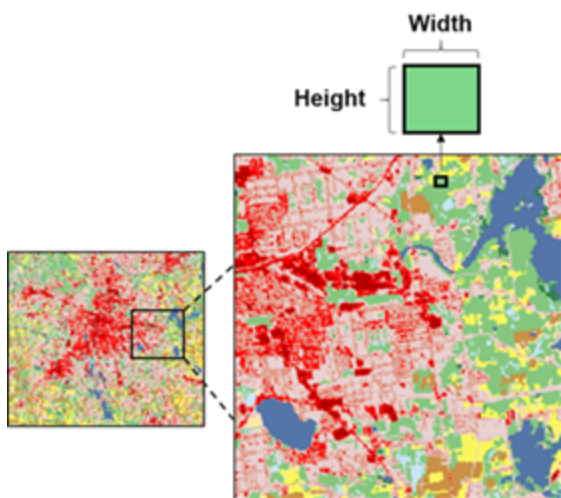
## Raster data considerations

---

When you perform analysis with raster data, you will consider things that you do not normally consider with vector data, such as cell size, masks, and NoData values. You can work with all the raster data considerations mentioned here as geoprocessing environment settings.

### Cell size

Cell size refers to the ground dimensions of a single cell in a raster, measured in map units. Cell size is determined at the point of data capture based on the scale and device. You must determine the best cell size for your data and analysis as it is often a parameter in analysis tools. With rasters of varying resolutions, your analysis results are only as accurate as the lowest resolution dataset.



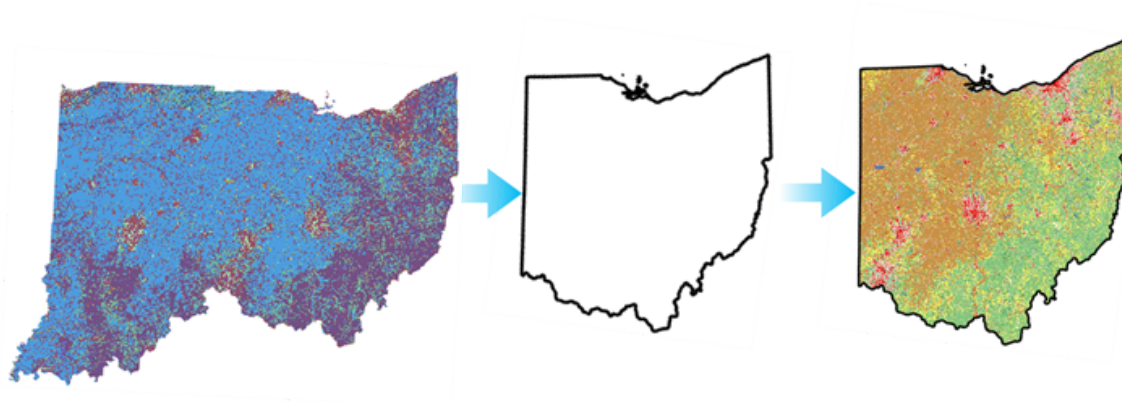
**Figure 2.1.** Each raster has a specific cell size. Cell size is often used synonymously with pixel size or resolution. The larger the cell size, the less detail, or resolution; the smaller the cell size, the more detail, or resolution.

## Raster data considerations (continued)

---

### Focusing your analysis using a mask

A mask is an area used to limit the cells on which analysis is performed, thus reducing map clutter, processing time, and other resources. You can set a raster or vector dataset as a mask in the environment settings before performing analysis. A mask is different from the extent.



**Figure 2.2.** The mask is the Ohio state boundary. The result of processing with that mask includes only data within the mask.

### Other key raster considerations

**NoData:** Most cells contain values, whether they are integer or floating point (decimal) values. Cells that do not have a recorded value are tagged as NoData. NoData does not equate to a zero value.



If rasters for your analysis contain NoData values, you can set ArcGIS Pro to ignore those cells or to estimate values for them.

**Extent:** The minimum bounding rectangle that defines an area of analysis. Extent could be another layer or the current extent of the map.

**Resampling:** The process of aggregating or interpolating new cell values when transforming rasters to a new coordinate space or cell size.



## Environment settings

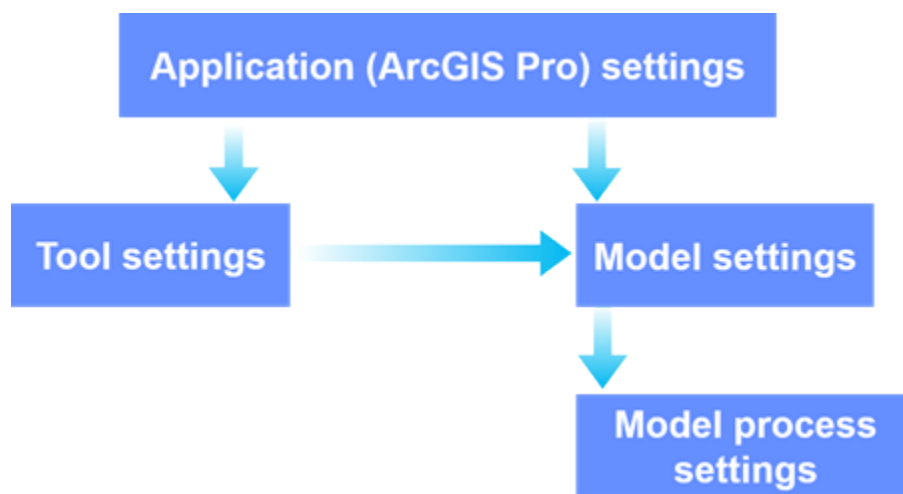
---

### What are environment settings and why should you use them?

Environment settings are background settings that directly affect tool outputs. Environment settings help ensure consistent analysis results.

### Environment setting hierarchy

You can set environment settings in several locations. Where you set an environment is important because your choice affects the hierarchical structure that controls how tools run. If you set an environment at the application level, it is applied to all tools. If you set an environment in the tool, model, or model process, then those choices override the application-level setting. If you run the model from the ModelBuilder, the application environment is passed down. If you run the model using its tool dialog box, the tool environment is passed down.



**Figure 2.3.** Environment settings hierarchy.

## Prepare data for analysis

When you have identified your analysis question and criteria, you should have a good idea of the required data. The second step of the analysis workflow is to explore data that you have and, if necessary, change or acquire new data to replace or supplement the existing data.

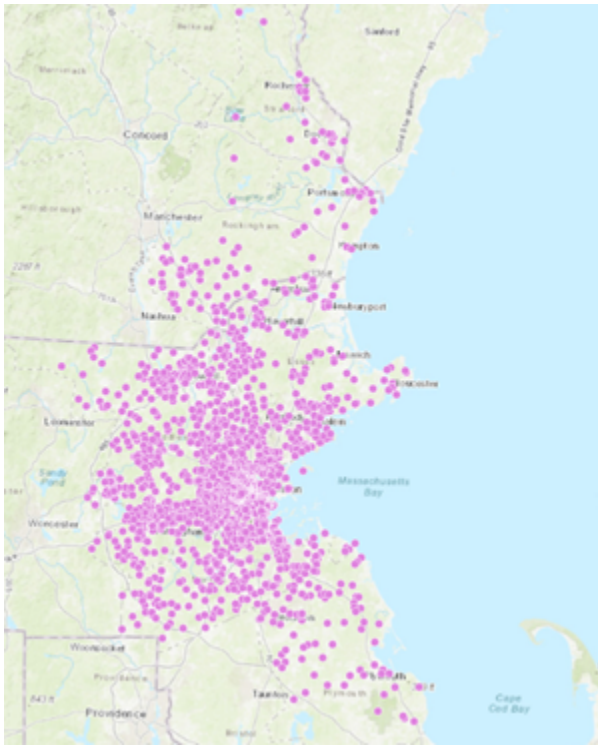
In this exercise, you will perform the following tasks:

- Change the coordinate system of a dataset.
- Display x,y data from a nonspatial table.
- Join attributes.
- Extract features and raster cells.

## Step 1: Set up an ArcGIS Pro project

You will open a project that has been created for you and use it for all course exercises.

- a Start ArcGIS Pro.
- b In the ArcGIS Sign In dialog box, click Enterprise Login.
- c For the organization's URL, type **trainingservices** and click Continue.
- d For Using, click Your Course Account.
- e Type the organizational account user name and password provided to you by your instructor and click Sign In.
- f In the bottom-left corner of the window, click Open Another Project.
- g Browse to **C:\EsriTraining\SNAP\SNAPCourse**, and double-click SNAPCourse.aprx.



The map displays customer locations in the Boston, Massachusetts, area. You will use the customer locations for several different analyses in the course.

- h In the Catalog pane, expand Folders, and then expand SNAP to see the course folder structure.

The course data is stored in several folders and geodatabases, but you will create outputs in the project geodatabase. By default, the output workspace is already set to the project geodatabase in the environment settings.

- i From the Analysis tab, in the Geoprocessing group, click Environments.

Workspace	
Current Workspace	SNAPCourse.gdb
Scratch Workspace	SNAPCourse.gdb

All geoprocessing outputs will be stored in the SNAPCourse geodatabase.



The scratch workspace differs from the current workspace in that it is designed for output data that you do not want to maintain. The primary purpose of the scratch workspace is for use in ModelBuilder and Python scripts.

- j Click OK.

## Step 2: Change the coordinate system for a feature class

Next, you will check the coordinate system of the Customers layer.

- a In the Contents pane, right-click Customers and choose Properties.
- b Click the Source tab.
- c Scroll down and expand Spatial Reference.

### ▼ Spatial Reference


Projected Coordinate System	NAD 1927 UTM Zone 19N
-----------------------------	-----------------------


The Customers layer is stored in NAD 1927 UTM Zone 19N. Your organization has standardized a coordinate system of NAD 1983 StatePlane FIPS 2001. For analysis, it is best to store all data in the same coordinate system to ensure consistent results.

- d Click OK.

Next, you will find the Project tool to reproject data. All licensed geoprocessing tools are available in the Geoprocessing pane. Some commonly used tools are located in the Analysis Gallery.

- e From the Analysis tab, in the Geoprocessing group, click Tools.
- f In the search field, type **project**.
- g Click Project (Data Management Tools) to open the tool.

- h In the Geoprocessing pane, set the following parameters:
  - Input Dataset Or Feature Class: Customers
  - Output Dataset Or Feature Class: **Customers**
  - Output Coordinate System: Click the Select Coordinate System button .
    - In the Coordinate system dialog box, type **1983 StatePlane Massachusetts** and press Enter.
    - Expand Projected Coordinate System, State Plane, and NAD 1983 (Meters), and then click NAD 1983 StatePlane Massachusetts FIPS 2001 (Meters) and click OK.
  - Geographic Transformation: Use the default setting.

Input Dataset or Feature Class	Customers
Output Dataset or Feature Class	Customers
Output Coordinate System	NAD_1983_StatePlane_Massachusetts_Mainland_FI
Geographic Transformation 	NAD_1927_To_NAD_1983_NADCON

- i Click Run.
- j In the Contents pane, double-click the first Customers layer to open its properties.
- k From the Source tab, view the Spatial Reference information.

Projected Coordinate System	NAD 1983 StatePlane Massachusetts Mainlanc
Projection	Lambert Conformal Conic
WKID	26986
Authority	EPSG
Linear Unit	Meter (1.0)

The updated coordinate system information indicates that your data is in the correct coordinate system for analysis. The coordinate system of the map is still set to NAD 1927, so you will change that.

- l Close the Layer Properties dialog box.
- m In the Contents pane, right-click Boston and choose Properties.
- n Click the Coordinate Systems tab.

Select the Coordinate System to view the available options.

Current XY Details Current Z

NAD 1927 UTM Zone 19N

XY Coordinate Systems Available

Favorites

Layers

- ▶ NAD 1983 StatePlane Massachusetts Mainland FIPS 2001
- ▶ **NAD 1927 UTM Zone 19N** ★

Customers

- o In the search field, type **1983 StatePlane Massachusetts** and press Enter.
- p Under Projected Coordinate System, expand State Plane and NAD 1983 (Meters), and then select NAD 1983 StatePlane Massachusetts FIPS 2001 (Meters).

Current XY Details

NAD 1983 StatePlane  
Massachusetts FIPS 2001 (Meters)

- q Click OK.
- r In the Contents pane, right-click the second Customers layer (the one in NAD27) and choose Remove.
- s Save the project.


### Step 3: Create a feature class from x,y coordinates

You have a nonspatial table that contains the coordinates for stores. You will use a tool called XY Table To Point to create a point feature class from the coordinates.

- a In the Contents pane, right-click Stores and choose Open.

X_CoordinateStore	Y_CoordinateStore
260646.359262	849426.718898
250355.862553	930693.41383
234848.780217	947347.501297
208604.802373	868336.315006
210536.458501	906299.556149
233178.484555	1002825.128011
232735.393551	899445.669823
203411.620062	929203.037232
252774.873788	879401.332402

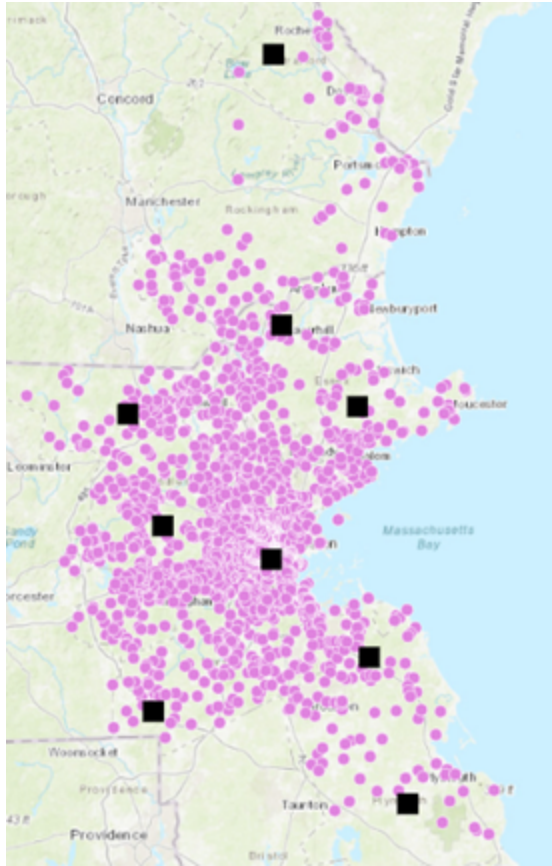
You will use the x- and y-coordinates to create the stores spatially.

- b** Close the table.
- c** At the top of the Geoprocessing pane, click the Back button .
- d** Search for **xy**, and find and open the XY Table To Point tool.
- e** In the Geoprocessing pane, set the following parameters:
  - Input Table: Stores
  - Output Feature Class: **BostonStores**
  - X Field: X\_CoordinateStore
  - Y Field: Y\_CoordinateStore
  - Coordinate System: From the drop-down list, choose Current Map [Boston]



The coordinate system should update to NAD\_1983\_StatePlane\_Massachusetts\_Mainland\_FIPS\_2001.

- f** Click Run.
- g** In the Contents pane, under the BostonStores layer, click the symbol and assign the Square 1 symbology to it.



The stores are now part of your geodatabase as their own feature class.

- h** Close the Symbology pane.
- i** In the Contents pane, right-click BostonStores and choose Attribute Table.

OBJECTID	Shape	STORE_NUM	X_CoordinateStore	Y_CoordinateStore	Latitude	Longitude
----------	-------	-----------	-------------------	-------------------	----------	-----------

BostonStores contains the same attributes as the original Stores table. However, some attributes are not present, such as the name and address.

- j** From the Contents pane, open StoresTable.

OBJECTID	CONAME	ADDR	CITY16	STATE	ZIP	STORE_NUM	NUMBER_EMP
1	FITNESSLAND-SOUTH	187 SUMMER ST # 5	KINGSTON	MA	02364	1	38
2	FITNESSLAND-ASBU...	307 LOWELL ST	ANDOVER	MA	01810	2	40
3	FITNESSLAND-HAVE...	2 WATER ST # 5	HAVERHILL	MA	01830	3	34
4	FITNESSLAND-FRAN...	750 UNION ST	FRANKLIN	MA	02038	4	65

The StoresTable contains address information and the number of employees for each store.

**1. How can you join the attributes from StoresTable to the BostonStores feature class?**

---



- k Save the project.

## Step 4: Enhance data using a table join

You want to include all store attributes in the BostonStores layer, so you will perform a table join. You will use the common field, STORE\_NUM, to join attributes from one table to another.

- a Make BostonStores the active table.
- b In the Contents pane, right-click BostonStores, point to Joins And Relates, and choose Add Join.
- c In the Geoprocessing pane, set the following parameters:
  - Layer Name Or Table View: BostonStores
  - Input Join Field: STORE\_NUM
  - Join Table: StoresTable
  - Output Join Field: STORE\_NUM
- d Click Run.

STORE_NUM	X_CoordinateStore	Y_CoordinateStore	Latitude	Longitude	OBJECTID	CONAME	ADDR	CITY16	STATE	ZIP
1	260646.359262	849426.718898	41.892838	-70.769205	1	FITNESSLAND-SOUTH	187 SUMMER ST # 5	KINGSTON	MA	02364
2	250355.862553	930693.41383	42.625183	-70.886155	2	FITNESSLAND-ASBU...	307 LOWELL ST	ANDOVER	MA	01810
3	234848.780217	947347.501297	42.77595	-71.074171	3	FITNESSLAND-HAVE...	2 WATER ST # 5	HAVERHILL	MA	01830
4	208604.802373	868336.315006	42.06538	-71.396031	4	FITNESSLAND-FRAN...	750 UNION ST	FRANKLIN	MA	02038

The attributes from StoresTable are added into BostonStores. Joins are stored within the project, but you could also export the layer that the join is based on into the geodatabase. You will keep the join as a virtual join for your analysis so that you can accomplish the same tasks without creating another feature class.

- e Close both tables.
- f Close the Boston map, and save the project.

## Step 5: Import a map file for a different study area

Next, you will import a map file that contains layers for an analysis that you will perform later. Map files contain the definition of maps that you have built, which others can import into their projects.

- a From the Insert tab, in the Project group, click Import Map.
- b Browse to `..\EsriTraining\SNAP\Prepare` and import Ohio.mapx.

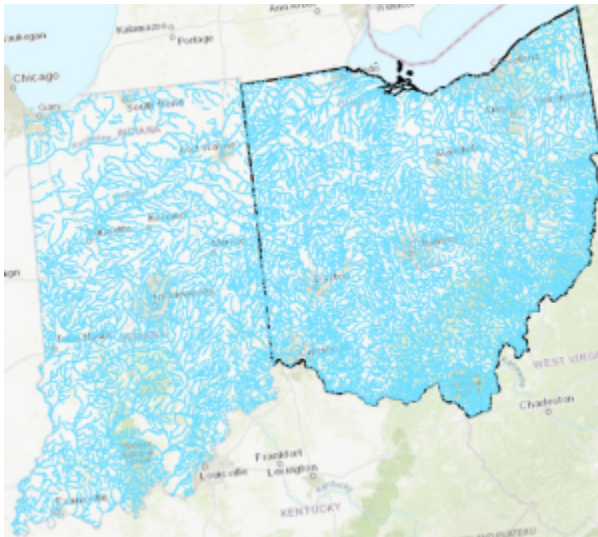


The map contains streams and land-use classifications for Ohio and Indiana, hydrologic unit boundaries for several states, and the state boundary of Ohio. Your analysis will focus on Ohio, so you will use the Ohio boundary layer to extract other features to narrow down the data.

## Step 6: Extract features using the Clip tool

Now, you will prepare the data by extracting the areas of interest (AOI). Extracting features in an AOI is a common data preparation step before analysis.

- a In the Contents pane, turn off Region5\_HUC8 and NLCD\_OhioInd, and turn on OhioIndStreams.



The Ohio state boundary will act as a "cookie cutter" to extract only the streams that are within it.

- b From the Analysis tab, in the Tools group, click Clip.
- c In the Geoprocessing pane, set the following parameters:

- Input Features: OhioIndStreams
- Clip Features: Ohio



You can also interactively draw the clip features for this parameter using the pencil icon.

- Output Feature Class: **OhioStreams**

- d Click Run.
- e In the Contents pane, remove OhioIndStreams.



You have extracted the streams for Ohio that you will use for overlay analysis. Next, you will extract the hydrologic unit boundaries for Ohio.

- f At the top of the Catalog pane, click the History tab.

All tools that you have run and their associated parameters are saved in the Geoprocessing history in each project. You can see the tools that you used in this exercise, including the Clip tool.

- g Double-click the Clip tool to open it and view its parameters.

You can quickly modify a parameter and rerun a tool from its history.

**h** Update the following parameters:

- Input Features: Region5\_HUC8
- Output Feature Class: **OhioHUC**



The Clip Features parameter (Ohio) remains the same.

**i** Run the Clip tool again.

**j** In the Contents pane, turn off OhioStreams.



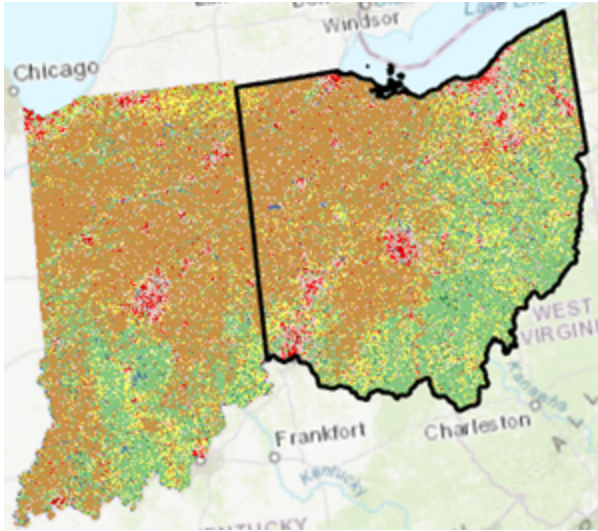
The hydrologic units are clipped to the Ohio state boundary.

**k** Save the project.

## Step 7: Extract raster data using a mask

You have successfully clipped vector features for streams and hydrologic unit boundaries within Ohio. Next, you will extract raster cells that fall within Ohio using a Spatial Analyst tool.

**a** In the Contents pane, make Ohio, NLCD\_OhioInd, and the basemap the only visible layers.



It is important to know the cell size of the input raster so that the output is the same. You can set cell size as an environment setting.

- b** From the Contents pane, open the properties for NLCD\_OhioInd.
- c** From the Source tab, expand Raster Information.

The cell size for this raster dataset is 30 meters.

- d** Close the Layer Properties dialog box.
- e** From the Analysis tab, click Environments.

In the Raster Analysis section, you can see that Cell Size is set to Maximum Of Inputs. This setting ensures that a smaller, higher-resolution cell size will not be applied to the output raster, which implies higher-quality data. For the data that you are working with, leaving the cell size environment as it is set will result in output rasters with 30-meter cell size.

- f** Close the Environments dialog box.

You want to extract the cells that fall within Ohio, but the Clip tool only works on vector data. When you want to extract and analyze raster data, you must use Spatial Analyst tools. Even though you are extracting cells in a raster dataset, you can use either a vector or raster dataset as the mask.

- g** In the Geoprocessing pane, click the Back button, and then click the Toolboxes tab, if necessary.
- h** Expand Spatial Analyst Tools, and then expand Extraction.



i Click Extract By Mask, and then set the following parameters:


- Input Raster: NLCD\_OhioInd
- Input Raster Or Feature Mask Data: Ohio
- Output Raster: **NLCD\_Ohio**

j Click Run.

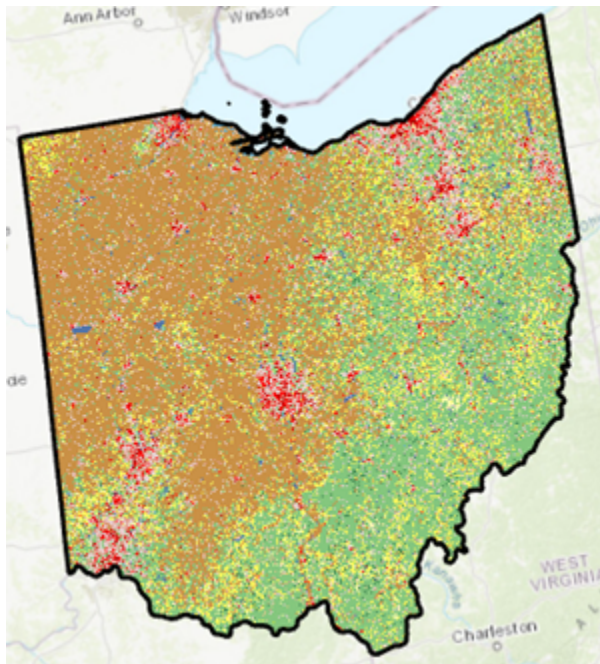
k In the Contents pane, turn off NLCD\_OhioInd.

The output raster does not have the standard NLCD symbology, so you will import it from a layer file.

l In the Contents pane, right-click NLCD\_Ohio and choose Symbology.

m In the top-right corner of the Symbology pane, click the Options menu button  and choose Import.

n Browse to `..\EsriTraining\SNAP\Prepare`, select `NLCD_OhioInd.lyrx`, and click OK.



Now, your new raster symbology matches the NLCD standard.

o Close all panes except the Contents and Catalog panes.

p In the Catalog pane, return to the Project tab.

q Close the Ohio map, save your project, and keep ArcGIS Pro open.



## Lesson review (continued)

---

**3. What should you consider when selecting an output cell size?**

---

---

---

---



# Answers to Lesson 2 questions

## Data properties (page 2-2)

What data properties are important to consider when performing analysis?

**Possible responses include the following:**

- Data format (tabular or spatial)
- Data source
- Quality
- Currency
- Extent
- Spatial reference
- Scale
- Raster resolution
- Attributes

## Exercise 2: Prepare data for analysis (page 2-6)

1. How can you join the attributes from StoresTable to the BostonStores feature class?

**Use the STORE\_NUM field in both tables as the common field and perform a table join.**



# 3

## Proximity analysis

Proximity analysis helps answer questions about the distances between features. It helps you understand details about features in close proximity to one another and features that are distant from one another.

ArcGIS Pro provides numerous proximity analysis tools that are designed to help answer various questions about proximal relationships. You will learn how ArcGIS Pro measures distance, the various data types on which you can use proximity tools, and how to apply several tools to answer spatial questions.

### Topics covered

---

Using proximity analysis

Measuring distance

Types of proximity analysis

Determining cost

## Using proximity in everyday life

---

How does proximity play a role in daily life? People think spatially every day. Many of your spatial thoughts pertain to proximity. How far away is the store? What is the best route to work?

- 1. What is the first thing that you do when figuring out how to get to a new place?**

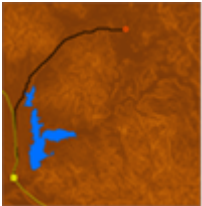

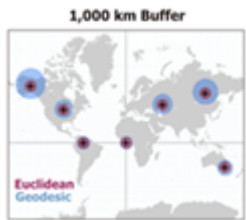
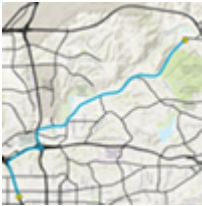
---

---

---

## Choosing the best distance measure

ArcGIS Pro tools calculate distance in four ways: cost, Euclidean, geodesic, and network.

Distance measurement	Description	
<b>Cost</b>	<ul style="list-style-type: none"> <li>• Determines unfavorable impact or impedance associated with moving over a geographic area</li> <li>• Finds the least impedance across a surface</li> </ul>	
<b>Euclidean (also referred to as planar)</b>	<ul style="list-style-type: none"> <li>• Straight-line distance on a flat map</li> </ul>	
<b>Geodesic</b>	<ul style="list-style-type: none"> <li>• Distance between points on the earth's surface that considers the earth's curvature</li> </ul>	
<b>Network</b>	<ul style="list-style-type: none"> <li>• Determines cost over a linear network</li> <li>• Travel time, best routes, and traffic conditions</li> </ul>	



## Ways to measure distance

---

The various ways to measure distance in proximity tools are all applicable in certain situations and with certain data. For each example, choose the best way to measure distance in proximity analysis.

### Scenario 1: Locating liquor stores near schools

Your study area is a small city. You want to buffer each school to see how many liquor stores are located within 3 miles. This information will help officials identify those liquor stores that students may try to patronize.

1. **Which distance measurement would you use when you run the Buffer tool on small-extent (large-scale) data?**
- 

### Scenario 2: Determining economic zones

You want to buffer coastlines for a continent by 200 miles to determine exclusive economic zones.

2. **Which distance measurement would you use when you run the Buffer tool on large-extent (small-scale) data?**
- 

### Scenario 3: Placing a pipeline

An oil company wants to run a pipeline for 2,000 kilometers over varying terrain to find the path that offers the least resistance.

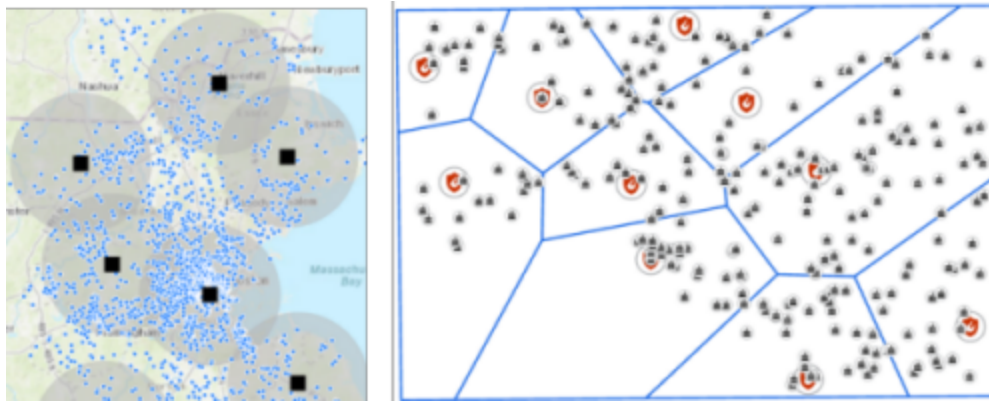
3. **Which proximity distance measurement is most suitable for finding the best path for a pipeline?**
-

## Outputs of proximity analysis

The proximity analysis tools in ArcGIS Pro create different kinds of outputs—either an expanded area or a numeric value.

### Area expanding

Some proximity analysis tools create polygon or raster data that represents a specific distance or a proximity zone, used for allocation.



**Figure 3.1.** Buffer and Thiessen polygons are examples of area-expanding proximity tools.

### Area-expanding proximity tools

Tool	Description
Buffer	Creates polygons around features representing specified distance.
Multiple Ring Buffer	Creates multiple polygons around features based on several specified distances.
Create Thiessen Polygons	Creates proximity zones around points indicating that all the area within a zone is closer to the point within it than to any other point.
Euclidean Allocation	Similar to Create Thiessen Polygons, but the result is a raster dataset that can be used for raster analysis.

## Outputs of proximity analysis (continued)

---

### Numeric value

The numeric values returned by some tools are distances from other features, and the x,y coordinates of the closest feature.

NEAR_FID	NEAR_DIST	NEAR_X	NEAR_Y
2	14299.041724	250355.8626	930693.4138
2	15583.098479	250355.8626	930693.4138

**Figure 3.2.** The Near tool returns a numeric value of the closest feature's ID, distance, and coordinates.

### Proximity tools that return numeric values

Tool	Description
Near	Determines the closest features in two different layers and appends the closest feature's ID and distance to input table. Adding the closest feature's coordinates is optional.
Generate Near Table	Determines the distances between features, within a specified search radius, and creates table of distances.

### Spatial Analyst toolbox

The Spatial Analyst toolbox contains the Distance toolset. The Distance toolset contains many distance analysis tools, such as tools for determining cost path and cost distance over a continuous surface.

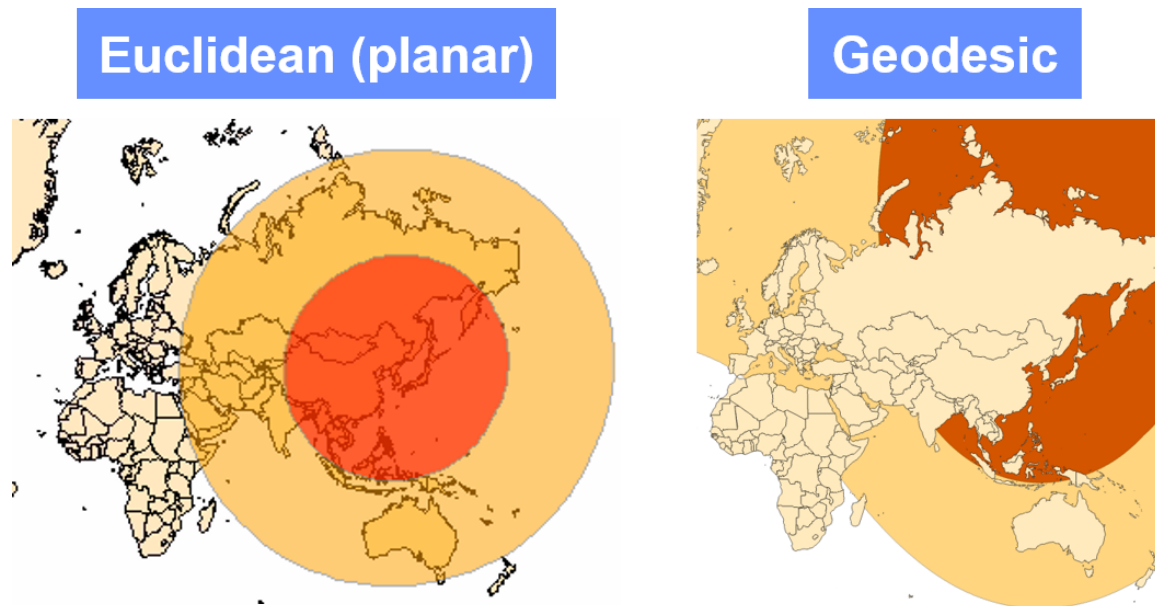


## Buffering using different distance measures

---

When buffering at a small scale, such as a world scale, you should use geodesic measurements. Geodesic measurements are the only true distance; they consider the curvature of the earth, unlike other proximity metrics like Euclidean or cost. When you buffer at a large scale, you can use planar because you are probably in a 2D coordinate system.

It is important to work with a projection that properly preserves distance at the given scale.



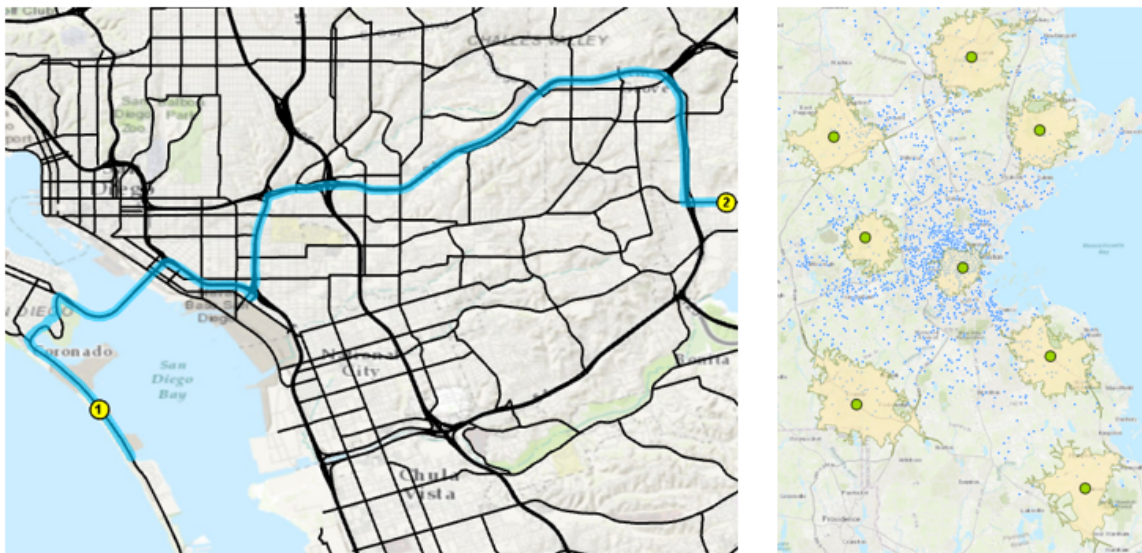
**Figure 3.3.** Shown is the difference between straight-line and geodesic distance measurements used for 5,000-kilometer and 10,000-kilometer buffers around North Korea.

## Measuring cost

---

Cost is another way in which distance is measured in ArcGIS Pro. Cost is the amount of unfavorable impact or impedance associated with moving across a geographic area. A common example of cost is time; it may cost more time to travel one route than another.

You can analyze cost using the Network Analyst extension. Network Analyst allows you to assess cost over a linear network, such as roads. You can use Network Analyst to determine the best routes for delivery companies to deliver packages, determine driving times, and allocate resources.



**Figure 3.4.** You can analyze cells in a raster dataset to create a least-cost path for transporting resources. You can also use Network Analyst to find the best route or driving times from specific locations.

You can also analyze surfaces to create a cost distance or cost-path surface. For example, assume that you have an elevation surface and want to go from point A to point B. The higher the elevation, the higher the cost of traveling that cell.



**[Esri Training courses: \*Creating Optimized Routes Using ArcGIS Pro, Creating an Origin-Destination Cost Matrix in ArcGIS Pro, and Finding the Closest Facilities Using ArcGIS Pro\*](#)**

## Analyze proximity

You have prepared store and customer data for Boston. You will use the data to perform proximity analysis to determine store customer relationships, distances, and driving times for each store based on an online service.

In this exercise, you will perform the following tasks:

- Use ArcGIS Pro proximity analysis tools.
- Add and calculate attributes.

## Step 1: Prepare the project

You will prepare your ArcGIS Pro project for analysis by adding layers and changing some of their display properties.

- a If necessary, restore the SNAPCourse project.
- b From the Catalog pane, expand Maps and open the Boston map.



If you did not finish the previous exercise, import the result map file named Ex2.mapx from **C:\EsriTraining\SNAP\Results\Ex02** to begin this exercise. Result maps are provided for all exercises in the Results folder.

- c Change the name of the Boston map to **Proximity**.
- d In the Contents pane, change the name of the BostonStores layer to **Stores**.
- e Remove the Stores and StoresTable stand-alone tables.

You will also change the color of the Customers layer to blue.


- f Click the Customers symbol, and then in the Symbology pane, click the Properties tab.
- g Update the Color to a dark blue of your choice, and then click Apply.

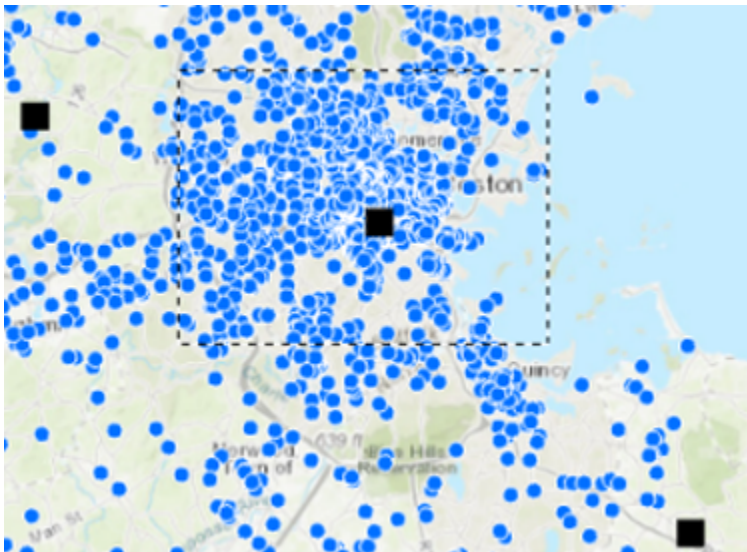


- h Close the Symbology pane.

## Step 2: Select features based on distance

One of the simplest forms of proximity analysis is to select features within a certain distance of other features.

- a At the top of the Contents pane, click the List By Selection button .
- b Right-click Stores and chose Make This The Only Selectable Layer.
- c From the Map tab, in the Selection group, click the Select tool.
- d Select the following store (the one closest to downtown Boston).



Next, you will select all customers within 5 miles of the chosen store.

- e From the Map tab, click Select By Location.

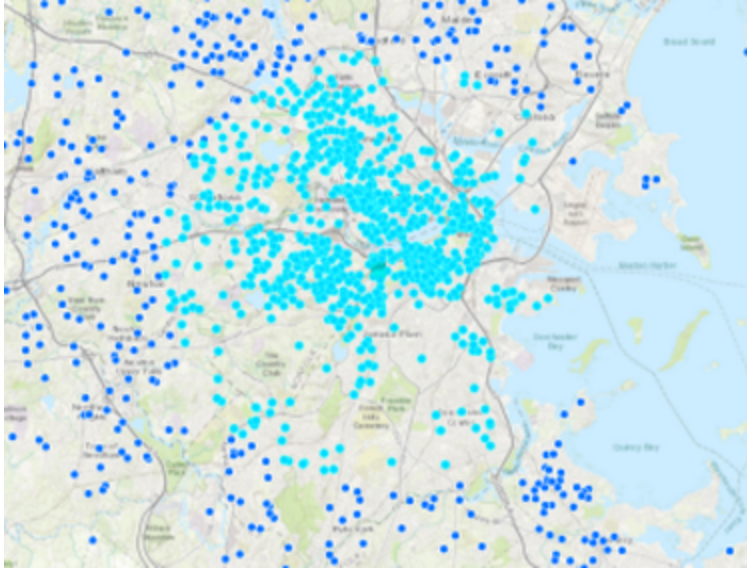
The Select Layer By Location tool opens in the Geoprocessing pane.

- f In the Geoprocessing pane, set the following parameters:

- Input Features: Customers
- Relationship: Within A Distance
- Selecting Features: Stores
- Search Distance: **5 Miles**

- g Click Run.

- h Zoom in to the selection.




For all geoprocessing tools, if a selection is present in an input layer, processing takes place only on the selected features.

You have located customers within 5 miles of a store. You could export these customers to their own feature class for further analysis or create a selection layer in the project.

- i** Make all layers selectable.

*Hint:* In the Contents pane, check the Customers box.

- j** At the top of the Contents pane, click the List By Drawing Order button .
- k** Save the project.

### Step 3: Create proximity zones

Next, you want to determine which store is closest to your customers, and then display that store on a map. When you identify the closest store, you can detect patterns (for example, where your customers are traveling from or where you might want to build a new store). When you need to answer questions about feature proximity or distance—to each other or to other points—you use Proximity tools.

- a** From the Map tab, in the Selection group, click Clear to clear the selection.
- b** Zoom to the extent of the Customers layer.
- c** In the Geoprocessing pane, click the Back button, and then click the Toolboxes tab, if necessary.
- d** Expand Analysis Tools, and then expand Proximity.

You have probably used a buffer to create a zone around a feature based on a specified distance. You will use a few other Proximity tools to analyze stores and customers. You want to create zones around each store that contain the closest customers. The zones indicate the store that customers will most likely travel to.

### 1. Is this operation an area-expanding or a distance-returned type?

---

The Create Thiessen Polygons tool creates proximity zones around points, indicating that a point is closer to another point within that polygon than to any other point.

**e** Open the Create Thiessen Polygons tool, and set the following parameters:

- Input Features: Stores
- Output Feature Class: **StoreZones**

**f** Click Run.







### 3. What do you think is causing the problem?

---

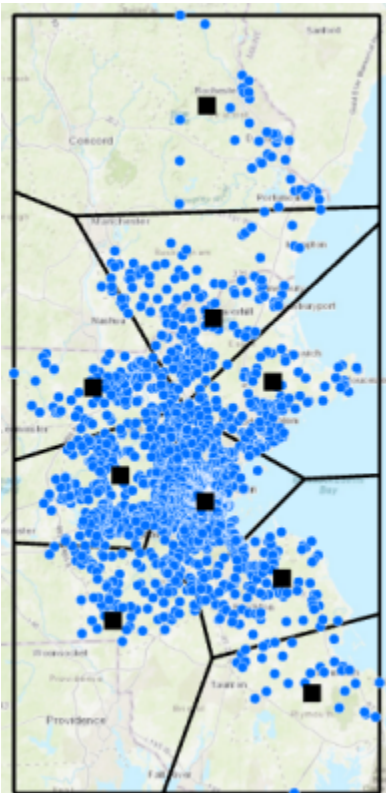
### 4. How can you modify the extent of geoprocessing outputs to better suit your analysis?

---

Next, you will run the tool again. This time, you will modify the output extent environment setting.

- g** Before you run the Create Thiessen Polygons tool again, on the Analysis tab, click Environments.
- h** Under Processing Extent, update the Extent to Customers, and then click OK.
- i** Run the tool.
- j** Change the StoreZones symbol to No Color and a solid black outline with a width of 2.

*Hint:* In the Symbology pane, from the Gallery tab, click Black Outline (2 pts).



Setting the extent environment to the same environment as the Customers layer allows the tool to incorporate all customers. You will use the StoreZones layer when you perform overlay analysis.

- k** Close the Symbology pane, and save the project.

## Step 4: Determine the closest store to each customer

You want to query your customers to see which stores are closest to them. You also want to know the distance between customers and the closest stores. In this step, you will use the Near tool to find the closest store to each customer.

- a In the Contents pane, make Stores, Customers, and the basemap the only visible layers.
- b Open the Customers attribute table, and scroll to the right.

After Latitude and Longitude, the table does not contain any additional fields.

- c Keep the table open.
- d In the Geoprocessing pane, click the Back button, and, if necessary, click the Toolboxes tab.
- e Expand Analysis Tools and Proximity, if necessary.
- f Open the Near tool and set the following parameters:

- Input Features: Customers
- Near Features: Stores
- Search Radius: **10** Miles
- Check the Location box



Checking the Location box will add the x,y coordinates of the closest feature as separate fields, NEAR\_X and NEAR\_Y.


- g Click Run.
- h In the Customers table, scroll to the right.

NEAR_FID	NEAR_DIST	NEAR_X	NEAR_Y
7	15888.077579	232735.3936	899445.6698
-1	-1	-1	-1
2	14299.041724	250355.8626	930693.4138
2	15583.098479	250355.8626	930693.4138
2	15623.141306	250355.8626	930693.4138
2	14552.279318	250355.8626	930693.4138

The Near tool added four fields to the Customers table.

Field name	Description
NEAR_FID	The feature ID of the closest feature to the input feature
NEAR_DIST	The distance, in map units, between the input and the closest feature
NEAR_X	X-coordinate of the closest feature
NEAR_Y	Y-coordinate of the closest feature










Some features have actual values, whereas others have a -1. A search distance of 10 miles was used. A -1 indicates that the feature falls outside the search radius. After you run the Near tool, and you query the Customers layer, you can determine which store is the closest, the x,y coordinates of the store, and the distance.

 Save the project.

## Step 5: Add and calculate a field

The NEAR\_DIST field indicates how many map units that the closest feature is to your input features. In this case, the map units are meters. You can add and calculate fields to show different units, such as miles or kilometers.

You will add a Miles field to the Customers table.

-  At the top of the table, click Add.
-  In the highlighted cell, for Field Name, type **Miles**.
-  Press Tab until you reach Data Type, and then choose Float.
-  Tab until you reach the Number Format cell.
-  In the cell, click the ellipses button.
-  In the Number Format window, for Category, choose Numeric.
-  Under Rounding, lower the Decimal Places to 2, and then click OK.
-  From the ribbon, on the Fields tab, in the Changes group, click Save.
-  Close the Fields view.

Next, you will calculate the Miles field.

- j** In the Customers table, right-click the Miles field and choose Calculate Field.

The Calculate Field tool opens in the Geoprocessing pane.

- k** In the Geoprocessing pane, for Fields, scroll down and double-click NEAR\_DIST to add it to the expression.

Miles =  
!NEAR\_DIST!

- l** After !NEAR\_DIST!, type /1609.

Miles =  
!NEAR\_DIST!/1609

- m** Click Run.

Miles
9.87
0
8.89
9.68
9.71

You can get a better idea of the distance using a more standard measurement.

- n** Close the table, and save the project.

## Step 6: Create desire lines

Another way to illustrate proximity in a map is to create desire lines between two sets of coordinates. In this example, your desire lines will lie between the stores and their nearest customers within a 10-mile radius. First, you will select only the customers who live within 10 miles of a store.

- a** From the Map tab, in the Selection group, click Select By Attributes.
- b** In the Geoprocessing pane, for Input Rows, ensure that Customers is chosen.
- c** Click Add Clause.
- d** Build the following expression: NEAR\_FID Is Not Equal To -1.

e Click Add.

Where NEAR\_FID is not equal to -1

f Click Run.

Out of 2,389 customers, 2,015 of them are selected.

g In the Geoprocessing pane, click the Back button, and then search for and open the **XY To Line** tool.

h Set the following parameters:

- Input Table: Customers
- Output Feature Class: **StoreCust10Miles**
- Start X Field: X\_Coordinate
- Start Y Field: Y\_Coordinate
- End X Field: NEAR\_X
- End Y Field: NEAR\_Y

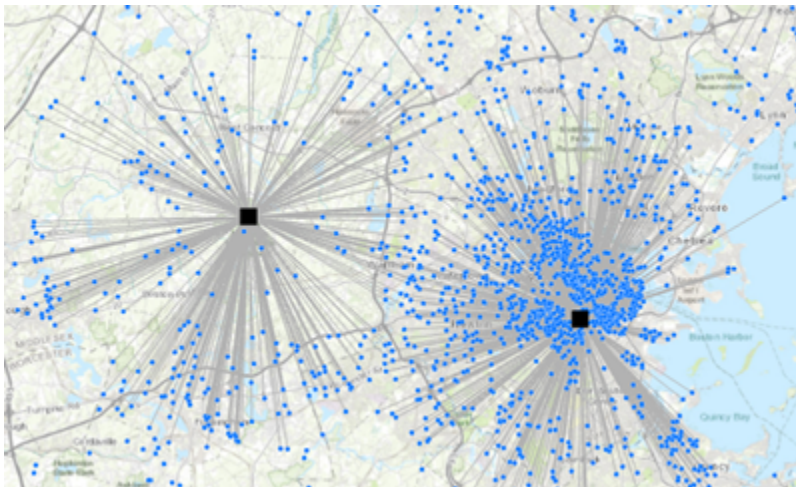
i Click Run.

j From the Map tab, clear the selection.

k In the Contents pane, for Customers, change the symbol size to 3 pt.

l For StoreCust10Miles, change the Color to a medium gray and the Line Width to **.75** pt.

m Zoom in to view the desire lines.



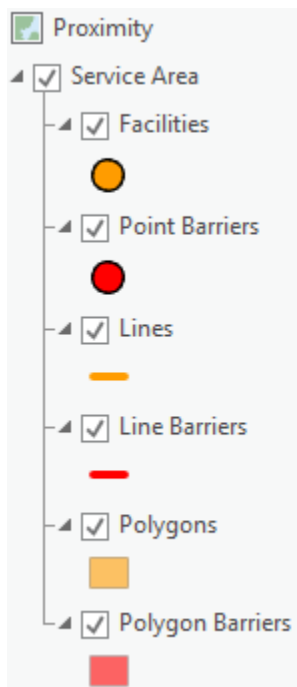
Now, you have a good visual representation of the relationship between stores and customers. The lines are beneficial for visualizing cannibalization (when many customers visit one store regardless of how close it is). You would see long lines from a store to customers even though those customers are much closer to another store. Visualizing the lines may help businesses identify potential stores for closing, or at least ask why customers choose to drive farther to shop at another store.

- n Save the project.

## Step 7: Create drive-time polygons

Next, you will use Network Analyst to create drive-time polygons from each store. First, you will add the stores as facilities to the Service Area layer.

- a In the Contents pane, turn off the StoreCust10Miles layer.
- b From the Analysis tab, in the Tools group, click Network Analysis and choose Service Area.



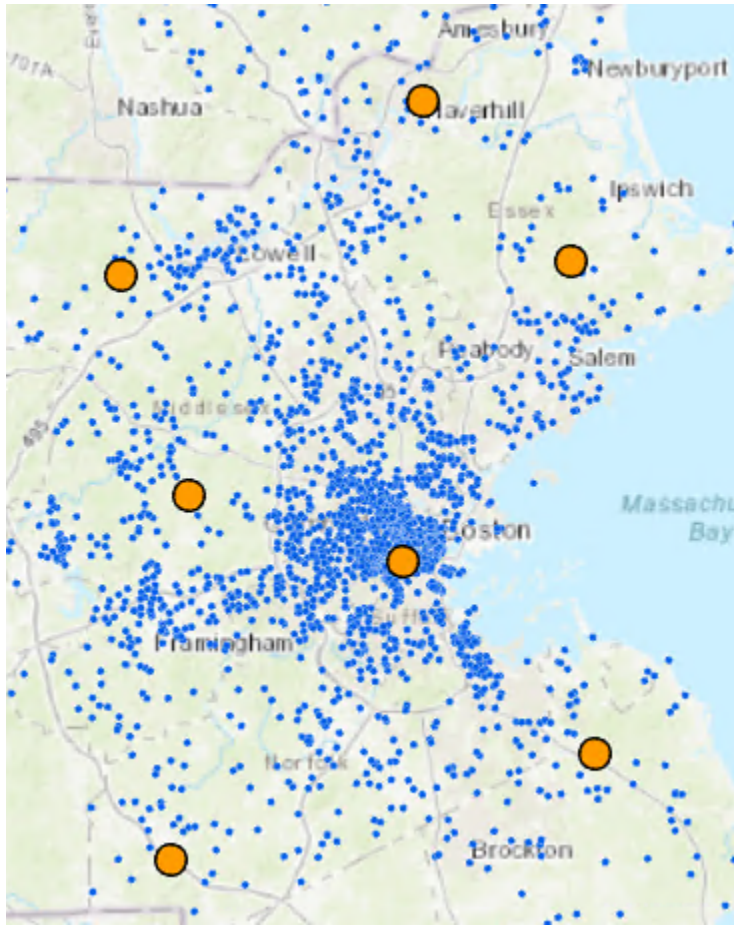
Network Analyst creates a Service Area group layer, adds it to the Contents pane, and adds a Service Area tab to the ribbon. Next, you will import the stores into the Network Analyst Facilities layer.

- c From the Service Area tab, in the Input Data group, click Import Facilities.

The Add Locations tool opens in the Geoprocessing pane.

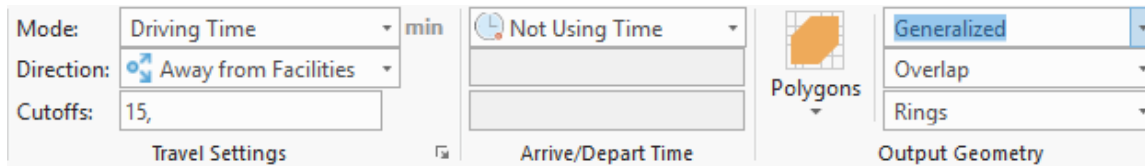
- d For Input Locations, choose Stores.

- e For Search Tolerance, choose **10 Miles**.
- f Click Run.
- g In the Contents pane, turn off the Stores layer, and then zoom to the Facilities layer.



The stores are added into the Facilities layer. Next, you will create drive-time polygons. Driving times give businesses an idea of which customers live within a designated time from a store.

- h In the Contents pane, click the Service Area group layer to select it.
- i On the Service Area tab, in the Travel Settings group, set Cutoffs to be only **15** for a 15-minute driving time.
- j In the Arrive/Depart Time group, verify that Not Using Time is selected.
- k In the Output Geometry group, click Standard Precision and choose Generalized.

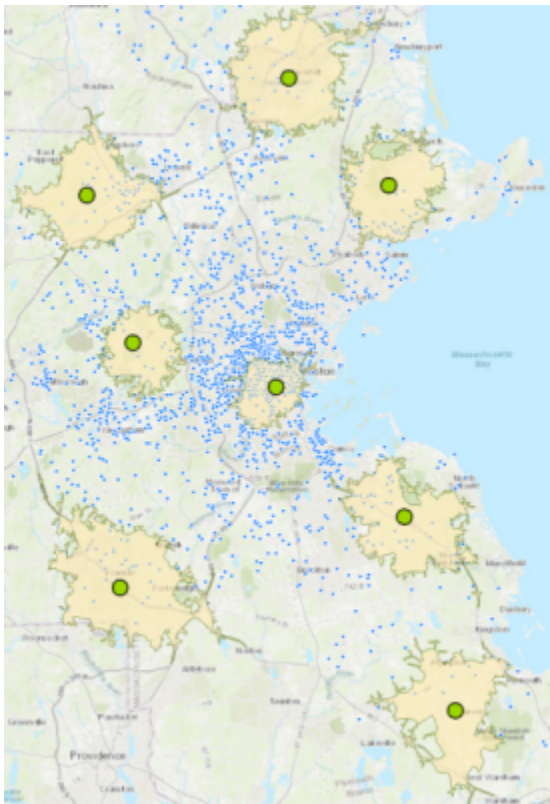


**i** In the Analysis group, click Run.



By default, Network Analyst uses an ArcGIS Online service to calculate driving times. If you have your own network dataset, you can specify to use that. Using the ArcGIS Online network dataset consumes credits.

**m** In the Contents pane, change the color of the Cutoff polygons to a beige, and then close the Symbology pane.



You used Network Analyst to create drive-time polygons, which offer a different perspective on proximity from that of a straight-line distance.

**n** Save the project.



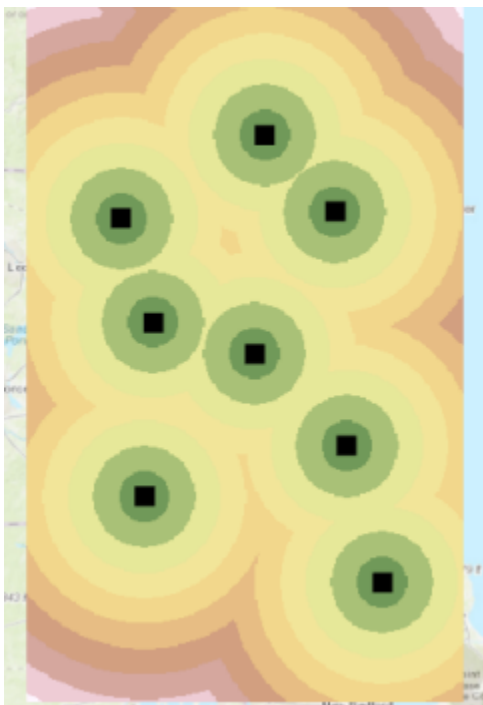
## Step 8: Create a distance surface

What if you were performing an analysis that required a distance surface? You may be using other raster data for site suitability modeling. You can create a distance using Spatial Analyst and the Euclidean Distance tool.

- a In the Contents pane, make Stores and the basemap the only visible layers.
- b In the Geoprocessing pane, click the Back button, and, if necessary, click the Toolboxes tab.
- c Expand Spatial Analyst Tools, if necessary, and then expand Distance.

You can use several Distance tools to produce a raster surface, including Euclidean Distance. The Euclidean Distance tool uses straight-line or geodesic distance from the inputs to create a distance surface. You can use the distance surface in suitability modeling, as you will later in the course.

- d Open the Euclidean Distance tool and set the following parameters:
  - Input Raster Or Feature Source Data: Stores
  - Output Distance Raster: **StoresDist**
  - Distance Method: Accept the default of Planar
- e Click Run.



The Contents pane now shows distance bands and a legend. Cells are given a value based on their straight-line distance from each store.

 Save the project, and keep ArcGIS Pro open.

## Lesson review

---

1. Explain the three ways that ArcGIS Pro measures proximity.

---

---

---

---

2. Explain the difference between using a straight-line distance and using cost.

---

---

---

---

---

# Answers to Lesson 3 questions

## Using proximity in everyday life (page 3-2)

1. What is the first thing that you do when figuring out how to get to a new place?

**Answers will vary based on personal experience.**

## Ways to measure distance (page 3-4)

### Scenario 1: Locating liquor stores near schools

1. Which distance measurement would you use when you run the Buffer tool on small-extent (large-scale) data?

**Euclidean (planar)**

### Scenario 2: Determining economic zones

2. Which distance measurement would you use when you run the Buffer tool on large-extent (small-scale) data?

**Geodesic**

### Scenario 3: Placing a pipeline

3. Which proximity distance measurement is most suitable for finding the best path for a pipeline?

**Cost**

## Exercise 3: Analyze proximity (page 3-9)

1. Is this operation an area-expanding or a distance-returned type?

**Area-expanding**

2. Is there anything about the result that you notice as a potential problem?

**Yes. Not all the customers are accounted for in the zones.**

3. What do you think is causing the problem?

**The output extent is causing the issue.**

## Answers to Lesson 3 questions (continued)

4. How can you modify the extent of geoprocessing outputs to better suit your analysis?

**You can change the environment setting.**



# 4

## Overlay analysis

Overlay is a fundamental and important type of spatial analysis. Overlay analysis is used to explore both spatial and attribute characteristics of combined data layers. More specifically, data is overlaid to answer questions about which geographic features are on top of other features (for example, what crimes are reported within which patrol areas). This lesson focuses on using overlay in analysis and how attributes of input layers are combined during the process. You will also learn how different tools manage the data's output extent when an overlay operation is performed (for example, whether all areas are preserved in the output or only the areas of overlap). Finally, you will learn about other tools that process the cells of a raster dataset based on how they overlap with other datasets.

### Topics covered

---

How overlay works

Overlay tools and choosing the appropriate one

Performing overlay analysis

## Introducing overlay

---

Overlay analysis can help you determine features that overlap. When you overlay one set of features with another, you create new information.

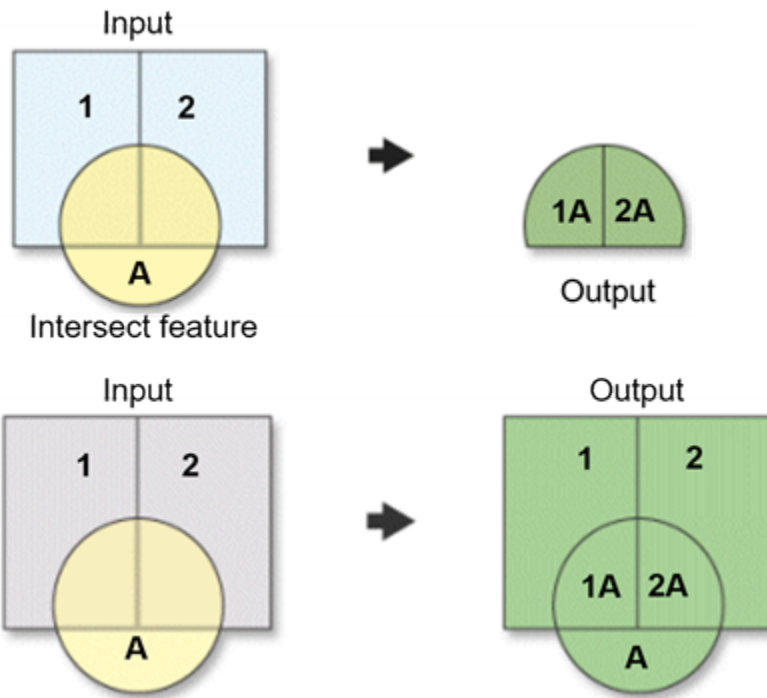


Figure 4.1. Overlay processes features and attributes into new information.

**What visual characteristics do you notice about the outputs?**



## How overlay works

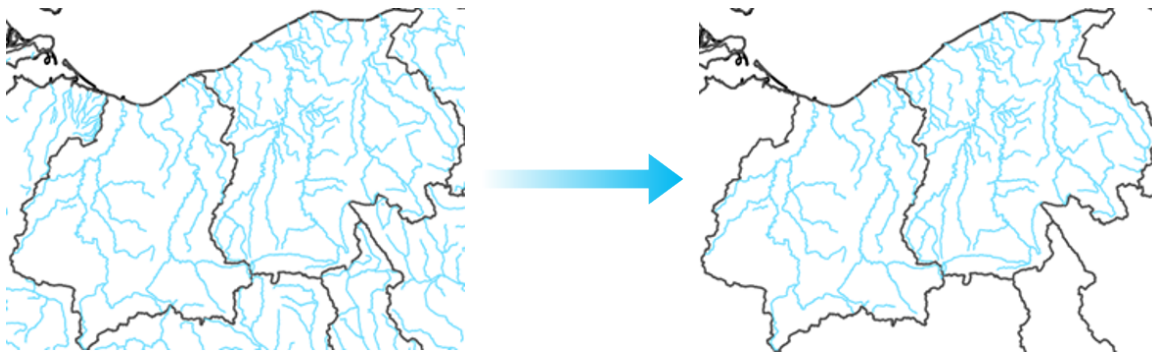
---

### What is overlay?

Overlay analysis is the geometric intersection of multiple datasets to combine, erase, modify, or update features in a new output dataset. As you learned earlier, the output dataset contains new information that combines existing information from the inputs. Overlay helps answer one of the basic questions in GIS: What is on top of what?

### Using overlay

Overlay allows all combinations of geometry. The extents of the inputs need not be identical. Overlay tools always result in the simplest geometry from all inputs. As in the following example, a polygon and line are the inputs and a line feature class is the result.



**Figure 4.2.** On the left, streams and watersheds are inputs to an overlay operation. On the right, the result is streams within specified watersheds.

Although an important result of overlay is the spatial dataset, you also get an attribute table that contains valuable information. All the overlay tools, except for Erase, produce an output feature class in which additional attributes are defined and populated. Having all attributes in one table allows you to quickly query a single feature to discover other attributes or apply symbology using the other attributes.

## How overlay works (continued)

Shape	Region	Basin	Subbasin
Polygon	Great Lakes	Western Lake Erie	Sandusky
Polygon	Great Lakes	Western Lake Erie	Lower Maumee

Shape	Name	LengthKM	FlowDir
Polyline	Silver Creek	2.4	1
Polyline	Sandusky River	5.3	2



Shape	Name	LengthKM	FlowDir	Region	Basin	Subbasin
Polyline	Silver Creek	2.4	1	Great Lakes	Western Lake Erie	Lower Maumee
Polyline	Sandusky River	5.3	2	Great Lakes	Western Lake Erie	Sandusky

**Figure 4.3.** On top, the tables for the two input datasets are shown. Below, the resulting table from performing overlay has attributes from both streams and the watersheds that overlap.

## Overlay tools

---

ArcGIS Pro contains several tools for performing overlay analysis. The tool that you use depends on the question that you want to answer, the types of features in your input data, and which features that you want to include in the output.

Tool name	Description
<b>Intersect</b>	<ul style="list-style-type: none"> <li>• Combines intersecting features (point, line, or polygon).</li> <li>• Only features that overlap each other are combined in the output. Attribute values from the input feature classes are copied to the output feature class.</li> <li>• Always outputs to simplest geometry type (for example, intersecting a point or line with polygon results in point or line).</li> </ul>
<b>Identity</b>	<ul style="list-style-type: none"> <li>• Combines features of any type (point, line, or polygon) with "identity" features, which must be polygons or have the same geometry as the input features.</li> <li>• Extent of output feature class has the same extent as the input feature class.</li> <li>• Any of the input features that overlap the identity features will get the attributes of those identity features.</li> <li>• Input features that do not overlap have null attributes.</li> </ul>
<b>Erase</b>	<ul style="list-style-type: none"> <li>• Creates a feature class by overlaying the input features with the polygons of the erase features.</li> <li>• Only those portions of the input features falling outside the erase feature's outside boundaries are copied to the output feature class.</li> </ul>

## Overlay tools (continued)

---

The following tools are not in the Overlay toolset but create information based on overlapping features.

Tool name	Description
<b>Summarize Within</b>	<ul style="list-style-type: none"><li>• Overlays a polygon layer with another layer to summarize the number of points, length of the lines, or area of the polygons within each polygon.</li><li>• Calculates summary statistics about the attributes of the features within the polygons.</li></ul>
<b>Tabulate Area</b>	<ul style="list-style-type: none"><li>• Calculates areas of features that fall within a zone.</li><li>• Zone is polygon feature class or raster dataset.</li><li>• Input can be raster or vector.</li><li>• Outputs a table.</li><li>• Requires Spatial Analyst.</li></ul>

## Choosing the appropriate tool

---

For each of the following examples, choose the appropriate overlay tool.

### Scenario 1: Drive-time polygons

You are overlaying customers and 15-minute drive-time polygons. You want all customers in the output feature class but want driving-time attributes appended to customers who overlap the polygons.

1. **Which overlay tool will output all customers but will only append attributes for customers who overlap 15-minute drive-time polygons?**
- 

### Scenario 2: Vacant parcels in a city

Given a layer of parcels in a county and a layer of city boundaries, you want to find the average value of vacant parcels within each city boundary.

2. **Which overlay tool would you use to find the average value of vacant parcels within each city boundary?**
- 

### Scenario 3: Flood zone shelters

You have a layer of schools that you can use for hurricane shelters and a flood zone layer. You do not want to use schools in the flood zone for shelters.

3. **Which overlay tool will create a layer containing only schools that fall outside the flood zone?**
-

## Perform overlay analysis

In this exercise, you will use several overlay tools to locate customers based on their spatial relationships with the drive-time polygons. You will also use Spatial Analyst tools to summarize the length of streams in watersheds and calculate the amount of each land-use classification in a raster.

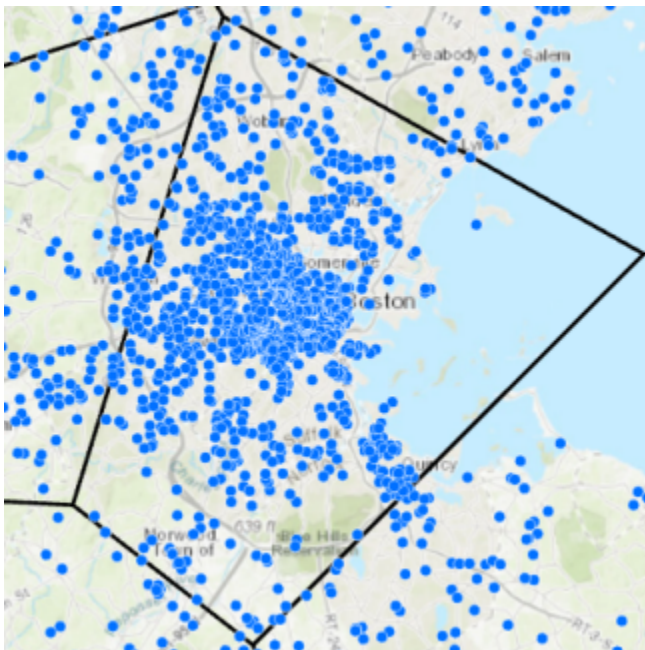
In this exercise, you will perform the following tasks:

- Make spatial selections based on overlap.
- Overlay features using the Intersect, Identity, and Erase tools.
- Summarize stream length in a watershed.
- Calculate land-use areas within a zone.

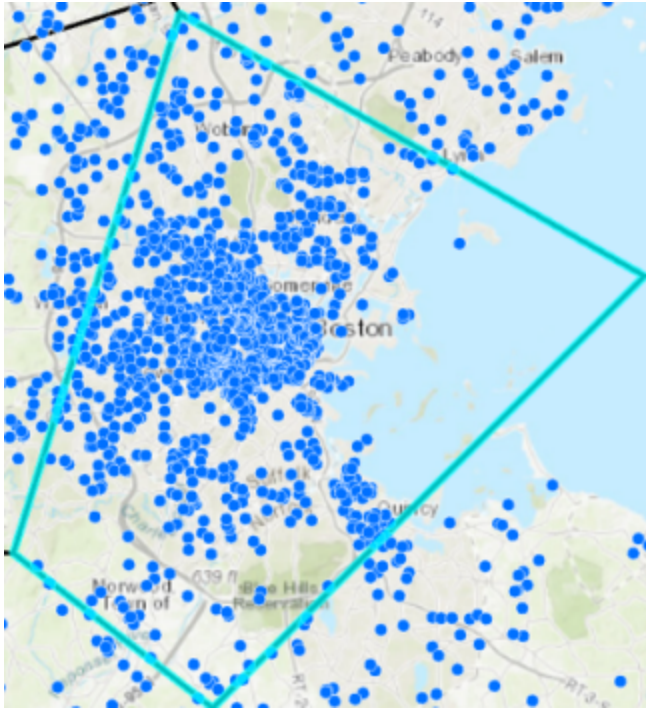
## Step 1: Make selections based on location

Earlier, you created Thiessen polygons. Now you will use them to select stores that fall within the polygons. One of the simplest forms of analysis based on spatial relationships is selecting features that overlap or are within other features.

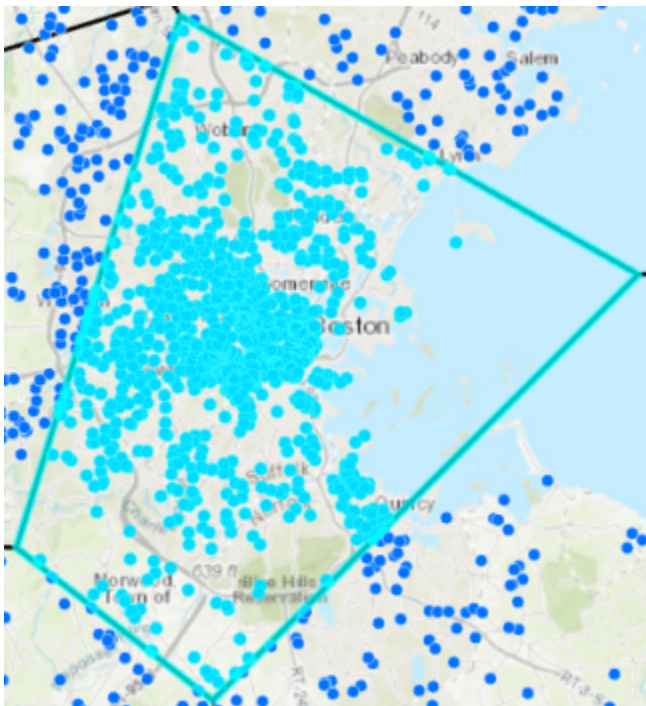
- a If necessary, restore the SNAPCourse project and the Proximity map.
- b Make Customers, StoreZones, and the basemap the only visible layers.
- c From the Contents pane, change the symbol size for the Customers layer to **6**, and then close the Symbology pane.
- d Zoom to the Thiessen polygon that contains the largest number of customers.



- e From the Map tab, in the Selection group, choose the Select tool.
- f Click in the Thiessen polygon to select it.



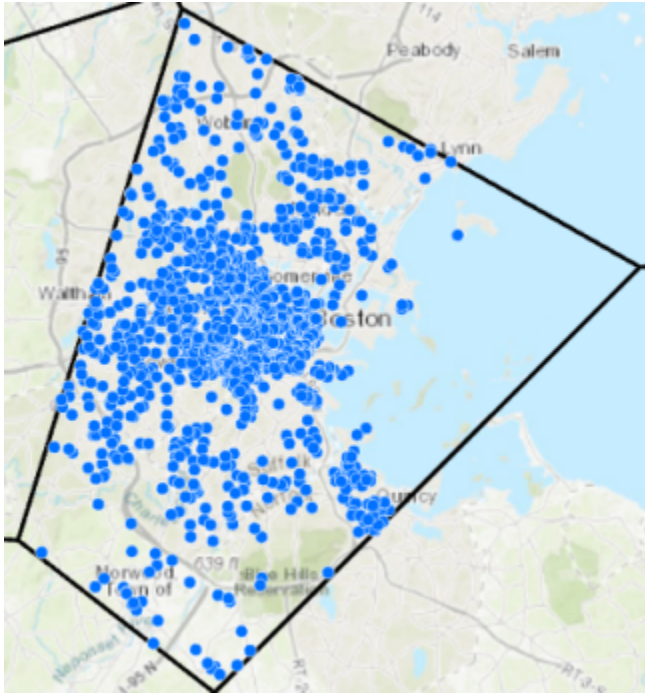
- g** From the Map tab, click Select By Location.
- h** In the Geoprocessing pane, set Selecting Features to StoreZones.
- i** Accept the other defaults and click Run.





Now that you have selected the customers based on the fact that they intersect the selected zone, you can export them to their own layer for further analysis.

- j** In the Contents pane, right-click the Customers layer, point to Selection, and choose Make Layer From Selected Features.
- k** Turn off the Customers layer and clear the selection.




You have isolated the customers who are closer to the store in the Thiessen polygon than to any other store.

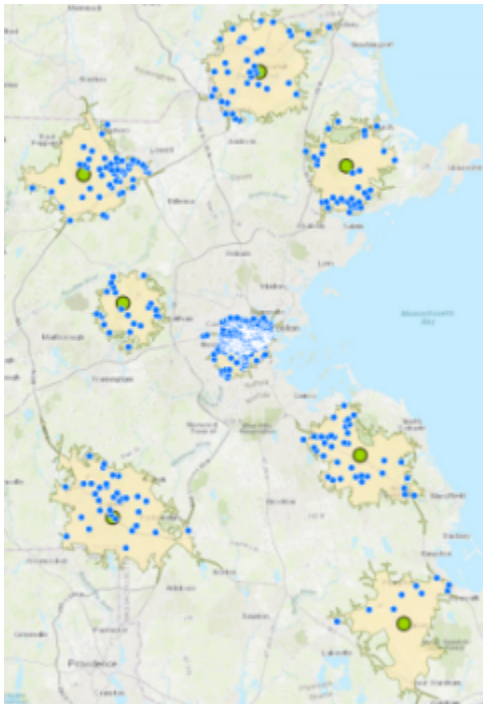
- l** Save your project.

## Step 2: Overlay customers and driving times using the Intersect tool

You want to show only the customers who live within the drive-time polygons. You could use a spatial query, but you want to store the attributes for customers and the drive-time polygons in the same layer. An advantage of using overlay tools is that they append attributes. In this step, you will intersect the customers with the drive-time polygons.

- a** Make Customers, Service Area, and the basemap the only visible layers.
- b** From the Analysis tab, in the Tools group, click Intersect.
- c** In the Geoprocessing pane, click the Environments tab.
- d** For Extent, click the As Specified Below down arrow and choose Customers.

- e Click the Parameters tab.
- f For Input Features, click the Add Many button , choose Customers and Service Area Polygons, and then click Add.
- g For Output Feature Class, type **CustDriveInt**.
- h Click Run.
- i In the Contents pane, turn off the Customers layer, and drag the CustDriveInt layer above the Service Area layer.



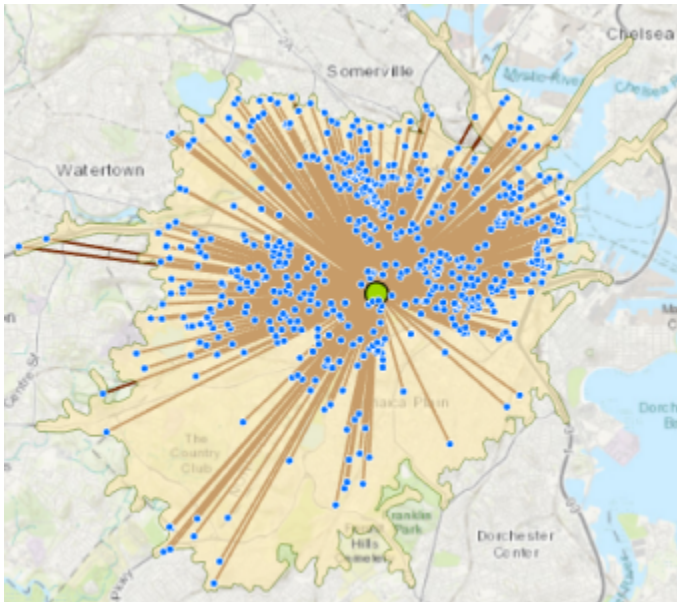
The Intersect tool creates a feature class containing only the customers who intersect the drive-time polygons. Further, the customers will also have drive-time attributes.

- j Open the CustDriveInt attribute table.

Now you can query a customer and see the ID of the store that is within the 15-minute driving time. You could create desire lines again, this time showing 15-minute driving times instead of 10 miles.

- k Close the attribute table.
- l From the Analysis tab, in the Geoprocessing group, click History.

- m Find and open the XY To Line tool, and then set the following parameters:
  - Input Table: CustDriveInt
  - Output Feature Class: **Cust15Min**
- n Ensure that the following fields are set with these parameters (they should have been previously set):
  - Start X Field: X\_Coordinate
  - Start Y Field: Y\_Coordinate
  - End X Field: NEAR\_X
  - End Y Field: NEAR\_Y
- o Click Run.
- p From the Contents pane, change the CustDriveInt layer symbol size to **4** pt, and then close the Symbology pane.
- q Zoom to the drive-time polygon in central Boston.



- r In the Contents pane, turn the StoreCust10Miles layer on and off to see the difference between 10 miles and 15 minutes.



You may need to zoom out.

You can use the results to determine where customers are coming from, how far they are willing to travel, and perhaps where new customers could exist. The way that you conceptualize distance, whether it be time or miles, can affect analysis results.

**s** Save the project.

### Step 3: Overlay customers and driving times using the Identity tool

What if you wanted to show all the stores but append only drive-time attributes to the stores that fall within the drive-time polygons? In this step, you will use the Identity overlay tool to append attributes for only overlapping features, while keeping all features in the output.

- a** In the Contents pane, make Customers, Service Area, and the basemap the only visible layers.
- b** In the Geoprocessing pane, search for and open the **Identity** tool.
- c** Set the following parameters:
  - Input Features: Customers
  - Identity Features: Service Area\Polygons
  - Output Feature Class: **CustIdentity**
- d** Click Run.

- e In the Contents pane, turn off the Customers layer.

All customer points are retained in the output. The difference is that features have attributes for customers and driving times.

- f Open the CustIdentity attribute table.
- g Scroll to the far right in the table.

FID_SAPolygons1	FacilityID	Name
-1	0	
-1	0	
-1	0	
-1	0	
-1	0	
-1	0	
1	1	Location 1 : 0 - 15
-1	0	
-1	0	
1	1	Location 1 : 0 - 15
1	1	Location 1 : 0 - 15
-1	0	
-1	0	
1	1	Location 1 : 0 - 15

**1. Why do only some of the points have FacilityID and Name values?**

---

- h Close the attribute table.
- i Use the Explore tool to zoom to the southernmost store.
- j Click a customer within the drive-time polygon, and then in the pop-up window, scroll all the way down.

Miles	3.89879
Kilometers	6.273154
FID_SAPolygons1	1
FacilityID	1
Name	Location 1 : 0 - 15
FromBreak	0
ToBreak	15

The customer has drive-time attributes.

- k** Click a customer outside the drive-time polygon.

Miles	5.342014
Kilometers	8.5953
FID_SAPolygons1	-1
FacilityID	0
Name	
FromBreak	0
ToBreak	0

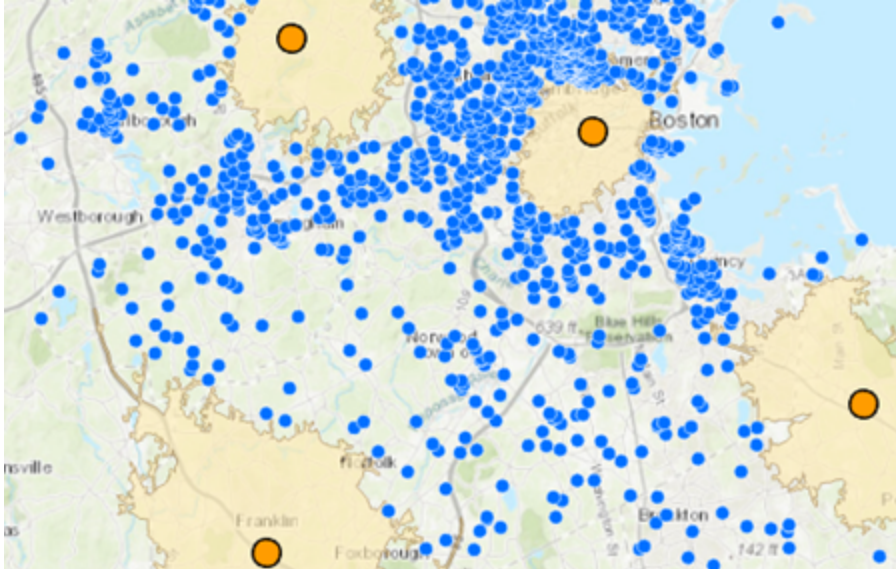
Now there are no drive-time attributes because the customer falls outside the polygon. Using the results of the Identity tool, you could symbolize different colors for a customer based on whether that customer is within 15 minutes of a store.

- l** Close the pop-up window, and then save the project.

#### Step 4: Remove customers within 15 miles

Suppose that you want to find a location for a new store. You must create a map of potential customers that excludes customers who are within 15 minutes of other stores. You can use an overlay tool called Erase to remove customers who overlap the drive-time polygons.

- a** In the Geoprocessing pane, search for and open the **Erase** tool.
- b** Set the following parameters:
- Input Features: Customers
  - Erase Features: Service Area\Polygons
  - Output Feature Class: **PotentialCustomers**
- c** Click Run.
- d** In the Contents pane, make PotentialCustomers, Service Area, and the basemap the only visible layers.



You may need to zoom out again.

Now you can see only the customers who are not within 15 minutes of other stores. These customer locations can help you identify the need for another store.

- e Save the project.

## Step 5: Summarize stream length in a watershed

Earlier, you prepared data for Ohio. Now you will use the data in overlay analysis to determine the length of streams in a watershed for conservation study.

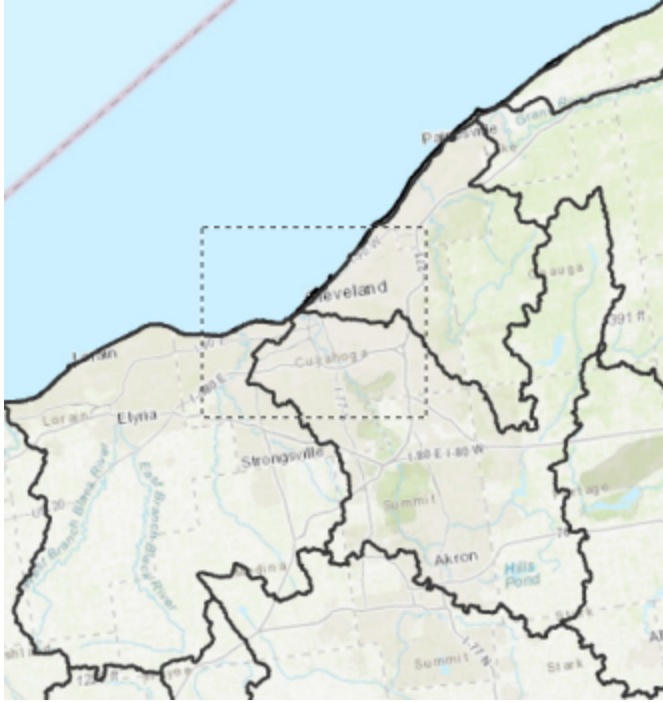
- a Open the Ohio map.
- b Make Ohio, OhioHUC, and the basemap the only visible layers.
- c Zoom to the Cleveland area in northeast Ohio.



Due to the number of stream features, you will focus your analysis on a subset of watersheds to save processing time.

- d From the Map tab, click the Select tool.
- e Select three watersheds by drawing a box that touches the following features:





You may want to update the OhioHUC symbology so that you can see it more clearly.

You will get a total length of streams within the three selected watersheds.

- f** In the Geoprocessing pane, click the Back button twice, and then click the Toolboxes tab.
- g** Expand Analysis Tools, and then expand Statistics.

You will use the Summarize Within tool to get the total length of streams in each watershed. Although the Summarize Within tool is not categorized as an overlay tool, it will summarize features based on streams overlapping a watershed.

- h** Open the Summarize Within tool and set the following parameters:

- Input Polygons: OhioHUC
- Input Summary Features: OhioStreams
- Output Feature Class: **StreamSummary**
- Field: LENGTHKM
- Statistic: Sum
- Shape Unit: Kilometers
- Click the Environments tab, and set Extent to Ohio.

- i** Click Run.

- j** In the Contents pane, turn off the Ohio and OhioHUC layers, and zoom out.





The three watersheds that you selected are the only features in the output feature class. The statistical result determined by amount of overlap is in the table.

- k** Open the StreamSummary attribute table, and scroll to the right.

Summarized length in KILOMETERS	Sum LENGTHKM
793.917692	794.465914
498.310994	498.446372
1001.090066	1001.908226

Now you have the total kilometers of streams in each watershed.

- l** Close the attribute table, and clear the selection.
- m** Save the project.

## Step 6: Calculate the amount of each land-use classification

Earlier, you extracted only the NLCD for Ohio using the Extract By Mask Spatial Analyst tool. Now, you will use that dataset to help determine the area of each land-use classification within each watershed.

- a** In the Contents pane, make OhioHUC and NLCD\_Ohio the only visible layers.
- b** In the Geoprocessing pane, return to the Toolboxes tab.
- c** Expand Spatial Analyst Tools, and then expand Zonal.

You will use a zonal analysis tool to tabulate the area of land use within each watershed polygon. The Zonal tools allow you to perform analysis where the output is a result of computations performed on all cells that belong to each input zone. In this case, the watershed polygons act as zones.

**d** Open the Tabulate Area tool and set the following parameters:

- Input Raster Or Feature Zone Data: OhioHUC
- Zone Field: HUC\_8
- Input Raster Or Feature Class Data: NLCD\_Ohio
- Class Field: Category
- Output Table: **LanduseArea**
- From the Environments tab, update Extent to Default

**e** Click Run.

**f** In the Contents pane, under Standalone Tables, open the LanduseArea table.

Now you have a table that has the areas of each land-use classification tabulated. You could join the LanduseArea table to the Ohio\_HUC layer to add the areas for all land-use types to the Ohio\_HUC layer.

**g** Close the LanduseArea table.

**h** Save the project, and leave ArcGIS Pro open.

## Lesson review

---

1. What is overlay analysis?

---

---

---

---

---

2. If you use the Intersect tool with streams and watersheds as the inputs, what would the resulting feature class contain?

---

---

---

---

---

---

---

# Answers to Lesson 4 questions

## Introducing overlay (page 4-2)

What visual characteristics do you notice about the outputs?

**Possible responses include the following:**

- **Extents are different.**
- **Attributes are combined where features overlap.**

## Choosing the appropriate tool (page 4-7)

### Scenario 1: Drive-time polygons

1. Which overlay tool will output all customers but will only append attributes for customers who overlap 15-minute drive-time polygons?

**Identity or Spatial Join**

### Scenario 2: Vacant parcels in a city

2. Which overlay tool would you use to find the average value of vacant parcels within each city boundary?

**Summarize Within**

### Scenario 3: Flood zone shelters

3. Which overlay tool will create a layer containing only schools that fall outside the flood zone?

**Erase**

## Exercise 4: Perform overlay analysis (page 4-8)

1. Why do only some of the points have FacilityID and Name values?

**Those points are the customers who overlap with driving times. The other points do not overlap with driving times.**

# 5

## Automating spatial analysis

Running geoprocessing tools one by one in succession can be a successful workflow for producing desired information products and results. But you may want a visual representation of your analysis that you can modify and rerun. You may also want a tool that can perform the same process over and over on multiple inputs. ModelBuilder allows you to chain together tools, automate workflows, and share tools so others can add their own data.

In this lesson, you will focus on ModelBuilder as a means of automation, but you will also see how Python and tasks can be used.

### Topics covered

---

Automating workflows in ArcGIS Pro

Building models and sharing them as tools

## Automating workflows

---

Earlier, you used individual geoprocessing tools to perform analysis and produce effective results. ArcGIS Pro enables you to automate your analysis—to set it to process multiple datasets at a time.

**Why would you automate your spatial analysis workflows?**

## Automation methods in ArcGIS Pro

You can perform spatial analysis by running individual geoprocessing tools in succession and produce satisfactory results. However, you may want to run the same tool on many inputs at one time to have a visual representation of your analysis.

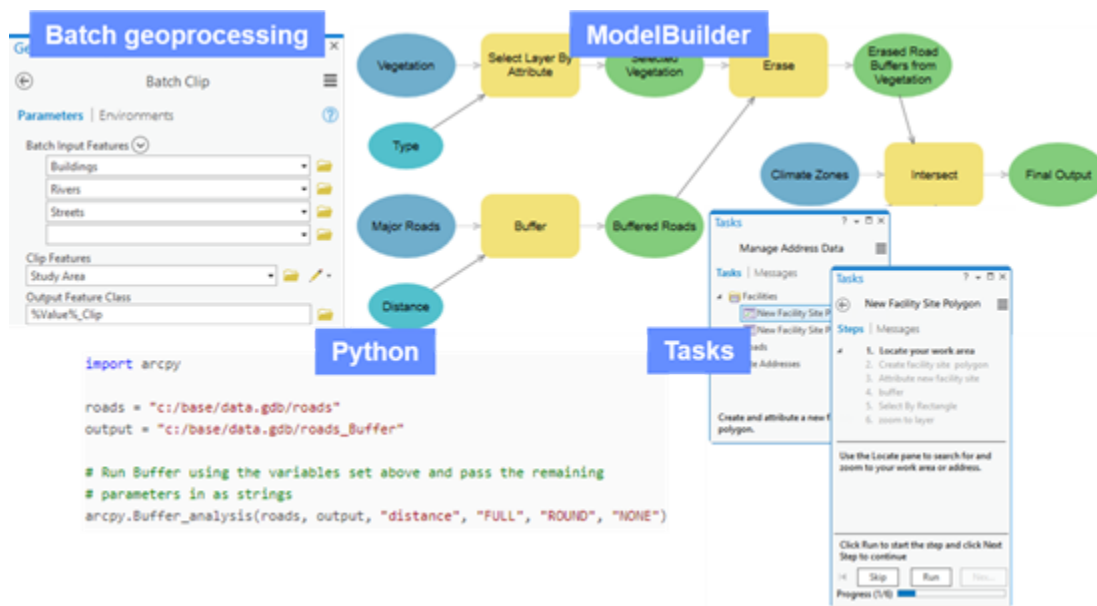


Figure 5.1. Ways to automate analysis and other operations in ArcGIS Pro.

Automation method	Description
<b>Batch geoprocessing</b>	Allows you to run the tool multiple times using many input datasets or different parameter settings. Makes it possible to run a given tool as many times as needed with very little interaction.
<b>ModelBuilder</b>	Visualize processes and geoprocessing tools chained together, frequently using the output of one process as the input to another process.
<b>Python</b>	The scripting language of ArcGIS. ArcGIS includes ArcPy, which gives you access to all geoprocessing tools, scripting functions, and specialized modules that help you automate a GIS process.

## Automation methods in ArcGIS Pro (continued)

---

Automation method	Description
<b>Tasks</b>	A set of preconfigured steps that guide you and others through a workflow or business process. You can use a task to implement a best-practice workflow, improve the efficiency of a workflow, or create a series of interactive tutorial steps.



**Esri Training course:** [Automating Workflows Using ArcGIS Pro Tasks](#)



**ArcGIS Pro Help:** [Create a new task](#)

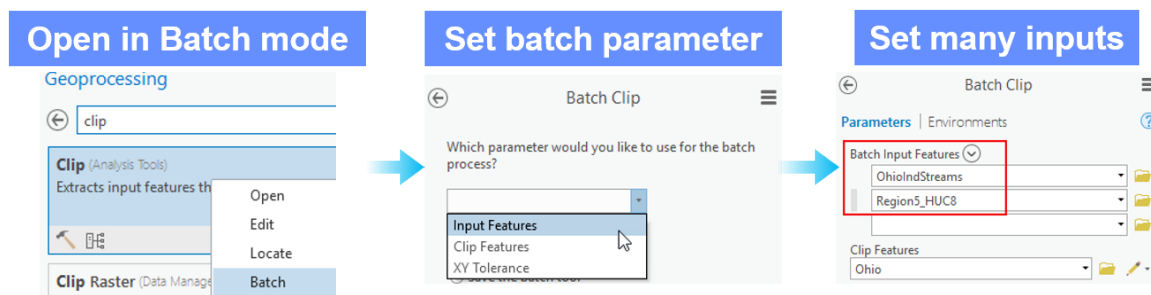


## Batch geoprocessing

You may have dozens of datasets that you want to clip to the same boundary. Without coding or creating a model, you could set up batch processing one time to execute the Clip tool on all inputs. Most tools have batch mode, but you can verify by right-clicking the tool. If Batch is not listed in the menu, then you cannot run the tool in batch mode.

To run a tool in batch mode, do the following:

1. Right-click the tool and choose Batch.
2. Set the parameter that you want to use for the batch process.
3. Choose to make a temporary batch tool or save it so that you can reuse it.
4. Run the batch tool and choose multiple options for the parameter that you specified.



**Figure 5.2.** After you choose Batch and set the batch parameter, you run the tool in batch mode.

## Build a model

You want to create a visual representation of your analysis that you can modify and rerun as needed. You will use ModelBuilder to chain tools together for the analysis of customers and stores that you performed earlier. The model will create desire lines from stores to customers.

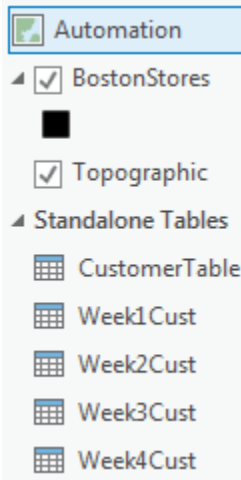
In this exercise, you will perform the following tasks:

- Create a model.
- Add tools to a model and set parameters.

## Step 1: Prepare ArcGIS Pro

First, you will prepare ArcGIS Pro by importing a map file.

- a If necessary, start ArcGIS Pro and restore the SNAPCourse project.
- b From the Insert tab, click Import Map.
- c Browse to **C:\EsriTraining\SNAP\Automation** and import Automation.mapx.



The map file contains one feature class of stores and five stand-alone tables. You will use the CustomerTable in the first exercise, and then you will use the other four tables in the second exercise.

- d Open CustomerTable.

CustomerTable contains many customer-specific attributes, such as name and address. It also contains x,y coordinates. The table is nonspatial, so you will use a geoprocessing tool to create a point feature class from the table.

- e Close the table.
- f Zoom to roughly a scale of 1:900,000, with the stores centered in the map.
- g From the Analysis tab, click Environments.
- h Under Processing Extent, set Extent to Current Display Extent, and then click OK.

## Step 2: Create a model

You will first create the empty model.

- a From the Analysis tab, click ModelBuilder.
- b In the Catalog pane, expand Toolboxes, and then expand SNAPCourse.tbx.
- c Right-click the model and choose Properties.
- d For Name, type **CustomersAndStoresAnalysis**.
- e For Label, type **Customers and Stores Analysis**, and then click OK.



Names cannot include spaces, but labels can include them. Spaces are not, however, required in label names.

- f From the ModelBuilder tab, click Save.

It is important to document your models so that other users understand what the model does.

- g In the Catalog pane, right-click the model and choose Edit Metadata.
- h For Tags, type **desire lines, stores, customers**.
- i For Summary, type: **The model automates adding customer points from a table, finding the closest store, and creating desire lines.**
- j From the Metadata tab, click Save.
- k Close the Item Description view.

### Step 3: Add the XY Table To Point tool

You will create a model to capture the overall analysis workflow that you worked with earlier in the course for stores and customers in Boston.

- a From the Geoprocessing pane, search for the **XY Table To Point** tool.
- b Drag the XY Table To Point tool into the empty model.



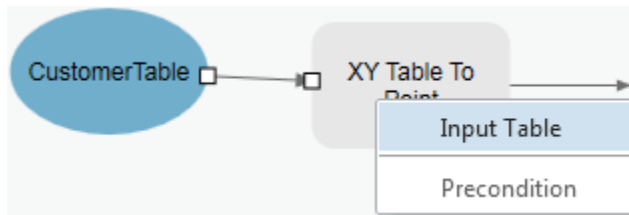
The tool is added along with an output element. You must specify an input table for the tool to be ready to run. You will use the CustomerTable as the input.

- c From the Contents pane, drag the CustomerTable into the model to the left of the tool.
- d Click in the white space to clear your selection of the element.

- e Click the blue input data element, and then drag a line to the tool.



- f Release the click, and then choose Input Table.



You set the input by connecting the model elements, but you can open the tool to set other parameters.

- g Double-click the XY Table To Point tool, and set the following parameters:

- Output Feature Class: **Customers**
- X Field: X\_Coordinate
- Y Field: Y\_Coordinate
- Coordinate System: BostonStores

- h Click OK.
- i From the ModelBuilder tab, click Save.

## Step 4: Add the Near tool

Next, you will add the Near tool to determine customers within a 10-mile radius.

- a In the Geoprocessing pane, search for **near**.
- b Right-click the Near tool and choose Add To Model.
- c If necessary, move the Near tool to the right of the green Customers element.
- d Connect the Customers output data element to the Near tool as Input Features.

- e Open the Near tool, and set the following parameters:
- Near Features: BostonStores
  - Search Radius: **10** Miles
  - Check the Location box

Input Features  
Customers

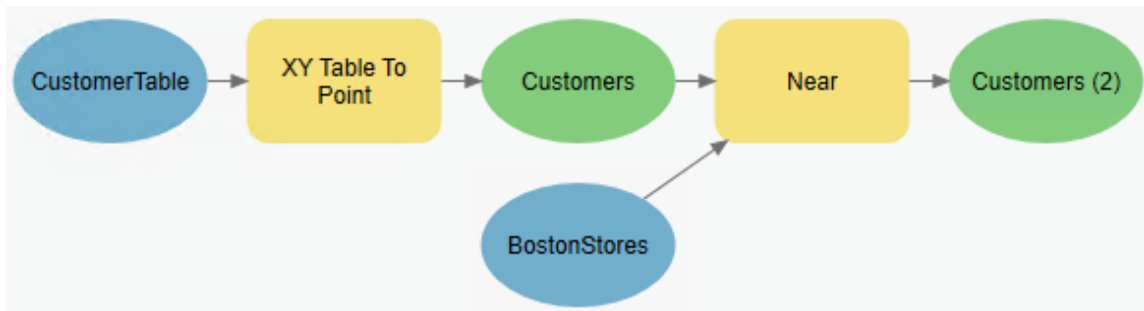
Near Features  
BostonStores

Search Radius  
10 Miles

Location  
 Angle

Method  
Planar

- f Click OK.
- g From the ModelBuilder tab, in the View group, click Auto Layout.



- h Save the model.

## Step 5: Add the Make Feature Layer tool

When you ran the analysis earlier, you selected the features that had a NEAR\_FID not equal to -1 because those features fell outside the 10-mile search radius. Rather than use the Select Layer By Attribute tool again, you will now use the Make Feature Layer tool, which works like a selection layer in ArcGIS Pro. You can build a query within the Make Feature Layer tool.

- a In the Geoprocessing pane, search for **make feature**.
- b Add the Make Feature Layer under the output of the Near tool.

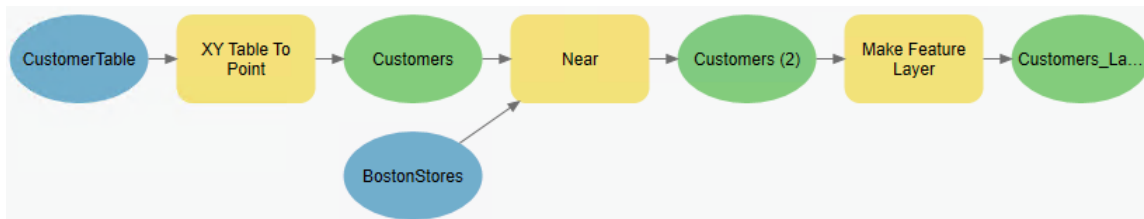
c For the Make Feature Layer tool, set the following parameters:

- Input Features: Customers (2)
- Click Add Clause
- Build the following clause: NEAR\_FID Is Not Equal To -1
- Click Add

d Click OK.

As you did earlier, you used the NEAR\_FID attribute to only process features within the 10-mile search radius.

e From the ModelBuilder tab, click Auto Layout.



f Save the model.

## Step 6: Add the XY To Line tool

Next, you will add the XY To Line tool to create the desire lines.

a In the Geoprocessing pane, click the Back button until you see the Favorites tab.

b Click Favorites, if necessary, and under Recent, add the XY To Line tool to the end of the model.



On the ModelBuilder tab, in the View group, model zoom tools are available to zoom in and out of the model so that you can place elements.

c Connect the output of the Make Feature Layer tool to the XY To Line tool as Input Table.

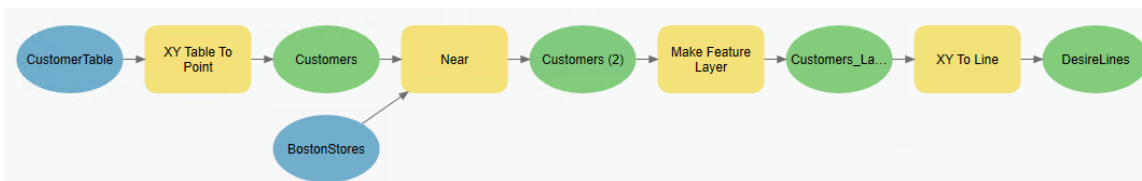
d Open the XY To Line tool, and set the following parameters:

- Output Feature Class: **DesireLines**
- Start X Field: X\_Coordinate
- Start Y Field: Y\_Coordinate
- End X Field: NEAR\_X
- End Y Field: NEAR\_Y
- Spatial Reference: Current Map [Automation]

Input Table	Customers_Layer
Output Feature Class	DesireLines
Start X Field	X_Coordinate
Start Y Field	Y_Coordinate
End X Field	NEAR_X
End Y Field	NEAR_Y
Line Type	Geodesic
ID	
Spatial Reference	NAD_1983_StatePlane_Massachusetts_Mainland_FIPS_201

e Click OK.

f From the ModelBuilder tab, click Auto Layout.




g Save the model.

## Step 7: Run the model

You will set several of the outputs to automatically add the display after the model runs.



- a In the model, right-click the Customers\_Layer output data element and then the DesireLines output data elements and choose Add To Display for both of them.
- b From the ModelBuilder tab, in the Run group, click Validate.  
 All the model elements are colored appropriately, so the model is ready to run. If any element had parameter issues, it would be gray.
- c From the ModelBuilder tab, click Run.
- d Close the progress window.
- e Activate the Automation view.
- f Zoom to central Boston.



You have created a model that performs the same analysis that you did earlier in the course. Next, you will automate this workflow to account for multiple inputs.

- g Save the project, and leave ArcGIS Pro open for the next exercise.

## Automating and sharing models

---

You can increase the power of your models through iteration, or the ability to process multiple datasets at one time. You can add an element called an iterator to enable bulk processing on items like feature classes or tables. Your model can become a powerful tool when it is given the ability to process many datasets at one time.

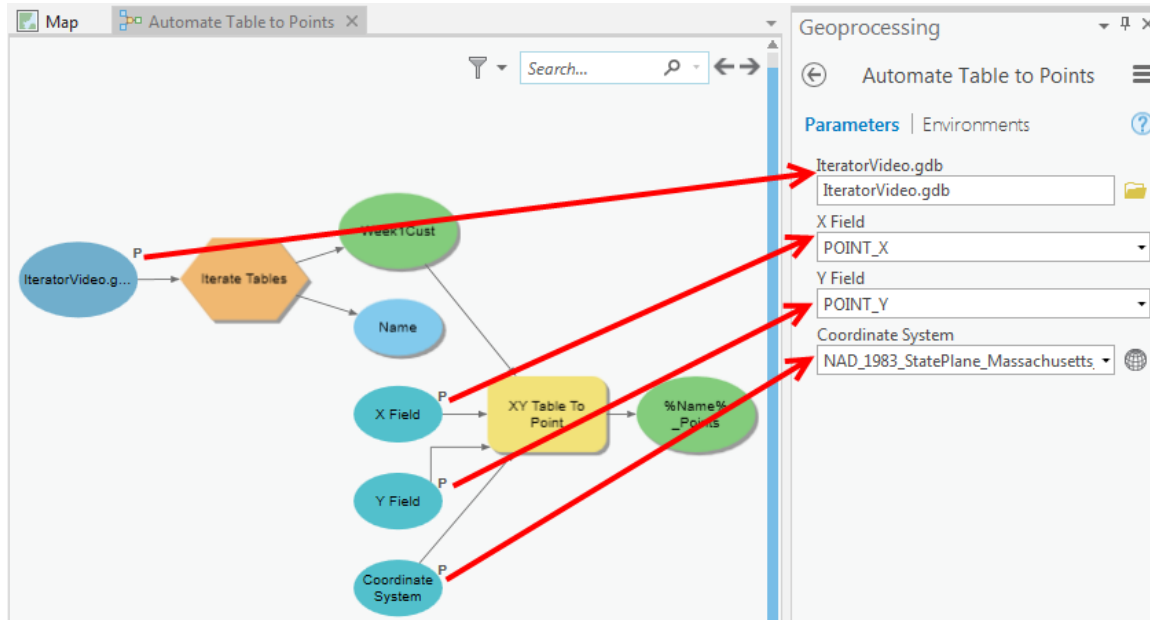
After you create and automate your model, you may want to share it as a tool. When you share a model, you should set model parameters for specific input data elements and individual tool parameters.

### 1. What does it mean to parameterize a model?

---

## Automating and sharing models (continued)

Automating a model by adding an iterator enables processing on multiple datasets at one time. You set model parameters if you plan to share your model as geoprocessing tool.



**Figure 5.3.** Setting model parameters allows users to add their own data and to choose tool parameters that meet their needs.

## Use a model to process multiple inputs

You are working as a business analyst, and you receive a customer report each week with new customers. You will use information about these new customers to compare by week and see where the new customers are coming from. Because you get a customer report every week, you want to create a tool that processes multiple tables at once to run the XY To Point, Near, and XY To Line tools. You already performed the analysis workflow using the tools. Next, you will set your model for multiple inputs using an iterator.

In this exercise, you will perform the following tasks:

- Add an iterator to a model.
- Set model parameters.

## Step 1: Prepare ArcGIS Pro and make a copy of a model

You will use the same map as you did before, but you will create a copy of the model and edit it.

- a Restore the ArcGIS Pro project, and activate the Automation map, if necessary.
- b Turn off all layers except the basemap.
- c In the Contents pane, open the Week1Cust table.

POINT_X	POINT_Y
176386.2267	993667.1819
222506.4483	1036245.7301
214050.2378	900176.8786

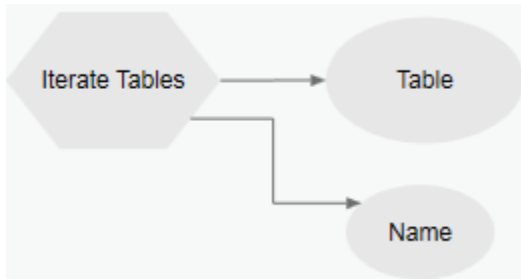
The table contains the x,y coordinates for customer locations. Each of the four tables that you added contains the same attributes. As you did before, you will use the x,y coordinates to map the customers.


- d Close the table.
- e Activate the Customers and Stores Analysis model.
- f From the ModelBuilder tab, in the Mode group, click Select to activate the tool, if necessary.
- g Drag a box around all model elements to select them (handles will appear around selected elements).
- h Right-click the selected elements and choose Copy.
- i From the ModelBuilder tab, in the Model group, click New.
- j Right-click in the blank model and choose Paste.
- k Clear the selection by clicking in the white space.
- l Select the initial CustomerTable blue input data element and press Delete.
- m From the ModelBuilder tab, in the Model group, click Properties.
- n Update the name for the model to **IterateTables**, and then update the label to **Iterate Tables** and click OK.
- o Save the model.

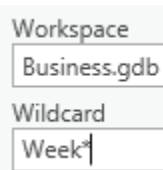
## Step 2: Add an iterator to a model

You will add an iterator to the model so that you can process multiple inputs.

- a From the ModelBuilder tab, in the Insert group, click Iterators and choose Iterate Tables.

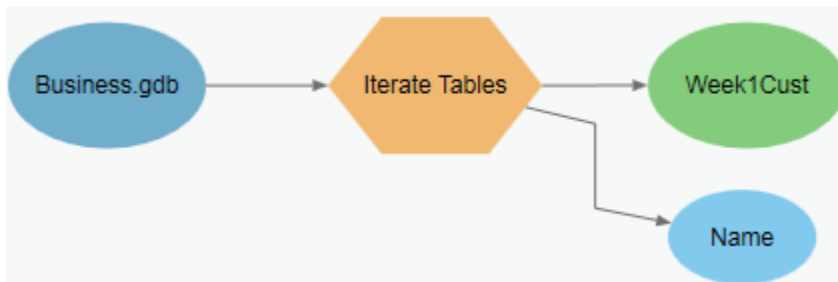


- b Drag all elements associated with the iterator to the left of the XY Table To Point tool.
- c Double-click the Iterate Tables tool to open it.
- d For Workspace, click the Browse button , browse to **C:\EsriTraining\SNAP\Automation**, select Business.gdb, and then click OK.
- e For Wildcard, type **Week\***.



The \* indicates that you want to process all tables that start with the word Week.

- f Click OK.



- g Connect the green output data element from the iterator to the XY Table To Point tool as Input Table.
- h Open the XY Table To Point tool.

- i For Output Feature Class, replace Customers with **%Name%\_Points**, and ensure that the output is being added to SNAPCourse.gdb.



Because you are using an iterator and will be processing multiple inputs, you must use an in-line variable for the output name.

- j For X Field, choose Point\_X, and for Y Field, choose Point\_Y.

Input Table  
Week1Cust:2

Output Feature Class  
%Name%\_Points

X Field  
POINT\_X

Y Field  
POINT\_Y

Z Field

Coordinate System  
NAD\_1983\_StatePlane\_Massachusetts\_Mainland\_FIPS\_201

- k Click OK.
- l Save the model.

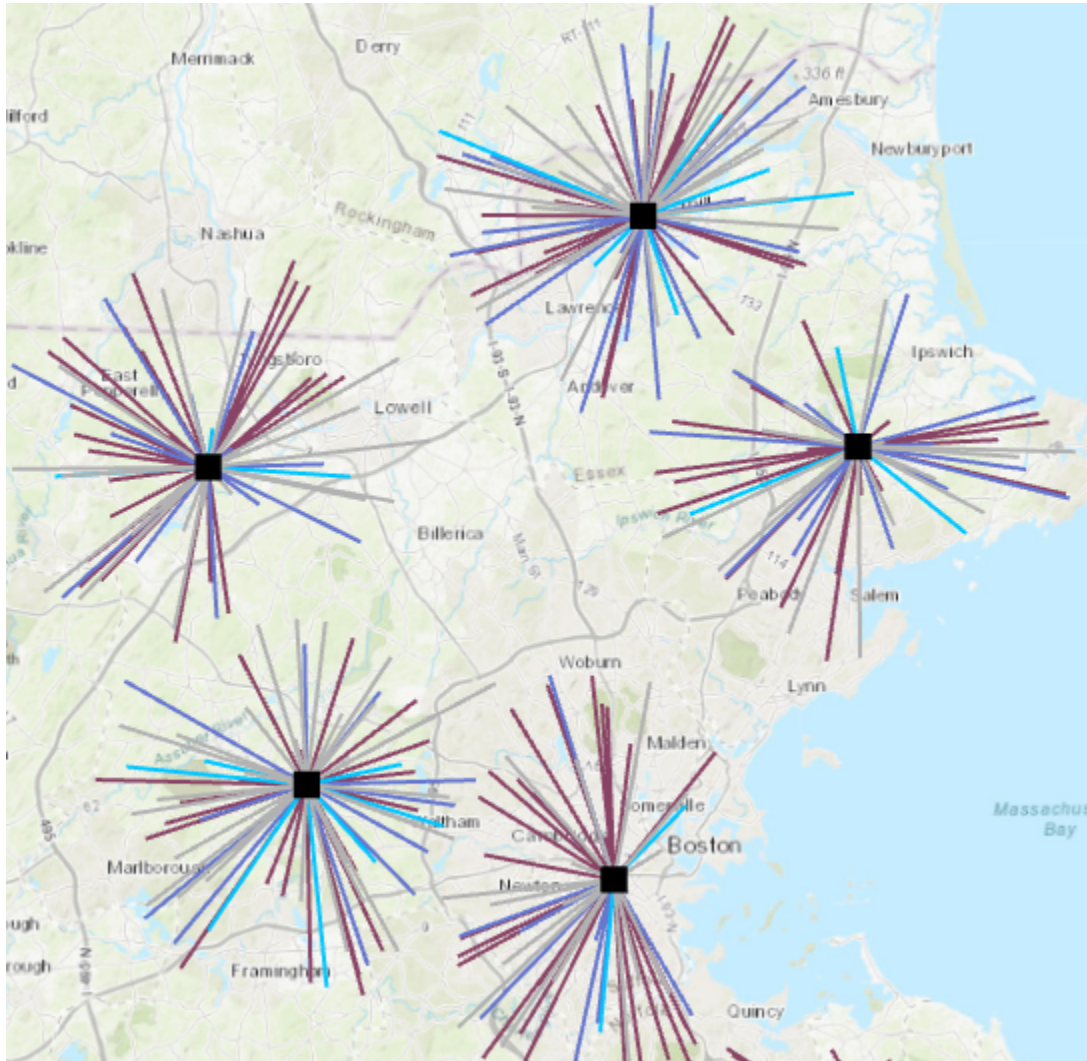
Some parameters for the XY To Line tool must change.

- m Open the XY To Line tool and change the Output Feature Class name to **%Name%\_DesireLines**.
- n For Start X Field, choose POINT\_X, and for Start Y Field, choose POINT\_Y.

You must use an in-line variable for the output name because four feature classes are being created. If you did not use a variable, each feature class would have the same name and be overwritten, and only one feature class would be created.

- o Click OK.
- p Right-click the intermediate data element named %Name%\_Points and ensure that Add To Display is not checked.
- q Right-click the final output of the model (%Name%\_DesireLines) and uncheck Add To Display.
- r From the ModelBuilder tab, click Validate, and then click Run.
- s Save the model, and then activate the Automation map.

- t** In the Catalog pane, expand Databases, and then expand SNAPCourse.gdb.
- u** Right-click the geodatabase and choose Refresh.
- v** Select and add each of the Week#Cust\_DesireLines feature classes to the map.
- w** Turn on the BostonStores layer.



You used ModelBuilder and iterators to automate an analysis workflow. You can run the same model on other input tables that you receive to map your customers.

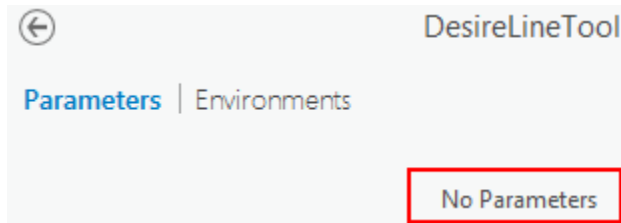
- x** Save the project.

### Step 3: Set model parameters

Next, you will add parameters to the model that allow you or others to run it as a tool and supply user-specified inputs.



- a Add a new model, and change its name and label to **DesireLineTool**.
- b Copy all model elements from the Iterate Tables model, and paste them into the DesireLineTool model.
- c Save the model.
- d In the Catalog pane, if necessary, expand Toolboxes and SNAPCourse.tbx, and then double-click the DesireLineTool model.



The model opens in a tool dialog box, which states that it has no parameters. Model parameters are required if you want to give users the ability to change inputs, outputs, or other tool properties.

1. **Which items in your model should be made available for users to provide their own data?**

---



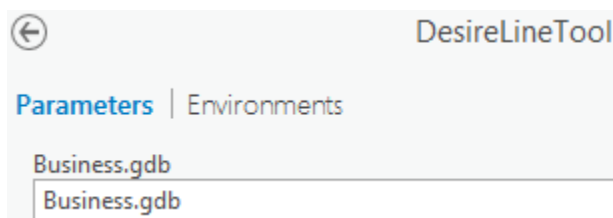
---

- e In the DesireLineTool model, click in the white space to clear your selection.
- f Right-click the first blue input data element (Business.gdb) and choose Parameter.



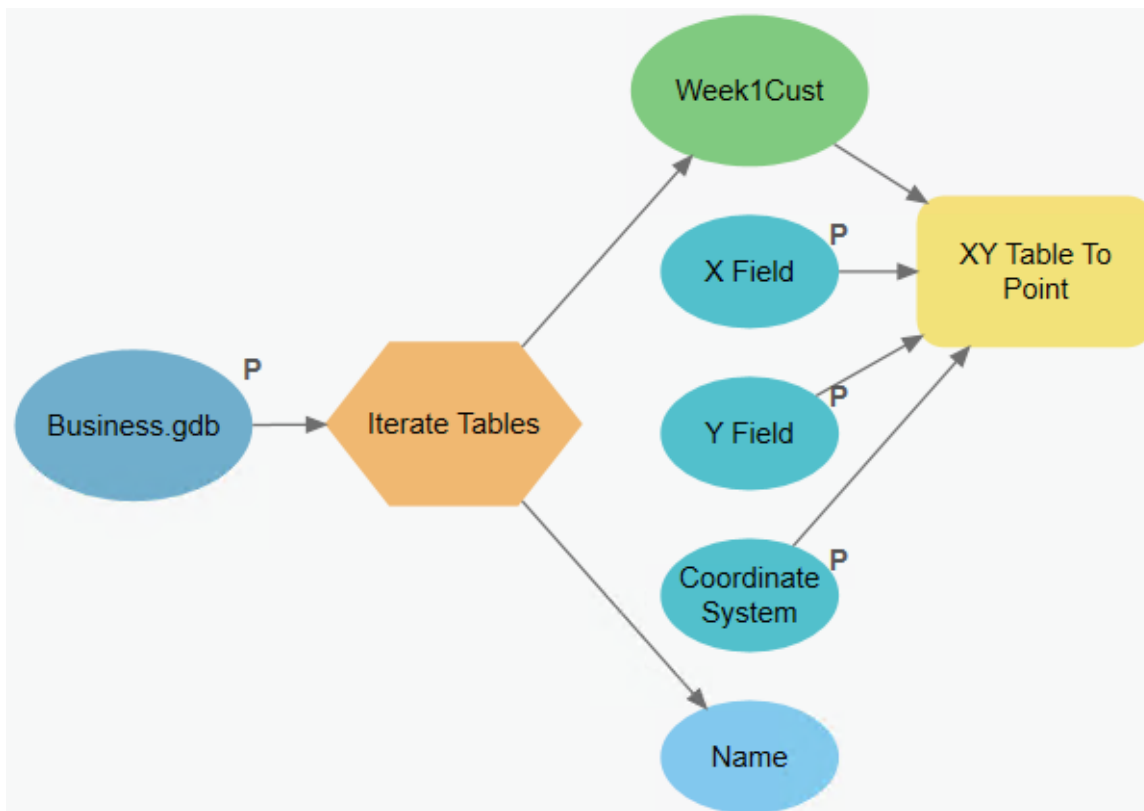
When you make a model element a parameter, a P is placed next to it.

- g Save the model.
- h In the Catalog pane, double-click DesireLineTool.

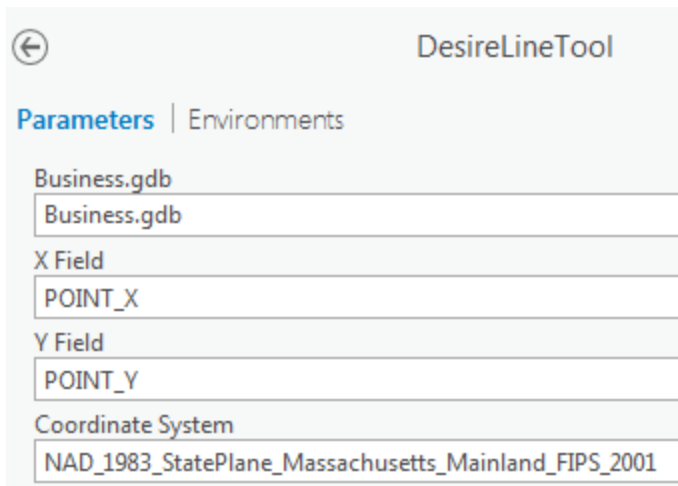


The parameter now displays in the tool dialog box, thus allowing users to input their own data. Next, you will make model parameters for the XY Table To Point tool.

- i Right-click the XY Table To Point tool, point to Create Variable, point to From Parameter, and choose X Field.
- j In the same manner, add model parameters for Y Field and Coordinate System.
- k From the ModelBuilder tab, in the View group, click Auto Layout.
- l Right-click each of the three variable elements and choose Parameter.



- m In the Catalog pane, double-click DesireLineTool.

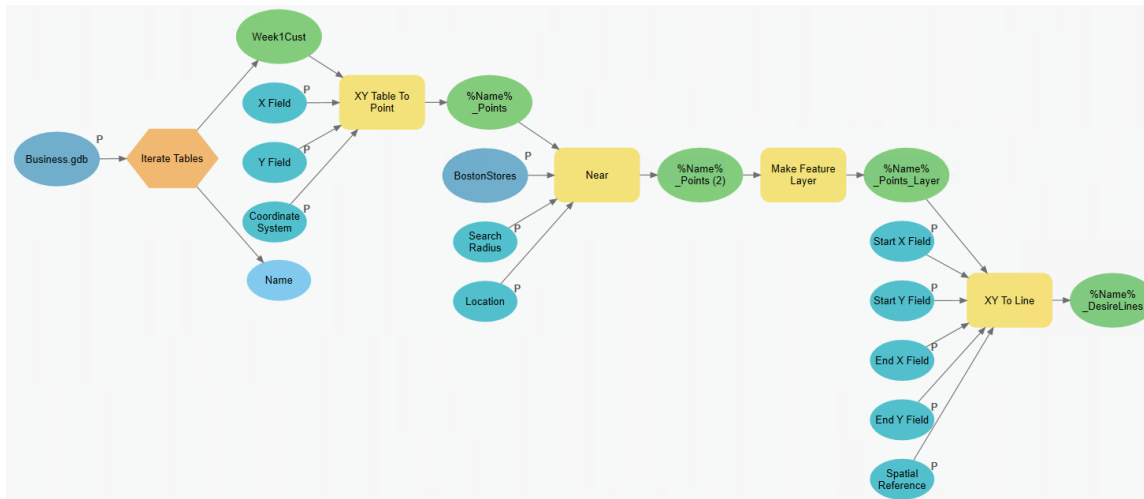


All model parameters that you set appear in the geoprocessing pane. You will continue adding parameters for most of the remaining elements.

- n Using the following table as a guide, create model parameters for the remaining elements.

Element	Parameters
BostonStores	Set as model parameter
Near	Search Radius, Location; set both as model parameters
XY To Line	Start X Field, Start Y Field, End X Field, End Y Field, Spatial Reference; set all as model parameters

- o From the ModelBuilder tab, choose Auto Layout.



- p** Save the model.

## Step 4: Change model element labels

Next, you will assign more meaningful names to each model parameter.

- a** In the model, right-click the Business.gdb input data element and choose Rename.
- b** Overwrite the existing name with **Workspace containing customer tables** and press Enter.
- c** Save the model.

- d Using the following table, rename each parameter.

Current name	Modified name
X Field	<b>Customer X</b>
Y Field	<b>Customer Y</b>
Coordinate System	<b>Spatial reference of each customer feature class</b>
BostonStores	<b>Near Features</b>
Start X Field	<b>Customer x-coordinate</b>
Start Y Field	<b>Customer y-coordinate</b>
End X Field	<b>Store X</b>
End Y Field	<b>Store Y</b>

- e Save the model.
- f In the Catalog pane, double-click the model.

DesireLineTool

**Parameters** Environments

Workspace containing customer tables  
Business.gdb

Customer X  
POINT\_X

Customer Y  
POINT\_Y

Spatial reference of each customer feature class  
NAD\_1983\_StatePlane\_Massachusetts\_Mainland\_ ▾

Near Features  
BostonStores ▾

Search Radius  
10 Miles

Location

Customer x-coordinate  
POINT\_X

Customer y-coordinates  
POINT\_Y

Store X  
NEAR\_X

Store Y  
NEAR\_Y

You have customized the labels of your parameters so that users know what the parameter is.

**g** Run the DesireLineTool.

You can make your tool accessible from the Analysis gallery so it is easier to find.

**h** In the Catalog pane, right-click the DesireLineTool model and choose Add To Analysis Gallery.



If you were to share your project as a project package, others could run the tool from the Analysis gallery.

**i** Close all models, saving them if prompted.

**j** Close the Geoprocessing pane.

- k Save the project, and leave ArcGIS Pro open.

## Lesson review

---

**1. What are the methods for automating processes in ArcGIS Pro?**

---

---

---

**2. Why would you set model parameters for your model elements and variables?**

---

---

---



# Answers to Lesson 5 questions

## Automating workflows (page 5-2)

Why would you automate your spatial analysis workflows?

**Possible responses include the following:**

- **To standardize the analysis**
- **To process many datasets at one time**
- **To share analysis results with others**
- **To repeat analysis with the same or different parameters**
- **To improve efficiency (no need to manually repeat workflows)**

## Automating and sharing models (page 5-14)

1. What does it mean to parameterize a model?

**To parameterize a model is to expose tool parameters so that when you use the model as a tool, you can add different input data.**

## Exercise 5B: Use a model to process multiple inputs (page 5-16)

1. Which items in your model should be made available for users to provide their own data?

**Input workspace, x,y coordinate fields, coordinate system, near features, search radius, location check box, input fields for XY To Line, output, and coordinate system**



# 6

## Creating surfaces using interpolation

Suppose that you want to model a feature as a continuous surface, but you only have data values for a finite number of points. For example, you want to create a precipitation map for an entire region, but you only have a few rain gauge locations recording observations in the area. How would you do it? Because surfaces represent continuous phenomena that have values at every point across their extent, you must interpolate the values for the unknown locations.

The word "interpolate" means to estimate a value that lies between two other values. From a GIS perspective, spatial interpolation refers to the process of estimating or predicting the unknown data values for specific locations using the known data values. Sample data is often collected at irregularly distributed locations, and attributes are sometimes difficult to consistently quantify. With GIS, you can use point samples to model complex surfaces that suit your specific needs and provide the information necessary to make informed and defensible decisions.

### Topics covered

---

Tobler's First Law of Geography

What is interpolation?

Interpolation methods and tools

## Tobler's First Law of Geography

---

The renowned geographer and cartographer Waldo Tobler formulated a statement known as the First Law of Geography:

*"Everything is related to everything else, but near things are more related than distant things."*

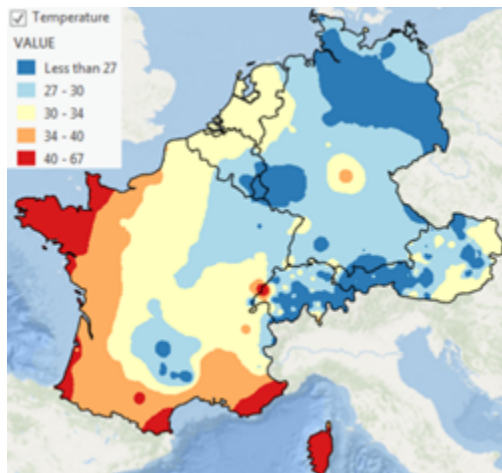
Tobler's First Law of Geography is the foundation for one of the most important concepts in spatial analysis: spatial autocorrelation. Spatial autocorrelation is a measure of the degree to which a set of spatial features and their associated data values tend to be clustered together in space (positive spatial autocorrelation) or randomly distributed (no autocorrelation). Spatial autocorrelation is an important concept used in interpolation and spatial statistics.

**What spatial phenomena might illustrate Tobler's Law?**

## What is interpolation?

---

When you watch the weather news coverage on TV, you probably see maps like the map in this image—with rainfall, snowfall, temperature, or wave height represented. The temperature values are represented as a surface. Surface data is commonly used in GIS to model continuous phenomena, like elevation, soil nutrient levels, air pollution, or temperature. Continuous phenomena do not have discrete x,y coordinates that define a boundary. However, the surface was derived from discrete points that contain a temperature value.



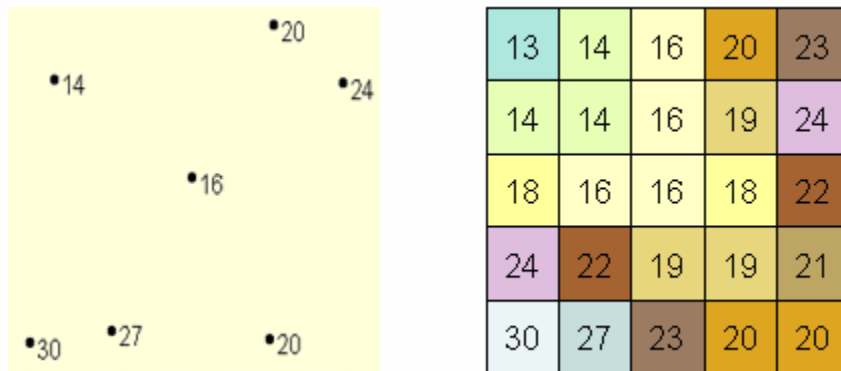
**Figure 6.1.** Weather monitoring stations are used throughout Europe to record temperature and other atmospheric conditions. A single monitoring station records values for only one location.

Data cannot be captured at every location, so a technique called spatial interpolation is used to estimate unknown values from known values. Your data must contain a value—such as elevation, precipitation, or another continuous variable.

## What is interpolation? (continued)

### General example

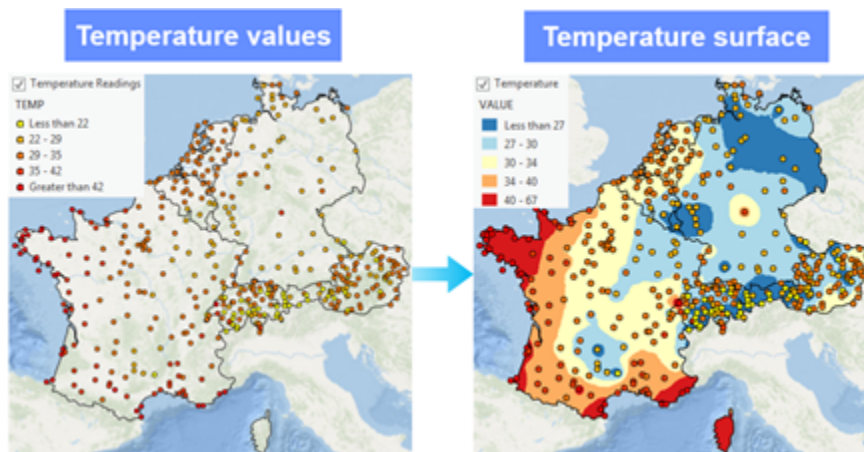
On the left are known values for some phenomenon. On the right is a surface created using the known values to predict values where no samples were recorded.



**Figure 6.2.** Imagine that the point values on the left are temperature readings. The surface on the right was created through interpolation to estimate unknown values.

### Weather map example

Individual points (recorded by monitoring stations) represent discrete temperature values at specific locations. The weather map of a continuous temperature surface can be derived from the individual values through interpolation.



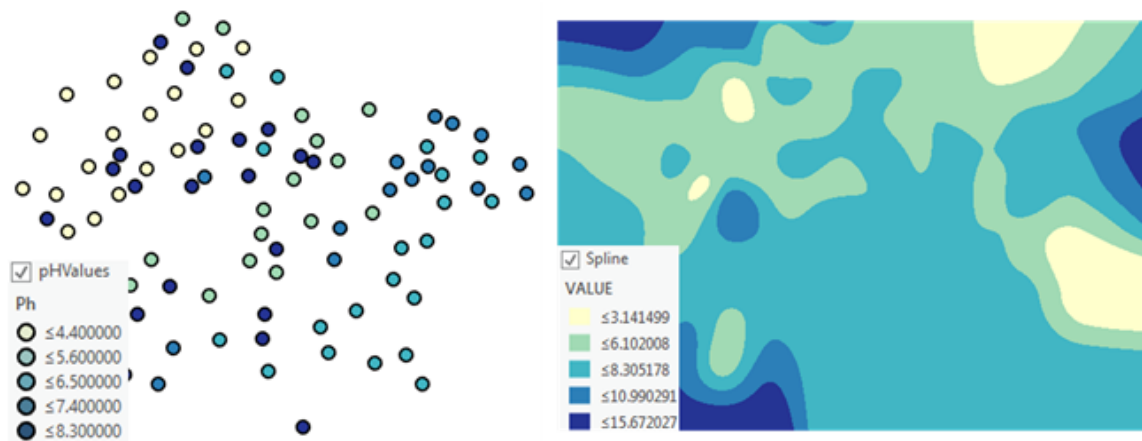
**Figure 6.3.** Temperature values at weather stations across Europe interpolated to a continuous weather surface.

## Interpolation methods

Interpolation methods can be either deterministic or geostatistical. All methods rely on the similarity of nearby sample points to create the surface, which is referred to as spatial dependence (or spatial autocorrelation).

### Deterministic

Deterministic methods use mathematical models (nonstatistical) for creating surfaces from measured points. These methods are "deterministic" because the spatial relationships in the measured points are determined by the initial data conditions and how the user specifies the model parameters.

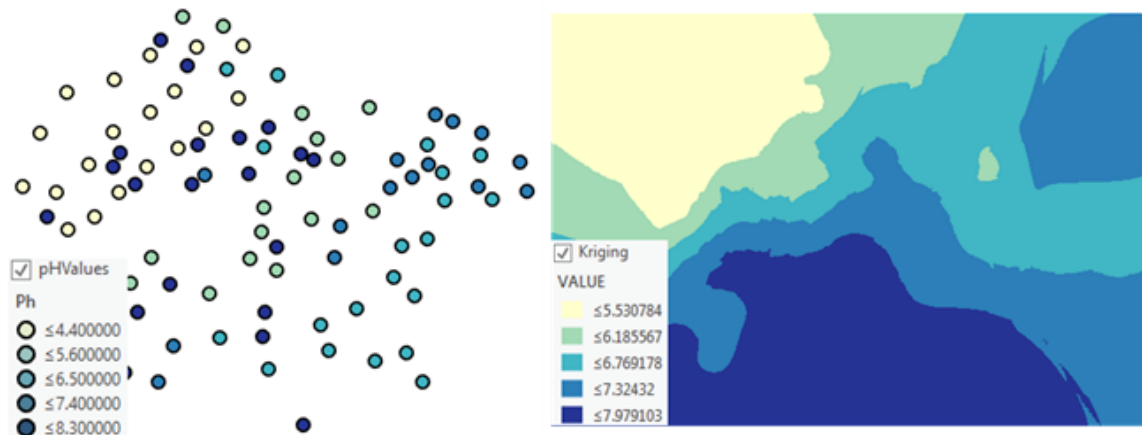


**Figure 6.4.** When using deterministic methods, no assumptions are made about the spatial statistical structure of variability in the data values. Also, uncertainty is not considered in the predictions.

## Interpolation methods (continued)

### Geostatistical

Geostatistical methods rely on both mathematical and statistical models to create output surfaces. The model parameters are estimated based on the spatial structure and statistical properties of the underlying data. Geostatistical methods assume that the data being modeled is subject to random variation and measurement error.



**Figure 6.5.** Geostatistical methods produce a prediction surface and a surface (not shown) showing estimates of prediction uncertainty.



[ArcGIS Pro Help: Deterministic methods for spatial interpolation](#)

[ArcGIS Pro Help: What are geostatistical interpolation techniques?](#)



## Interpolation tools

---

Each interpolation method estimates unknown values from known values. However, the ways in which methods work are different. The following list includes commonly used deterministic and geostatistical interpolation tools.

Tool	Description
<b>Inverse Distance Weighted (IDW)</b>	Uses the measured values surrounding the prediction location to predict a value for any unsampled location. Predicted values are based on the assumption that things that are close to one another are more alike than things that are farther apart.
<b>Natural Neighbors</b>	Finds the closest subset of input samples to a query point and applies weights to them based on proportionate areas to interpolate a value.
<b>Spline</b>	Estimates values to minimize overall surface curvature, resulting in a smooth surface that passes exactly through the input points.
<b>Kriging</b>	An advanced geostatistical procedure that generates an estimated surface from a scattered set of points with measured values of an attribute of interest.
<b>Empirical Bayesian Kriging</b>	A geostatistical interpolation method that automates the most difficult manual aspects of building a valid model.

Interpolation tools are located in the Spatial Analyst, Geostatistical Analyst, and 3D Analyst toolboxes, and all require extensions.

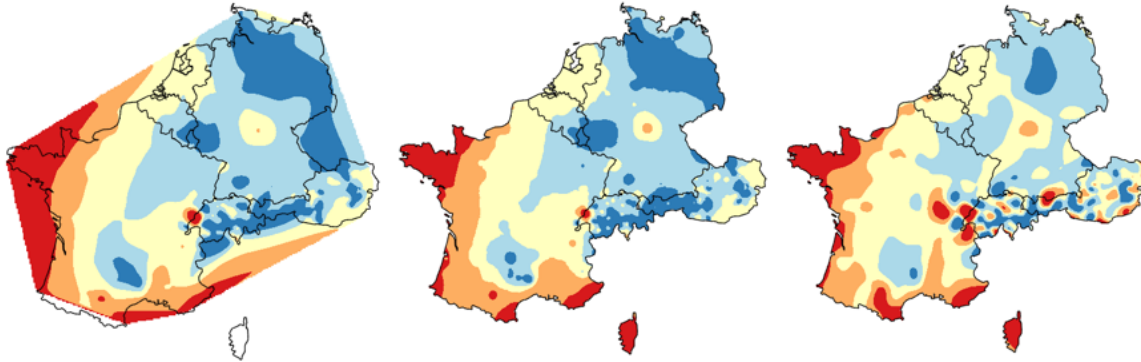


**ArcGIS Pro Help:** *Classification trees of the interpolation methods offered in Geostatistical Analyst*

## Deterministic interpolation

---

Deterministic interpolation techniques create surfaces from measured points, based on either the extent of similarity (inverse distance weighted) or the degree of smoothing.



**Figure 6.6.** Compare results from three deterministic interpolators.

## Deterministic interpolation (continued)

---

### How do you know which surface is best?

Geostatistical interpolation uses validation, but most deterministic interpolators do not. You can validate deterministic surfaces using the following methods:

- Manually validate the surfaces using the Explore tool, and compare the cell values with the sample point values in the same location.
- Create the surface on a subset of points, thus withholding some sample points. After you create the surface on the subset, explore how well the interpolator estimated values where the withheld points are located. You can create the subset manually using the Subset Features tool in the Geostatistical Analyst toolbox to create training and testing data, and then use the GA Layer To Points tool to perform the validation.
- Use the Cross Validation tool in the Geostatistical Analyst toolbox. If you use a deterministic interpolator from the Geostatistical Analyst toolbox, you can perform cross validation on it using a geoprocessing tool.



**ArcGIS Pro Help:** [Subset Features](#)

**ArcGIS Pro Help:** [GA Layer To Points](#)

**ArcGIS Pro Help:** [Performing cross-validation and validation](#)

## Interpolate surfaces

The U.S. Environmental Protection Agency is responsible for monitoring atmospheric ozone concentration in California. Ozone concentration is measured at monitoring stations throughout the state. The concentration levels of ozone are known for all the stations, but the ozone values for other unmonitored locations in California are also of interest. However, it is too costly and impractical to put monitoring stations everywhere. In this exercise, you will use interpolation to create continuous surfaces from the ozone sample points.

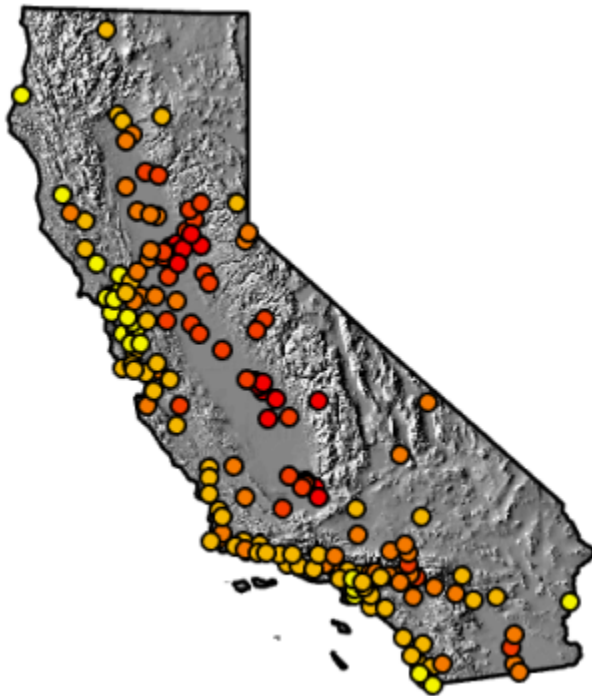
In this exercise, you will perform the following tasks:

- Use interpolation to create surfaces.
- Validate surfaces.

## Step 1: Examine data

You will open the class project and add data.

- a If necessary, start ArcGIS Pro and open the SNAPCourse project.
- b Close any open views.
- c From the Insert tab, in the Project group, click Import Map.
- d Browse to **C:\EsriTraining\SNAP\Interpolation** and import Interpolation.mapx.



You can clearly see visual clusters of high and low ozone concentrations. The clusters indicate that interpolation is a good option for creating a continuous surface. Next, you will explore the attributes for the sample points.

- e In the Contents pane, right-click Samples and open its attribute table.

OBJECTID	Shape	SITE_NAME	BASIN_NAME	ELEVATION	OZONE
1	Point	El Capitan Beach	South Central Coast	24	0.045
2	Point	Fresno-Drummond S...	San Joaquin Valley	89	0.106
3	Point	San Diego-Overland...	San Diego	135	0.04
4	Point	Ventura-Emma Woo...	South Central Coast	3	0.041
5	Point	Visalia-N Church Stre...	San Joaquin Valley	97	0.09
6	Point	Carmel Valley-Ford R...	North Central Coast	131	0.053

The table contains the name of the monitoring station, its elevation, and the ozone measurement. Currently, the Samples layer is symbolized using the OZONE field. You will use the OZONE field when you interpolate a continuous surface from the sample points.

- f Close the table.
- g In the Contents pane, turn off the Samples and Hillshade layers.

## Step 2: Set geoprocessing environments

Before you interpolate, you will ensure that the correct environments are set.

- a From the Analysis tab, in the Geoprocessing group, click Environments.
- b Set the following environments:
  - Extent: Same As Layer - Hillshade
  - Cell Size: Same As Layer - Hillshade
  - Mask: StateBoundary
- c Click OK.

## Step 3: Interpolate using the Natural Neighbor tool

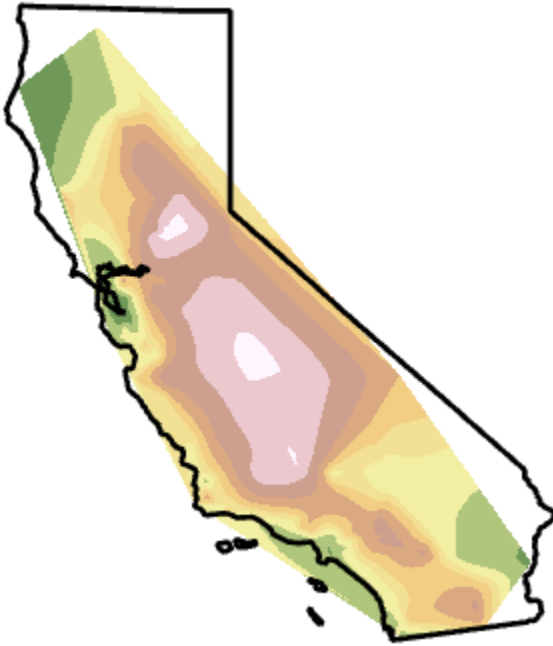
You will compare interpolation tools on the ozone sample points. First, you will use the Natural Neighbor tool.

- a From the Analysis tab, click Tools.
- b In the Geoprocessing pane, click the Toolboxes tab.
- c Expand Spatial Analyst Tools, and then expand Interpolation.

d Open the Natural Neighbor tool, and set the following parameters:

- Input Point Features: Samples
- Z Value Field: OZONE
- Output Raster: **OzoneNN**

e Click Run.



There are no sample points in some areas, and the Natural Neighbor tool creates surfaces that pass through only the sample points. You set the analysis mask, but Natural Neighbor does not honor an analysis mask.

f In the Contents pane, turn the Samples layer on and off to view the points with the surface.

g When you are finished, ensure that the Samples and OzoneNN layers are turned off.

## Step 4: Interpolate using the Spline tool

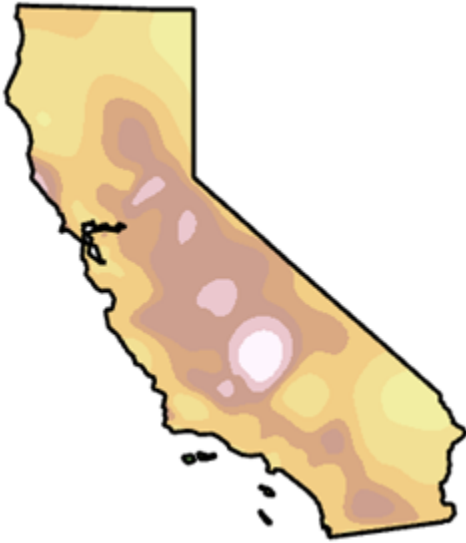
Next, you will use the Spline tool to interpolate an ozone surface using the regularized method.

a In the top-left corner of the Geoprocessing pane, click the Back button .

**b** From the Interpolation toolbox, open the Spline tool, and set the following parameters:

- Input Point Features: Samples
- Z Value Field: OZONE
- Output Raster: **OzoneSpline**

**c** Click Run.



**d** Turn the Samples layer on and off to view it with the surface.

Next, you will run the Spline tool using the tension method. With the Tension spline type, higher values entered for the weight parameter result in somewhat coarser surfaces that nonetheless closely conform to the control points.

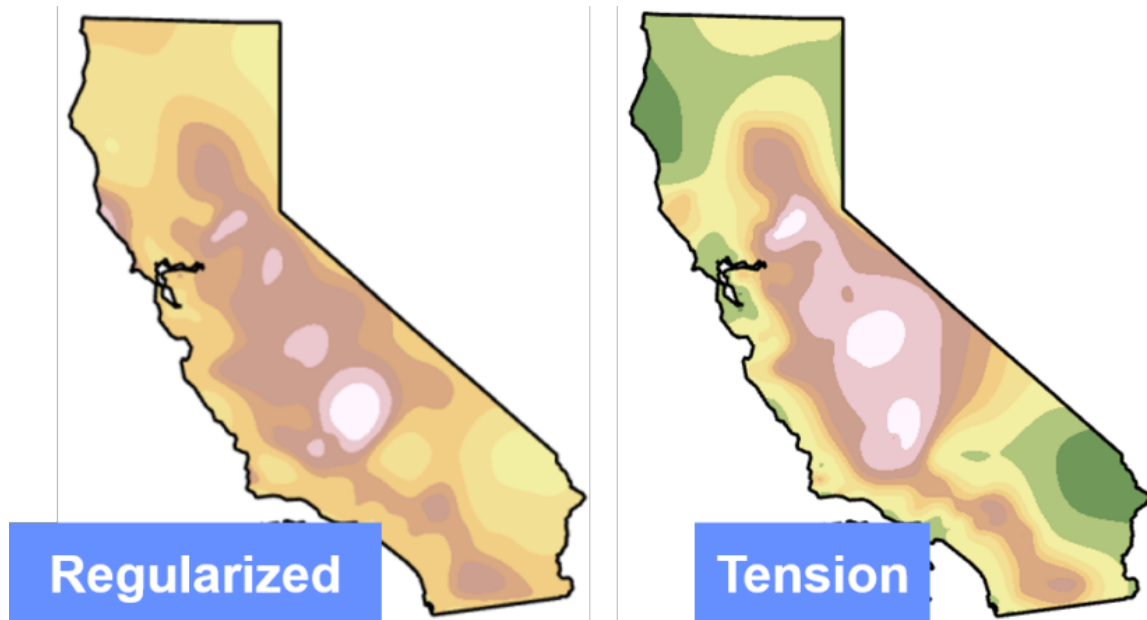
**e** In the Geoprocessing pane, change the Output Raster name to **OzoneSplineTension**.

**f** Change the Spline Type to Tension, and then click Run.

**g** In the Contents pane, ensure that the OzoneSplineTension layer is selected.

**h** From the Appearance tab, in the Effects group, use the Swipe tool to compare results from the regularized and tension methods.





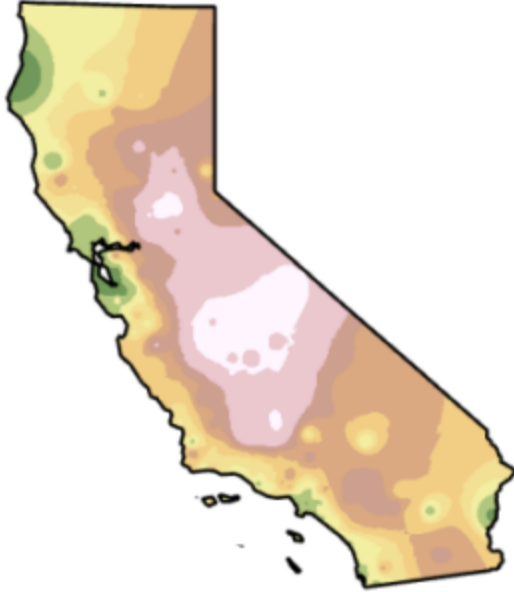
Regularized spline creates a smoother surface than the tension method does.

- i** In the Contents pane, turn off all interpolated surface layers.

## Step 5: Interpolate using inverse distance weighted interpolation

Next, you will interpolate using an inverse distance weighted (IDW) tool and compare the results with other surfaces.

- a** In the Geoprocessing pane, click the Back button, and from the Interpolation toolbox, open the IDW tool.
- b** Set the following parameters:
  - Input Point Features: Samples
  - Z Value Field: OZONE
  - Output Raster: **OzoneIDW**
- c** Click Run.



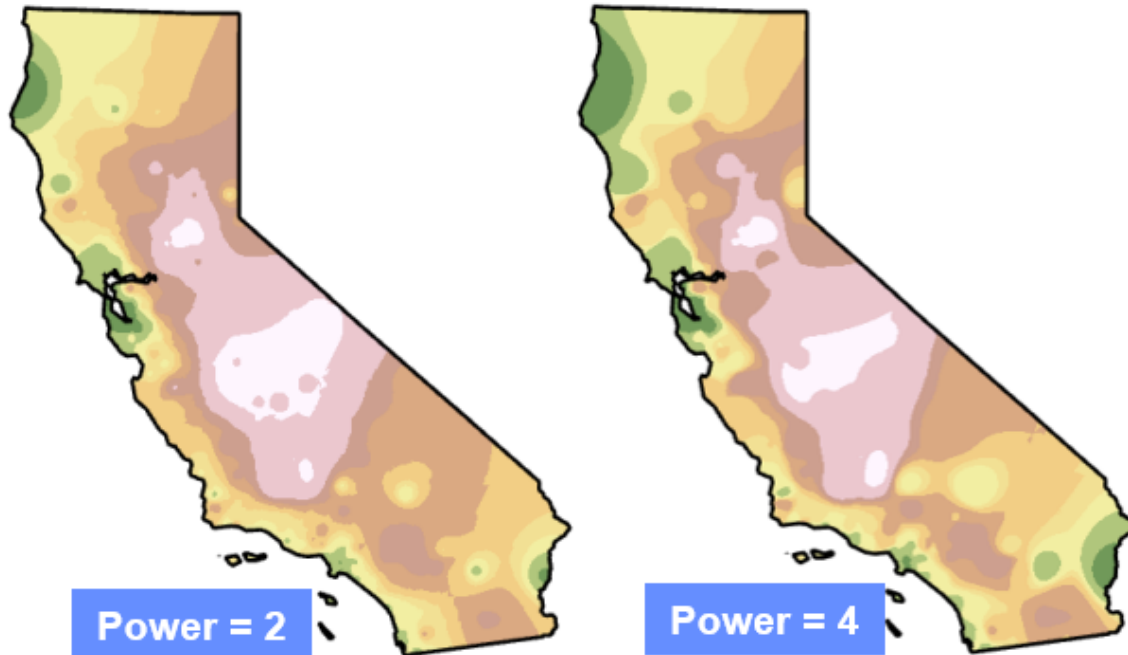
Next, you will modify the Power parameter. The Power parameter allows you to control the significance of known points on the interpolated values based on their distance from the output point. By defining a higher Power value, more emphasis is placed on the nearest points. Thus, nearby data will have the most influence on the estimated values.

**d** In the Geoprocessing pane, change the following parameters for the IDW tool:

- Output Raster: **OzoneIDWP**
- Power: **4**

**e** Run the tool again.

**f** In the Contents pane, compare the IDW results by turning the layers off and on.



- g** Save the project.

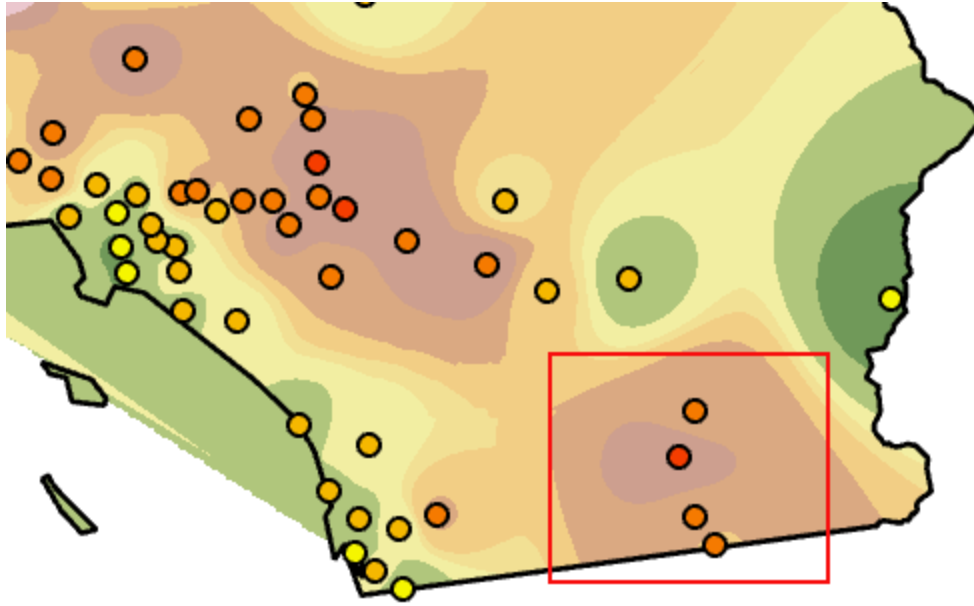
## Step 6: Examine interpolated values

How do you know which of the surfaces that you created is the best at estimating ozone values? In this step, you will explore the results of your interpolation by comparing estimated values with known values.

- a** In the Contents pane, turn on all layers except the basemap and Hillshade.
- b** From the Map tab, in the Navigate group, click the Explore tool to activate it.
- c** Click the Explore tool down arrow and choose Selected In Contents.
- d** In the Contents pane, select the Samples layer.

Now you can base which attributes are displayed by the Explore tool on the selected layer.

- e** Use the Explore tool to zoom to the southern part of California.



You will focus on the sample points specified in the graphic to see which tool predicted values more closely to the recorded value.



f With the Explore tool, click the second point from the top.

A screenshot of a data popup window titled "Samples - Westmorland-W 1st Street". The window displays the following data for the selected sample point:

OBJECTID	108
SITE_NAME	Westmorland-W 1st Street
BASIN_NAME	Salton Sea
ELEVATION	-32
OZONE	0.071

The recorded ozone measurement from the monitoring station is 0.071.

- g** In the Contents pane, select OzoneNN.
- h** With the Explore tool, click the same point.

The Natural Neighbor tool interpolated this part of the surface to a value of 0.070. Next, you will get values for the Spline and IDW tool surfaces.

- i** Using the skills that you have learned, get the interpolated value at the same point for the other interpolated layers.



Closing the pop-up window is not necessary. Simply change the selected layer, and then click the point again. If the pop-up window opens over the point, zoom out so that you can see the point and the pop-up window together or move the pop-up.

The predicted values for the sample point are all close or exact. Normally, you would perform this operation on several points throughout the study area to determine which tool predicted the best. You could also determine the best surface by withholding some sample points from the interpolation. You could perform the interpolation on a subset of points, and then compare the predicted values with the values of the withheld points.

- j** Save your project.
- k** If you have time and want to explore the interpolated surfaces further, proceed to the challenge step.

## Challenge: Challenge step



*10 minutes*

- a** Interpolate `..\EsriTraining\SNAP\Interpolation\CaliOzone.gdb\SubsetPoints` using the Natural Neighbor, Spline, and IDW tools (all with default settings) to create three surfaces.
- b** When you have created the surfaces, add `..\EsriTraining\SNAP\Interpolation\CaliOzone.gdb\WithHeldPoints` to the map.
- c** Use the Explore tool to see how well the interpolation tools predicted values that were excluded from the interpolation.

## Lesson review

---

**1. Describe interpolation.**

---

---

---

---

---

**2. What is deterministic interpolation?**

---

---

---

---

**3. What are some ways in which you can validate surfaces created using interpolation?**

---

---

---

---

---

# Answers to Lesson 6 questions

## Tobler's First Law of Geography (page 6-2)

What spatial phenomena might illustrate Tobler's Law?

**Possible responses include the following:**

- Houses in one development are more similar in value than in neighborhoods farther away.
- If it is snowing where you are, it is likely to be snowing 100 feet away from where you are but maybe not 100 miles from where you are.

# Exercise 6 challenge solution

- a Turn off all layers except the StateBoundary layer.
- b In the Catalog pane, expand Folders, and then expand Snap and Interpolation.
- c Expand the CaliOzone geodatabase, and add SubsetPoints to the map.

SubsetPoints has fewer sample points in it than the initial Samples layer. You will use the same interpolation tools as you did earlier, but this time, you will use the SubsetPoints feature class as the input.

- d Locate and run the Natural Neighbor tool with the following parameters:

- Input Point Features: SubsetPoints
- Z Value Field: OZONE
- Output Raster: **NNSubset**

- e Locate and run the Spline tool with the following parameters:

- Input Point Features: Samples
- Z Value Field: OZONE
- Output Raster: **SplineSubset**

- f Locate and run the IDW tool with the following parameters:

- Input Point Features: Samples
- Z Value Field: OZONE
- Output Raster: **IDWSubset**

- g From the Catalog pane, in the CaliOzone geodatabase, add WithHeldPoints to the map.

- h Zoom to the following points:



# Exercise 6 challenge solution

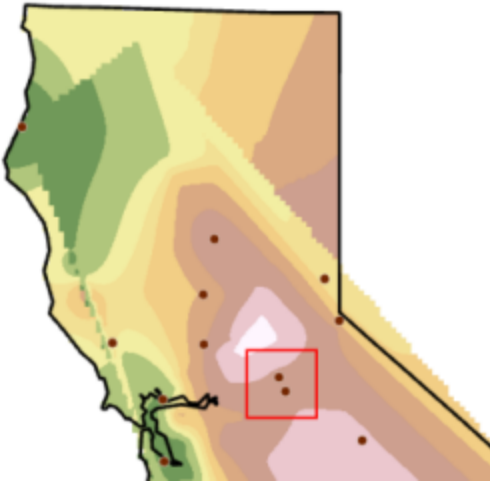


Figure 6.7.

- i In the Contents pane, ensure that WithHeldPoints is selected.
- j In the map, using the Explore tool, click the top point.

WithHeldPoints - Jackson-Clinton Road	
OBJECTID	13
SITE_NAME	Jackson-Clinton Road
BASIN_NAME	Mountain Counties
ELEVATION	377
OZONE	0.074

Figure 6.8.

- k In the Contents pane, select NNSubset, and then click the point.
  - l Repeat the process of selecting SplineSubset and then IDWSubset and clicking the point to get the predicted values.
1. Which interpolation tool best predicted the value of the withheld point?

**IDW, with a value of 0.074**

As you can see, manually validating a surface is time-consuming. Later in the course, you will use geostatistical tools to validate surfaces.

- m Save your project, and keep ArcGIS Pro open.



# 7

## Suitability modeling

Suitability modeling is a type of analysis that locates places that are "suitable" or favorable for certain phenomenon. An example would be the best locations for a wind farm in Colorado. You will learn about a standard workflow for performing suitability modeling that you can apply to any data or analyses and use it to solve a problem.

### Topics covered

---

- Suitability modeling workflow

- Differences between raster and vector overlay

- Deriving surfaces from other sources

- Levels of measurement

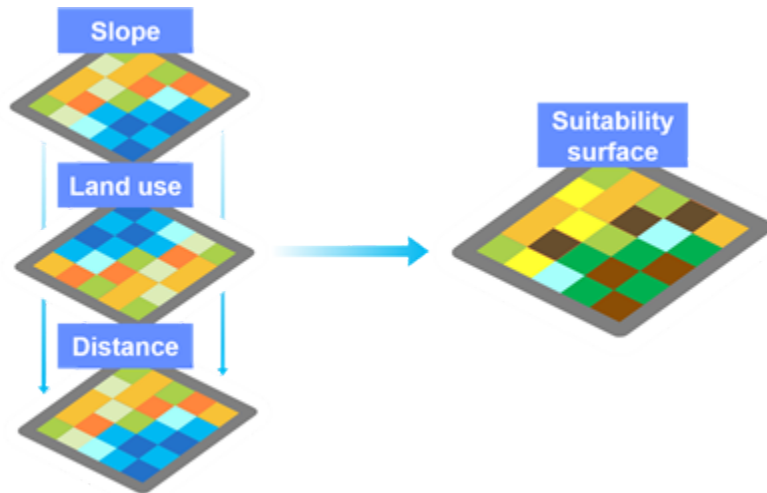
- Transforming values to a common scale

- Types of overlay analysis

## What is suitability modeling?

---

Suitability modeling is the process of combining multiple datasets, usually raster, together into one layer with the intention of finding optimal locations for various phenomena. ArcGIS Pro has several raster overlay tools, such as Weighted Overlay and Weighted Sum, that allow you to weight layers based on their relative importance to a suitability modeling scenario.

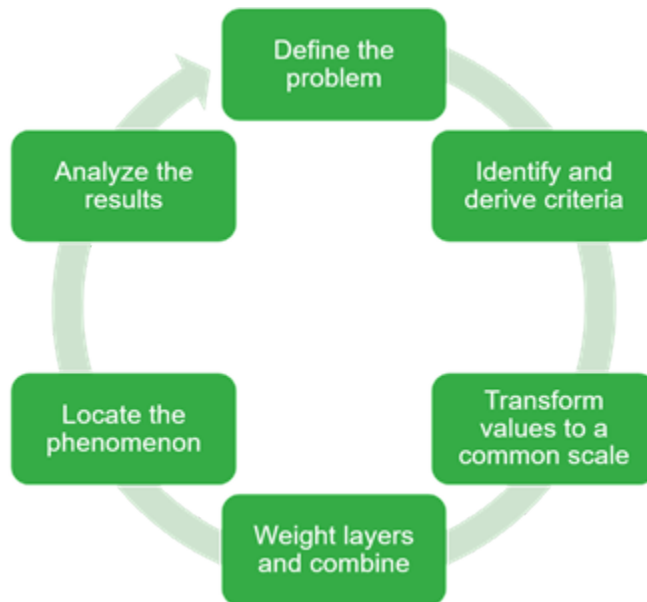


**Figure 7.1.** Distance, land use, and slope rasters are combined into a single raster that contains cells suitable for a vineyard. Vineyards are successful on certain slopes, on certain land uses, and at certain distances from roads.

## Suitability modeling workflow

---

When performing suitability modeling, you can guide your analysis using a standard workflow.



**Figure 7.2.** You can follow the suitability modeling workflow using any analysis problems and datasets.

**Define the problem:** State the problem that you are trying to solve, such as finding the best location for a wind energy facility.

**Identify and derive criteria:** Criteria are the conditions that geographic areas must meet to be considered suitable. For example, daily average wind speed must be at least 25 mph.

**Transform values to a common scale:** When dealing with datasets containing different measures and ranges, you must transform data values so that you can rank them on the same scale.

**Weight layers and combine:** Suitability modeling involves weighting layers based on their relative importance to the problem that you are trying to solve. After you weight the layers, you use tools to combine them onto a suitability surface.

**Locate the phenomenon:** With the resulting suitability surface, you can dig deeper to locate best sites or regions that are most suitable.

**Analyze the results:** Explore the findings, perhaps alter some tool parameters, and rerun the analysis to get the best result.

## Evaluating analysis criteria

---

The first step of the suitability modeling workflow is to define the problem. You must determine the proper climate conditions for growing crops. You will consider phenomena like temperature, elevation, slope, and distance from roads to find the most suitable places. The datasets of interest become the analysis criteria.

**What types of data can you identify in the criteria, and can you use vector overlay tools like Intersect and Union with them?**

## Choosing vector or raster overlay

---

Based on the data and scenario provided, determine whether you would use raster or vector analysis tools.

### Scenario 1: Wind power

You want to determine the best locations in the North Sea where wind could be effectively harvested as an alternative energy source. Your criteria are as follows:

- Must have good wind potential
- Must be within 100 kilometers of major European ports
- Must be away from high shipping areas
- Must be in water less than 60 meters deep
- Must be at least 100 kilometers away from marine protected areas

Potential datasets include wind, ports, shipping lanes, bathymetry, and nature preserve boundary.

1. **Would you use raster or vector overlay to determine suitable ocean locations to harvest wind power? Why?**
- 

### Scenario 2: New store location

You want to determine the best locations for a new shopping center in a city. Your criteria are as follows:

- Must be near major roads and highways
- Must be in demographic areas that will provide customers
- Must be at least 5 miles from competitors
- Must be commercially zoned

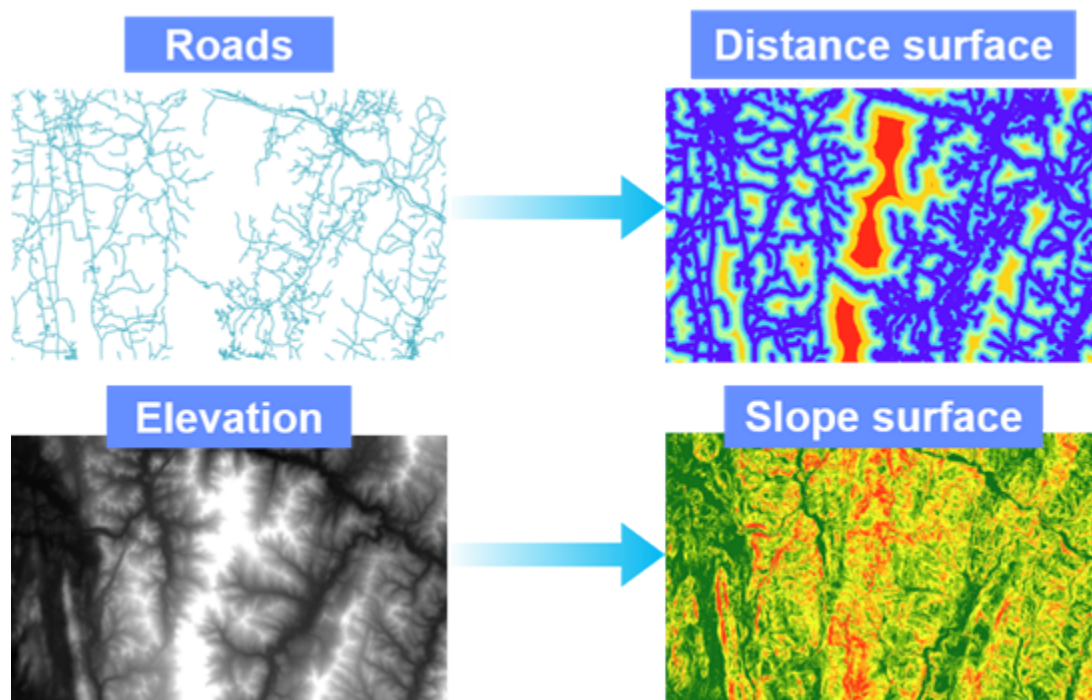
Potential datasets include roads, census blocks, competitor stores, and zoning.

2. **Would you use raster or vector overlay to determine the most suitable locations for a shopping center? Why?**
-

## Deriving surfaces from other sources

---

In raster overlay analysis, you work with surfaces. A surface is a geographic phenomenon represented as a set of continuous data (such as elevation, geological boundaries, or air pollution). Surfaces do not have discrete x,y coordinates for phenomena because the data being modeled does not have set boundaries and is more continuous over the landscape. After determining the criteria that you need for your analysis, you may not always have the data that is required to model these criteria. For example, you might need slope but have only elevation, or need distance and have only roads.



**Figure 7.3.** You can derive surfaces from vector or raster data using geoprocessing tools or raster functions.



## Raster functions and geoprocessing tools

---

In ArcGIS Pro, you can create raster data in two main ways: raster functions and geoprocessing tools.

### Raster functions

Using a raster function is a quick way to process and analyze rasters in ArcGIS Pro. You can apply a raster function to raster datasets, mosaic datasets, or image services that are in your map. The resulting virtual layers are stored in your current project. You can apply system functions for data management, visualization, and analysis.

Raster functions do not create permanent data; they process only the pixels that are visible on your screen, creating virtual layers in the map, which saves disk space and results in fast processing. If you want to save the result of a raster function, you can export it to a geodatabase raster.

### Geoprocessing tools

To build a geodatabase with rasters, you would use geoprocessing tools. Some raster functions and geoprocessing tools are similar, such as Hillshade and Slope, and whether you use a function or a tool depends on the output that you want. You can add geoprocessing tools to models, but you cannot add raster functions. However, you can create a function chain, which is similar to a model.

## Levels of measurement

---

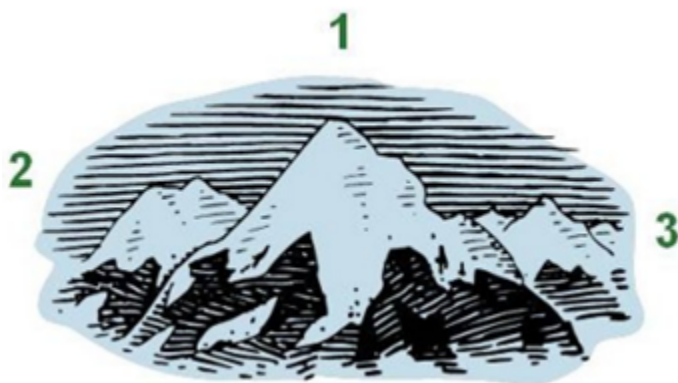
There are various types of data in GIS—nominal, ordinal, interval, and ratio—that are referred to as levels of measurement. Each type allows various mathematical operations to be performed on it. An understanding of levels of measurement is vital in weighted suitability modeling because you are essentially taking nominal (land use), interval (temperature), or ratio (distance to roads) and transforming it into interval or ratio data.

Nominal data supports the relational operation equality (=).



**Figure 7.4.** Nominal data is a name or description, such as the peak names shown in the image.

Ordinal data supports the relational operators *equal to* (=), *not equal to* ( $\neq$ ), *greater than* (>), *less than* (<), *greater than or equal to* ( $\geq$ ), and *less than or equal to* ( $\leq$ ).



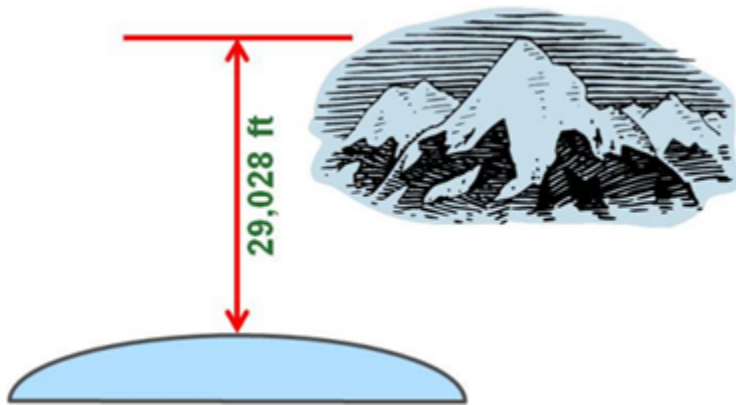
**Figure 7.5.** Ordinal measurements determine importance, such as 1st, 2nd, and 3rd place in a race, or which peak is higher than another. You cannot add, subtract, multiply, or divide the numbers.

Interval measurements capture values that are measurable on an interval scale, such as temperature or elevation. Interval data has an arbitrary zero point (for example, 0 degrees F does

## Levels of measurement (continued)

---

not imply "no temperature"). Interval data supports all relational operations supported by nominal and ordinal measurements, and the mathematical operations of addition and subtraction.



**Figure 7.6.** The elevation of Mt. Everest is an example of interval data or measurement. The zero point (sea level) is arbitrary.

Ratio scales are used for many measurements in the physical sciences and engineering, such as mass, length, time, height, and energy. All relational operators can use a variable measured at the ratio level, and all necessary mathematical operators (+, -,  $\times$ , /). For example, height can be represented as ratio data because one object, such as a building, can be twice as tall as another. Ratio data has an absolute zero point. For example, a county with zero population implies the complete absence of people.



**Figure 7.7.** The height of a mountain from an absolute zero point is an example of ratio data or measurement.

## Transforming values to a common scale

When you perform a weighted overlay, you combine many surfaces that contain different ranges of values and may be in different units of measure. You must address this issue of differing values and ranges before you overlay the rasters. Based on the type of data that you have, you may reclassify the data values manually or use a tool that automates the process.

### Reclassify

Reclassification involves manual assignment of data values into discrete classes. Reclassify is best used for discrete data that will have distinct class breaks. You can enter the values manually or load them from a table. The output raster will contain only the values of your suitability scale and its cells distributed to the various classes as specified by the user.

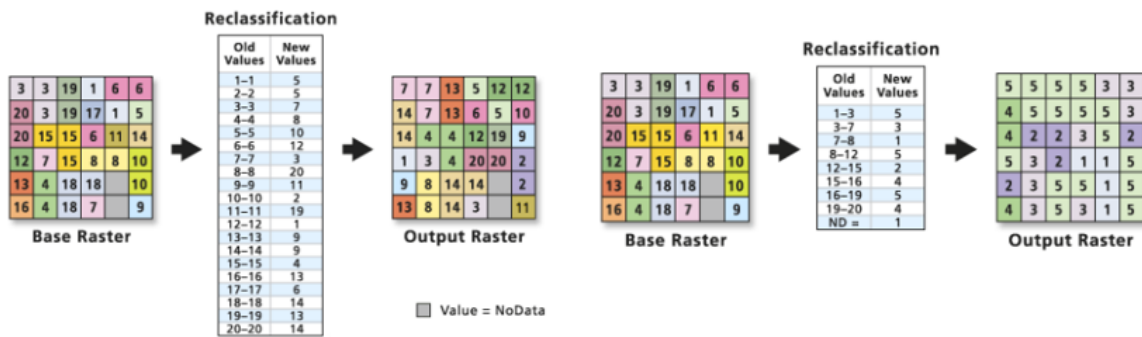


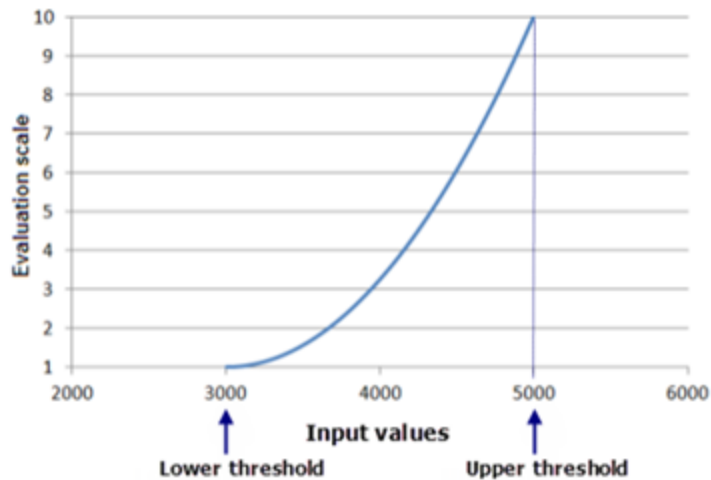
Figure 7.8. You can reclassify using individual values (shown in the example on the left) or ranges of values (shown in the example on the right).

### Rescale by function

Another method available to transform data values is called rescale by function. Many times, the suitability changes continuously with the changing values of the criterion and often does so in a nonlinear manner. For example, cell locations close to existing roads may be the most preferred in a housing suitability model because the cost of getting power to those locations is cheaper. As the distance from a road increases, the cost of getting power to those locations may increase exponentially. As a result, the suitability for farther locations may decrease dramatically. Rescale by function is a better option than reclassification for continuous data.

## Transforming values to a common scale (continued)

---



**Figure 7.9.** A function used to transform data values onto the desired suitability scale. Notice that there are no distinct class breaks.

### Using rescale by function

- Use if you have a complete understanding of the distribution of values in your data and if that distribution is consistent with the phenomenon being modeled.
- Should be applied only to continuous data.
- A way of automating the assignment of continuous data to a suitability scale based on the values in your data.
- There are many functions to choose from, so pick the one that best captures the phenomenon being studied.

## Build a model and classify data to a common scale

You will build a model to derive surfaces from various sources and to classify and transform data values to a common suitability scale. Then, in the second exercise, you will overlay the surfaces.

In this exercise, you will perform the following tasks:

- Build a suitability model.
- Derive a surface from raster and vector sources.
- Reclassify and rescale data onto a common scale.

## Step 1: Prepare a project and set environments

You will prepare the project and evaluate the analysis criteria.

- a If necessary, start ArcGIS Pro and restore the course project.
- b Close any open views.
- c From the Insert tab, click Import Map.
- d Browse to **C:\EsriTraining\SNAP\RasterAnalysis** and import BearSuitability.mapx.

The map displays streams, roads, land use, and elevation for an area in Vermont. For raster analysis, it is important to verify the cell size of your data.

- e In the Contents pane, right-click Elevation and choose Properties.
- f Click the Source tab, and then expand Raster Information.

### ▼ Raster Information

Columns	1439
Rows	900
Number of Bands	1
Cell Size X	30
Cell Size Y	30

- g Click OK.
- h Using the same steps that you just performed, find the cell size for the LandUse raster.

In this case, the cell sizes match for the input rasters. You will set the output cell size for other rasters that you create to match the cell size of the inputs, which is 30 meters. Next, you will set several environments for the analysis.

- i Close the Layer Properties dialog box.
- j From the Analysis tab, in the Geoprocessing group, click Environments.
- k Set the following environments:
  - Extent: Same As Layer - Elevation
  - Cell Size: Same As Layer - Elevation
  - Mask: Highlight StateBoundary and press Delete to clear the mask

Raster Analysis	
Cell Size	C:\EsriTraining\SNAP\RasterAnal ▾
Cell Size Projection Method	Convert units
Mask	▾

- l** Click OK.
- m** Save the project.

## Step 2: Create a model

You will perform the analysis in ModelBuilder, so you will first create a model.

- a** From the Analysis tab, in the Geoprocessing group, click ModelBuilder.
- b** Update the model name to **BearSuitability** and its label to **Bear Suitability**.

*Hint:* Catalog pane > Toolboxes > SNAPCourse.tbx > Right-click Model > Properties

- c** From the ModelBuilder tab, click Save.

## Step 3: Add input layers and Euclidean Distance tools

You have identified four input data sources (two raster and two vector) that you will use to derive surfaces for overlay analysis. You will begin by adding the four input data layers to the model.

- a** From the Contents pane, drag each of the layers, excluding the basemap, into the model and arrange them one under each other as follows:
  - LandUse
  - Roads
  - Streams
  - Elevation





You will use two vector layers, Streams and Roads, as inputs into the Euclidean Distance tool to create distance surfaces for the raster analysis.

- b** In the Geoprocessing pane, search for **euclid**.
- c** Drag the Euclidean Distance tool into the model and place it to the right of Streams.
- d** Connect Streams to Euclidean Distance as Input Raster Or Feature Source Data.
- e** Open the Euclidean Distance tool, and change the Output Distance Raster name to **StreamDist**.

Input raster or feature source data	
Streams:1	
Output distance raster	
StreamDist	
Maximum distance	<input type="text"/>
Output cell size	
Elevation	
Output direction raster	
<input type="text"/>	
Distance Method	
Planar	

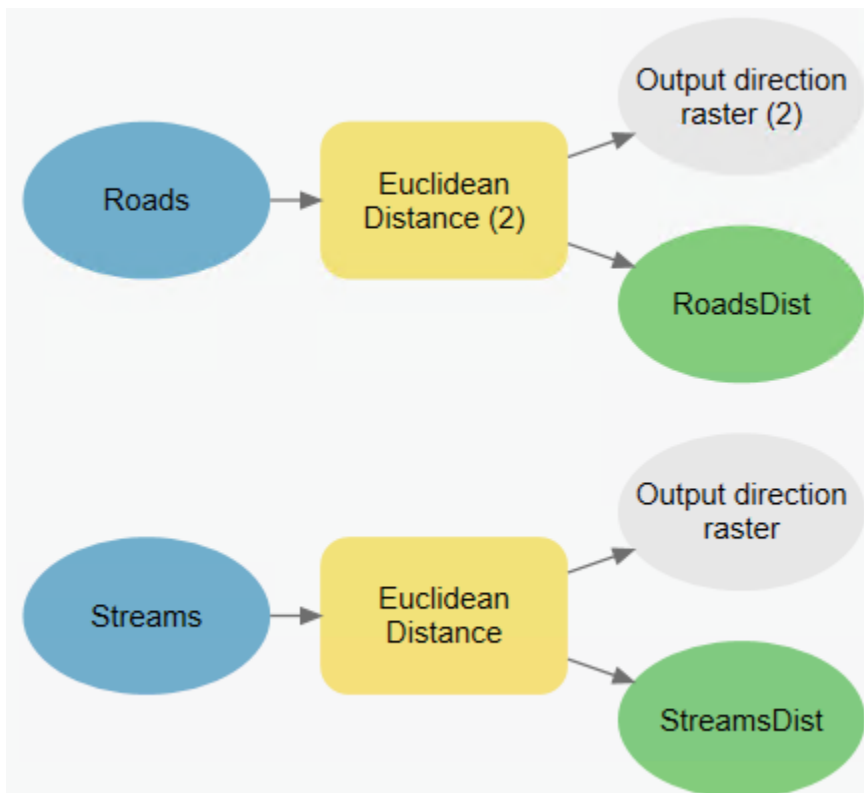
- f** Click OK.
- g** From the ModelBuilder tab, click Auto Layout.

Next, you will add another Euclidean Distance tool to create the distance surface for the Roads layer.

- h From the Geoprocessing pane, add another Euclidean Distance tool to the model to the right of Roads.
- i Connect Roads to Euclidean Distance (2) as Input Raster Or Feature Source Data.
- j Open the Euclidean Distance (2) tool, and change the Output Distance Raster name to **RoadsDist**.

Input raster or feature source data
Roads:1
Output distance raster
RoadsDist
Maximum distance
Output cell size
Elevation
Output direction raster
Distance Method
Planar

- k Click OK.
- l From the ModelBuilder tab, click Auto Layout.



- m Save the model.

## Step 4: Add the Slope tool and set parameters

You will use the Spatial Analyst Slope tool to derive a slope surface from the Elevation layer.

- a From the Geoprocessing pane, search for **slope**.
- b Drag Slope (Spatial Analyst Tools) into the model to the right of Elevation.
- c Connect Elevation to Slope as Input Raster.
- d Open the Slope tool, and for Output Raster, type **Slope**.
- e Accept the remaining defaults.

Input raster  
Elevation:1

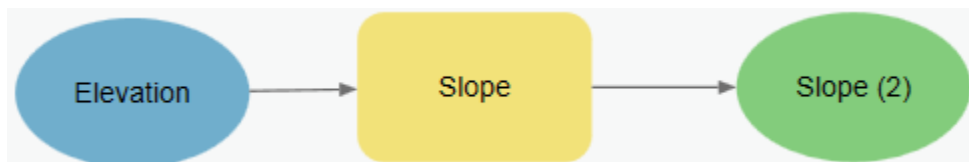
Output raster  
Slope

Output measurement  
Degree

Method  
Planar

Z factor  
0.304800609601219

- f Click OK.
- g Click Auto Layout.



- h Save the model.

## Step 5: Reclassify land-use values

Next, you will reclassify the values for land use. The most suitable land-use types for bear habitats are forested areas. You will look at the table for land-use codes, and then reclassify the raster using a saved remap file.

- a In the Contents pane, open the LandUse layer attribute table.

Rowid	VALUE	COUNT	CLASS_NAMES
1	1	858	Developed, High Inte...
2	2	4956	Developed, Medium...
3	3	23416	Developed, Low Inte...
4	4	152919	Cultivated Crops/Pas...
5	5	7848	Grassland/Herbaceous

Integer-based raster datasets can have attribute tables. The CLASS\_NAMES field contains descriptions for each of the cell values. You can analyze the table to determine which land-use codes are most suitable for bear habitats.

1. Which values should receive the most suitable class of 5 when you reclassify?

---

2. Which values do you consider to be highly suitable for bear habitats, but not as good as values 6, 7, and 8?

---

3. Which values are least suitable for bear habitats?

---

b Close the table.

You will now reclassify the LandUse raster.

c Add a **Reclassify** (Spatial Analyst Tools) tool to the model, placing it to the right of the LandUse data element.

d Connect the LandUse element to the Reclassify tool as Input Raster.

e Open the Reclassify tool, and ensure that Reclass Field is set to VALUE.

f Click Reverse New Values.

g In the Value and New columns, type the following values and classes, pressing Enter after each row to add another row.



After you press Enter, you must double-click in the Value cell to type, and then press Tab to move to the New cell.

Value	New
1	1
2	1
3	2
4	3
5	4
6	9
7	7
8	8
9	10
10	9
11	2
12	3

- h** For Output Raster, type **LandUseRcl**, and then click OK.
- i** Apply auto layout and save the model.

## Step 6: Rescale the roads distance surface

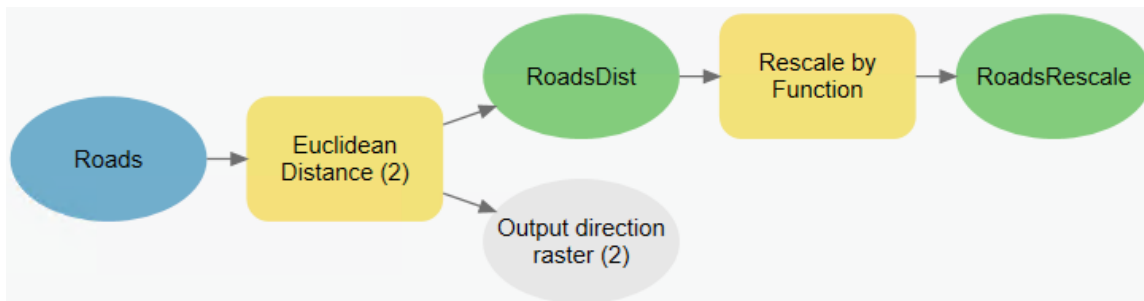
Sometimes, manually setting class breaks is not the best way to put cells into a suitability scale. For example, you may not know at which distance from roads that bear habitats will exist. In cases

where you are not sure of the class breaks, you can let ArcGIS Pro determine them using a transformation function. Next, you will use Rescale By Function to rescale the RoadsDist layer.

- a In the Geoprocessing pane, search for **rescale**.
- b Add Rescale By Function (Spatial Analyst Tools) to the model next to RoadsDist.
- c Connect RoadsDist to the Rescale By Function tool as Input Raster.
- d Open the Rescale By Function tool, and set the following parameters:
  - Output Raster: **RoadsRescale**
  - Transformation Function: Linear
  - Accept the default scale of 1 to 10

The Linear transformation function is best used when the preferences for values increase or decrease at a constant linear rate. For example, the most suitable bear habitats are farther away from roads.

- e Click OK.
- f Apply auto layout.



- g Save the model.

## Step 7: Rescale the stream distance surface

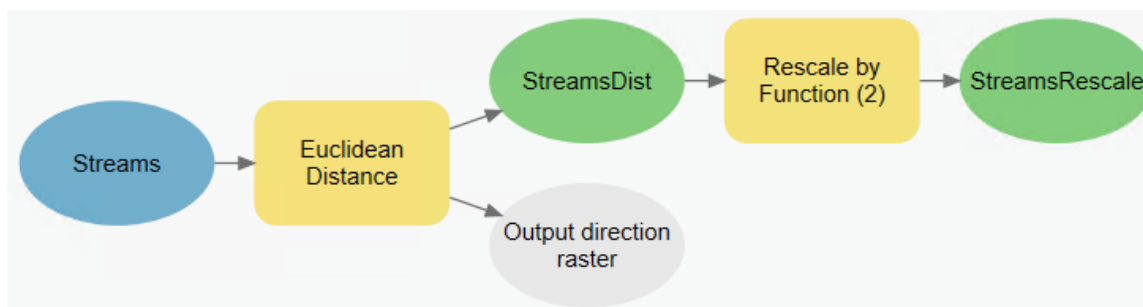
You will use a transformation function to rescale the StreamsDist layer, as you are unsure at which distances from streams that bears do or do not inhabit.

- a In the Geoprocessing pane, add the Rescale By Function tool to the model, placing it to the right of StreamDist.
- b Connect StreamDist to Rescale By Function (2) as Input Raster.

- c Open the Rescale By Function (2) tool, and set the following parameters:
  - Output Raster: **StreamsRescale**
  - Transformation Function: Small
  - Accept the default 1 to 10 scale

Typically, bears like to be closer to streams for food and water. The Small transformation function is used when the smaller input values are more preferred.

- d Click OK.
- e Apply auto layout and save the model.



## Step 8: Rescale the slope surface

In this step, you will rescale the slope surface.

- a Add a Rescale By Function tool to the model, placing it to the right of the Slope (2) data element.
- b Connect Slope (2) to Rescale By Function as Input Raster.
- c Open the Rescale By Function (3) tool, and set the following parameters:
  - Output Raster **SlopeRescale**
  - Transformation Function: Logistic Decay
  - Accept the default 1 to 10 scale

The Logistic Decay function is best used when the lower input values are more preferred, and as the input values increase, the preferences rapidly decrease.

- d Click OK.
- e Apply auto layout.

- f With the Explore tool active, hold the Shift key and select SlopeRescale, RoadsRescale, and StreamsRescale, and then right-click and choose Add To Display.
- g Save the model, and then save the project.

## Step 9: Run the model

You will now run the model to see the results of reclassify and rescale by function.

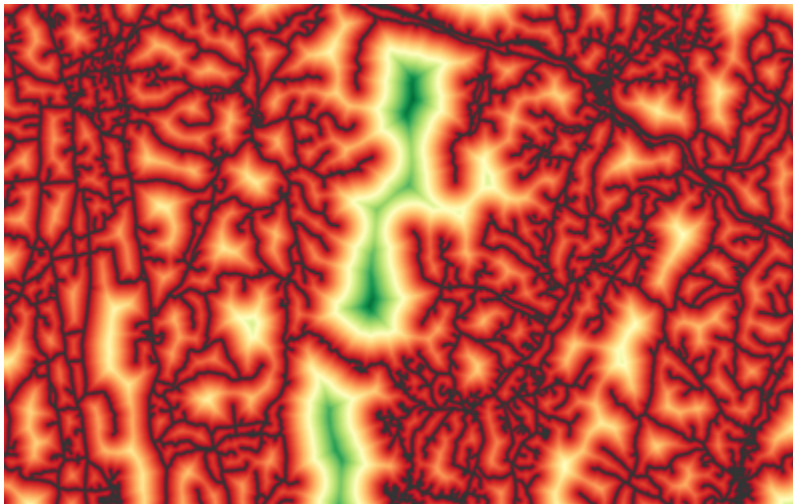
- a Right-click each of the final outputs (LandUseRcl, SlopeRescale, RoadsRescale, and StreamsRescale) and choose Add To Display.
- b From the ModelBuilder tab, click Validate, and then click Run.
- c Activate the Bear Suitability map view.

The legends for the layers that you rescaled are different from the legend for LandUseRcl in that they are stretched rather than having distinct class breaks.

- d Turn off and on each of the resulting layers to compare them.

In the rescaled layers, green indicates more suitable and red indicates less suitable areas. Consider distance from roads: The linear function that you used makes smaller distances less suitable and greater distances more suitable.

- e In the Contents pane, make Roads and RoadsRescale the only visible layers.



By viewing the layers together, you can see how the layer was rescaled. Where there are no roads, it is most suitable for bear habitats.

- f Turn off Roads and RoadsRescale.



- g View only Streams and StreamsRescale.

Streams cover most the study area, so if you consider only distance from streams, there are many suitable areas.

- h In the Contents pane, remove SlopeRescale, StreamsRescale, RoadsRescale, and LandUseRcl.
- i In the Bear Suitability model, right-click each of the final green output data elements and uncheck Add To Display.
- j Validate and save the model.
- k Save the project.

You have added the necessary tools to derive surfaces, reclassify the land-use raster, and rescale the continuous surfaces. You have also explored the results of the model up to this point. In the next exercise, you will overlay the reclassified and rescaled layers to create a suitability surface.

## Types of raster overlay

### Binary

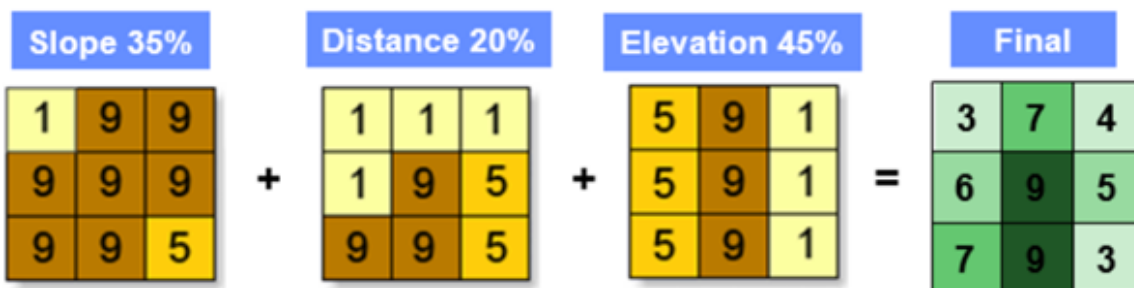
In binary overlay, each cell is assigned a value of 0 or 1, based on whether it meets all the analysis criteria. Zero (0) indicates that the cell does not meet all the criteria, and one (1) indicates that a cell meets all the criteria.



**Figure 7.10.** For each input surface, a 1 indicates that a cell is suitable. The final result reveals that only two cells are suitable based on the criteria; all other cells (for example, cells with a value of 0) are considered unsuitable.

### Weighted

In weighted overlay, values are manually reclassified onto a common scale (for example, 1 to 5 or 1 to 10). Layers are also weighted based on their influence to the particular analysis scenario, which is a key component to the suitability modeling workflow. Weighting allows stakeholders to assign relative importance to certain layers in the analysis. Weights can be a highly subjective component to your analysis, unless they are determined in a proper manner (for example, using the Delphi method). Altering weights can change the results of your analysis. You have the option to choose weights as relative percentages that sum to 1.



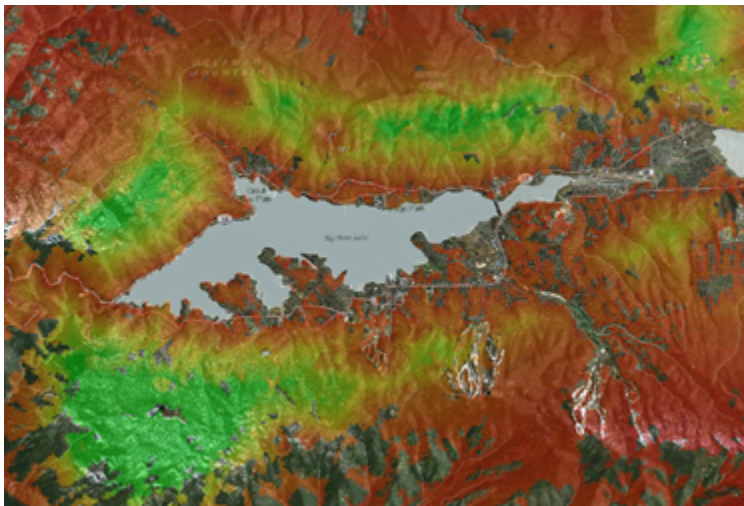
**Figure 7.11.** In this example, 1 indicates a cell that is not suitable and 9 indicates the most suitable cells. The remaining values indicate varying degrees of suitability.

## Types of raster overlay (continued)

---

### Fuzzy overlay

Fuzzy overlay is based on fuzzy logic. The basic premise behind fuzzy logic is that there are inaccuracies in attributes and in the geometry of spatial data. Cells are assigned values that represent their membership to a set of suitable locations. These membership values range from zero to 1, with zero indicating non-membership to a set (unsuitable) and 1 indicating membership to a set (suitable). Fuzzy overlay is best suited for analyzing data that does not adhere to discrete polygons and boundaries, such as landslides or disease outbreaks.



**Figure 7.12.** This image is the result of running fuzzy overlay to find bald eagle habitats near Big Bear Lake, California. Red cells are least suitable and green are more suitable.



**Esri Training course: [Using Raster Data for Site Selection](#)**

## The Raster Calculator

---

Binary analysis determines good or bad sites by assigning a value of 0 or 1, with 0 being unsuitable and 1 being suitable. The Raster Calculator is a commonly used Spatial Analyst tool for performing map algebra. Map algebra is a powerful language for raster analysis that allows you to perform various mathematical, logical, relational, and other types of operations on raster data. In this example, you are testing criteria within each raster and then creating a raster that identifies the cells in which all the criteria have been met.

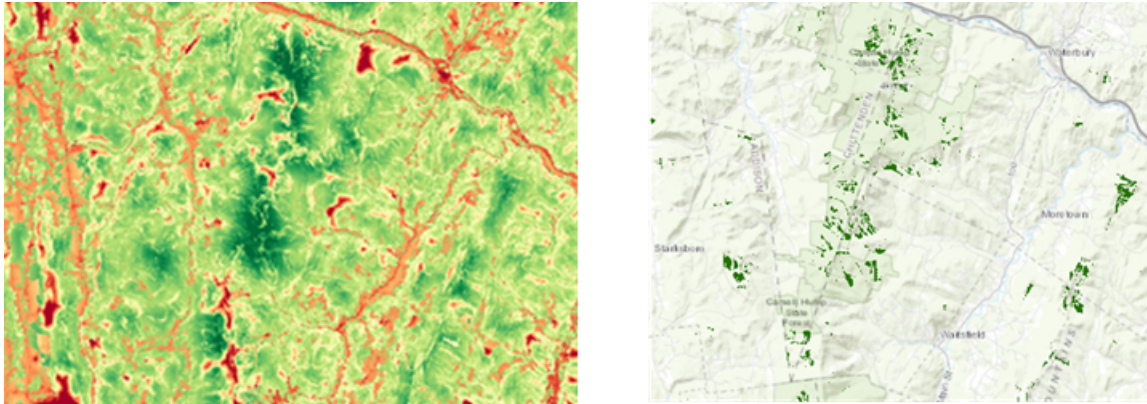


Figure 7.13. Raster Calculator expression for binary overlay.

## Locating and analyzing results

---

After you have transformed values to a common suitability scale and combined the layers, your output is a suitability surface. Within the suitability surface are values within your chosen suitability scale (1 to 5, 1 to 10, and so on). You may want to show only the best sites that range from 8 to 10.



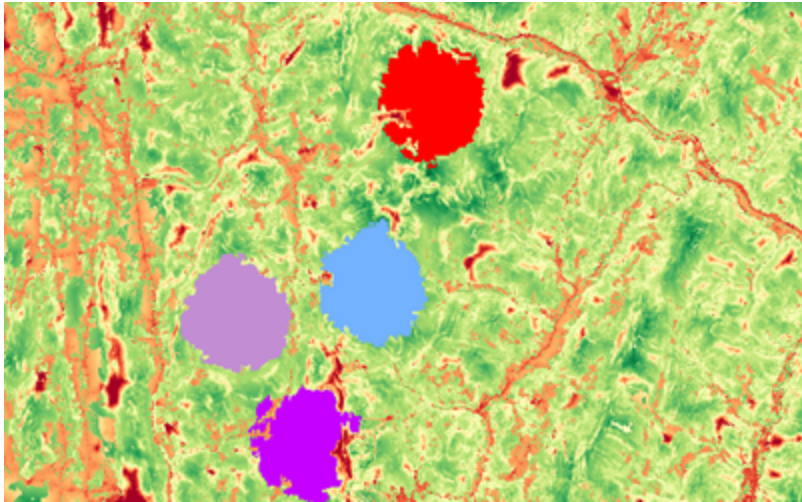
**Figure 7.14.** On the left, a suitability surface with values ranging from 1 to 10. On the right, the result of using map algebra in the Raster Calculator to only show cells with a value of 8 or higher.

## Locating regions

When you create a suitability surface, it may be difficult to know which areas are the most suitable. Often, the most suitable areas are not contiguous or are isolated from one another. By incorporating a tool called Locate Regions, you can avoid arbitrary assignment of the most suitable locations. The Locate Regions tool is part of the final step in any suitability modeling workflow. Locate Regions identifies the best regions, or groups of contiguous cells, in the suitability surface that meet your desired suitability criteria and other spatial constraints.

## Locating and analyzing results (continued)

---



**Figure 7.15.** Suitability regions created by the Locate Regions tool.

Locate Regions is often used along with the Cost Connectivity tool to select and then connect the best available regions in the least-cost way.

## Exploring data sources

---

The analysis that you perform in the next exercise will determine the most suitable places for bear habitats in northern Vermont, near Lake Champlain. Before you perform the analysis, you will use ArcGIS Pro to evaluate the criteria and data.

For this analysis, the criteria for suitable bear habitats are as follows:

- Must be close to streams
- Must be far away from roads
- Must be in forested areas
- Must be on terrain with relatively flat slopes

### Instructions

- a If necessary, start ArcGIS Pro and restore the course project.
- b In the Catalog pane, browse to the SNAP folder.
- c Expand RasterAnalysis and BearSuitability.gdb.
- d Answer the following questions in your workbook.

1. **Does the BearSuitability geodatabase contain the necessary datasets for each criterion?**

---

2. **Which tool will you use on Streams and Roads to create distance surfaces that you can use in raster overlay?**

---

3. **How will you fulfill the data requirement for slope?**

---

4. **Do you have to derive a surface from the LandUse raster or can you use it as an input as is?**

---

## Perform suitability modeling

You have built a model and classified and transformed data to a common scale. Now, you are ready to overlay the rasters to locate suitable regions for bear habitats in Vermont.

In this exercise, you will perform the following tasks:

- Overlay rasters.
- Locate suitability regions.



## Step 1: Overlay input rasters

During the overlay process, you assign weights to input rasters to indicate their relative importance to the analysis. For example, being close to a stream may be more important than being far from roads, so you would weight the streams higher.

- a If necessary, restore the ArcGIS Pro project and view the Bear Suitability model.
- b In the Geoprocessing pane, search for **weighted**.

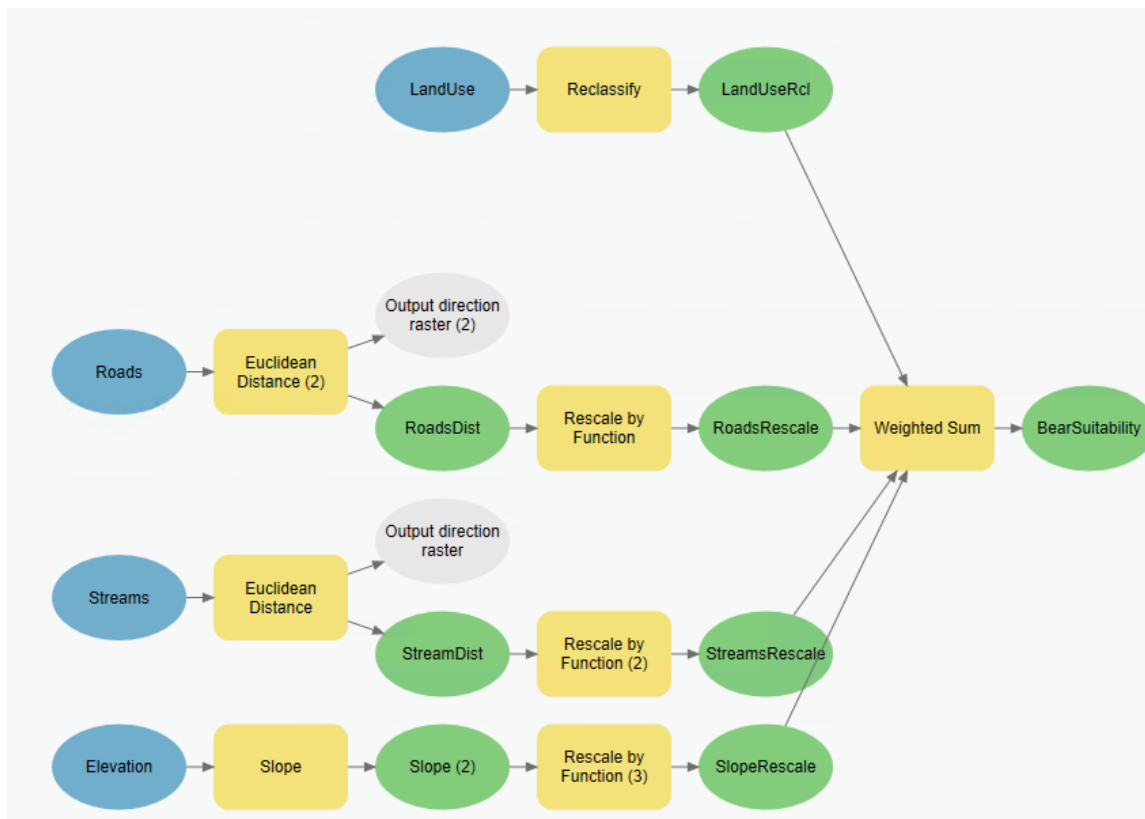
There are several raster overlay tools, such as Weighted Overlay and Weighted Sum. You will use Weighted Sum because it allows you to input floating point rasters (for example, the outputs of the Rescale By Function tool) and Weighted Overlay does not.

- c Add Weighted Sum (Spatial Analyst Tools) to the right of the model.
- d Connect the LandUseRcl, RoadsRescale, SlopeRescale, and StreamsRescale green data elements to Weighted Sum as Input Rasters.
- e Apply auto layout to arrange the elements.
- f Open the Weighted Sum tool.
- g For each input layer, assign the following weights:

Input raster	Weight
LandUseRcl	<b>0.25</b>
SlopeRescale	<b>0.20</b>
RoadsRescale	<b>0.20</b>
StreamsRescale	<b>0.35</b>

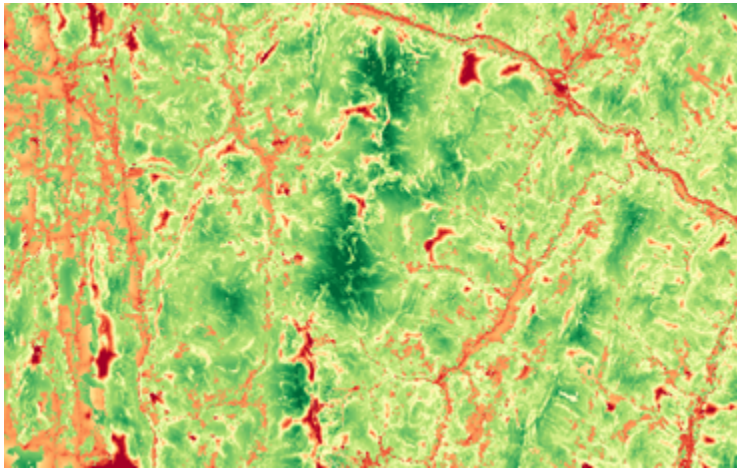
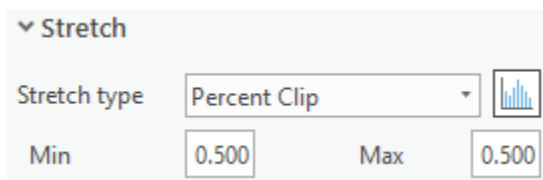
Raster	LandUseRcl	
Field	VALUE	
Weight	0.25	
Raster	SlopeRescale	
Field	VALUE	
Weight	0.2	
Raster	RoadsRescale	
Field	VALUE	
Weight	0.2	
Raster	StreamsRescale	
Field	VALUE	
Weight	0.35	

- h** For Output Raster, type **BearSuitability**.
- i** Click OK.
- j** Apply auto layout and save the model.



- k** Right-click the BearSuitability output element and choose Add To Display.

- l Validate and run the model.
- m Activate the Bear Suitability map view.
- n In the Contents pane, turn off all layers except BearSuitability.
- o Right-click BearSuitability and choose Symbology.
- p In the Symbology pane, change the Stretch Type to Percent Clip.
- q For Min and Max, replace the current values with **0.5** for each.



In many cases, you can assume that most pixel values fall within an upper and lower limit. Therefore, it is reasonable to trim off the extreme values. You can do this statistically by defining either a standard deviation or a clipping percent.

The suitability surface illustrates the suitability ranges for the entire study area. From the surface, you can get a better idea where it is suitable for bear habitats.

- r In the Contents pane, turn input layers on and off to evaluate the results.

You can see that the cells are red, or less suitable, where there are roads. You can also see that it is more suitable for bear habitats where the elevation is lower, due to less steep slopes.

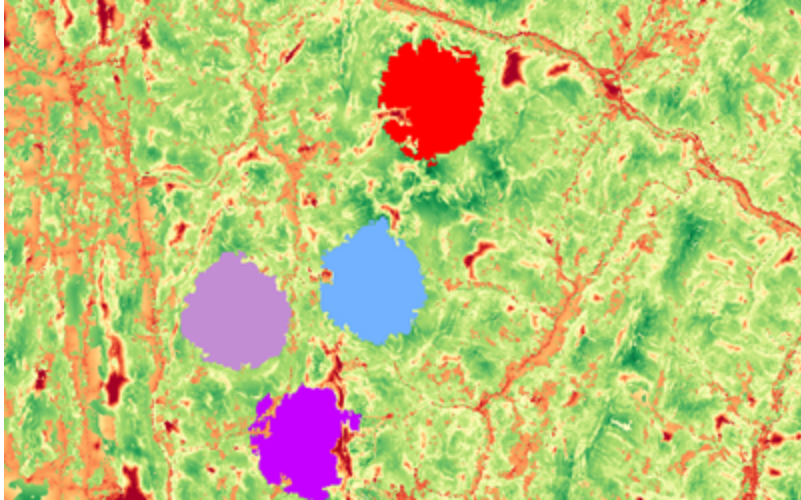
- s Close the Symbology pane, and save the project.

## Step 2: Create regions

Next, you will create regions for your suitable bear habitats. For example, the most preferred bear habitat to conserve may include four habitat patches (regions) to maintain a viable population, with each region being approximately 50 contiguous acres. To support breeding opportunities within the group, the regions should be close enough to one another so that they can be feasibly connected through wildlife corridors.

- a View the Bear Suitability model.
- b In the Geoprocessing pane, search for **locate regions**.
- c Add the Locate Regions tool to the model next to the BearSuitability output element.
- d Connect BearSuitability to Locate Regions as Input Raster.
- e Open the Locate Regions tool, and set the following parameters:
  - Total Area: **100**
  - Output Raster: **BearPatches**
  - Number Of Regions: **4**
  - Region Shape: Circle
  - Region Minimum Area: **4**
  - Region Maximum Area: **26**
  - Minimum Distance Between Regions: **1.5**
  - Maximum Distance Between Regions: **6**
- f Click OK.
- g Right-click the BearPatches data element and choose Add To Display.
- h Apply auto layout and save the model.
- i In the model, right-click Locate Regions and choose Run.
- j Activate the Bear Suitability map view.
- k In the Contents pane, for BearPatches, set the fill for the zero value to No Color.

*Hint:* Right-click the zero value color swatch > No Color



Now, you have four bear suitability regions.

- 1 In the Contents pane, toggle off and on BearPatches so that you can see how they compare to the original suitability surface.

You have used raster data to perform suitability modeling to locate the bear habitat regions based on a set of criteria.

- m Save the project, and leave ArcGIS Pro open.

## Lesson review

---

1. What is the difference between binary and weighted overlay?

---

---

---

---

2. Explain the difference between the Reclassify tool and the Rescale By Function tool.

---

---

---

---

---

---

# Answers to Lesson 7 questions

## Evaluating analysis criteria (page 7-4)

What types of data can you identify in the criteria, and can you use vector overlay tools like Intersect and Union with them?

**Temperature, slope, and elevation are continuous data types stored in raster format.**

**Roads are stored in vector format, but you would use Euclidean Distance to create a raster for suitability modeling with the other data.**

**You cannot use regular overlay tools on continuous (raster) data. You must use Spatial Analyst tools to analyze raster data.**

## Choosing vector or raster overlay (page 7-5)

### Scenario 1: Wind power

1. Would you use raster or vector overlay to determine suitable ocean locations to harvest wind power? Why?

**You would use raster because the data types are continuous.**

### Scenario 2: New store location

2. Would you use raster or vector overlay to determine the most suitable locations for a shopping center? Why?

**You would use vector because all the data types are discrete.**

## Exercise 7A: Build a model and classify data to a common scale (page 7-12)

1. Which values should receive the most suitable class of 5 when you reclassify?

**Values 6, 7, and 8, which represent deciduous, evergreen, and mixed forest**

2. Which values do you consider to be highly suitable for bear habitats, but not as good as values 6, 7, and 8?

**Values 9 and 10, which are scrub/shrub and forested wetlands**

# Answers to Lesson 7 questions (continued)

3. Which values are least suitable for bear habitats?

**Values 1, 2, and 3, which represent developed areas**

## Exploring data sources (page 7-29)

1. Does the BearSuitability geodatabase contain the necessary datasets for each criterion?

**The geodatabase contains relevant data to start from, but you will have to derive data from these sources for the analysis.**

2. Which tool will you use on Streams and Roads to create distance surfaces that you can use in raster overlay?

**Use the Euclidean Distance tool.**

3. How will you fulfill the data requirement for slope?

**Use the elevation to create a slope surface.**

4. Do you have to derive a surface from the LandUse raster or can you use it as an input as is?

**You can use it as an input as is.**



When you look at a map, your mind will naturally try to identify patterns, trends, and spatial relationships.

Spatial statistics extend these natural processes by quantifying spatial distributions and spatial relationships. Spatial statistics allow you to supplement the subjective perspective of your data with concrete numbers and statistics. Statistics help with enhancing communication, fostering consensus, facilitating problem-solving through analysis, promoting decision making, and providing mechanisms for evaluating the impacts of those decisions. In this lesson, you will focus on the most intuitive and commonly used spatial statistics solutions.

## Topics covered

---

Spatial patterns

What are spatial statistics?

Types of spatial statistics

Data distributions

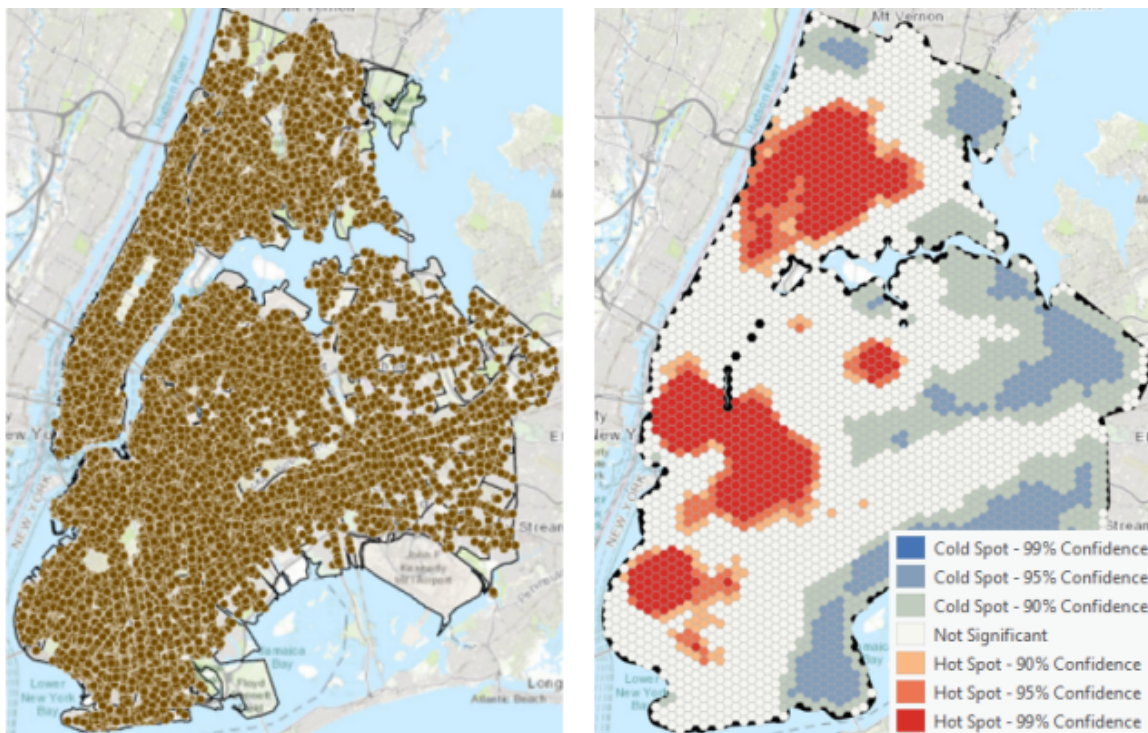
Clusters and outliers

Spatial statistics tools

## Spatial patterns

A spatial pattern may lead to questions about the possible processes that create the pattern. Most spatial phenomena exhibit some type of pattern that is probably influenced by some other factor. For example, an animal species may migrate along the same path every year because there is plenty of food and water and few predators.

Have you worked with datasets containing thousands of points, like you see in the map on the left? Did you feel unsure about where to begin to better understand patterns?



**Figure 8.1.** It is difficult to distinguish spatial patterns from a map of points, as shown on the left. The result of running a spatial statistics tool, shown on the right, shows you clusters of high and low counts of graffiti incidents.

The red clusters in the map on the right are called hot spots. Hot spots are statistically significant spatial clusters of high values. You will use hot spots to distinguish spatial patterns.

**1. What have you done to busy, subjective maps to show them in a meaningful way?**

---



---



---

## What are spatial statistics?

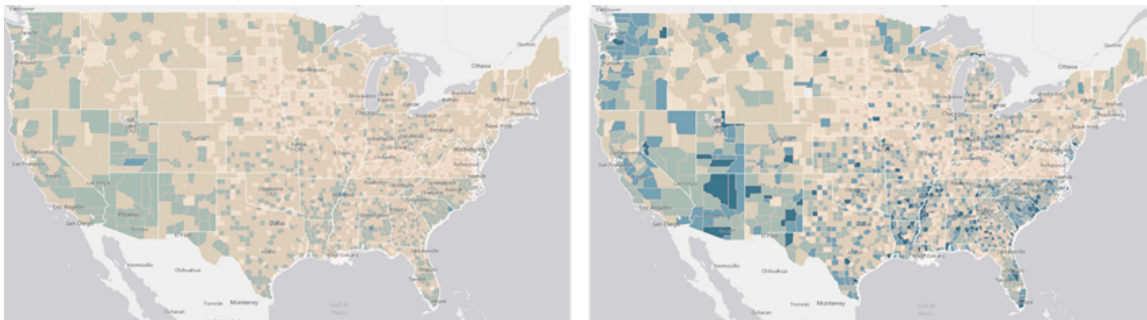
---

Spatial statistics are the application of tools and methods that use space and spatial relationships (such as distance, area, length, orientation, centrality, coincidence, connectivity, orientation) directly in their mathematical computations.

Spatial statistics allow you to do the following tasks:

- Minimize the subjectivity inherent in human visual interpretation of maps and spatial data.
- Identify and quantify patterns and trends in data that may not be revealed in visual analysis.
- Answer questions more confidently and make important decisions using more than simple visual analysis.

The following maps are based on graduated symbols.

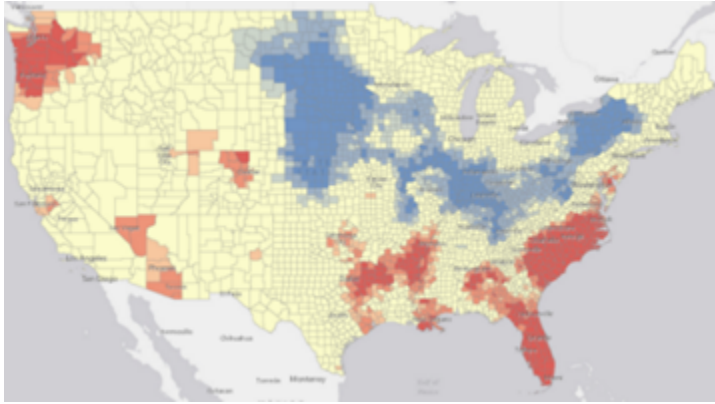


**Figure 8.2.** Maps displayed using a graduated color classification can be subjective and highlight things that the mapmaker wants you to notice. Both maps have the same crime index data but are visualized using different classification methods.

## What are spatial statistics? (continued)

---

The following map is based on statistically significant hot and cold spots.



**Figure 8.3.** The same data as the previous maps is now visualized using a hot spot map. The red and blue colors indicate the results of a statistical test for spatial clustering, rather than the data values of the attribute mapped in the graduated color map.

## Types of spatial statistics

---

### Descriptive statistics

Descriptive statistics return a summary about your data, whether the result is quantitative like a summary statistic (mean, sum, and so on) or visual, such as a graph or feature class. In GIS, descriptive statistics commonly measure central tendency, dispersion or concentration, or orientation of spatial phenomenon.



**Figure 8.4.** On the left, the Mean Center tool locates the geographic center for a sample of points. On the right, the Directional Distribution tool uses standard deviational ellipses to show directional trends of incidents by day and night.

## Types of spatial statistics (continued)

### Inferential statistics

In classical statistics, inferential statistics infer something about the entire population based on the distribution of values in a sample. A common example of inferential statistics is predicting the outcome of an election based on polls. Inferential statistic tests begin by identifying a null hypothesis. The null hypothesis for pattern analysis tools is complete spatial randomness in either the location of features or the values associated with those features. Some spatial statistics tools return statistics that indicate the degree of statistical significance, which in turn provides you with a degree of confidence in rejecting or not rejecting the null hypothesis of complete spatial randomness.

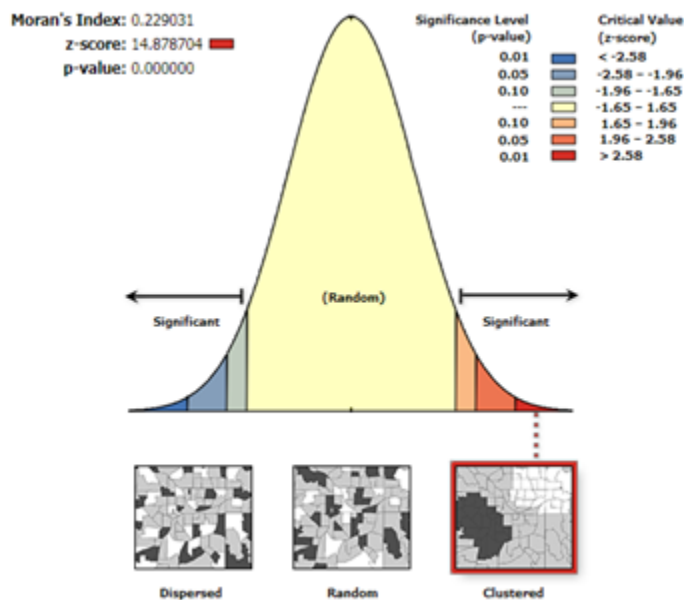


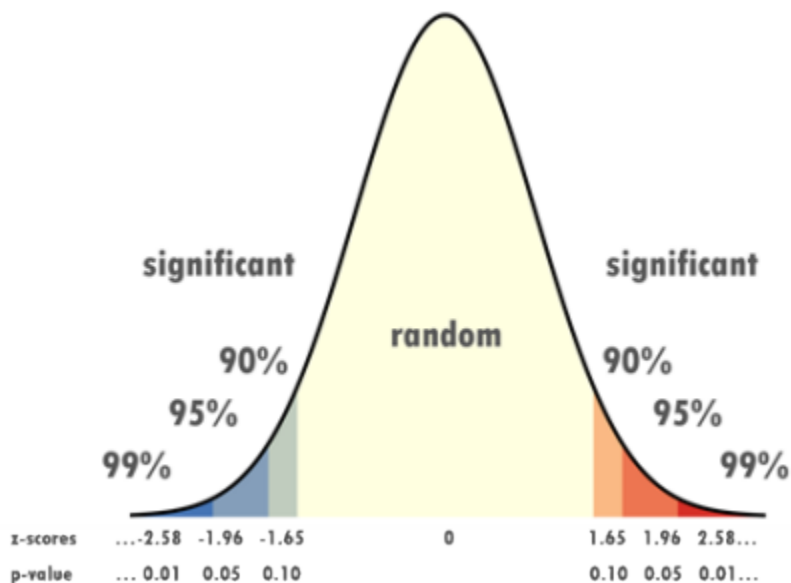
Figure 8.5. The statistics returned from spatial autocorrelation.



**Esri Press:** [The Esri Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics](#)

## Interpreting inferential statistics

Given the null hypothesis of complete spatial randomness, many of these spatial statistics tools compare the observed spatial distribution in data to that of a theoretical random spatial distribution and calculate common statistical significance tests from this comparison. You can visually represent your randomization null hypothesis using the standard normal distribution (with a mean of zero and a standard deviation of 1) and use it to interpret two common outputs from these tools: z-scores and p-values.



**Figure 8.6.** A normal distribution chart looks like this example, with most the data falling around the mean and the highest and lowest values occurring in the tails. When this graphic is used to interpret the results of spatial statistics tools, output values in the tails indicate that it is unlikely that the observed spatial pattern is the result of random chance.

### Z-scores and p-values

In classical statistics, a z-score represents a deviation from the mean in a data distribution. A p-value (confidence level) dictates the critical z-score for rejecting or not rejecting the null hypothesis that the value is statistically significantly different from the mean of the distribution.

In spatial statistics, the z-score and p-value may be interpreted differently. When you run a spatial statistics tool, the resulting statistics are then compared with the expected value of that statistic under the null hypothesis of complete spatial randomness. This comparison results in the z-scores that you see in the outputs.

## Interpreting inferential statistics (continued)

---

For example, assume that you run a tool and receive a z-score of 2.16. For a p-value of 0.05, this z-score exceeds the critical value of 1.96, meaning that you can reject the null hypothesis of complete spatial randomness. In other words, there is less than 5 percent likelihood that the observed pattern is the result of random chance.

- Cold Spot - 99% Confidence
- Cold Spot - 95% Confidence
- Cold Spot - 90% Confidence
- Not Significant
- Hot Spot - 90% Confidence
- Hot Spot - 95% Confidence
- Hot Spot - 99% Confidence

**Figure 8.7.** You can compare the legend with the chart to see which features are statistically significant.



**ArcGIS Pro Help:** *What is a z-score? What is a p-value?*



## Descriptive versus inferential

---

Based on the result provided, determine whether the statistical tool used is descriptive or inferential.

### Scenario 1: Directional trend returned

The Directional Distribution tool creates an ellipse that represents the orientation for a sample set of points.

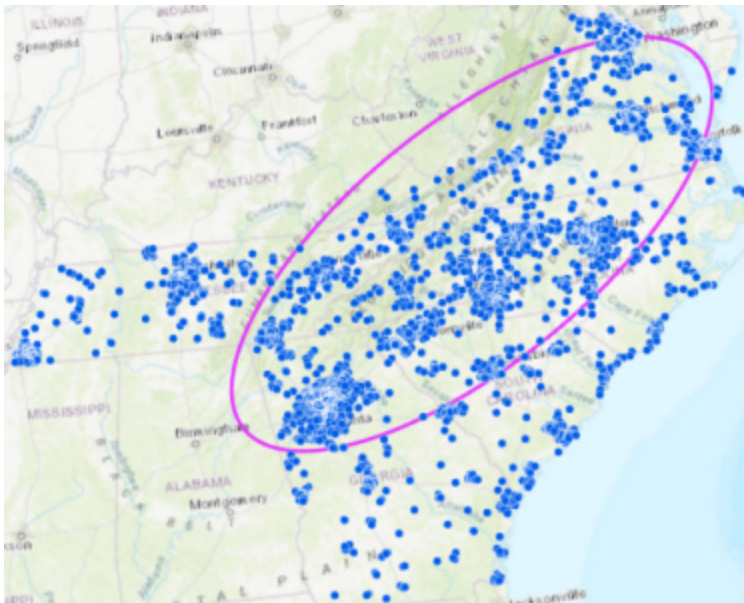


Figure 8.8. The ellipse shows directional trend of the points.

1. Does the Directional Distribution tool describe your data or make statistical inferences based on the data provided?
-

## Descriptive versus inferential (continued)

### Scenario 2: Chart and statistics returned

The Average Nearest Neighbor tool measures the distance between each feature and its nearest neighbor's location. It then averages all these nearest neighbor distances. If the average distance is less than the average for a hypothetical random distribution, the distribution of the features being analyzed is considered clustered. If the average distance is greater than a hypothetical random distribution, the features are considered dispersed. Average Nearest Neighbor uses only feature location and does not consider the values of those features.

The Average Nearest Neighbor tool was run to create a report and provide statistics about a distribution of point features.

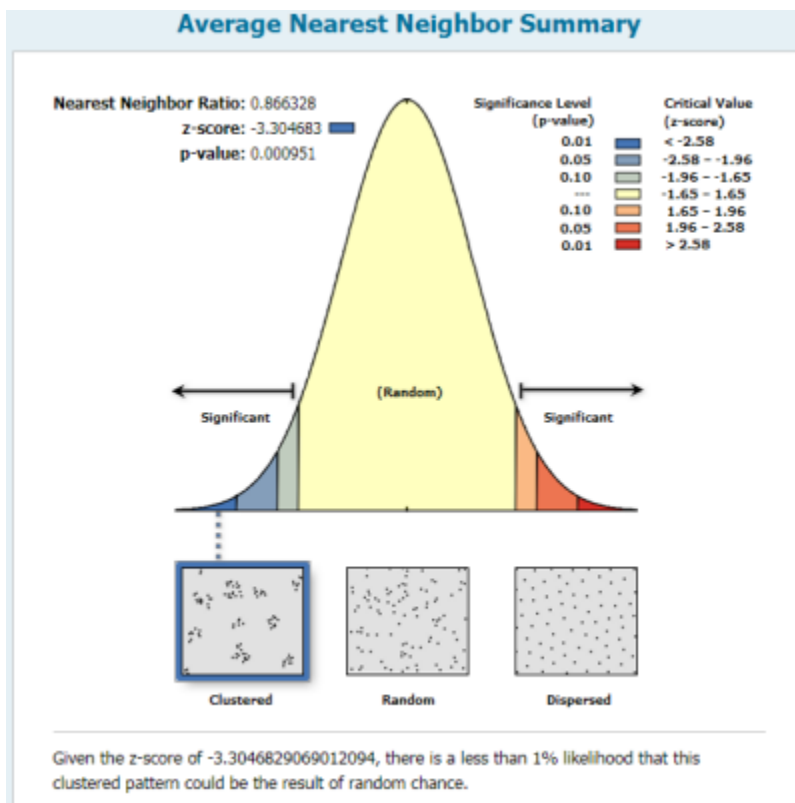


Figure 8.9. Result of the Average Nearest Neighbor tool.

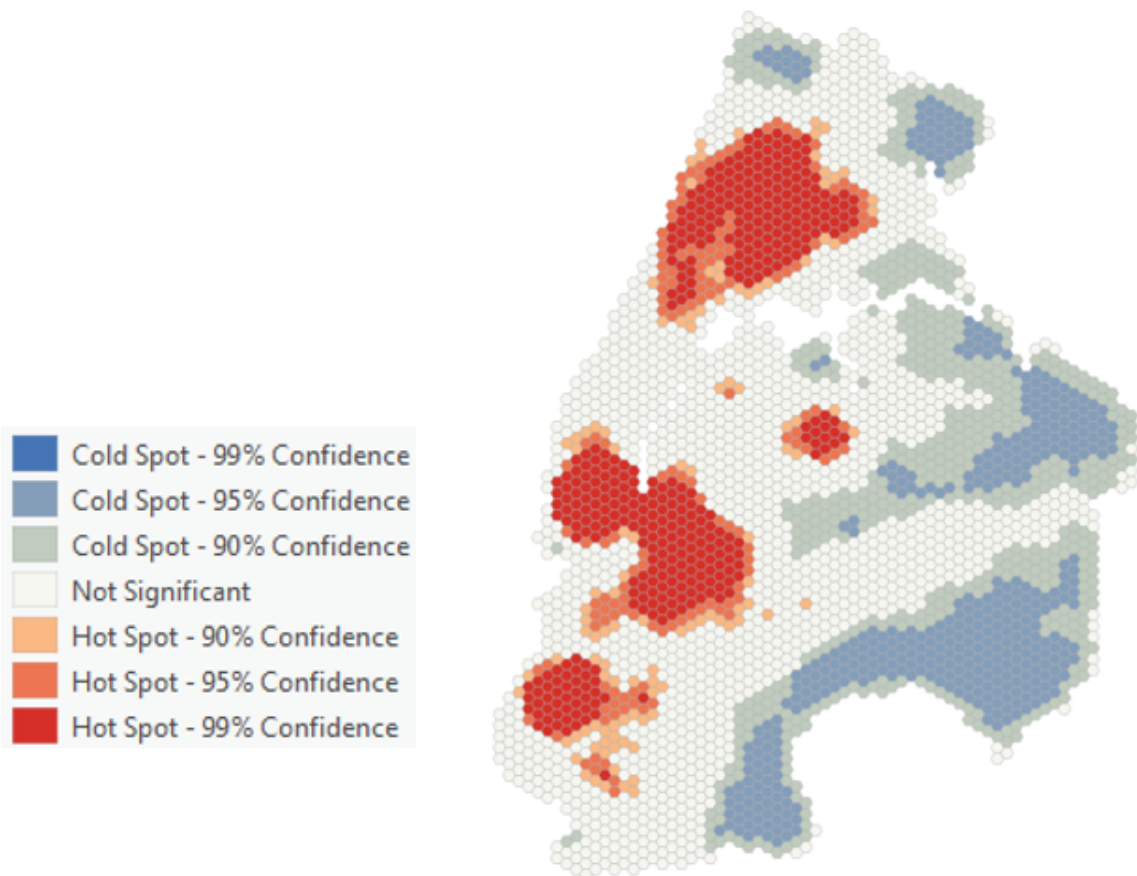
- Does the Average Nearest Neighbor tool describe your data or make statistical inferences about it based on the data provided?

## Descriptive versus inferential (continued)

---

### Scenario 3: Map of hot and cold spots returned

The Optimized Hot Spot tool creates a result that identifies statistically significant hot and cold spots and the statistics that go along with it. P-values indicate the probability that the observed spatial pattern is the result of a random process. You can tell by comparing the legend and map where there are statistically significant hot and cold spots and the confidence in which you can rely on these being true.

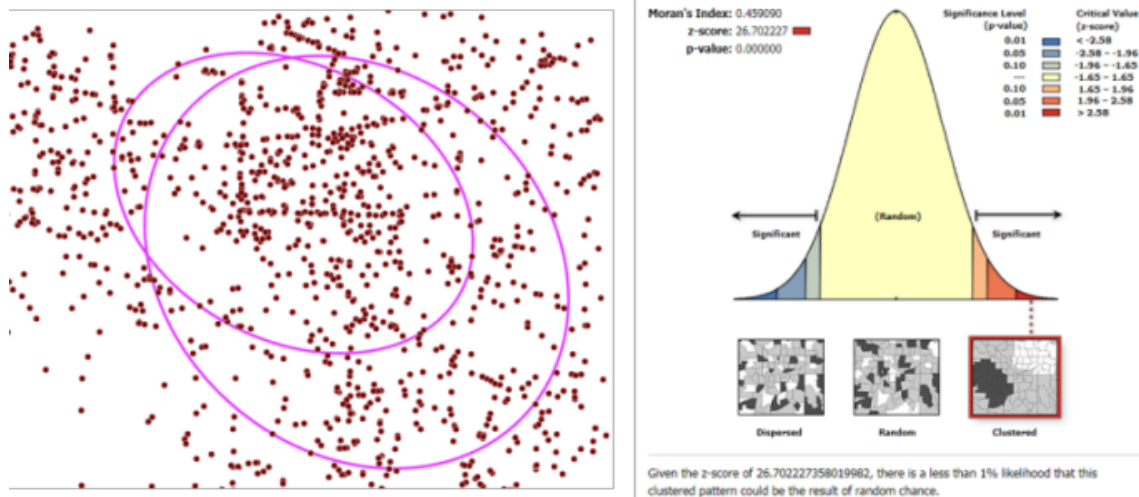


**Figure 8.10.** Results of the Optimized Hot Spot tool.

3. Does the Optimized Hot Spot tool describe your data or make statistical inferences about it based on the data provided?
-

## Spatial statistics tools

Spatial statistics may seem daunting, but running the tools can quickly provide valuable information that you can use to understand your data and make informed decisions. Two common tools used are the Directional Distribution tool and the Spatial Autocorrelation tool.



**Figure 8.11.** Results of running the Directional Distribution tool on the left and the Spatial Autocorrelation tool on the right.

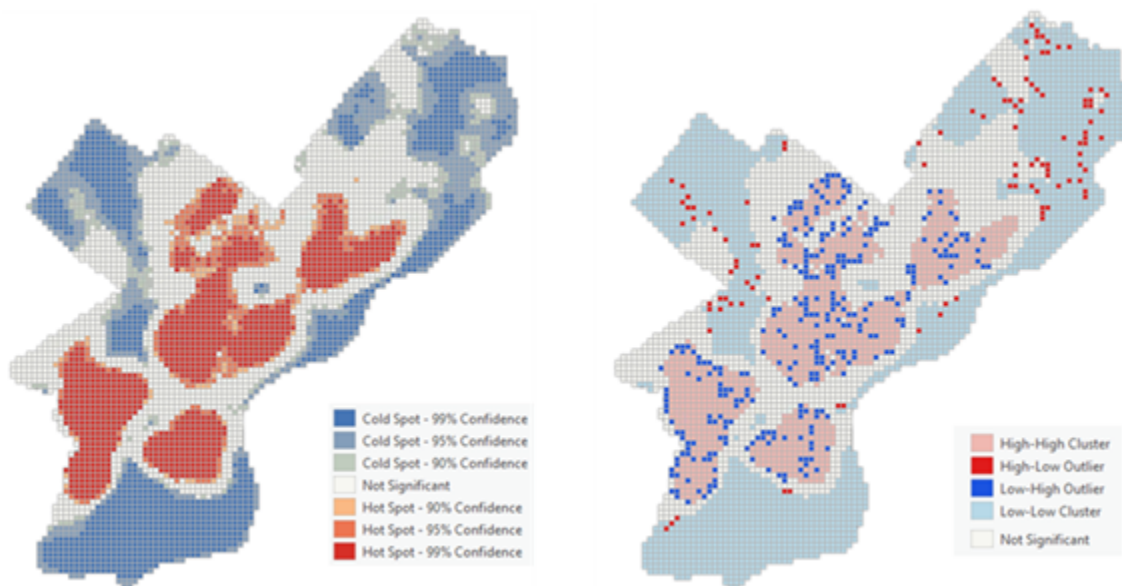
- Directional Distribution tool: Creates standard deviational ellipses to summarize the spatial characteristics of geographic features, such as central tendency, dispersion, and directional trends.
- Spatial Autocorrelation tool: Measures spatial autocorrelation based on both feature locations and feature values simultaneously. Given a set of features and an associated attribute, it evaluates whether the pattern expressed is clustered, dispersed, or random.

## Clusters and outliers

Clusters occur when phenomena are found in close proximity to one another. Clusters also occur when groups of features with similarly high or low values are in close proximity, or based on the degree of similarity among feature attributes. Clusters in your data can identify the locations of hot spots, cold spots, outliers, and similar features. Finding locations of clusters and spatial patterns in your data can lead to powerful discoveries and fuel further research questions and analysis.



**Figure 8.12.** This graphic illustrates the different ways in which your data can be spatially distributed. To the left, the data is dispersed, and there is no clustering observed. As you move to the right, the data gets more clustered.



**Figure 8.13.** On the left, results of the Optimized Hot Spot tool show statistically significant hot and cold spots (clusters). On the right, results of the Optimized Outlier Analysis tool identify areas where there are low outliers within hot spots and high outliers within cold spots.

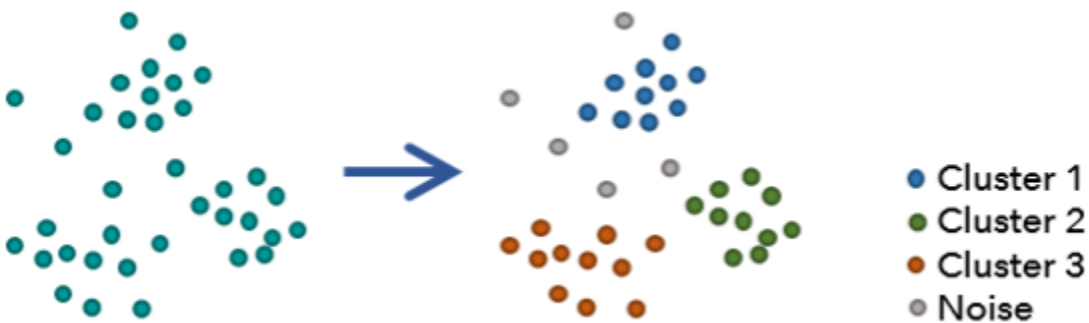
Cluster analysis also may find unusual or extreme data values, called outliers, where one or a few features may have values that are very different from nearby features. In data analysis, outliers can potentially have a strong effect on results, so they must be analyzed carefully to determine if they represent valid or erroneous data.

## Clusters and outliers (continued)

---

### Density-based clustering

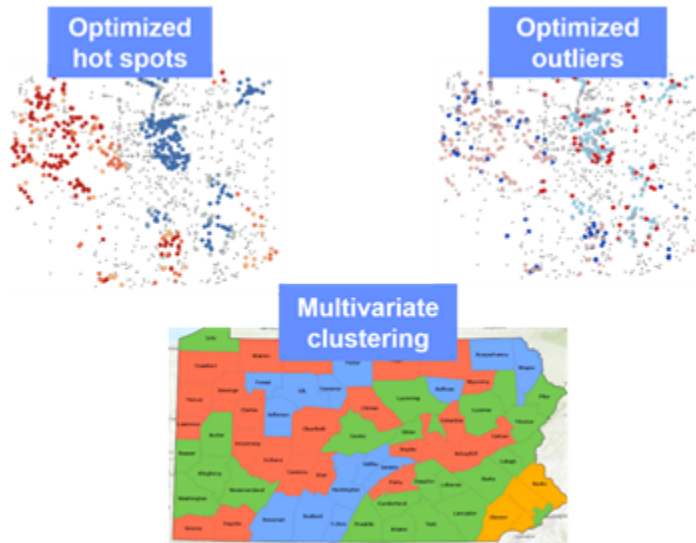
Another powerful clustering tool is Density Based Clustering. The Density Based Clustering tool finds clusters of point features within surrounding noise based on their spatial distribution. Do not confuse density-based clustering with density analysis tools, like Kernel Density. Density analysis tools take known quantities of a phenomenon and spread it across the landscape.



**Figure 8.14.** Density-based clustering can identify natural clusters in your data.

## Clustering tools

---



**Figure 8.15.** Hot spot, outlier, and multivariate clustering tools in action.

### Hot spot analysis

There are two main hot spot analysis tools in ArcGIS Pro: Hot Spot Analysis (Getis-Ord  $G_i^*$ ) and Optimized Hot Spot Analysis. Both tools execute the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool, but the Optimized Hot Spot Analysis tool interrogates your data to find optimal parameters. Each hot spot tool returns z-scores and p-values that indicate where features with either high or low values cluster spatially. To be a statistically significant hot spot, a feature will have a high value and be surrounded by other features with high values, as well.

The  $G_i^*$  statistic returned for each feature in the dataset is a z-score. For statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (hot

## Clustering tools (continued)

---

spot). For statistically significant negative z-scores, the smaller the z-score is, the more intense the clustering of low values (cold spot).

- You would use the Hot Spot Analysis (Getis-Ord-Gi\*) tool when you want full control over every parameter option.
- You would use the Optimized Hot Spot Analysis tool when you want the tool to interrogate your data to determine optimal parameter values.

### Optimized Outlier Analysis tool

The Optimized Outlier Analysis tool executes the Cluster and Outlier Analysis (Anselin Local Moran's I) tool using parameters derived from characteristics of your input data. The optimized version interrogates your data to obtain the settings that will yield optimal analysis results.

This tool identifies statistically significant spatial clusters of high values (hot spots) and low values (cold spots), as well as high and low outliers within your dataset.

### Multivariate Clustering tool

The Multivariate Clustering tool finds natural clusters of features based solely on feature attribute values. Given the number of clusters to create, it will look for a solution where all the features within each cluster are as similar as possible and all the clusters themselves are as different as possible. Another tool called Spatially Constrained Multivariate Clustering allows you to set cluster size and force spatial contiguity in the clusters to focus your analysis even further.



## Use spatial statistics to explore data

You will use various descriptive and inferential statistics tools to quantify patterns and relationships in ozone sample data.

In this exercise, you will perform the following tasks:

- Create standard deviational ellipses.
- Use the Average Nearest Neighbor tool.
- Use the Spatial Autocorrelation tool.
- Perform hot spot analysis.

## Step 1: Prepare ArcGIS Pro

In this step, you will prepare the project and set analysis environments.

- a If necessary, start ArcGIS Pro and open the SNAPCourse project.
- b Close all open views, saving if prompted.
- c From the Insert tab, click Import Map.
- d Browse to **C:\EsriTraining\SNAP\Statistics** and add Spatial Stats.mapx.



The map displays the same ozone points that you worked with in the interpolation lesson.

- e From the Analysis tab, click Environments.
- f Set the following environments:
  - Extent: Same As Layer - StateBoundary
  - Mask: StateBoundary
- g Click OK.
- h Save the project.

## Step 2: Locate directional trends in data

In this step, you will use the Directional Distribution tool to identify trends in the ozone data.

- a** If necessary, open the Geoprocessing pane.

*Hint:* Analysis tab > Tools

- b** In the Geoprocessing pane, type **direction**.

- c** Open the Directional Distribution (Standard Deviational Ellipse) tool, and set the following parameters:

- Input Feature Class: Samples
- Output Ellipse Feature Class: **Ellipses**
- Ellipse Size: 1 Standard Deviation
- Weight Field: OZONE

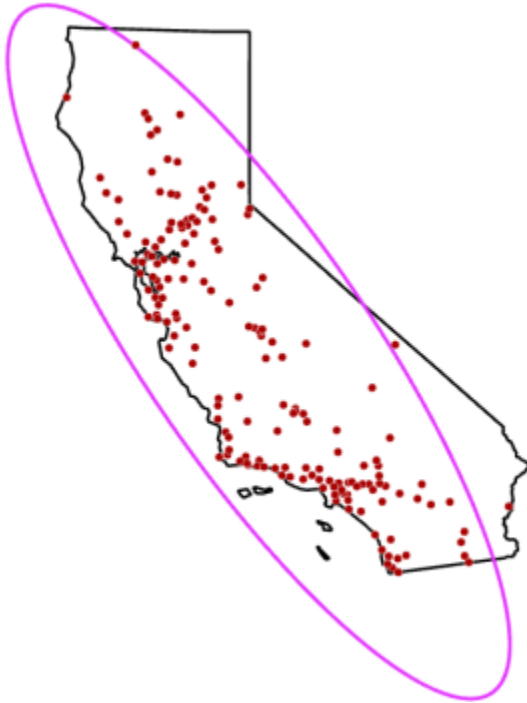
- d** Click Run.



When the underlying spatial pattern of features is concentrated toward the center with fewer features toward the periphery, one standard deviational ellipse polygon will cover approximately 63 percent of the features.

The ellipse indicates a directional trend that is based on the orientation of the sample points. Next, you will run the same tool using a different ellipse size to account for more sample points.

- e In the Contents pane, turn off the Ellipses layer.
- f In the Geoprocessing pane, modify the following parameters for the Directional Distribution tool:
  - Output Feature Class: **Ellipses2**
  - Ellipse Size: 2 Standard Deviations
- g Click Run.



Choosing two standard deviations and including more of the sample points results in a larger ellipse with a similar directional trend as the other ellipse for one standard deviation.

- h Turn off the Ellipses2 layer.


### Step 3: Run the Average Nearest Neighbor tool

Next, you will use the Average Nearest Neighbor tool to determine the degree of clustering or dispersion in feature locations.

- a In the Contents pane, open the StateBoundary layer attribute table.

The Average Nearest Neighbor tool uses area as a parameter, so you will copy the area from the StateBoundary layer.

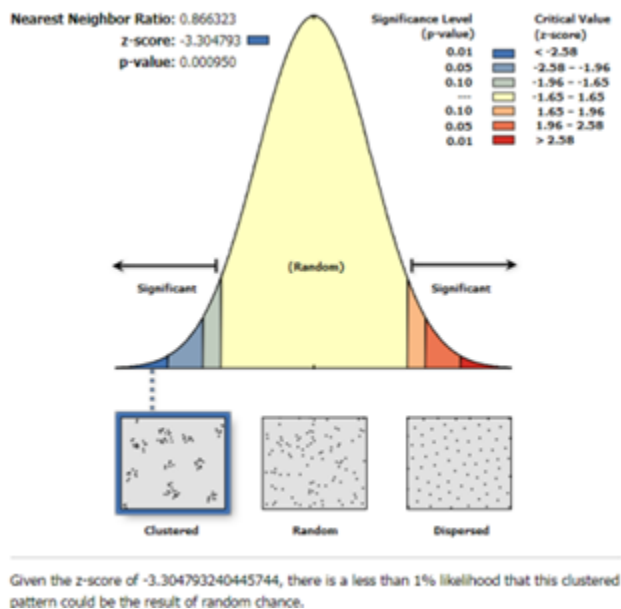
- b In the table, double-click the value for AREA, right-click it, and choose Copy.

- c Close the table.
- d In the Geoprocessing pane, click the Back button .
- e In the search field, type **average**.
- f Open the Average Nearest Neighbor tool, and set the following parameters:
  - Input Feature Class: Samples
  - Check the Generate Report box
  - Area: Right-click in the cell and choose Paste
- g Click Run.



The output from the Average Nearest Neighbor tool is a report, not a feature class.

- h Open File Explorer, browse to `..\EsriTraining\SNAP\SNAPCourse`, and double-click the Nearest Neighbor HTML file to view the report.



The null hypothesis states that the locations of the features are randomly distributed. However, the z-score (-3.30) returned from the Average Nearest Neighbor tool is statistically significant at a confidence level of 99 percent, meaning that there is less than a 1 percent likelihood that the spatial pattern of ozone locations is the result of random chance.

- i Close the report, minimize File Explorer, and return to ArcGIS Pro.

## Step 4: Run the Spatial Autocorrelation tool

The Average Nearest Neighbor tool only looks at feature locations to determine if they are clustered or dispersed. If you want to determine if there is clustering in the attribute values at those locations, you must run the Spatial Autocorrelation (Moran's I) tool.

- a In the Geoprocessing pane, go back to the search results and search for **spatial**.
- b Open the Spatial Autocorrelation (Global Moran's I) tool, and set the following parameters:
  - Input Feature Class: Samples
  - Input Field: OZONE
  - Check the Generate Report box
  - Conceptualization Of Spatial Relationships: Fixed Distance Band
  - Standardization: Row

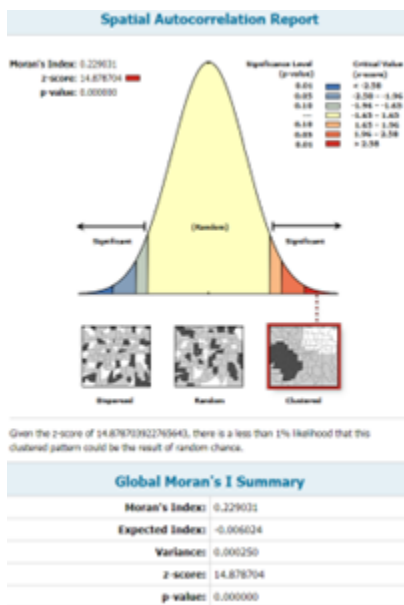
- c Click Run.



The warning is due to not providing a distance band or threshold distance. The software calculates a default value of 154,019 meters, or roughly 95 miles, to ensure that every feature has at least one neighbor.

If you were investigating crime clusters, a distance of 800 meters might work because that is the average size of a city block. When choosing the appropriate distance threshold, it is often based on common sense, theory, and the field that you are working in.

- d In File Explorer, open the MoransI\_Result HTML file.



The high z-score indicates that you can reject the null hypothesis and that the spatial distribution of high values and low values in the data is more spatially clustered than would be expected if the underlying spatial processes were random.

- e Close the report, minimize File Explorer, and return to ArcGIS Pro.

## Step 5: Run the Hot Spot Analysis tool

In this step, you will perform hot spot analysis to determine where the clusters are located.

- a In the Geoprocessing pane, search for **hot**.
- b Open the Hot Spot Analysis (Getis-Ord  $G_i^*$ ) tool, and set the following parameters:
  - Input Feature Class: Samples
  - Input Field: OZONE
  - Output Feature Class: **HotSpots**
  - Distance Band Or Threshold Distance: **150000**



150000 is the value that the Spatial Autocorrelation tool used as the threshold distance.

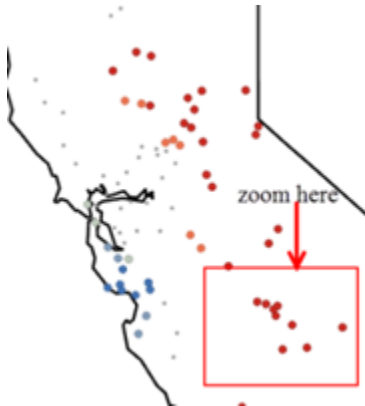
- c Accept all remaining default settings and click Run.
- d In the Contents pane, turn off the Samples layer.



Statistically significant hot spots of high ozone values are in north-central California. These hot spots indicate higher ozone values surrounded by other higher ozone values.

- e From the map tab, in the Navigate group, click the Explore down arrow and choose Topmost Layer.

- f** Zoom to the hot spots in north-central California.



- g** Click a few of the red points.

The OZONE values are all close to one another and are relatively high, around 0.08 to +1.0. Another benefit of performing hot spot analysis is seeing where there are clusters of low values, or cold spots.

- h** Zoom to the southern part of California where there are cold spots (blue points).

Most of the ozone values for the cold spots are around 0.05 or lower.

- i** Close the pop-up window.

- j** In the Contents pane, right-click HotSpots and choose Attribute Table.

OZONE	GiZScore Fixed 150000	GiPValue Fixed 150000
0.045	-0.010205	0.991858
0.106	5.917053	0
0.04	-1.849451	0.064393
0.041	-1.762682	0.077954
0.09	7.453489	0
0.053	-2.546165	0.010891
0.031	-2.385811	0.017041

Each point is assigned a value for z-score and p-value, and those values are added to the attribute table. When you have a combination of a high z-score and a low p-value, it indicates spatial clustering of high values. The lower or negative z-scores in combination with a lower p-value indicate spatial clustering of low values.

- k** Close the table.

Next, you will use the Optimized Hot Spot Analysis tool and let ArcGIS Pro determine the optimal parameters.

- l** In the Geoprocessing pane, return to the search results.



**m** Open the Optimized Hot Spot Analysis tool, and set the following parameters:

- Input Features: Samples
- Output Features: **OptimizedOzone**
- Analysis Field: OZONE

**n** Click Run.

The results are similar to that of the Hot Spot Analysis tool, but there are some differences if you turn the two layers off and on.

## Step 6: Create a density surface

In this step, you will compare the results of hot spot analysis with a density surface, or heat map.

**a** In the Contents pane, turn off the OptimizedOzone and HotSpots layers.

**b** In the Geoprocessing pane, search for **kernel**.

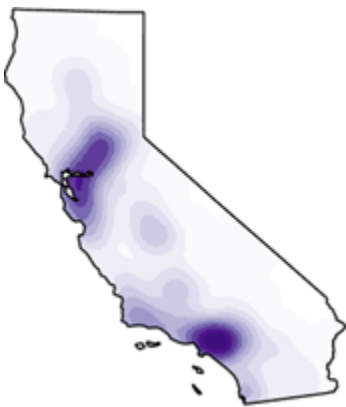
**c** Open the Kernel Density tool, and then at the top of the pane, click the Environments tab.

**d** Set the Cell Size to Maximum Of Inputs.

**e** At the top of the pane, click Parameters, and then set the following parameters:

- Input Point Or Polyline Features: Samples
- Population Field: NONE
- Output Raster: **Density**

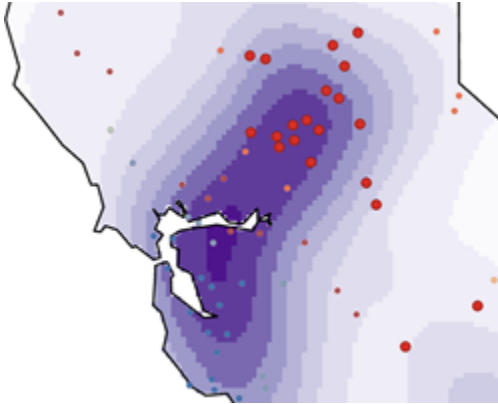
**f** Accept all remaining default settings and click Run.



**g** In the Contents pane, turn on the Samples layer.

The density surface aligns well with the sample points, as it should, because you did not use a population field and only used the point location to create the surface. However, the symbology of the heat map makes it difficult to interpret patterns.

- h** In the Contents pane, turn on the OptimizedOzone layer and turn off the Samples layer.
- i** In the OptimizedOzone layer legend, click the symbol for Hot Spot - 99% Confidence.
- j** At the top of the Symbology pane, click the Properties tab.
- k** Change Size to **15**, and then click Apply.
- l** Close the Symbology pane.
- m** In the map, zoom in on the northernmost part of California, where there is a high density of sample points.



The hot spots at a 99% confidence interval are not all located in the high-density area of the points. The density surface is symbolized based on density of features, not ozone values. Comparing a density surface to a statistical result demonstrates how spatial statistics quantify spatial patterns and how heat maps do not.

- n** Save the project, and then close the map view and continue to the next exercise.

## Perform clustering and outlier analysis

Clustering is an important concept that can give great insights into your data, its relationships with other data, and potential reasons about why it behaves a certain way. You will use the Spatial Statistics clustering tools to locate natural clusters in your data based on geographic location, perform a hot spot analysis on incident points to determine clustering of incidents, and run outlier analysis to verify patterns that you have visually analyzed.

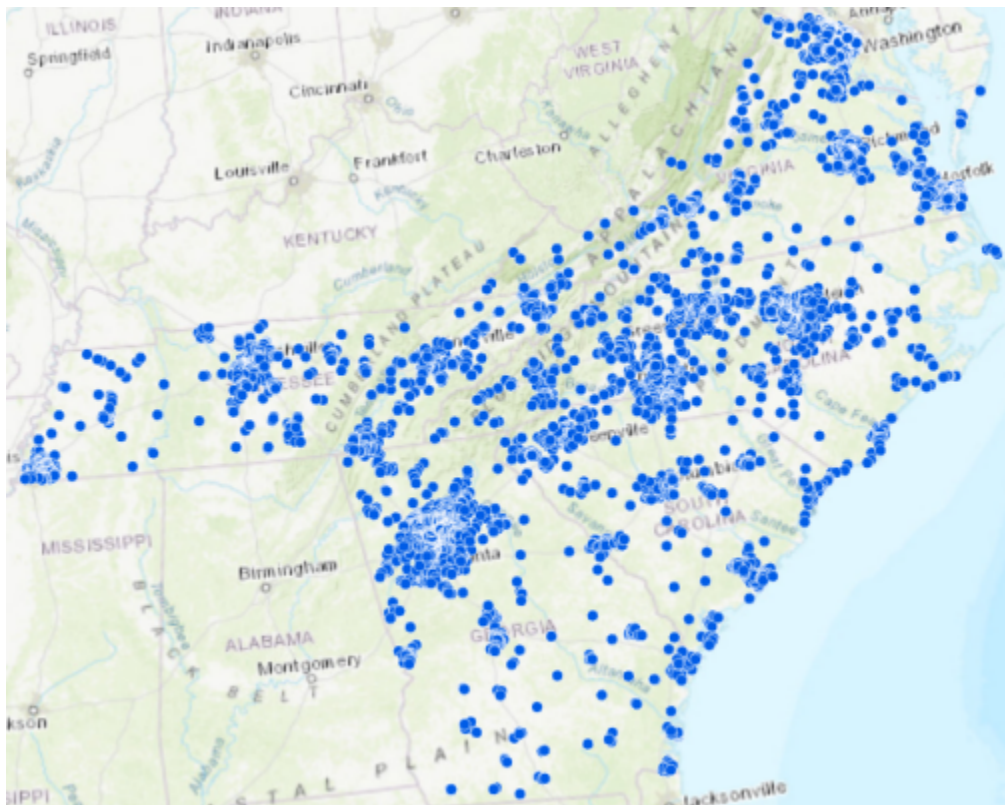
In this exercise, you will perform the following tasks:

- Perform density-based clustering analysis.
- Perform optimized hot spot analysis.
- Perform optimized outlier analysis.

## Step 1: Prepare the project

In this step, you will add a map file to your class project.

- a If necessary, restore the SNAPCourse project.
- b From the Insert tab, click Import Map.
- c Browse to **C:\EsriTraining\SNAP\Statistics** and open Clustering.mapx.



From a quick visual analysis, you can tell that more applicants are in the larger cities, like Atlanta. You will use density-based clustering tools to validate your initial insights and locate natural clusters in the data.

## Step 2: Perform density-based clustering

In this step, you will use clustering tools to analyze college applicant locations to determine several potential areas in the southeast to have career fairs.

- a In the Geoprocessing pane, search for **density**.

**b** Open the Density-Based Clustering tool and set the following parameters:

- Input Point Features: Applicants
- Output Features: **Clusters**
- Clustering Method: Self-Adjusting (HDBSCAN)



Some clustering methods require the user to input a search distance. The Self-Adjusting (HDBSCAN) method is a data-driven approach where the software determines the best search distance for the input features.

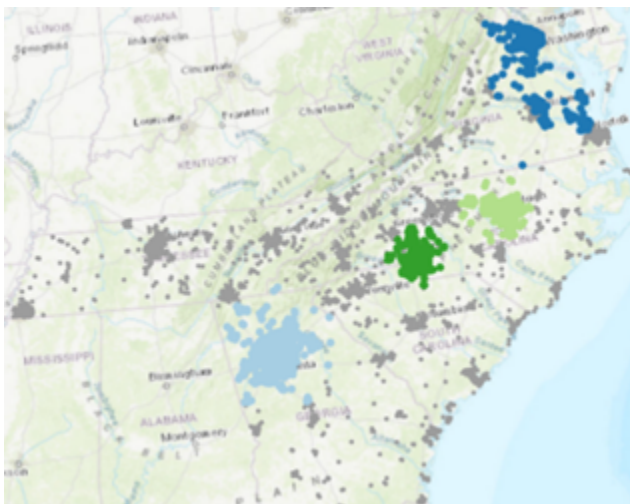
- Minimum Features Per Cluster: **500**



An ideal number of people to attend a career fair is 500.

**c** Click Run.

**d** In the Contents pane, turn off the Applicants layer.



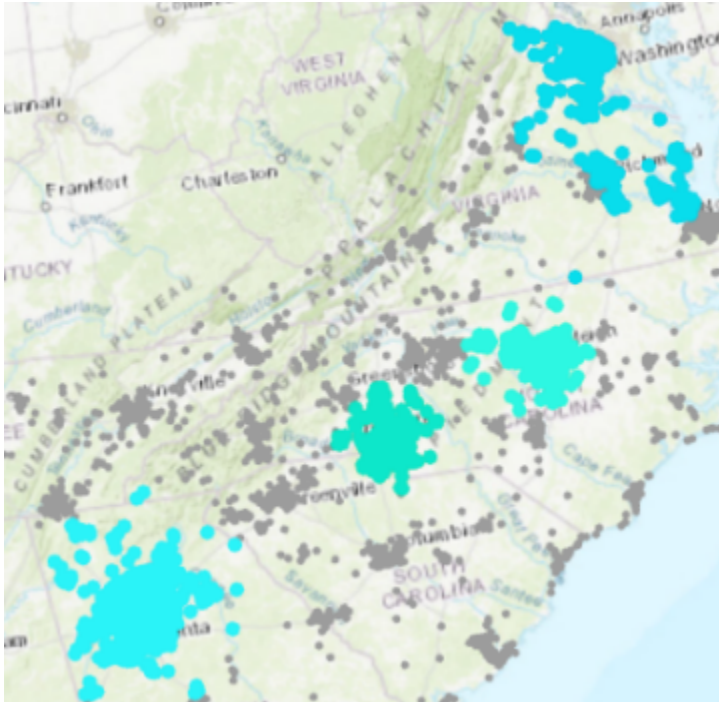
All the colored areas represent clustering, whereas the gray locations represent "noise" in your data.

**e** In the Contents pane, open the Clusters layer attribute table.

Source ID	Cluster ID	Probability	Outlier
1	2	0.822509	0.177491
2	2	0.892063	0.107937
3	-1	0	0.576296
4	-1	0	0.091964
5	-1	0	0.113664
6	-1	0	0.116846
7	-1	0	0.145086
8	1	0.287308	0.712692

Each feature is assigned a Cluster ID value. If a feature is part of a valid cluster, then the value is positive, whereas a -1 indicates that the point location is noise. ArcGIS Pro symbolizes the output layer using the Cluster ID attribute. Next, you will find the best locations for career fairs using the clusters and Mean Center, a descriptive statistical tool.

- f Close the table.
- g From the Map tab, in the Selection group, click Select By Attributes.
- h In the Geoprocessing pane, ensure that Input Rows is set to Clusters.
- i Click Add Clause.
- j Build the following clause: Cluster ID Is Not Equal To -1.
- k Click Add, and then run the tool.



All the features that are part of a natural cluster are selected. Next, you will use the Mean Center tool on the selected points to locate the best places for career fairs.

- l In the Geoprocessing pane, search for and open the **Mean Center** tool.

**m** Set the following parameters:

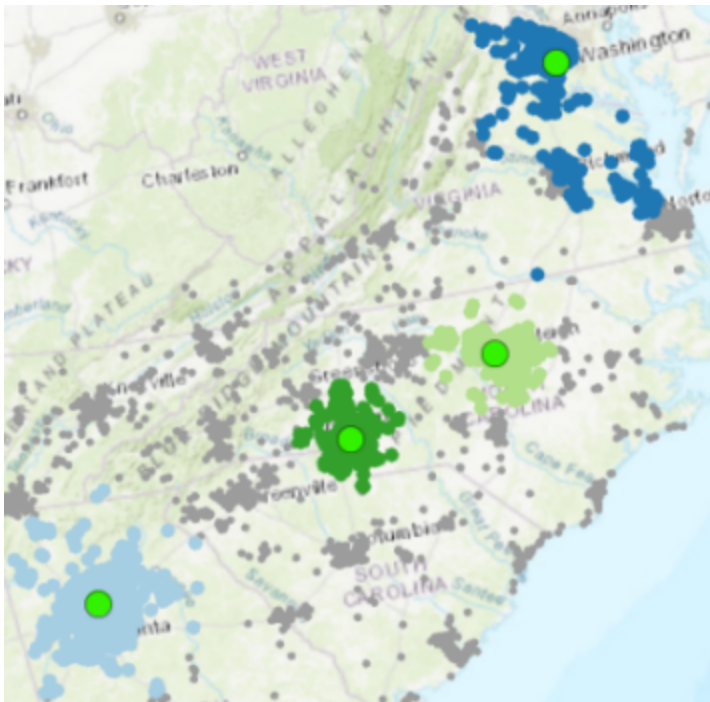
- Input Feature Class: Clusters
- Output Feature Class: **CareerFairs**
- Case Field: Cluster ID



Geoprocessing tools honor the selected set.

**n** Click Run.

**o** Clear the selected features.



The Mean Center tool located the geographic center of each cluster, thus identifying some possible areas to explore for locating the career fairs.

**p** Save the project.

### Step 3: Perform optimized hot spot analysis

In this step, you will use cluster and outlier tools to gain insights about graffiti locations in New York.

**a** From the Insert tab, click Import Map.

**b** Browse to `..\EsriTraining\SNAP\Statistics` and import New York.mapx.





There are more than 120,000 graffiti locations in these New York neighborhoods. With more than 120,000 points, how can you acquire useful information from the data? Using spatial statistics tools is a great place to start. You will first perform an optimized hot spot analysis to identify hot and cold spots based solely on location.

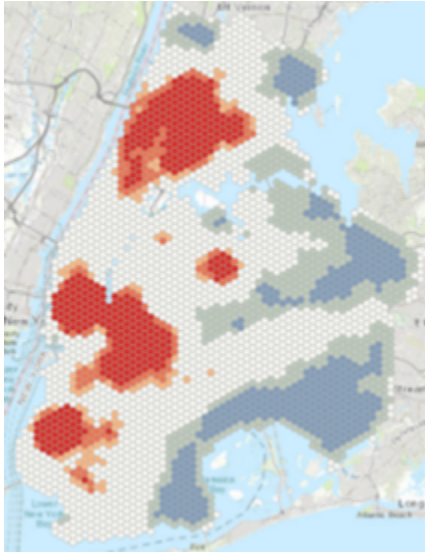
- c From the Analysis tab, click Environments.
- d Set Extent to Same As Layer - Boundary.
- e Clear the Mask, and then click OK.
- f Search for and open the **Optimized Hot Spot Analysis** tool, and then set the following parameters:
  - Input Features: Graffiti
  - Output Features: **OptimizedHS**
  - Analysis Field: Leave this field blank, as you will perform the analysis on the point locations rather than an attribute at each location
  - Incident Data Aggregation Method: Count Incidents Within Hexagon Grid
  - Bounding Polygons Defining Where Incidents Are Possible: Boundary
  - Expand Override Settings:
    - Cell Size: **500** Meters



If you do not override the cell size, the cell size will be determined using the Average Nearest Neighbor tool, which will be roughly 430 meters. You can let the software calculate certain settings or override them if you are unfamiliar with the data.

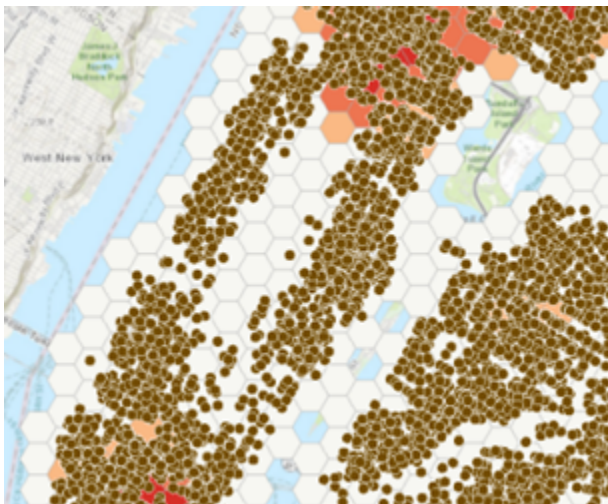


- g** Click Run.
- h** In the Contents pane, turn off the Graffiti layer.



There are statistically significant hot spots and some statistically significant cold spots. You may want to examine the hot spots further to determine the causes for more graffiti in those areas and then establish a remediation plan.

- i** In the Contents pane, turn on the Graffiti layer.
- j** Zoom in to the area, as seen in the following graphic:



The non-significant areas often have no graffiti points in them. For example, the large area in the center has no points.

## 1. Why might points not exist here?

---

You will use the basemap layer to gain a better understanding of why some locations have no graffiti points.

- k** In the Contents pane, turn off the OptimizedHS layer.



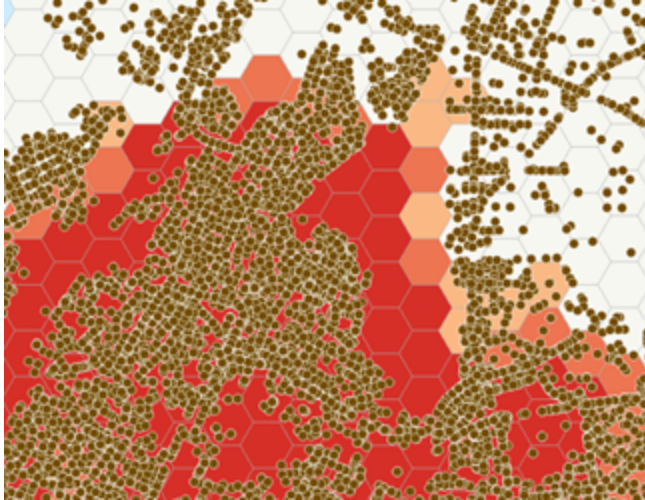
There are no graffiti incidents reported in Central Park, possibly due to more security and a lack of structures or objects to paint or draw on.

- l** Save the project.

### Step 4: Perform optimized outlier analysis

You have used the Spatial Autocorrelation tool to quantify patterns on a global level, meaning one statistic for an entire area. In this step, you will use a local statistic to determine where there are cold spots within hot spots.

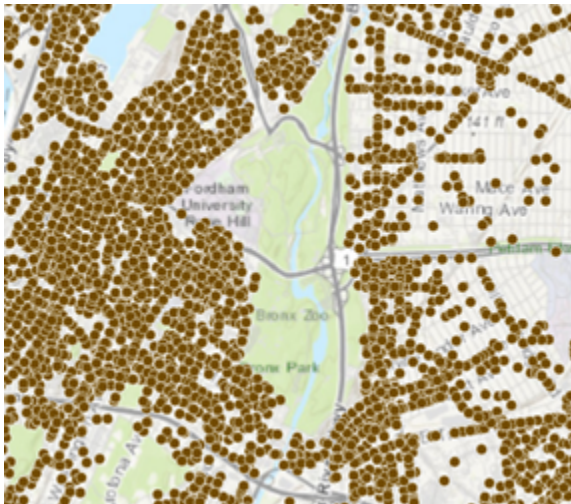
- a** In the Contents pane, turn on the OptimizedHS layer.
- b** With the Explore tool, zoom to the large hot spot that is located farthest north.



## 2. Why might there be no actual graffiti points within a hot spot?

---

- c In the Contents pane, ensure that the OptimizedHS layer is selected, and then click the Appearance tab.
- d In the Effects group, click Swipe.
- e Click in the map, and drag the Optimized layer to see the basemap.



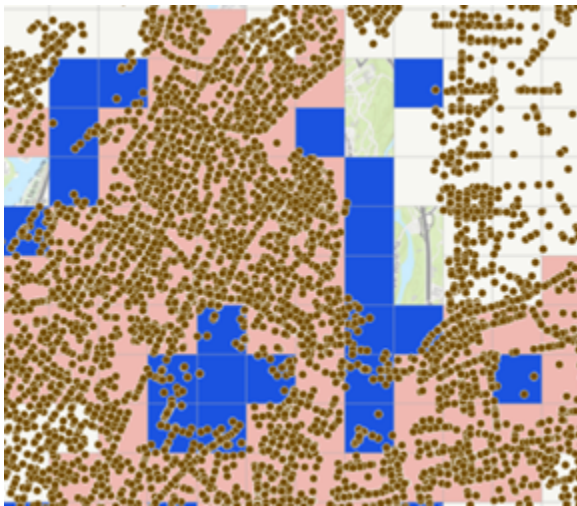
The Bronx Zoo is in the location where there are few to no graffiti points, yet it exists within a hot spot. The reason there are hot spots even though there are no points is because the red hex bins are red based on their neighbors. You will use a local outlier tool to explore this area further.

f Search for and open the **Optimized Outlier Analysis** tool, and then set the following parameters:

- Input Features: Graffiti
- Output Features: **Outliers**
- Expand Override Settings:
  - Cell Size:**500** Meters

g Click Run.

h In the Contents pane, turn off the OptimizedHS layer.



The Optimized Outlier tool validates your findings that the area of the zoo contains spatial outliers of graffiti incidents. The blue fishnets indicate a low number of graffiti incidents per bin, surrounded primarily by fishnets with high numbers of incidents.

i Save the project, and leave ArcGIS open.

## Lesson review

---

1. Explain how spatial statistics remove subjectivity from your data.

---

---

---

---

---

---

2. Briefly describe descriptive and inferential statistics.

---

---

---

---

---

---

# Answers to Lesson 8 questions

## Spatial patterns (page 8-2)

1. What have you done to busy, subjective maps to show them in a meaningful way?

**Answers will vary based on personal experience.**

## Descriptive versus inferential (page 8-9)

### Scenario 1: Directional trend returned

1. Does the Directional Distribution tool describe your data or make statistical inferences based on the data provided?

**It describes the data.**

### Scenario 2: Chart and statistics returned

2. Does the Average Nearest Neighbor tool describe your data or make statistical inferences about it based on the data provided?

**It makes statistical inferences about the data.**

### Scenario 3: Map of hot and cold spots returned

3. Does the Optimized Hot Spot tool describe your data or make statistical inferences about it based on the data provided?

**It makes statistical inferences about the data.**

## Exercise 8B: Perform clustering and outlier analysis (page 8-27)

1. Why might points not exist here?

**This area could be where there are no places to create graffiti.**

2. Why might there be no actual graffiti points within a hot spot?

**There could be underlying geographic factors, as seen in the Central Park example.**

Incorporating time into your spatial analysis allows you to focus on how spatial patterns in your data may vary or change over time. By analyzing data over time, you may detect trends or patterns that you would not otherwise detect had you analyzed the data for the entire time period. Understanding how patterns have changed over time can help you determine how they might change in the future and better prepare you for these changes. ArcGIS Pro has tools for analyzing data in space and time.

## Topics covered

---

Temporal analysis

Emerging hot spots

Space-time analysis workflow



## Incorporating time into your analysis

---

You have performed many analyses in the course and each one answered one specific question: Where is something?

You have answered questions such as these:

- Where are customers within 15-minute drive times of each store?
- Where do streams and watersheds overlap?
- Where are the most suitable locations for bear habitats?
- Where are there natural clusters of college applicants?
- Where are best locations for a career fair?
- Where should an outdoor retailer target its marketing?
- Where are there hot spots of ozone concentrations and graffiti incidents?

GIS is often used to find where things are located, but you can add another factor to your analysis that can give you more information about data and its patterns: time.

**How can incorporating time into your data improve your analysis results?**



## Temporal analysis

Proximity, overlay, and statistical analysis help you determine the "where" questions about your data and look at the spatial variations in your data. Time-based analysis, or temporal analysis, adds another dimension to your analyses and can help answer the "when" questions about your data. In GIS, temporal analysis refers to an analysis that involves a time attribute. Temporal analysis is useful for studying the variation, or changes, in data over time at the same location.

### Spatial variation versus temporal variation

In the following example, researchers are collecting ozone readings at monitoring stations throughout California. To gain an understanding of the spatial variability of the ozone measurements across the landscape, you would capture many samples throughout the entire state and compare the results.

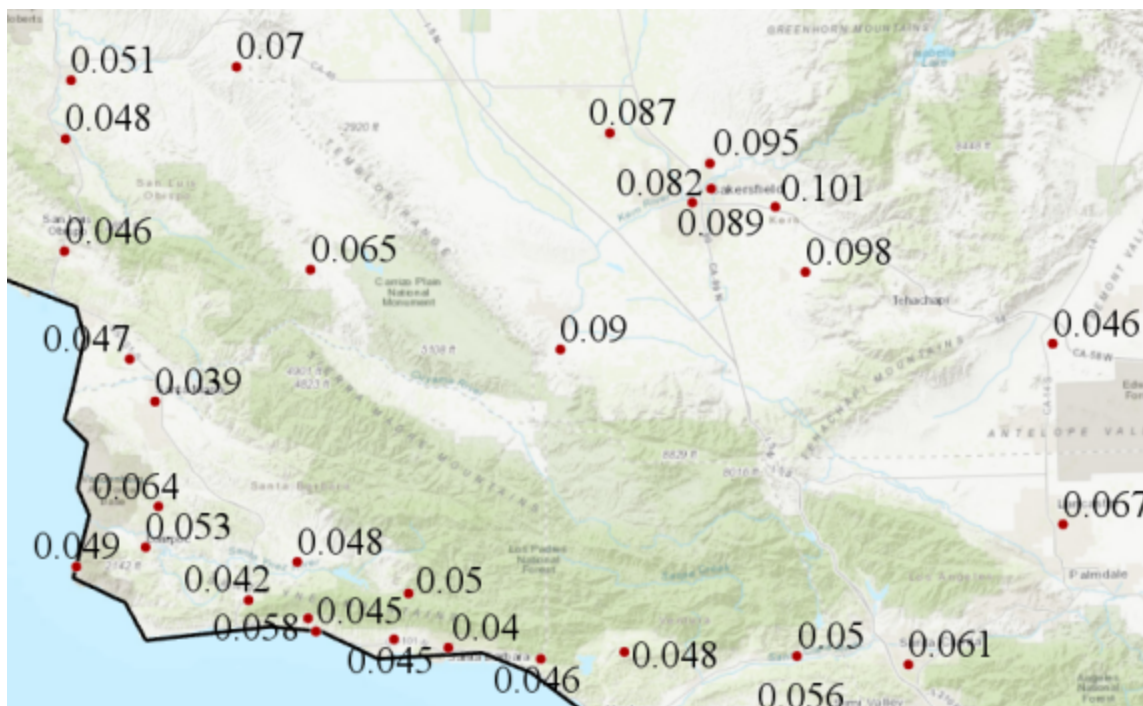
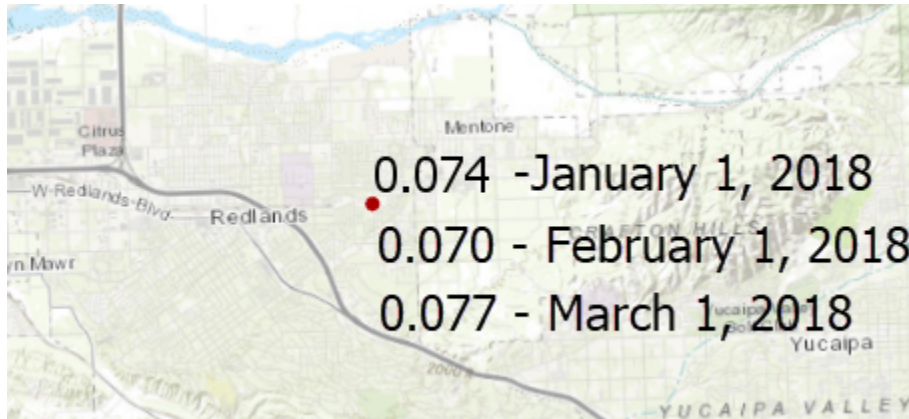


Figure 9.1. Spatial variation of ozone readings at monitoring stations.

## Temporal analysis (continued)

---

To gain an understanding of the temporal variability of the ozone measurements, you would capture ozone readings at the same location and measure the variance in the samples over time.



**Figure 9.2.** Temporal variance from one monitoring station.

Benefits of adding a temporal component to your analysis include the following:

- Reveal patterns that may not be detected when visualizing data over the full time period.
- Determine if an event occurs more frequently during certain times of day, week, month, hour, or minute.
- Focus efforts on more recent occurrences.

## Explore data

Exploratory data analysis is a big part of analysis. You can view data in a map, view attributes, or make charts to discover new information about your data that will help in your analysis. You will examine historical tornado data from Texas and do some exploratory analysis using time.

In this exercise, you will perform the following task:

- Perform exploratory data analysis using a chart.

## Step 1: Use a chart to explore data

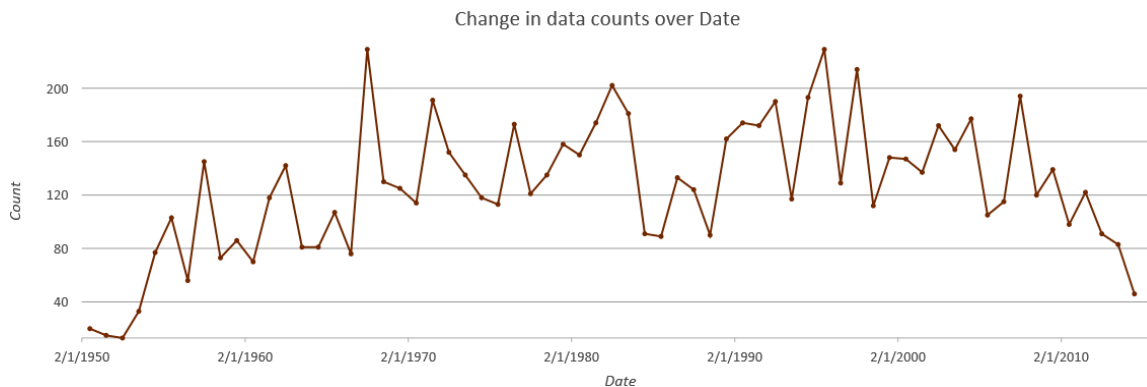
Charts are powerful visualization and exploration tools to help you learn more about your data. You will use a line chart to explore tornado data in Texas.

- a If necessary, start ArcGIS Pro and open the SNAPCourse project.
- b From the Insert tab, choose Import Map.
- c Browse to **C:\EsriTraining\SNAP\SpaceTime** and open the Texas.mapx map file.
- d In the Contents pane, open the Tornado Start Points layer attribute table and explore the attributes.

### 1. Which attribute can you use to analyze the temporal aspect of the tornado points?

---

- e Close the table.
- f In the Contents pane, ensure that Tornado Start Points is selected, and then click the Data tab.
- g In the Visualize group, click Create Chart and choose Line Chart.
- h In the Chart Properties pane, update Date Or Number to Date.
- i Under Time Binning Options, set Interval Size to **1** Years, and then click away from the setting to see the change in the chart.



### 2. What patterns can you determine about tornadoes in Texas?

---



---

**3. How can using 70 years' worth of data influence a hot spot analysis?**

---

- i** Close the Texas map, and keep ArcGIS Pro open for the next exercise.

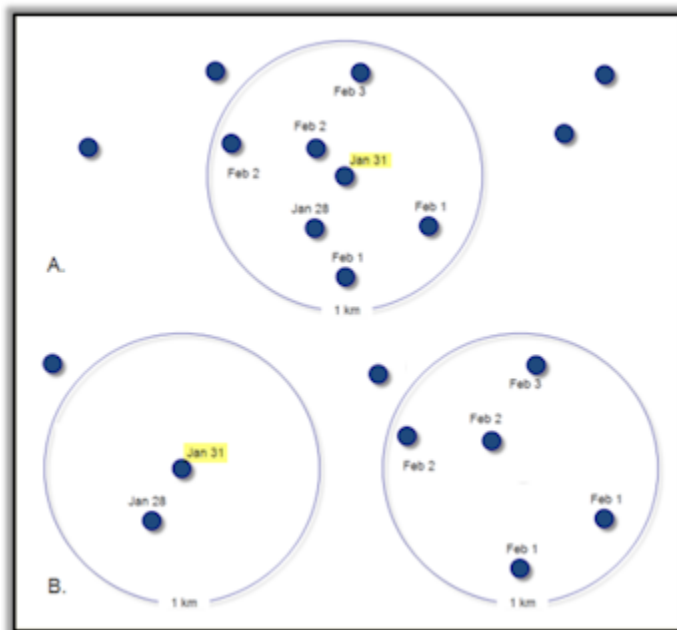
## Space-time analysis

Everything happens within the context of space (location) and time. GIS can analyze spatial patterns well, but spatial patterns may change over time. If you are analyzing your data in space only, then you may only be getting half the story. You can create and analyze time snapshots and perform true space-time analysis.

### Time snapshots

Time is often analyzed as a time snapshot, or arbitrary groupings based on time. For example, you may have data that spans one year and break it up into 12 layers, each representing a month, or you could add a Month attribute to the table and categorize by month. While breaking up layers by time snapshots allows you to visualize temporal trends over the course of the year, you may be arbitrarily breaking up data that is truly related in space and time, and possibly missing important patterns and trends.

In the following image, the data is broken up into snapshots for January and February, but notice the dates associated with each point. Each incident occurred within a few days of each other, but by arbitrarily separating them into separate month bins, you may miss a potential pattern.



**Figure 9.3.** Time snapshots can disjoin related data.

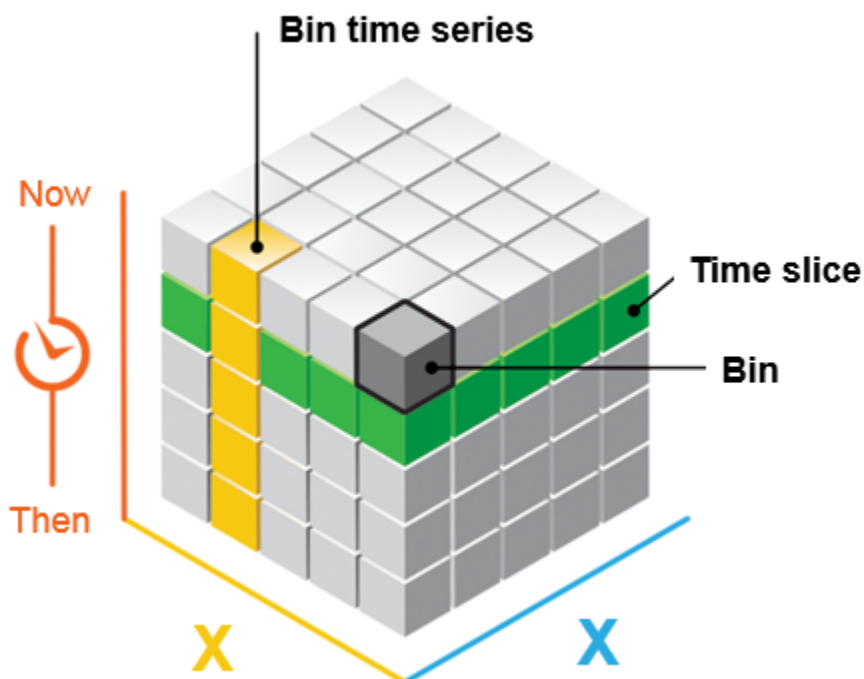
## Space-time analysis (continued)

### Space-time pattern mining

In the previous graphic, six features fall within a 1-kilometer and a seven-day space-time window of the feature labeled Jan 31. However, only one feature will be included as a temporal neighbor if the data is analyzed using monthly snapshots. The separation of two to three days between dates may not mean as much as the proximity of the incidents.

True space-time analysis considers each incident in relation to incidents near to it in both space and time, and it is not dependent on arbitrary categories, such as months or days.

In the following graphic, the time slice, also referred to as the time step interval, may seem similar to a time snapshot, but it is not. A time step interval is a moving window of time that considers all its neighbors in time and space, while a time snapshot does not consider other features.



**Figure 9.4.** Each point is put into a bin time series, with the bottom being older and the top being more recent events.

## Space-time analysis (continued)

---

### Creating a space-time cube

Creating a space-time cube allows you to visualize and analyze your spatiotemporal data in the form of time-series analysis, integrated spatial and temporal pattern analysis, and powerful 2D and 3D visualization techniques.

Two tools for creating a space-time cube are the Create Space Time Cube By Aggregating Points tool and the Create Space Time Cube From Defined Locations tool. Both tools take time-stamped features and structure them into a netCDF (network Common Data Form) data cube by generating space-time bins with either aggregated incident points or defined features with associated spatiotemporal attributes. NetCDF is a file format for storing multidimensional scientific data (variables), such as temperature, humidity, pressure, wind speed, and direction.

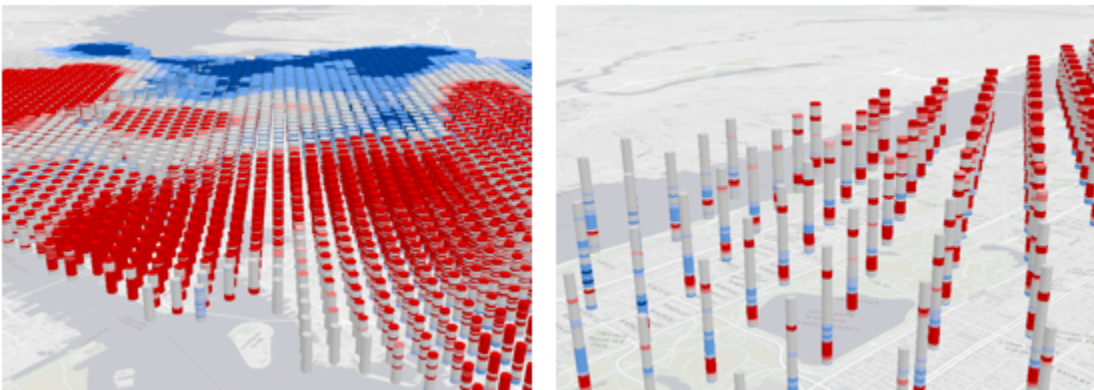


Figure 9.5. A space-time cube viewed in 3D.

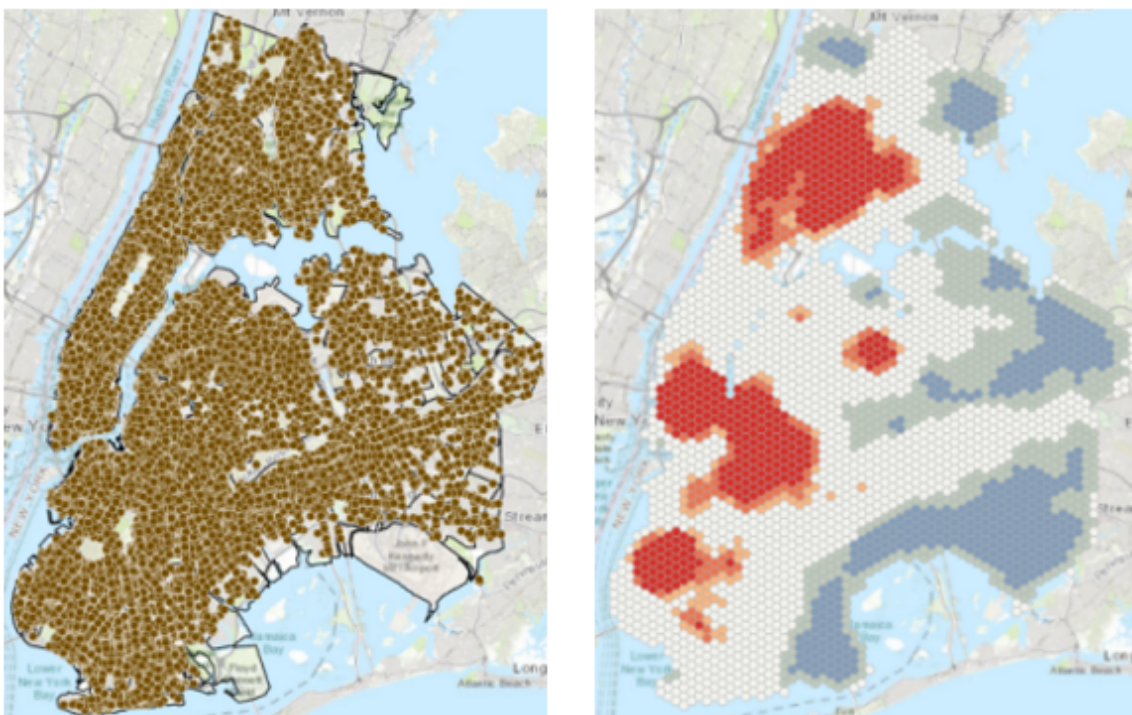


**ArcGIS Pro Help:** *Why hexagons?*



## Emerging hot spot analysis

Earlier, you examined a dataset containing more than 120,000 graffiti incident locations. Because of the sheer number of points, it was difficult to visually identify any spatial patterns. You ran the Optimized Hot Spot Analysis tool and its result showed statistically significant hot and cold spots. From the optimized map and statistical results, you can identify clusters. You know where the hot and cold spots are, but you want to add time into the analysis to see how graffiti patterns have changed to help narrow your areas of interest. You cannot put resources everywhere, so by narrowing the focus to current hot spots, you may be more efficient at prevention or mitigation.

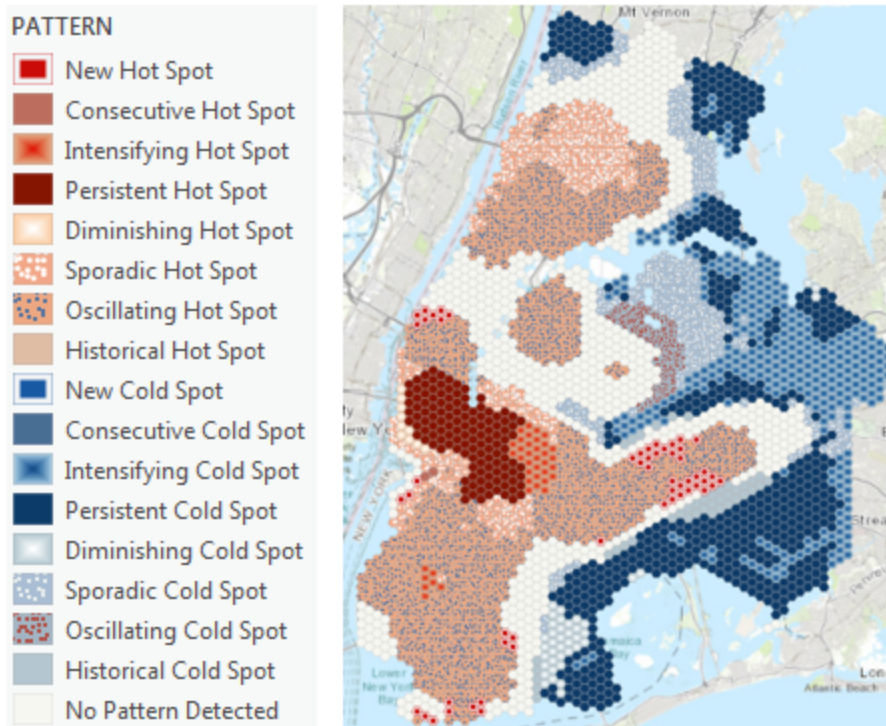


**Figure 9.6.** Hot spots created from thousands of points indicate clusters where there are statistically significantly higher numbers of graffiti incidents occurring than the remainder of the study area. But does this map tell the whole story?

You can use a tool called Create Space Time Cube By Aggregating Points to create a netCDF file (network Common Data Form—a file format for storing multidimensional scientific data) containing the space-time cube. Then you can add the space-time cube into the Emerging Hot Spot Analysis tool to get the result in the following graphic. The following result uses a space-time cube based on eight years of data broken up into three-month time intervals to reflect the seasons, as there may be seasonal variance to graffiti occurrences. The legend for the layer provides useful descriptions of the symbology in the map. For example, a new hot spot may be an area of interest as it was never a hot spot before the most recent time interval.

## Emerging hot spot analysis (continued)

---



**Figure 9.7.** Emerging Hot Spot Analysis results show you varying degrees of hot and cold spots based on a time interval.

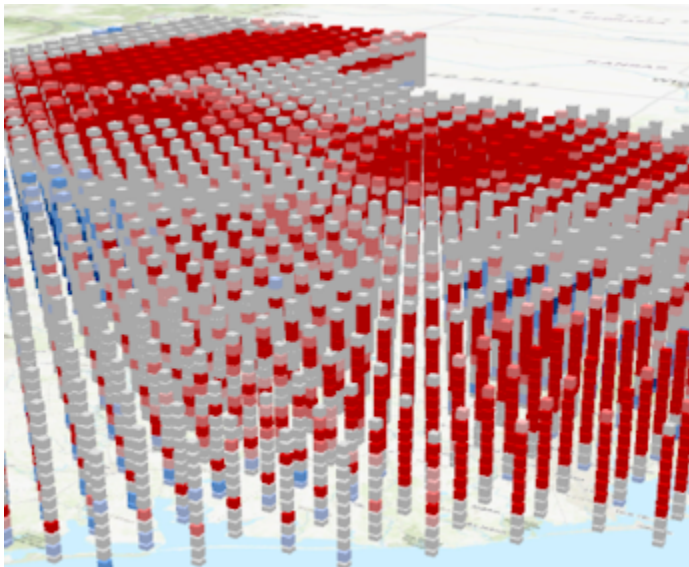
## Space-time analysis workflow

---

To perform true space-time analysis, you perform the following operations:

- Run the Create Space Time Cube By Aggregating Points tool to create a netCDF dataset.
- Run the Emerging Hot Spot Analysis tool.
- Run the Visualize Space Time Cube in 3D tool.

You can run the Visualize Space Time Cube In 3D tool in any order with the Emerging Hot Spot Analysis tool, but you have to create the space-time cube before either. The space-time cube netCDF file is used as an input in both the Emerging Hot Spot Analysis tool and the Visualize Space Time Cube In 3D tool.



**Figure 9.8.** Visualizing the space time cube in 3D.

## Space-time analysis workflow (continued)

---

Why visualize the space-time cube in 3D?

- Better understand the structure of the space-time cube and how the process of aggregation into the cube works.
- Offer insights into the results of Emerging Hot Spot Analysis and Local Outlier Analysis, providing evidence that can help you understand the result categories themselves.
- Additionally, visualizing summary fields and variables can help you understand how confident that you can be in subsequent analyses by displaying the spatial pattern of empty bins that had to be estimated.

## Explore space-time pattern mining tools

Earlier, you used the Optimized Hot Spot Analysis tool to evaluate more than 120,000 graffiti incidents and show where there are hot spots, cold spots, and areas with no statistically significant clusters of graffiti. However, the data spans an eight-year time period, and you would like to dive deeper into the analysis by incorporating a third dimension: time. Your optimized hot spot map shows you where there are hot spots and cold spots of graffiti incidents for eight years of cumulative data without consideration of how consistent they have been or if they have changed in location over time. By factoring in time, you can narrow down areas of interest and focus on a few problem spots to help reduce graffiti. In this exercise, you will use ArcGIS Pro space-time pattern mining tools to further explore the incidents by factoring in when the incident occurred.

In this exercise, you will perform the following tasks:

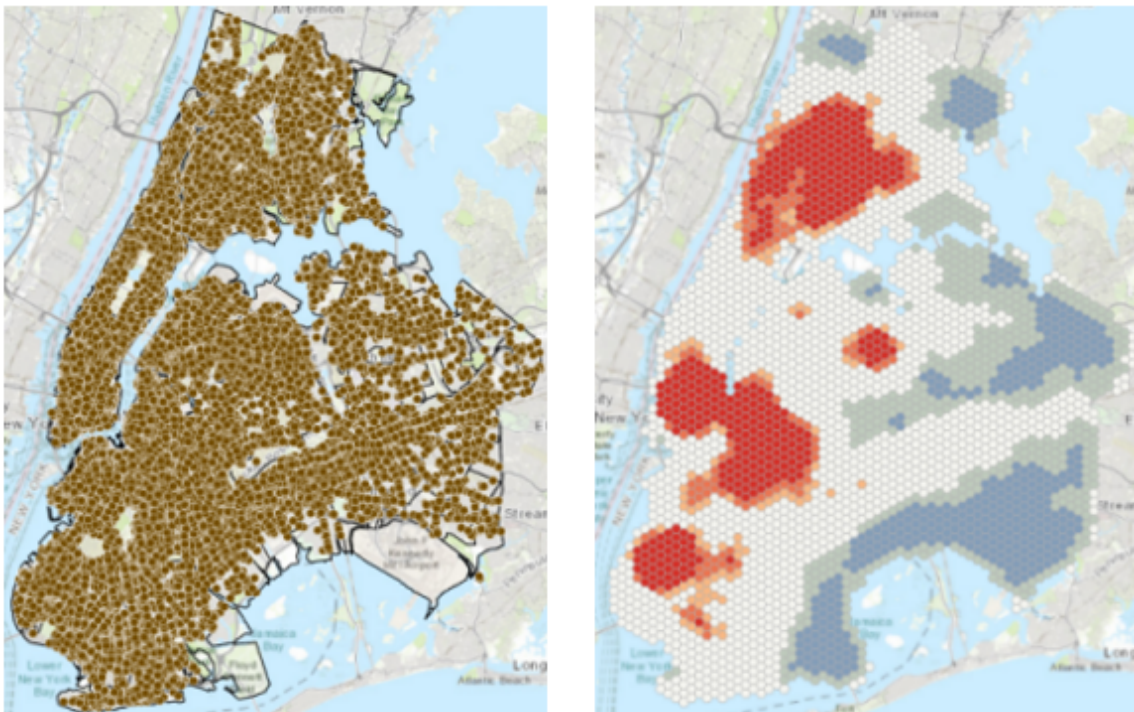
- Explore temporal data using a chart.
- Create a space-time cube.
- Perform emerging hot spot analysis.
- View the space-time cube in 3D.



## Step 1: Explore data using charts

You will view your hot spot map from the previous exercise (or open the available map provided in the Results file in your data), and then perform some exploratory analysis using charts and a time attribute.

- a If necessary, start ArcGIS Pro and open the SNAPCourse project.
- b View the New York map.
- c In the Contents pane, make Graffiti, OptimizedHS, and the basemap the only visible layers.



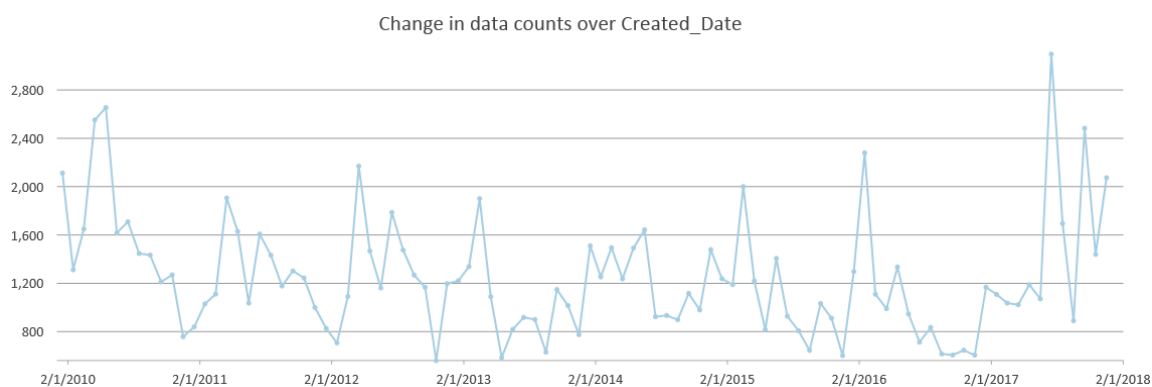
You will incorporate time into your analysis to determine if there is more to the story regarding the graffiti incidents. There are 120,878 points that span over eight years. You are more concerned about incidents that happened in the last few months to a year rather than ones that happened eight years ago. Space-time analysis tools can locate temporal hot spots that you can focus on.

- d Turn off the OptimizedHS layer.
- e From the Map tab, click the Explore tool.
- f In the Contents pane, select the Graffiti layer.
- g In the map, click a graffiti point.

OBJECTID	2444
Unique_Key	27186052
Created_Date	1/18/2014 9:48:52 AM
Agency	DSNY

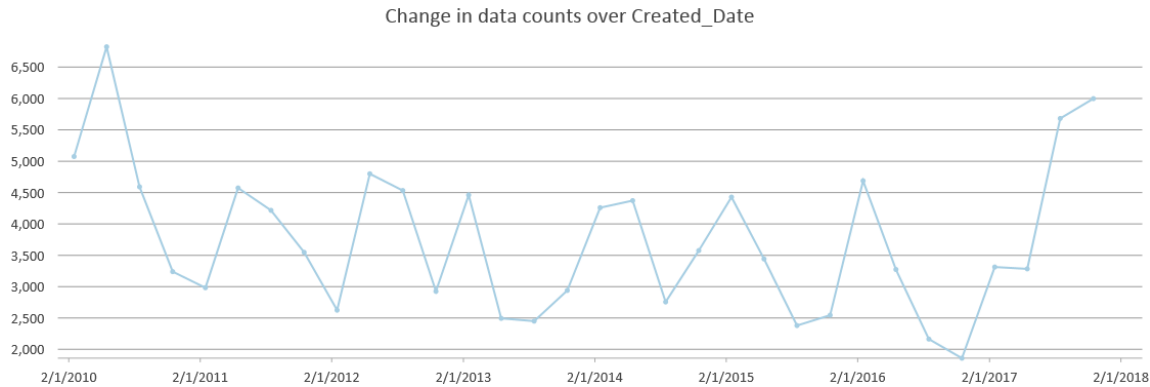
You likely clicked a different point in the map, but you will notice the Created\_Date field. Each point has a date and time associated with it that you will use to further investigate the spatial patterns in the graffiti incidents. Exploratory data analysis is always recommended before statistical analysis, as many statistical methods require an understanding of data distribution, presence of outliers, spatial autocorrelation, and other factors. You will display the temporal trend in the graffiti incidents using a line chart.

- h** Close the pop-up window.
- i** With the Graffiti layer still selected, click the Data tab.
- j** In the Visualize group, click Create Chart and choose Line Chart.
- k** In the Chart Properties pane, for Date Or Number, choose Created\_Date.



The chart shows all eight years of data using a one-month interval size. You can see that there are natural dips and spikes in incidents. One notable pattern is that in the spring of each year, there is a large spike in incidents. You are not trying to determine why graffiti occurs; rather, you are trying to get a sense of temporal patterns. The interval size used in the chart depends on the data that you have. For example, you could show data that spans 50 years in five-year intervals, but if the data is for one day, you may show it in increments of several hours. For the incidents, you will change the interval to three months, based on seasons.

- l** In the Chart Properties pane, change the Interval Size to 3 Months.



When you view the chart using a three-month time interval based on season length, you can still see a temporal pattern that indicates graffiti spikes at times and dips at times. Temporal variation in graffiti incidents may be caused by seasonal changes, as spring and summer weather may be more common times to create graffiti. The chart shows you that there is temporal variation in graffiti incidents, but by incorporating time using the space-time pattern mining tools, you can learn much more.

**m** Close the chart and the Chart Properties pane.

## Step 2: Create a space-time cube

In this step, you will use the ArcGIS Pro space-time pattern mining tools to create a space-time cube.

**a** In the Geoprocessing pane, click the Back button until you see the list of recently used tools.

*Hint:* If you closed the Geoprocessing pane: Analysis tab > Tools.

**b** Click the Toolboxes tab.

**c** Expand Space Time Pattern Mining Tools.

**d** Open the Create Space Time Cube By Aggregating Points tool, and set the following parameters:

- Input Features: Graffiti
- Output Space Time Cube: **Cube\_3Month**



The output has a .nc extension to represent netCDF.

- Time Field: Created\_Date
- Time Step Interval: **3** Months





The Time Step Interval parameter is not a time snapshot. By setting the time step interval to 3 months, you will be able to compare each season to the previous season, and so on.

- Aggregation Shape Type: Hexagon Grid



Hexagon grids are a good alternative to a fishnet grid. Hexagon grids reduce sampling bias and represent patterns in your data more naturally than a fishnet grid. Finding neighbors is also easier because the length of contact is the same on each side.

- Distance Interval: **500** Meters



You used 500 meters when you ran the Optimized Hot Spot Analysis tool, so you will use the same value here.

- e Click Run.
- f At the bottom of the Geoprocessing pane, point to the green box to view the messages for the tool.



You can also click View Details.

- g Scroll down in the messages to view the statistics about your space-time cube file.

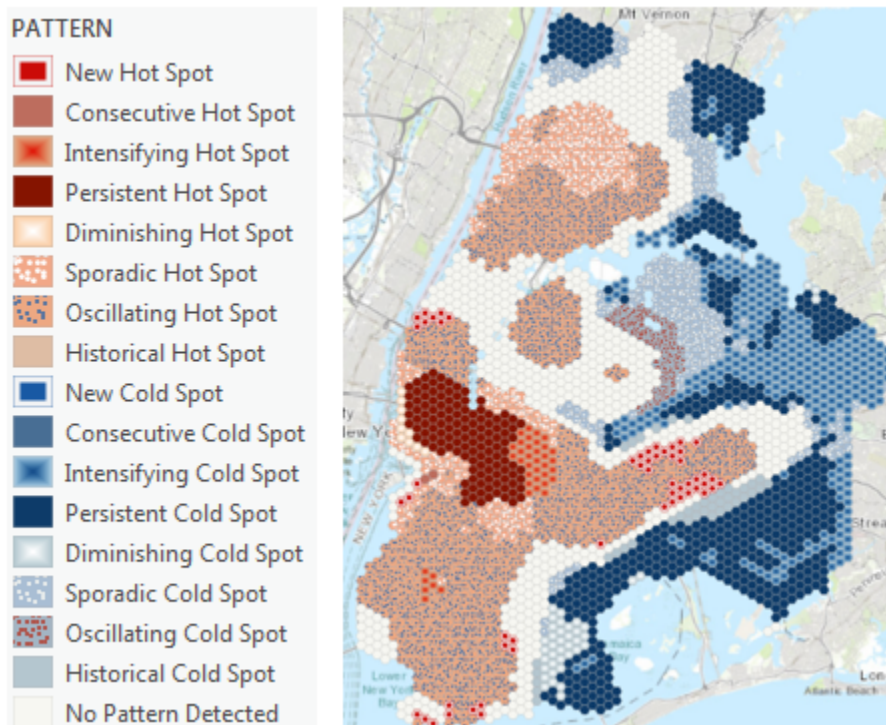
No visual result is created, but you will input the netCDF file that you created into the Emerging Hot Spot Analysis tool.

### Step 3: Run the Emerging Hot Spot Analysis tool

In this step, you will run the Emerging Hot Spot Analysis tool to identify trends in clustering or values in your space-time cube.

- a In the Geoprocessing pane, click the Back button.
- b In the Space Time Pattern Mining Tools toolbox, open the Emerging Hot Spot Analysis tool.

- c Set the following parameters:
  - Input Space Time Cube: **C:\EsriTraining\SNAP\SNAPCourse\Cube\_3Month.nc**
  - Analysis Variable: COUNT
  - Output Features: **EmergingHS**
  - Conceptualization Of Spatial Relationships: Fixed Distance
  - Polygon Analysis Mask: Boundary
- d Accept the remaining defaults, and click Run.
- e In the Contents pane, make EmergingHS and the basemap the only visible layers.



Running the Emerging Hot Spot Analysis tool provides valuable information about your data. You can match the map symbology with the legend and description to gain a good understanding of the areas that you should focus on. For example, any persistent, intensifying, or new hot spots may be of interest, while sporadic hot spots may not be a concern. The results narrow down the focus even more from the result of the Optimized Hot Spot Analysis tool because you incorporated time.

How are the descriptions in the legend created? You will find information in the ArcGIS Pro Help documentation.

- f Open ArcGIS Pro Help.

*Hint:* In the upper-right corner of ArcGIS Pro, click the View Help question mark.

- g** Search for **emerging hot spot**.
- h** Open the *How Emerging Hot Spot Analysis works* result.

The table provides descriptions of all the categories in the emerging hot spot result. You can see why certain locations in your map are new hot spots—they were a statistically significant hot spot for the final time step and were never a hot spot before. The information that you can discover from the space-time pattern mining tools may reveal patterns and trends that were not visual to the human eye, or that were not apparent when analyzing the data over the entire time period.

- i** Briefly read the descriptions for Intensifying Hot Spot and Persistent Hot Spot.
- j** Minimize the ArcGIS Pro Help window.

You started with eight years' worth of data and more than 120,000 points. Now, you have narrowed down your analysis focus to two or three smaller areas.

- k** Save the project.

## Step 4: Visualize a space-time cube in 3D

In this step, you will visualize the results of your space-time analysis in 3D.

- a** From the View tab, in the View group, click Convert and choose To Local Scene.
- b** In the Geoprocessing pane, return to the Toolboxes tab.
- c** Under Space Time Pattern Mining Tools, expand Utilities.
- d** Open the Visualize Space Time Cube In 3D tool, and set the following parameters:

- Input Space Time Cube: `..\EsriTraining\SNAP\SNAPCourse\Cube_3Month.nc`
- Cube Variable: COUNT
- Display Theme: Hot And Cold Spot Results

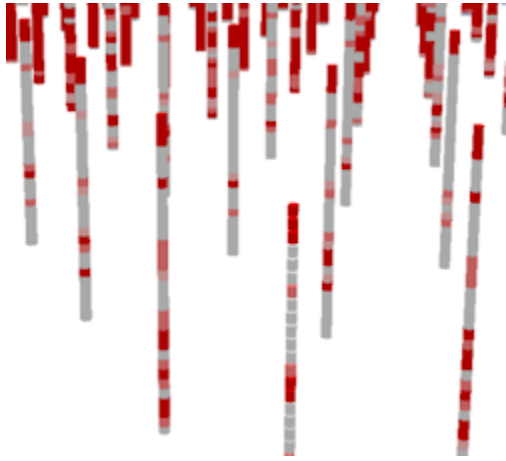


The Display Theme parameter determines the type of information shown in the legend.

- Output Features: **GraffitiCube3D**
- e** Click Run.

The GraffitiCube3D layer is automatically added into the 3D scene.

- f In the Contents pane, in 2D Layers, turn off all layers, including the basemap.
- g From the Map tab, click the Explore tool, if necessary.
- h Click the roller wheel of the mouse, and tilt the 3D view.
- i Zoom in so that you can see the hexagon bins.




Visualizing the space-time cube in 3D allows you to see how hot and cold spots have changed over time. Each hexagon bin represents a three-month time interval, with the oldest time intervals on the bottom and the most recent on the top.

You can enable time on the layer to see each time interval display using the Time Slider.

- j In the Contents pane, right-click GraffitiCube3D and choose Properties.
- k Click the Time tab.
- l For Layer Time, choose Each Feature Has Start And End Time Fields.

The remaining information will be automatically populated.

Layer Time	Each feature has start and end time fields
Start Time Field	Start Date
End Time Field	End Date
Time Extent	10/13/2009 3:45:45 PM  - 1/13/2018 3:45:44 PM

- m Click OK.

A time slider is now available at the top of your map.

- n If necessary, zoom out so that you can see all the hexagon bins.
- o Point to the time slider, and then click the play button to watch each hexagon bin's three-month time interval appear.
- p Close the New York\_3D scene.

You have used space-time analysis tools to dig deeper into your data to better understand it.

- q Save the project, and leave ArcGIS Pro open.

## Lesson review

---

1. Explain how adding time can enhance your analysis results.

---

---

---

---

2. Explain spatial and temporal variance.

---

---

---

---

3. Differentiate between analyzing time snapshots of data and true space-time analysis.

---

---

---

---

---

---

# Answers to Lesson 9 questions

## Incorporating time into your analysis (page 9-2)

How can incorporating time into your data improve your analysis results?

**A possible response includes the following: *When* things occur could change the way that you approach decision making. For example, you may not give incidents that happened in a location five years ago much attention, but you could focus on recent incidents.**

## Exercise 9A: Explore data (page 9-5)

1. Which attribute can you use to analyze the temporal aspect of the tornado points?

**You can use the Date attribute.**

2. What patterns can you determine about tornadoes in Texas?

**The main observation is that the data is noisy and you cannot determine much, but ArcGIS Pro has tools that can quantify patterns and make sense of the noise.**

3. How can using 70 years' worth of data influence a hot spot analysis?

**Hot spot analysis will show clusters, but what you will not see are the possible changes in those hot or cold spots over time.**





Most of the GIS analyses that you have performed in the course (including proximity, overlay, and statistical hot spot analyses) determine where phenomena are located. You may want to further explore the observed spatial patterns to determine why a phenomenon is occurring. When you understand what contributes to a phenomenon's occurring, you are better equipped to offer data-driven solutions to mitigate or prevent a problem.

In this lesson, you will learn about regression—how to use it for explanatory analysis and some of the statistics that help you locate a valid model. You will also learn a workflow for finding the best regression model, and use ArcGIS Pro to perform ordinary least squares (OLS) regression.

## Topics covered

---

What is regression?

Regression equation

Ordinary least squares (OLS) regression

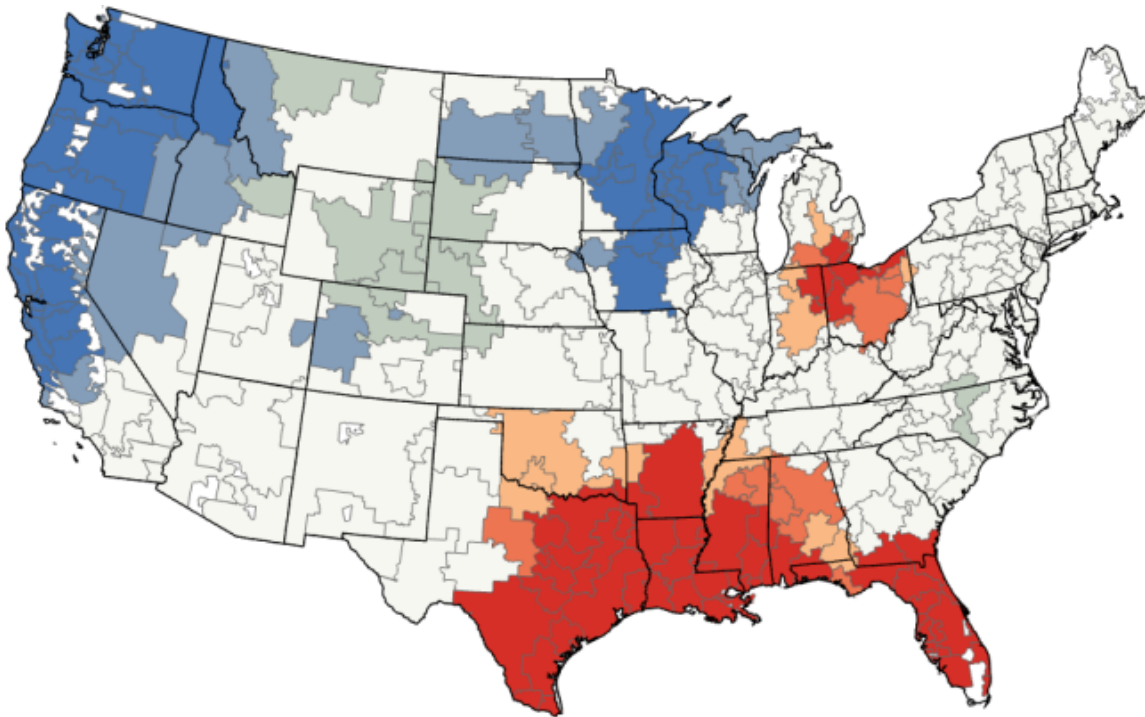
Six statistical checks

Finding a properly specified regression model

## Explaining spatial patterns

---

Most GIS analysis results tell you where something is located—for example, the best site for a new store, spatial clustering of various phenomena, or areas where a specific disease is more prominent than other diseases. Knowing where something occurs is beneficial, but understanding what contributes to spatial phenomena can help solve a problem. ArcGIS Pro provides statistical analysis tools that allow you to model spatial relationships to help you understand the factors causing spatial patterns and predict future patterns based on current data and trends.



**Figure 10.1.** In this map, optimized hot spot analysis indicates the location of statistically significant hot spots of higher Medicare spending in hospital referral regions. But why might this spatial pattern exist?

When you look at this map of Medicare spending, you may wonder why there is a hot spot for spending in the southern and southeastern states. ArcGIS Pro can help identify factors that contribute to or cause spatial patterns.

## Causes of spatial patterns

---

You will consider two situations to determine the factors that might have caused certain phenomena.

### Scenario 1: Medicare spending

Your data indicates that Medicare spending in a particular region is higher than expected. You want to analyze the patterns and identify possible causes of these high Medicare costs.

- 1. What factors might contribute to higher Medicare spending?**
- 

### Scenario 2: Graffiti

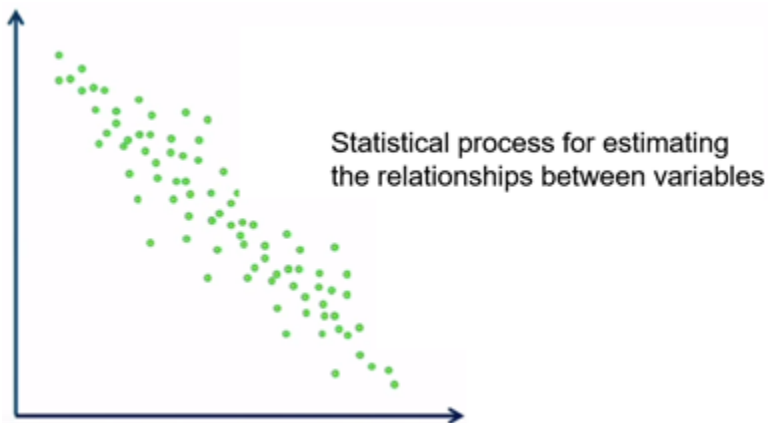
In your job as a GIS analyst for a police department, you notice a high volume of graffiti in the city. You want to determine the possible causes of graffiti in certain areas.

- 2. What factors might contribute to more graffiti in certain areas of a city?**
-

## What is regression?

---

Regression is a statistical method for evaluating the relationship between variables. By evaluating that relationship, it is possible to hypothesize about the causes of a pattern. Using regression analysis, you can model, examine, and explore spatial relationships to better understand factors behind spatial patterns. When you have quantified the factors that contribute to a phenomenon, you can make better decisions.



**Figure 10.2.** Regression evaluates relationships between variables.

You can perform several types of regression analysis in ArcGIS Pro:

- **Continuous (Gaussian):** The variable that you are modeling is continuous. This model performs ordinary least squares (OLS) regression.
- **Binary (logistic):** The variable that you are modeling represents presence or absence. This can be either conventional 1s and 0s, or continuous data that has been recoded based on some threshold value.
- **Count (Poisson):** The variable that you are modeling is discrete and represents events (for example, crime counts, disease incidents, or traffic accidents).

## What is regression? (continued)

---

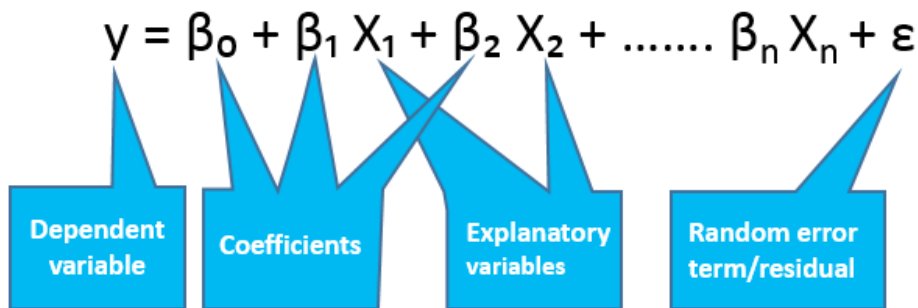
### Benefits of regression

- Explore correlations: Does higher Medicare spending translate to better health or better-quality health care?
- Predict unknown values: How many claims for heat-related illness are expected given current weather forecasts?
- Understand or explain key factors that contribute to a process: Why are test scores higher in certain parts of the country?

When you understand key factors that contribute to a process, you can be confident that the relationships that you find are real, and you can use information to guide decision making.

## Regression equation

The regression equation is the core of regression analysis. It provides a context for understanding terms used in regression analysis. ArcGIS Pro regression tools create the equation based on parameters that you set for those tools. In regression analysis, there is a dependent variable and one or more independent variables thought to influence or contribute to the dependent variable. Regression is used to predict the value of the dependent variable that you are trying to model or to determine the degree to which an independent variable is important to your model.



**Figure 10.3.** The OLS regression equation. The equation assumes that all relationships are linear.

**Dependent variable:** The variable representing the process being predicted or modeled, such as test scores, foreclosures, or Medicare spending. The dependent variable is also called the response variable.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

**Figure 10.4.** Dependent variable.

**Independent variable:** One variable or a set of variables used to explain or predict the dependent variable values. Independent variables are often called explanatory variables.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

**Figure 10.5.** Independent variables attempt to explain the dependent variable.

**Coefficients:** Values associated with each independent variable in a regression equation, representing the strength and direction of the relationship between the independent and the dependent variable. Coefficients indicate the strength of the relationship; the larger the

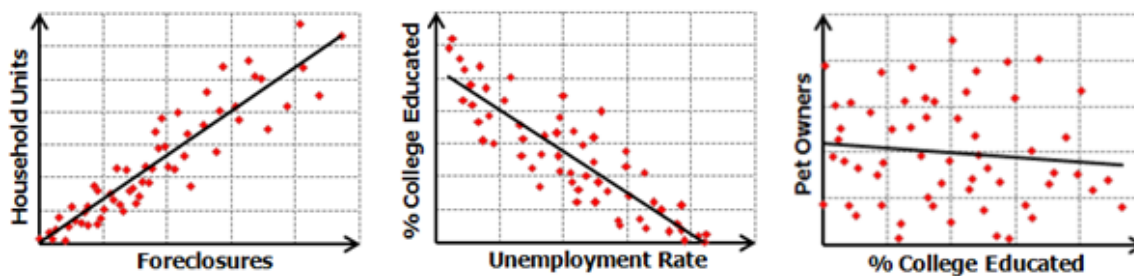
## Regression equation (continued)

coefficient (relative to the units of the independent variable that it is associated with), the stronger the relationship.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

**Figure 10.6.** Each independent variable has a coefficient. For example,  $\text{spending} = b_0 + b_1(\text{distance}) + b_2(\text{imaging events}) + b_3(\text{hospital beds}) + e$

Coefficients can indicate positive, negative, or no relationship between the dependent and independent variables.



**Figure 10.7.** Scatter plot: A scatter plot is a type of mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. In each example, there is one dependent variable and one independent variable (univariate).

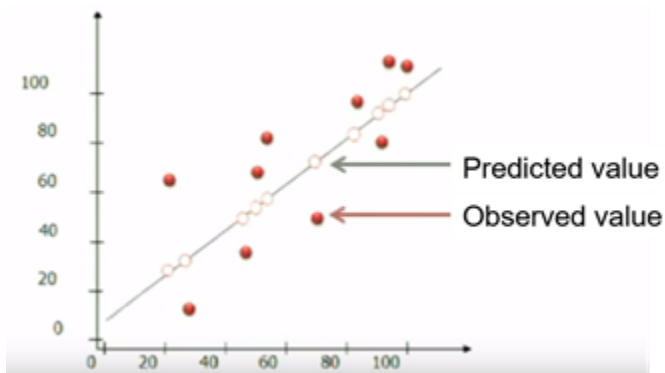
The equation also has a y-intercept. The y-intercept ( $b_0$ ) is the expected value for the dependent variable if all the independent variables are zero.

- |                 |  |
|-----------------|--|
| <b>Positive</b> | As the independent variable value increases, so does the dependent variable's value. For example, the positive relationship in the chart shows that as the number of household units on the y-axis increases, so do foreclosures (x-axis). |
| <b>Negative</b> | As the value of the independent variable increases, the dependent variable's value decreases. For example, as the percentage of college educated people increases (y-axis), the unemployment rate (x-axis) decreases.                      |
| <b>None</b>     | A flat line in a scatter plot indicates no relationship.   |

## Regression equation (continued)

---

**Residual:** The over- and under-predictions (errors) in the model, or the differences between actual observed values and predicted values.



**Figure 10.8.** Distribution of the residuals can indicate whether you have found all key variables. The magnitude of the difference between the observed and predicted values is one measure of model fit.

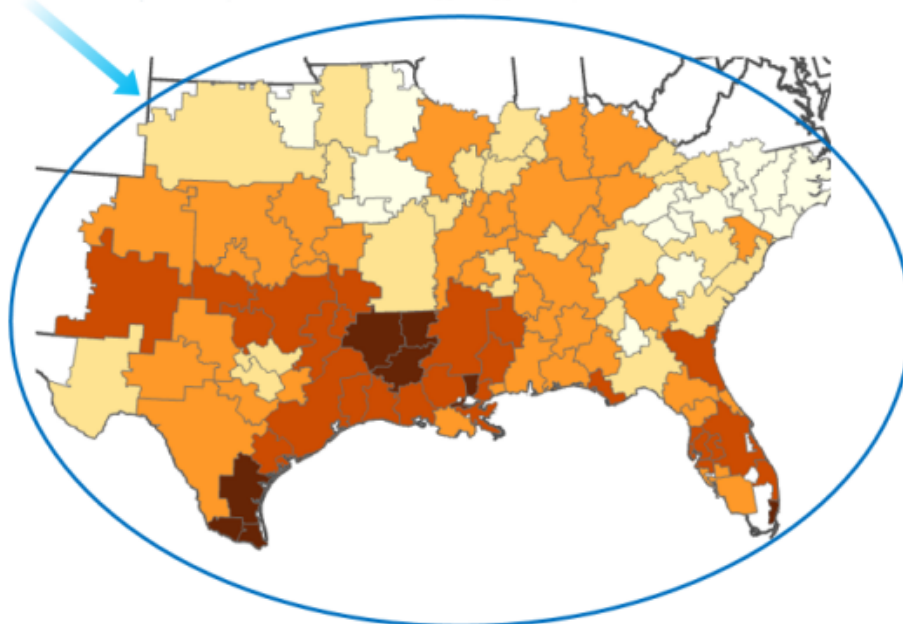


## OLS regression

---

OLS is the best known of all regression techniques, and you can access it in the Generalized Linear Regression (GLR) tool using the continuous (Gaussian) model type. OLS is widely used outside GIS, and it is the proper start point for all spatial regression analyses. It provides a global model of the dependent variable or the process that you are trying to explain or predict. Global means that a single regression equation to represent a process is applied to all the features in the study area, thus assuming that relationships are fixed. You evaluate the OLS summary that contains various diagnostics, including how well the model is performing and how each independent variable is helping the model.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$



**Figure 10.9.** OLS is a global regression model that applies one regression equation to all features.

Being a global regression model, OLS creates one equation. Each variable has a single coefficient, and the relationships between data variables are fixed across geographic space. This process is referred to as stationarity. OLS is global and assumes stationarity, meaning you could move all the points to different locations and the regression equation would be the same. Another type of regression analysis that you will use later accounts for spatial variation in your variable's relationships (nonstationarity).

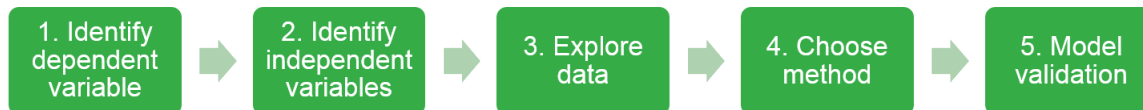
## OLS regression (continued)

---

### OLS workflow for explanatory analysis

The regression workflow can be described as follows:

1. Identify the process that you want to explain or predict, as well as the data variable that represents it.
2. Select variables that represent the factors influencing the process.
3. Explore and analyze data (descriptive stats, univariate, bivariate).
4. Choose the method (for example, OLS) and specify the model based on what you learned about data relationships in step 3.
5. Validate and evaluate the model; perform six checks.



**Figure 10.10.** The regression workflow is iterative and can require a lot of work to properly specify an OLS model.

Depending on what happens in step 5, you may have several different options:

- Try new data variables or transform existing ones.
- Turn to a spatial regression method.
- Start over entirely.

## Checkpoint

---

1. Which of the following options describes the dependent variable in the regression equation?
  - a. Represents the strength and type of relationship between phenomena
  - b. The process that you want to predict or model
  - c. A factor contributing to a process
  - d. The over- and under-predictions in the model
2. Which statement about OLS regression is correct?
  - a. OLS is used primarily to locate clusters or hot spots.
  - b. OLS is a local regression model.
  - c. OLS attempts to explain which variables explain a phenomenon, and to what degree.
  - d. OLS accounts for spatial variation in a variable's relationships.

## Interpreting OLS diagnostics

---

Performing OLS regression analysis is more in depth than running a tool. You must evaluate the statistical results to determine if the variables that you selected explain the variance in the dependent variable. If your variables meet all six requirements, then you have found a properly specified model. Watch the video to see how to interpret the OLS results, and then answer the following questions.

1. **What should you do if the probability associated with a coefficient is not statistically significant?**

---

2. **What is the adjusted R-squared statistic, and how does it indicate model performance?**

---

---

3. **What does it mean if the OLS residuals are spatially clustered? What should you do to solve the problem?**

---

---

## Interpreting OLS diagnostics (continued)

---

When you perform OLS regression, you must ensure that your model passes the six checks. In this video, you learned how to find the results and to perform the six OLS checks. A model that passes all six checks is properly specified.

```

Summary of OLS Results
Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE Robust_t Robust_Pr [b] VIF [c]
Intercept -22.530277 35.134815 -0.641252 0.523394 35.357649 -0.637211 0.526008 -----
HS_COUNT 16.176920 6.796062 2.380337 0.019942* 7.270324 2.225062 0.029207* 1.516463
TOTPOP_CY -0.001897 0.000975 -1.945521 0.055617 0.001184 -1.602205 0.113494 2.402808
ACSSNAP 0.129231 0.012140 10.644910 0.000000* 0.016204 7.975317 0.000000* 3.103987
COLLEGE_P 0.277226 0.884232 0.313522 0.754795 1.053614 0.263119 0.793213 1.694377
OLS Diagnostics
Input Features: Analysis Layer Dependent Variable: JOIN_COUNT
Number of Observations: 77 Akaike's Information Criterion (AICc) [d]: 972.174925
Multiple R-Squared [d]: 0.828914 Adjusted R-Squared [d]: 0.819410
Joint F-Statistic [e]: 87.210520 Prob(>F), (4,72) degrees of freedom: 0.000000*
Joint Wald Statistic [e]: 201.354183 Prob(>chi-squared), (4) degrees of freedom: 0.000000*
Koenker (BP) Statistic [f]: 18.042309 Prob(>chi-squared), (4) degrees of freedom: 0.001211*
Jarque-Bera Statistic [g]: 8.431853 Prob(>chi-squared), (2) degrees of freedom: 0.014759*

```

Figure 10.11. OLS report.

## Six OLS checks

---

The workflow for regression is to locate your dependent and independent variables, run OLS, and then perform checks on the statistics in the OLS report. The six checks should be performed to determine whether the variables result in a usable model. Your goal is to find a properly specified model, or one that you can trust to explain the process represented by your dependent variable. After you run OLS, you manually perform the six checks in any order.

Check	Description
<b>1. Do coefficients have expected relationship?</b>	When the sign associated with the coefficient is negative, the relationship is negative (for example, the larger the distance from the urban core, the smaller the number of residential burglaries). When the sign is positive, the relationship is positive (for example, the larger the population, the larger the number of residential burglaries).
<b>2. Ensure that each independent variable is statistically significant.</b>	Statistically significant coefficients, indicated by an asterisk (*), are important to the model. The absence of an asterisk could mean that the variable is not significant in modeling the dependent variable. OLS performs a statistical test to compute a probability (p-value) for each coefficient. The null hypothesis is that the variable is not helping the model. Small p-values reflect small probabilities and suggest that the coefficient is important to the model.
<b>3. Residuals should not be clustered in location or in value.</b>	After you run OLS, you will see a message in the geoprocessing results suggesting that you run the Moran's I tool to ensure that your residuals are not spatially autocorrelated. Statistically significant spatial autocorrelation (clustering of residuals) can be a symptom of misspecification, which is the wrong type of regression model. It occurs when one or more key variables are missing.

## Six OLS checks (continued)

Check	Description
<p><b>4. Verify that residuals are normally distributed using the Jarque-Bera test.</b></p>	<p>A properly specified model has residuals that are normally distributed with a mean of zero. One example of a biased model is one that does a good job of predicting high values but performs poorly when predicting low values. A biased model might be the result of outliers within or non-linear relationships between the data variables. If the Jarque-Bera statistic (test) is statistically significant (it has an asterisk next to the p-value), the model is biased and you cannot trust the model.</p>
<p><b>5. Are all VIF values lower than 7.5?</b></p>	<p>The variance inflation factor (VIF) should be less than 7.5. A VIF over 7.5 for an independent variable indicates variable redundancy (multicollinearity). At least one of the variables with a VIF above 7.5 should be removed.</p>
<p><b>6. Evaluate model performance.</b></p>	<p>Adjusted R-squared is a statistic derived from the regression equation to quantify model performance (how well explanatory variables model the dependent variable). Adjusted R-squared is the percentage of variance in the dependent variable that is explained by variance in the independent variables. Akaike's information criterion (AIC) is an estimator of the relative quality of statistical models for a set of data, and is useful for comparing multiple models that use the same dependent variable. A higher adjusted R-squared value and a lower AIC value for the same dependent variable in different models indicate a better model.</p>

## Six OLS checks (continued)

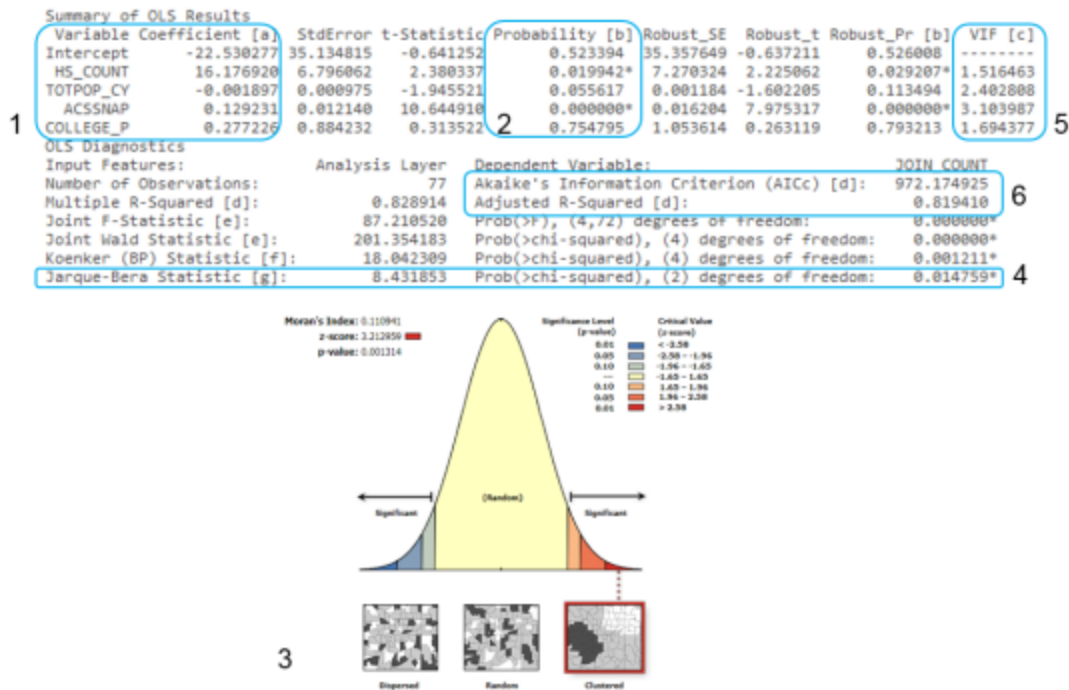


Figure 10.12. Match the number in the table with the number in the diagram to see where each of the six checks is in the OLS diagnostic report.



**ArcGIS Pro Help: Regression analysis basics**

**ArcGIS Pro Help: What they don't tell you about regression analysis**



## OLS reports

---

Examine the following OLS reports that use the same dependent variable yet with different combinations of independent variables, and answer the questions.

### Scenario 1: Removing variables

```

Summary of OLS Results
Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE Robust_t Robust_Pr [b] VIF [c]
Intercept -31.161352 40.485266 -0.769696 0.443994 44.340190 -0.702779 0.484457 -----
LIQ_COUNT -9.218742 1.742996 -5.289022 0.000001* 2.717869 -3.391900 0.001135* 3.342344
VACANT_CY 0.157268 0.020546 7.654436 0.000000* 0.033464 4.699612 0.000013* 2.744522
RENTER_CY -0.003765 0.007990 -0.471219 0.638914 0.009262 -0.406475 0.685605 8.203650
TOTPOP_CY 0.005321 0.001734 3.068299 0.003036* 0.002140 2.486879 0.015202* 3.798446
OLS Diagnostics
Input Features: Analysis Layer Dependent Variable: JOIN_COUNT
Number of Observations: 77 Akaike's Information Criterion (AICc) [d]: 1025.576000
Multiple R-Squared [d]: 0.657701 Adjusted R-Squared [d]: 0.638685
Joint F-Statistic [e]: 34.585627 Prob(>F), (4,72) degrees of freedom: 0.000000*
Joint Wald Statistic [e]: 61.729396 Prob(>chi-squared), (4) degrees of freedom: 0.000000*
Koenker (BP) Statistic [f]: 27.676150 Prob(>chi-squared), (4) degrees of freedom: 0.000015*
Jarque-Bera Statistic [g]: 15.680360 Prob(>chi-squared), (2) degrees of freedom: 0.000394*

```

Figure 10.13. OLS report.

#### 1. Should you remove any variables from this model? Why?

---



---

## OLS reports (continued)

### Scenario 2: Model performance

Summary of OLS Results										
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]		
Intercept	73.927486	31.577301	2.341159	0.022032*	28.209599	2.620650	0.010725*	-----		
HS_COUNT	16.779943	5.497572	3.052246	0.003197*	5.509878	3.045429	0.003261*	1.498258		
LIQ_COUNT	-0.932117	0.915857	-1.017754	0.312246	0.797908	-1.168202	0.246632	2.787620		
DIVINDX_CY	-2.368100	0.448616	-5.278680	0.000002*	0.396909	-5.966352	0.000000*	1.136477		
VACANT_CY	0.033591	0.014816	2.267251	0.026418*	0.013401	2.506642	0.014476*	4.310877		
ACSSNAP	0.097357	0.009531	10.214353	0.000000*	0.012230	7.960264	0.000000*	2.888744		
OLS Diagnostics										
Input Features:		Analysis Layer		Dependent Variable:				JOIN_COUNT		
Number of Observations:		77		Akaike's Information Criterion (AICc) [d]:	941.797496					
Multiple R-Squared [d]:	0.888259			Adjusted R-Squared [d]:	0.880390					
Joint F-Statistic [e]:	112.879839			Prob(>F), (5,71) degrees of freedom:	0.000000*					
Joint Wald Statistic [e]:	270.939955			Prob(>chi-squared), (5) degrees of freedom:	0.000000*					
Koenker (BP) Statistic [f]:	26.018961			Prob(>chi-squared), (5) degrees of freedom:	0.000088*					
Jarque-Bera Statistic [g]:	9.084599			Prob(>chi-squared), (2) degrees of freedom:	0.010649*					

Summary of OLS Results										
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]		
Intercept	-31.161352	40.485266	-0.769696	0.443994	44.340190	-0.702779	0.484457	-----		
LIQ_COUNT	-9.218742	1.742996	-5.289022	0.000001*	2.717869	-3.391900	0.001135*	3.342344		
VACANT_CY	0.157268	0.020546	7.654436	0.000000*	0.033464	4.699612	0.000013*	2.744522		
RENTER_CY	-0.003765	0.007990	-0.471219	0.638914	0.009262	-0.406475	0.685605	8.203650		
TOTPOP_CY	0.005321	0.001734	3.068299	0.003036*	0.002140	2.486879	0.015202*	3.798446		
OLS Diagnostics										
Input Features:		Analysis Layer		Dependent Variable:				JOIN_COUNT		
Number of Observations:		77		Akaike's Information Criterion (AICc) [d]:	1025.576000					
Multiple R-Squared [d]:	0.657701			Adjusted R-Squared [d]:	0.638685					
Joint F-Statistic [e]:	34.585627			Prob(>F), (4,72) degrees of freedom:	0.000000*					
Joint Wald Statistic [e]:	61.729396			Prob(>chi-squared), (4) degrees of freedom:	0.000000*					
Koenker (BP) Statistic [f]:	27.676150			Prob(>chi-squared), (4) degrees of freedom:	0.000015*					
Jarque-Bera Statistic [g]:	15.680360			Prob(>chi-squared), (2) degrees of freedom:	0.000394*					

Figure 10.14. Compare OLS reports.

- These OLS reports are for modeling total crime in a city but use different independent variables. Which model is better and why?

---



---



---

## Exploratory regression

---

Exploratory regression evaluates all possible combinations of independent variables and automatically selects the models that meet the requirements and assumptions of OLS. Several of the six checks are performed during exploratory regression. When you perform exploratory regression, do not randomly add all variables and then try to pick the best combination. Base your independent variables on logical reasoning and research. Exploratory regression gives you an idea of good variable combinations, but you should still perform OLS and your statistical checks.

### Exploratory regression trade-offs

Most regression analyses use the deductive approach (questions > hypothesis > data > analysis > result). Exploratory regression uses a data-driven, inductive approach (data > questions > hypothesis > analysis > result). On one hand, mining data with exploratory regression reveals new data relationships and allows you to generate new questions and hypotheses. On the other hand, it might result in data overfitting, or report models that meet the assumptions yet do not actually reflect real-world processes.

## Exploratory regression (continued)

---



**Figure 10.15.** The goal of exploratory regression is finding properly specified OLS models, given the variables that you provide.



**ArcGIS Pro Help:** *How Exploratory Regression works*

## Find a properly specified regression model

You will use OLS regression to analyze Medicare spending and explain the factors that contribute to higher spending.

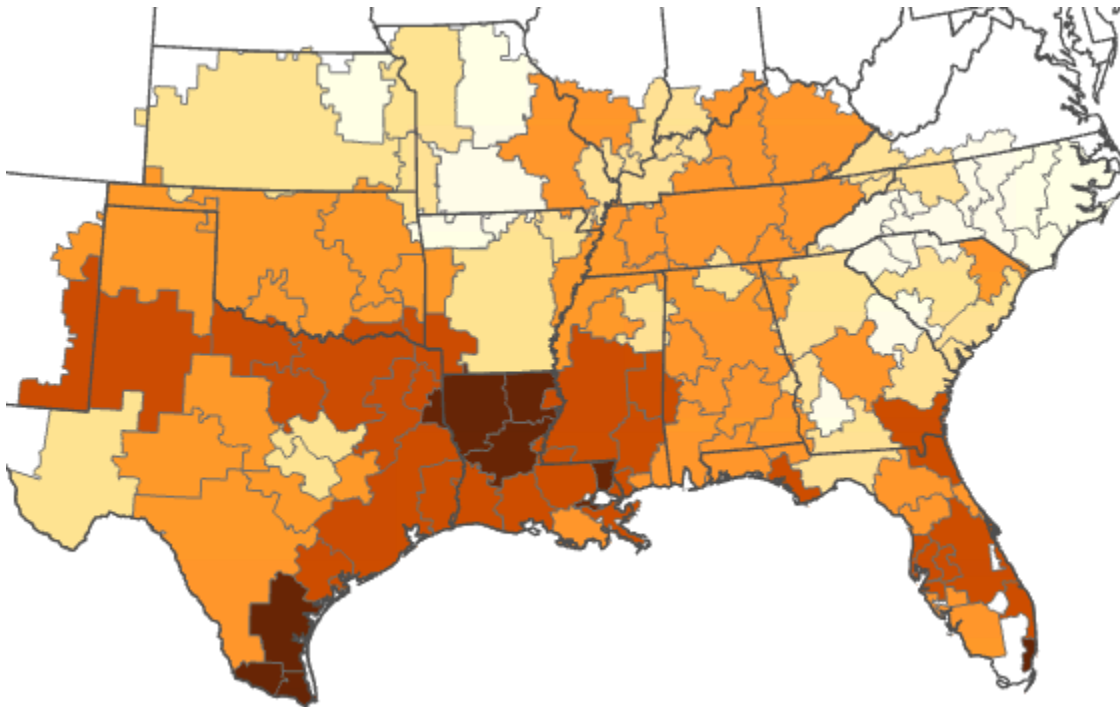
In this exercise, you will perform the following tasks:

- Run OLS regression.
- Perform the six statistical checks.

## Step 1: Set up ArcGIS Pro

First, you will set up the project for the exercise.

- a If necessary, start ArcGIS Pro and open the SNAPCourse project.
- b From the Insert tab, click Import Map.
- c Browse to **C:\EsriTraining\SNAP\Regression** and add Regression Analysis.mapx to the project.



The map is symbolized using graduated colors based on the total Medicare spending per hospital referral region (HRR). Lighter colors indicate less spending, and darker colors indicate more spending. You will use OLS regression to analyze potential causes for higher Medicare spending.

- d Save the project.

## Step 2: Perform exploratory data analysis

You will perform exploratory data analysis by looking at available attributes and creating a scatter plot to show the relationship among variables. To perform OLS regression, you must have key variables to analyze. The variables are attributes within the layer's table.

- a In the Contents pane, open the Study Area attribute table.

The Total Costs 2010 variable contains the total cost of Medicare spending per HRR. According to the regression workflow, the process or phenomenon that you are trying to understand is Medicare spending costs, so Total Costs 2010 will be the dependent variable. Other attributes, such as number of hospital beds, readmission rate, and number of emergency visits, may be independent variables that explain costs.

**b** Scroll through the table to view the attributes, and then close it.

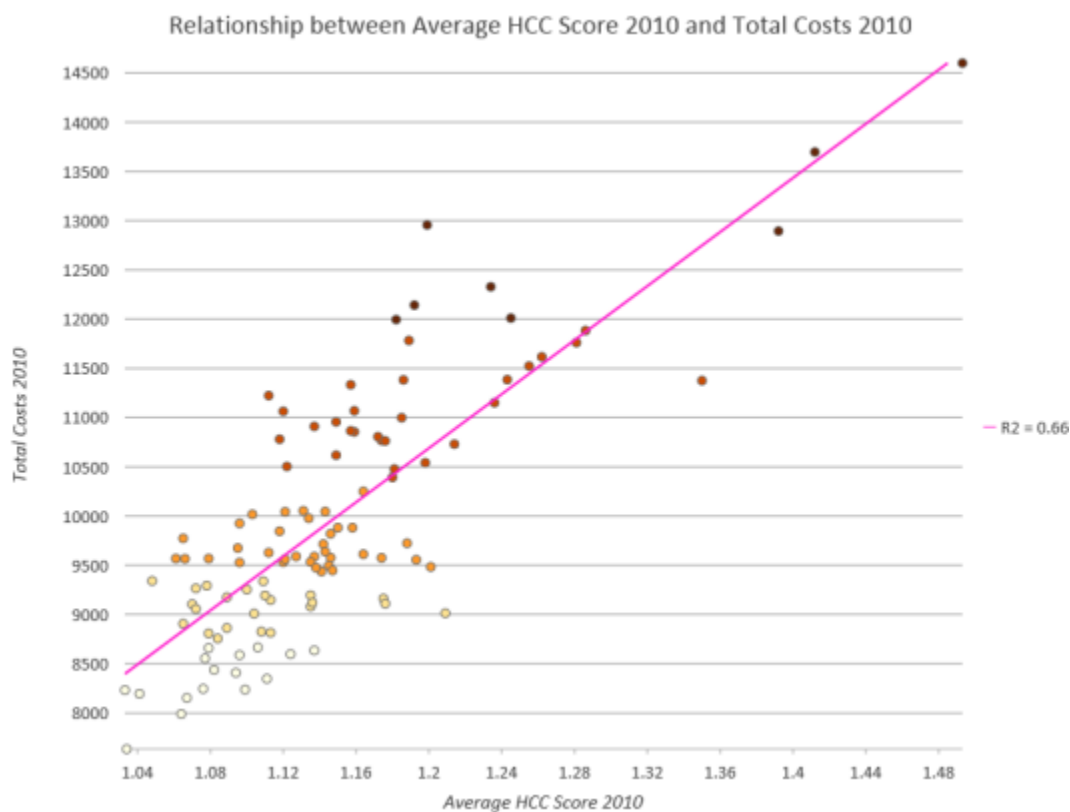
In your first analysis, you will test how the hierarchical condition category (HCC) score (a measure of a population's overall health) contributes to spending. The HCC score measures the prevalence of chronic health conditions. Before you use HCC as the independent variable to help explain total costs, you will create a scatter plot to explore the relationships between total costs and HCC score.

**c** In the Contents pane, verify that Study Area is selected.

**d** From the Data tab, click Create Chart, and then choose Scatter Plot.

**e** In the Chart Properties pane, for X-Axis Number, choose Average HCC Score 2010.

**f** For Y-Axis Number, choose Total Costs 2010.



There is a positive relationship between the HCC score and total costs (the HCC score increases as total costs increase). The R-squared of 0.66 indicates that the HCC score explains 66 percent of

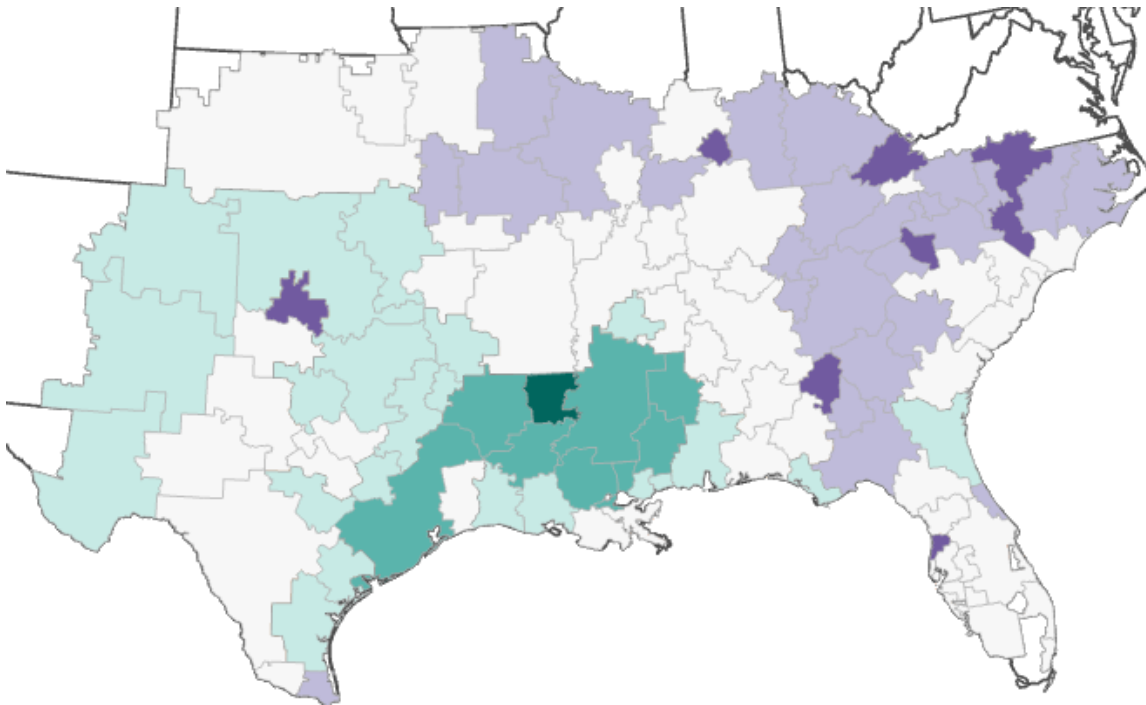
the variance in total costs, a solid value. With an R-squared value of 0.66, you may want to include the HCC score variable in your OLS model. Although a scatter plot shows you the relationship and an R-squared value, you cannot view and evaluate other key diagnostics to determine if the model is properly specified.

**g** Close the chart and the Chart Properties pane.

### Step 3: Use the Generalized Linear Regression tool to test for higher spending factors

In this step, you will test how well the HCC index score explains higher spending.

- a** From the Geoprocessing pane, search for **generalized**.
- b** Open the Generalized Linear Regression (GLR) tool, and set the following parameters:
  - Input Features: Study Area
  - Dependent Variable: Total Costs 2010
  - Model Type: Continuous (Gaussian)
  - Explanatory Variables: Average HCC Score 2010
  - Output Features: **GLRContinuous**
- c** Click Run.







In your map, the States layer may be covered by the GLRContinuous layer.

The results of the GLR tool are displayed in the map and symbolized using the standardized residuals. After running regression, evaluate the report to see how well the model performed and explains variance in the dependent variable.

- d At the bottom of the Geoprocessing pane, in the green box, click View Details.
- e Expand the Generalized Linear Regression report, both horizontally and vertically, so that you can see the entire report.

```
----- Summary of GLR Results [Model Type: Continuous (Gaussian/OLS)] -----
-----
Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE Robust t Robust_Pr [b]
Intercept -5780.317179 1089.153986 -5.307162 0.000001* 877.090014 -6.590335 0.000000*
AVERAGE_HCC_SCORE_2010 13725.329285 948.497894 14.470595 0.000000* 773.693345 17.740012 0.000000*
-----
----- GLR Diagnostics -----
-----
Input Features: Study Area Dependent Variable: TOTAL_COSTS_2010
Number of Observations: 110 Akaike's Information Criterion (AICc) [d]: 1770.650750
Multiple R-Squared [d]: 0.659733 Adjusted R-Squared [d]: 0.656583
Joint F-Statistic [e]: 209.398132 Prob(>F), (1,108) degrees of freedom: 0.000000*
Joint Wald Statistic [e]: 314.708028 Prob(>chi-squared), (1) degrees of freedom: 0.000000*
Koenker (BP) Statistic [f]: 1.042391 Prob(>chi-squared), (1) degrees of freedom: 0.307266
Jarque-Bera Statistic [g]: 2.909458 Prob(>chi-squared), (2) degrees of freedom: 0.233464
-----
```

1. What does the adjusted R-squared value tell you about HCC score and Medicare spending?

---

2. What does the AIC value of 1770 tell you about the HCC score and Medicare spending?

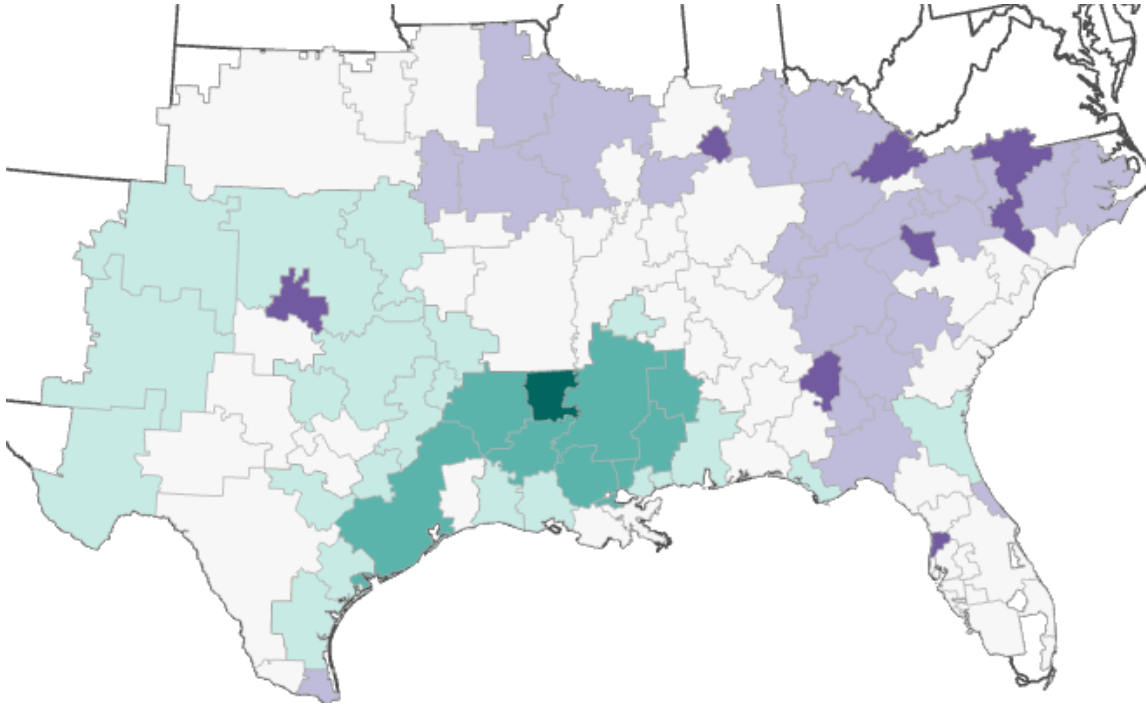
---

An adjusted R-squared of 0.65 is suitable. If this model was properly specified, it would explain about 65 percent of the variation in Medicare spending. However, HCC index scores may not be the only part of the Medicare spending story in this area. Several other OLS regression assumptions should be met before you have a properly specified model. You will wait until you have created the full OLS model using several independent variables to do the six checks, but first, you will explore the spatial output of OLS.

- f Close the report.

## Step 4: Evaluate the spatial output from the GLR tool

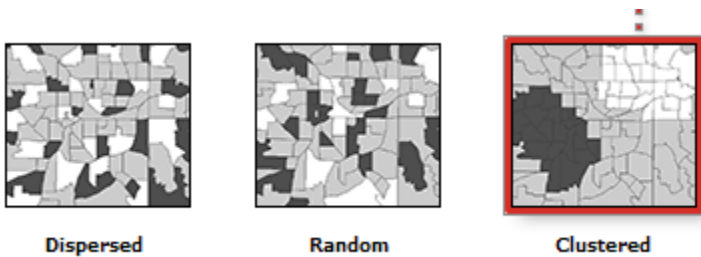
OLS provides an in-depth report of the regression diagnostics and also provides an output layer of the residuals. Residuals are the over- and under-predictions in the model. If you recall from the video, the OLS residuals should be spatially random in a properly specified regression model.



### 3. What can you extrapolate about the residuals from the map?

You can use the Spatial Autocorrelation tool to validate your visual analysis of the residuals.

- a In the Geoprocessing pane, click the Back button.
- b Search for and open the **Spatial Autocorrelation** tool, and set the following parameters:
  - Input Feature Class: GLRContinuous
  - Input Field: Std Residual
  - Generate Report: Check the box
- c Click Run.
- d If necessary, in File Explorer, browse to `..\EsriTraining\SNAP\SNAPCourse` and open the most recent Moran's I report.



Spatial autocorrelation validates visual observation of residual clustering, which suggests that your regression model may be missing key variables.

- e Close the report, and minimize File Explorer.

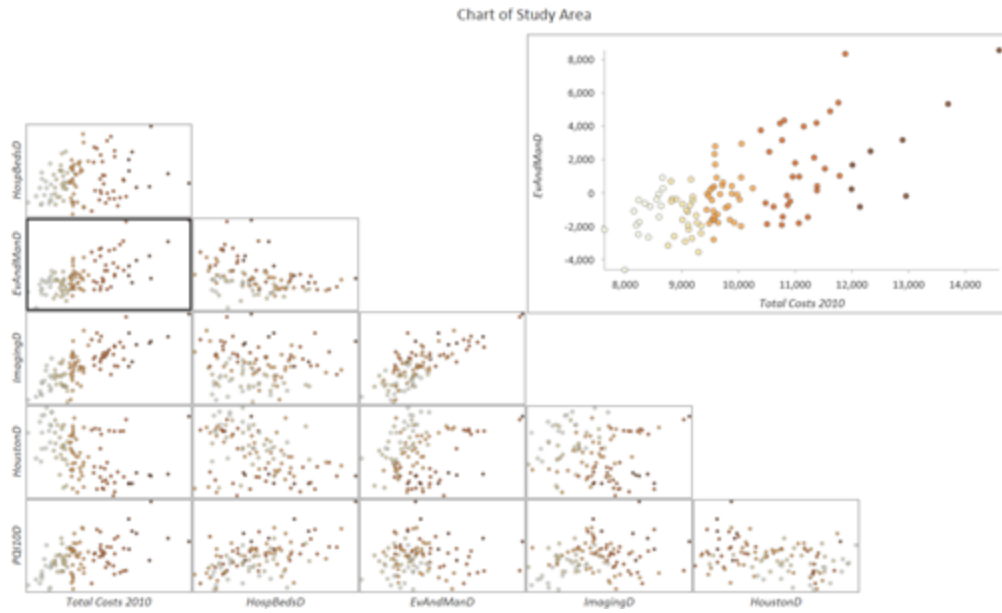
## Step 5: Create a scatter plot matrix

Next, you will create a scatter plot matrix to visualize the variable relationships as a means to choose the best variables for regression analysis. You have an idea or theory as to which factors contribute to higher Medicare spending and want to visualize those relationships. A scatter plot matrix is a good exploratory step before running regression on multiple variables. Like any analysis, it is good to know the data before proceeding.

- a In the Contents pane, select StudyArea.
- b From the Data tab, click Create Chart and choose Scatter Plot Matrix.

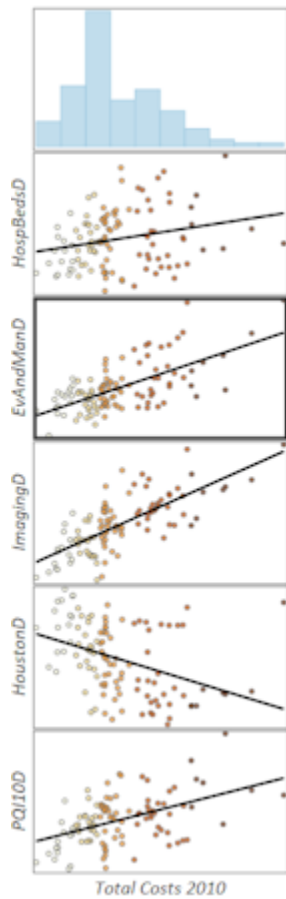
You will use Total Costs 2010 as the dependent variable because you want to determine the factors that contribute to it. You will choose several independent variables that previous research and theory have indicated are strong factors that contribute to higher Medicare spending. You will analyze the number of hospital beds, evaluation and management costs, total imaging events (MRI, CAT scan), distance to Houston, and dehydration rates.

- c In the Chart Properties pane, for Numeric Fields, select the following options:
  - Total Costs 2010
  - HospBedsD
  - EvAndManD
  - ImagingD
  - HoustonD
  - PQI10D
- d Expand the scatter plot matrix view vertically.



You will focus on the column on the far left that has the dependent variable of Total Costs 2010 on the y-axis and the dependent variables on the x-axis.

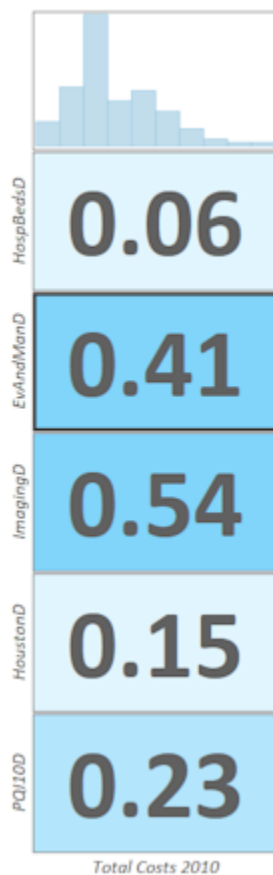
- e In the Chart Properties pane, check the Show Histogram and Show Linear Trend boxes.



The histogram shows the distribution of the variables, and the linear trends indicate positive, negative, or no relationship between the variables.

For the most part, the relationships between all the variables and Total Costs 2010 are positive. A positive relationship makes sense because as there are more hospital beds, evaluations and management of a hospital, and imaging events; a greater distance from Houston; and more dehydration, there will be more costs. These relationships are expected and indicate that these variables may be strong factors in higher costs.

**f** In the Chart Properties pane, check the Show As R2 box.



The  $R^2$ , or adjusted R-squared, is an indicator of how well the independent variables model the dependent variable. So, how well do the five variables chosen explain total Medicare costs? Viewing the adjusted R-squared value is another exploratory measure that you can perform before you run a regression tool.

**g** In the Chart Properties pane, uncheck Show As R2.

You can also look at the relationships between independent variables. A strong positive correlation indicates that the variables may be telling the same part of the story (multicollinearity). You could always remove independent variables that tell the same story to reduce redundancy.

- h Close the scatter plot matrix.

## Step 6: Run the GLR tool on multiple dependent variables

Next, you will create a properly specified regression model. The independent variables selected for this model are a result of the iterative process of running OLS many times to find the best combination of variables.

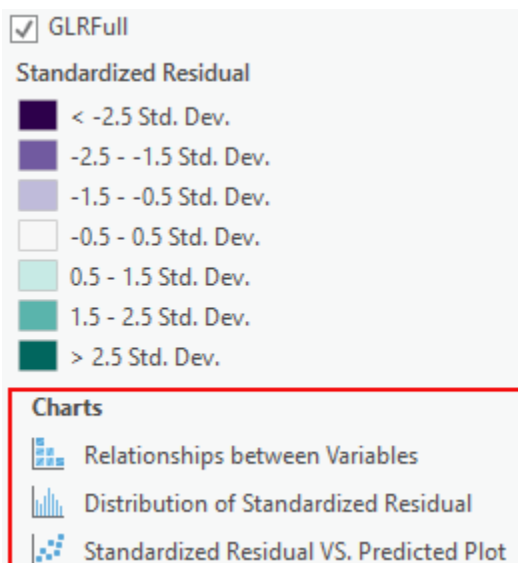
- a At the top of the Catalog pane, click the History tab.
- b Double-click the Generalized Linear Regression tool that you ran earlier.
- c Modify only the following parameters (leaving the others as they are):
  - Explanatory Variables: HospBedsD, EvAndManD, ImagingD, HoustonD, and PQI10D (ensure that Average HCC Score 2010 is unchecked)
  - Output Feature Class: **GLRFull**
- d Click Run.

The spatial output is added to the map. Next, you will use the results to perform the six checks and validate the model.

## Step 7: Perform OLS checks

You will use the statistical results of the GLR tool and some charts to properly specify your regression model.

- a In the Contents pane, find the charts created with the output from the GLR tool.



- b** Double-click each chart to view the results.

The Relationships Between Variables chart shows a scatter plot matrix similar to the one that you viewed before you ran the GLR tool. The Distribution Of Standard Residual chart shows how the residuals from the model compare to a normal distribution. The Standardized Residual VS. Predicted Plot shows the standardized residuals plotted against the standardized predicted values. No patterns should be present if the model fits well. Do you see a pattern in the chart?

- c** Close all open charts and the Chart Properties pane.
- d** At the bottom of the Geoprocessing pane, click View Details.
- e** Expand the Generalized Linear Regression report so that you can see the entire report.

```
----- Summary of GLR Results [Model Type: Continuous (Gaussian/OLS)] -----
-----
Variable Coefficient [a] StdError t-Statistic Probability [b] Robust_SE Robust_t Robust_Pr [b] VIF [c]
Intercept 9947.038596 44.461950 223.720250 0.000000* 43.232352 230.083219 0.000000* -----
HOSPBEDSD 453.838027 94.410182 4.807088 0.000006* 95.160978 4.769161 0.000007* 1.789789
EVANDMAND 0.399674 0.036395 10.981628 0.000000* 0.041993 9.517526 0.000000* 3.476159
IMAGINGD 0.359023 0.165592 2.168123 0.032424* 0.161761 2.219469 0.028620* 3.044556
HOUSTOND -0.001145 0.000115 -9.999029 0.000000* 0.000104 -10.976894 0.000000* 1.329077
PQI10D 1.041130 0.220645 4.718573 0.000009* 0.213298 4.881106 0.000005* 1.528101
-----
----- GLR Diagnostics -----
-----
Input Features: Study Area Dependent Variable: TOTAL_COSTS_2010
Number of Observations: 110 Akaike's Information Criterion (AICc) [d]: 1672.966945
Multiple R-Squared [d]: 0.870841 Adjusted R-Squared [d]: 0.864631
Joint F-Statistic [e]: 140.241872 Prob(>F), (5,104) degrees of freedom: 0.000000*
Joint Wald Statistic [e]: 556.069919 Prob(>chi-squared), (5) degrees of freedom: 0.000000*
Koenker (BP) Statistic [f]: 27.470483 Prob(>chi-squared), (5) degrees of freedom: 0.000046*
Jarque-Bera Statistic [g]: 1.591597 Prob(>chi-squared), (2) degrees of freedom: 0.451221
-----
```

*Check 1: Are the independent variables helping the model?*

- 4. What does the OLS summary indicate about the statistical significance for each variable?**
- 

- 5. What does it mean when a coefficient's probability has an asterisk next to it?**
- 

*Check 2: Do coefficients have an expected relationship?*

Number of hospital beds shows a positive relationship with Medicare spending costs.

- 6. Should the number of hospital beds contribute to higher spending?**
-

Most the other variables have a positive relationship. For example, higher evaluation and management costs in an HRR, more MRI and other imaging events, and a higher dehydration rate contribute to more spending.

There is one negative relationship: the distance to Houston. The distance to Houston variable measures how far each referral region is from Houston, as Houston has one of the largest medical complexes in the world. This variable is spatial. When you are not finding a properly specified model, including a spatial variable can sometimes help capture the nonstationarity (regional variation) in the data relationships.

*Check 3: Are the independent variables redundant?*

**7. What is the statistical check for determining if variables are redundant?**

---

**8. Are any of the independent variables redundant?**

---

*Check 4: Is the model biased?*

**9. Which of the OLS checks assesses normality in the distribution (nonspatial) of residual values? (*Hint: Look in the GLR Diagnostics section.*)**

---

**10. Is this model biased, and why?**

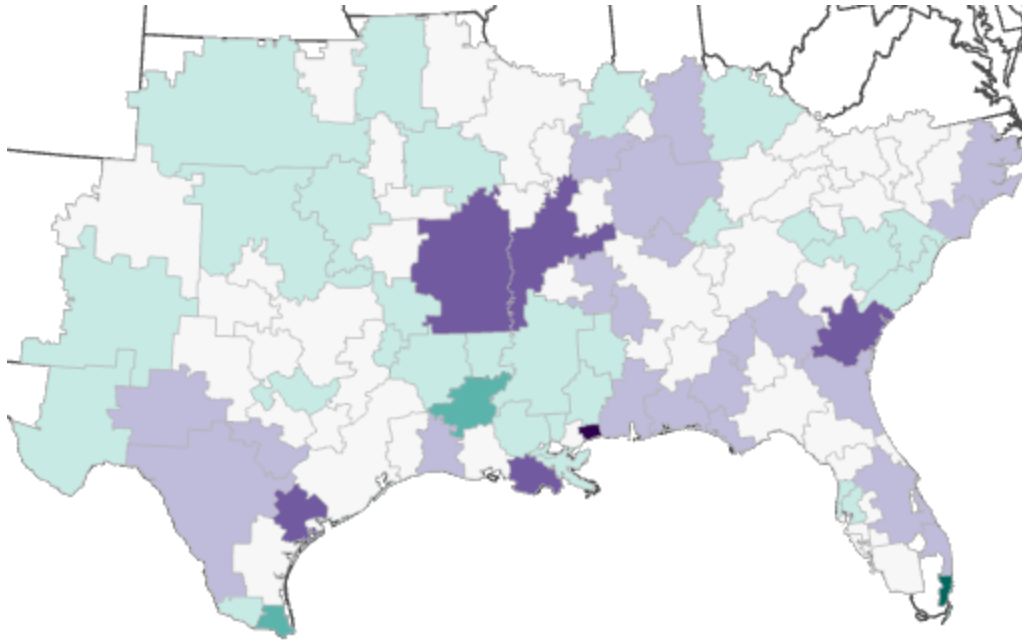
---

*Check 5: Is there spatial clustering in the OLS residuals?*

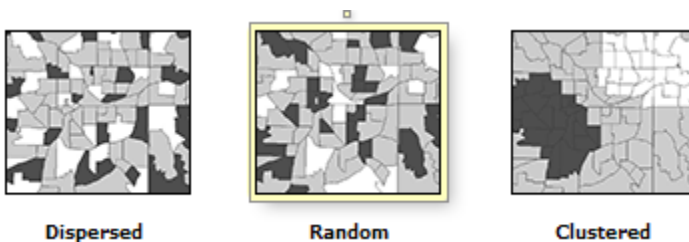
**11. Based on the spatial OLS output of the residuals, do you think that the residuals are clustered?**

---





- f** Leave the Generalized Linear Regression report open.
- g** Search for and open the **Spatial Autocorrelation** tool, and set the following parameters:
  - Input Feature Class: GLRFull
  - Input Field: Std Residual
  - Generate Report: Check the box
- h** Click Run.
- i** If necessary, in File Explorer, browse to `..\EsriTraining\SNAP\SNAPCourse` and open the most recent Moran's I report.



Spatial autocorrelation validates your initial visual analysis of randomly distributed residuals.

- j** Close the Spatial Autocorrelation Report, minimize File Explorer, and return to the Generalized Linear Regression report.

*Check 6: Is the model performing well?*

**12. What are the two main measures of model performance?**

---

**13. What information can you obtain from the AIC and adjusted R-squared values in comparison to the first model that tested only the HCC score variable?**

---

---

---

Based on the six OLS checks, you have a properly specified regression model, or one that you can trust with explaining total Medicare costs in this study area. As you know, OLS is a global regression model that applies one equation to all features and has fixed relationships. What if Medicare spending changed based on location? Most spatial relationships are not static, and you can improve your regression model by incorporating varying relationships. You will work with a different regression tool in the next lesson that incorporates spatial variation of variables.

- k** Leave the Generalized Linear Regression report open, as you will use it in the next exercise.
- i** Save the project and leave ArcGIS Pro open.

## Lesson review

---

1. Explain the key components of the regression equation.

---

---

---

---

2. Explain what the AIC value represents and how to use it.

---

---

---

---

3. You want to perform OLS regression analysis but do not have key independent variables in the attribute table. What can you do to get the required information in the table?

---

---

---

---

---

---

---

## Enriching data for analysis

---

You can enrich your data by adding demographic and landscape facts about the people and places that surround or are inside your data locations. Enriched data comes from demographic data that Esri curates annually and is available through ArcGIS Online or locally installed Business Analyst data. The output feature class from the Enrich tool is a duplicate of your input with new attribute fields added to the table. You can then use the attributes for any operation that uses attributes, including as variables in regression analysis. This tool requires an ArcGIS Online organizational account and consumes credits, or a locally installed Business Analyst dataset.

- ▾  2018 Key Demographic Indicators (Esri)
  - 2018 Total Population
  - 2018 Population in Group Quarters
  - 2018 Total Households
  - 2018 Average Household Size
  - 2018 Total Housing Units
  - 2018 Owner Occupied HUs
  - 2018 Renter Occupied HUs
  - 2018 Vacant Housing Units
  - 2010-2018 Growth Rate: Population
  - 2010-2018 Growth Rate: Households
  - 2010-2018 Growth Rate: Families

**Figure 10.16.** Enrich Layer tool.

# Answers to Lesson 10 questions

## Causes of spatial patterns (page 10-3)

### Scenario 1: Medicare spending

1. What factors might contribute to higher Medicare spending?

**Answers may vary, but some potential causes include overall health of the area, location, climate of area, and demographics.**

### Scenario 2: Graffiti

2. What factors might contribute to more graffiti in certain areas of a city?

**Answers may vary, but some potential causes include time of year, weather, income, and availability of structures.**

# Answers to Lesson 10 questions (continued)

## Checkpoint (page 10-11)

1. Which of the following options describes the dependent variable in the regression equation?
  - a. Represents the strength and type of relationship between phenomena
  - b. The process that you want to predict or model**
  - c. A factor contributing to a process
  - d. The over- and under-predictions in the model
2. Which statement about OLS regression is correct?
  - a. OLS is used primarily to locate clusters or hot spots.
  - b. OLS is a local regression model.
  - c. OLS attempts to explain which variables explain a phenomenon, and to what degree.**
  - d. OLS accounts for spatial variation in a variable's relationships.

## Interpreting OLS diagnostics (page 10-12)

1. What should you do if the probability associated with a coefficient is not statistically significant?  
**If the coefficient is not statistically significant, it is not helping your model and can therefore be removed.**
2. What is the adjusted R-squared statistic, and how does it indicate model performance?  
**Adjusted R-squared is a statistic that quantifies model performance. The value returned is a percentage that shows how much of the variation in the dependent variable is explained by the independent variables in the model.**
3. What does it mean if the OLS residuals are spatially clustered? What should you do to solve the problem?  
**Spatially clustered residuals indicate that you may be missing key variables. Try adding other independent variables to the model until the residuals are not spatially clustered.**

# Answers to Lesson 10 questions (continued)

## OLS reports (page 10-17)

### Scenario 1: Removing variables

1. Should you remove any variables from this model? Why?

**Yes. RENTER\_CY does not have a statistically significant probability associated with its coefficient and the VIF is over 7.5, which indicate that the variable is not helping the model and is redundant.**

### Scenario 2: Model performance

2. These OLS reports are for modeling total crime in a city but use different independent variables. Which model is better and why?

**The model on top has a higher adjusted R-squared and a lower AIC, indicating that the model is performing better than the model at the bottom. The model at the bottom has an independent variable with a VIF over 7.5, indicating variable redundancy.**

### Exercise 10: Find a properly specified regression model (page 10-21)

1. What does the adjusted R-squared value tell you about HCC score and Medicare spending?

**A value of 0.65 indicates that the HCC score explains 65 percent of Medicare spending.**

2. What does the AIC value of 1770 tell you about the HCC score and Medicare spending?

**Nothing. When you have other regression models using Total Costs 2010 as the dependent variable, you can compare AIC values.**

3. What can you extrapolate about the residuals from the map?

**Visually, there appears to be some spatial clustering of over-predictions (red) and under-predictions (blue).**

4. What does the OLS summary indicate about the statistical significance for each variable?

**The variables have an asterisk (\*) and are therefore statistically significant, helping the model.**

# Answers to Lesson 10 questions (continued)

5. What does it mean when a coefficient's probability has an asterisk next to it?

**There is a low probability that the variable is not helping the model.**

6. Should the number of hospital beds contribute to higher spending?

**Yes. More beds allow for a greater number of patients in a hospital, which could lead to more spending.**

7. What is the statistical check for determining if variables are redundant?

**Variance inflation factor (VIF) values over 7.5.**

8. Are any of the independent variables redundant?

**No, they are all under a VIF of 7.5.**

9. Which of the OLS checks assesses normality in the distribution (nonspatial) of residual values? (*Hint: Look in the GLR Diagnostics section.*)

**Jarque-Bera**

10. Is this model biased, and why?

**No, the Jarque-Bera value is not statistically significant (no asterisk [\*]).**

11. Based on the spatial OLS output of the residuals, do you think that the residuals are clustered?

**No, they do not appear to be clustered.**

12. What are the two main measures of model performance?

**Adjusted R-squared and AIC are the two main measures.**

13. What information can you obtain from the AIC and adjusted R-squared values in comparison to the first model that tested only the HCC score variable?

**Adjusted R-squared increased from 0.65 to 0.86, indicating that the variables explain 86 percent of the Medicare spending story. Further, the AIC value decreased from 1770 to 1672, indicating a better model for the dependent variable.**



You have used OLS, a global regression method, to create a regression model that is applied to all features in the study area. Tobler's First Law of Geography states: "Everything is related to everything else, but near things are more related than distant things." With Tobler's Law in mind, you may speculate that spatial relationships vary over a study area. OLS regression assumes that the relationships between your variables are static over space, but another type of regression analysis, called geographically weighted regression (GWR), allows for these variable relationships for change over space. In this lesson, you will use GWR to see if your model improves by allowing the data relationships to vary spatially.

## Topics covered

---

How relationships change over space

When to use GWR

## How relationships change over space

---

Earlier, you learned several important terms and concepts related to OLS regression. Using what you have learned about dependent and independent variables, coefficients, and R-squared values, assess the following situation and answer the questions.

### Scenario 1: Housing price prediction

In a city planning meeting in which housing prices were discussed, an attendee suggested that prices could be predicted by the size and age of the house—the bigger and newer the house, the higher its value.

- 1. Is OLS, a global regression model, an appropriate choice to make these predictions? Why?**

---

---

---

---

---

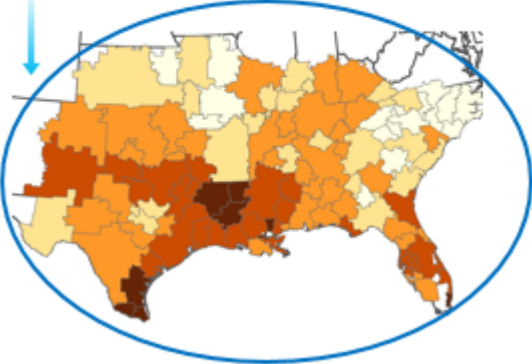
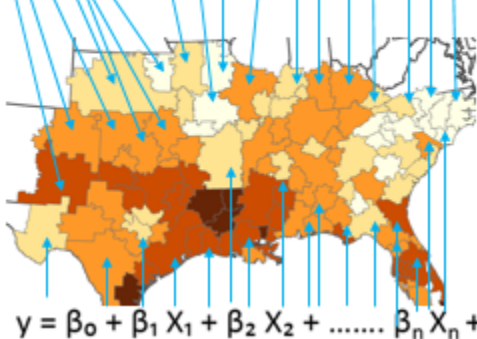
---

---

---

## GWR characteristics

You have learned that OLS regression constructs one equation for all features in the study area. As such, it is considered a global regression model, which assumes that the relationships between your data variables are static over space. Another type of regression, GWR, is a local regression model in that it constructs a single equation for each feature in the study area using only its neighboring features. As a result, GWR allows for variable relationships to change over space. When you find a properly specified model using OLS, you can use the same variables in GWR and potentially improve your results.

Ordinary least squares	Geographically weighted
<ul style="list-style-type: none"> <li>• Global regression model</li> <li>• Fixed relationships</li> <li>• One equation for all features</li> </ul>	<ul style="list-style-type: none"> <li>• Local regression model</li> <li>• Varying relationships</li> <li>• One equation for each feature</li> </ul>
$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n + \epsilon$ 	$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n + \epsilon$  $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n + \epsilon$

### What could contribute to spatially varying relationships?

Tobler's First Law of Geography indicates that relationships may be different over space. When relationships among variables change over space, it is called nonstationarity.

## When to use GWR

There are three reasons to use GWR:

- You have found a properly specified OLS model and you want to know if allowing for spatially varying relationships improves model performance. Use the same variables as you did in the OLS model.
- You want to predict alternative or future values using a model calibrated with existing values.
- The OLS diagnostics indicate statistically significant nonstationarity.

It is important to find a properly specified model using OLS regression. OLS has strong diagnostics that you can use to perform the six checks to find the best combination of variables. GWR has some diagnostics, but not all of them are suitable for finding a properly specified model. A recommended approach is to find a properly specified model using OLS, and then use the same variables in GWR.

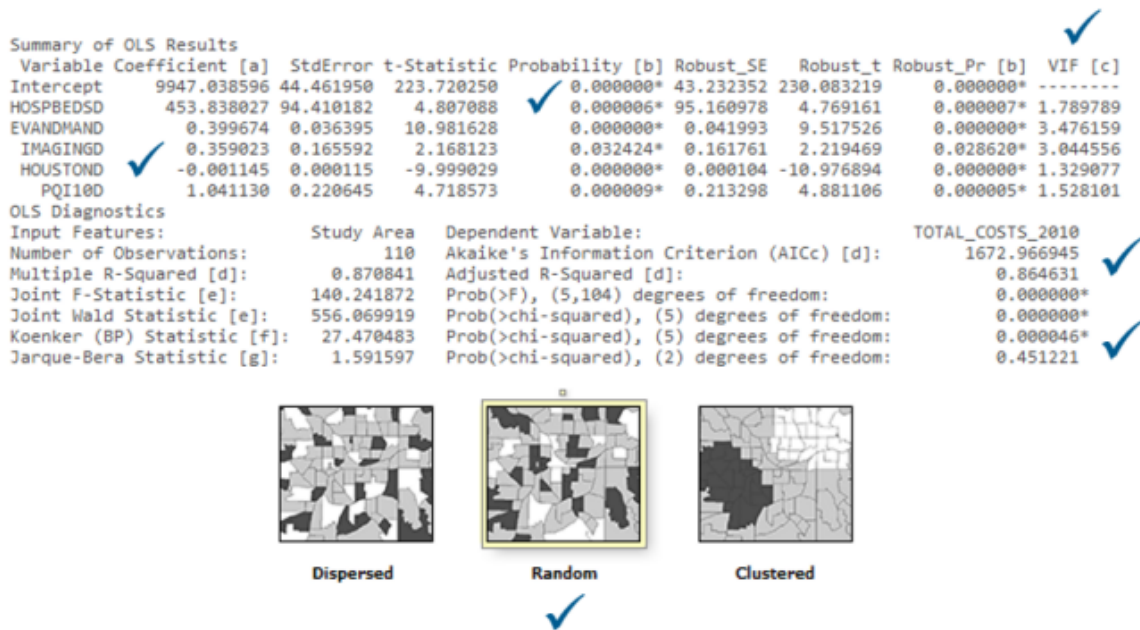


Figure 11.1. Perform six checks using OLS before running GWR.

In the OLS diagnostic report, there is a statistic called the Koenker statistic. The Koenker (BP) statistic is a value that indicates whether the independent variables in the model have a consistent relationship to the dependent variable both in geographic space and in data space. A statistically significant Koenker value, or an asterisk (\*) next to it, indicates that the modeled relationships are not consistent.

## When to use GWR (continued)

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	9947.038596	44.461950	223.720250	0.000000*	43.232352	230.083219	0.000000*	-----
HOSPBEDSD	453.838027	94.410182	4.807088	0.000006*	95.160978	4.769161	0.000007*	1.789789
EVANDMAND	0.399674	0.036395	10.981628	0.000000*	0.041993	9.517526	0.000000*	3.476159
IMAGINGD	0.359023	0.165592	2.168123	0.032424*	0.161761	2.219469	0.028620*	3.044556
HOUSTOND	-0.001145	0.000115	-9.999029	0.000000*	0.000104	-10.976894	0.000000*	1.329077
PQI10D	1.041130	0.220645	4.718573	0.000009*	0.213298	4.881106	0.000005*	1.528101

OLS Diagnostics

Input Features:	Study Area	Dependent Variable:	TOTAL_COST1_2010
Number of Observations:	110	Akaike's Information Criterion (AICc) [d]:	1672.968945
Multiple R-Squared [d]:	0.870841	Adjusted R-Squared [d]:	0.864631
Joint F-Statistic [e]:	140.241872	Prob(>F), (5,104) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	556.060010	Prob(>chi-squared), (5) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	27.470483	Prob(>chi-squared), (5) degrees of freedom:	0.000046*
Jarque-Bera Statistic [g]:	1.591597	Prob(>chi-squared), (2) degrees of freedom:	0.451221



**Figure 11.2.** If the Koenker test is statistically significant, use the Robust Probability values to determine coefficient significance.

Two main factors cause a statistically significant Koenker test:

- **Nonstationarity:** When relationships between the dependent variable and independent variables are not consistent across geographic space. In other words, the model performs differently in different parts of the study area. For example, imaging events may be a strong predictor for Medicare costs in one area but not in another area.
- **Heteroscedasticity:** When the relationships between the dependent variable and independent variables are not consistent across data space. In other words, the model performs differently based on a the range of values in the independent variable or variables. For example, the model may perform better for areas with lower Medicare spending than areas with higher Medicare spending.

## GWR in action

---

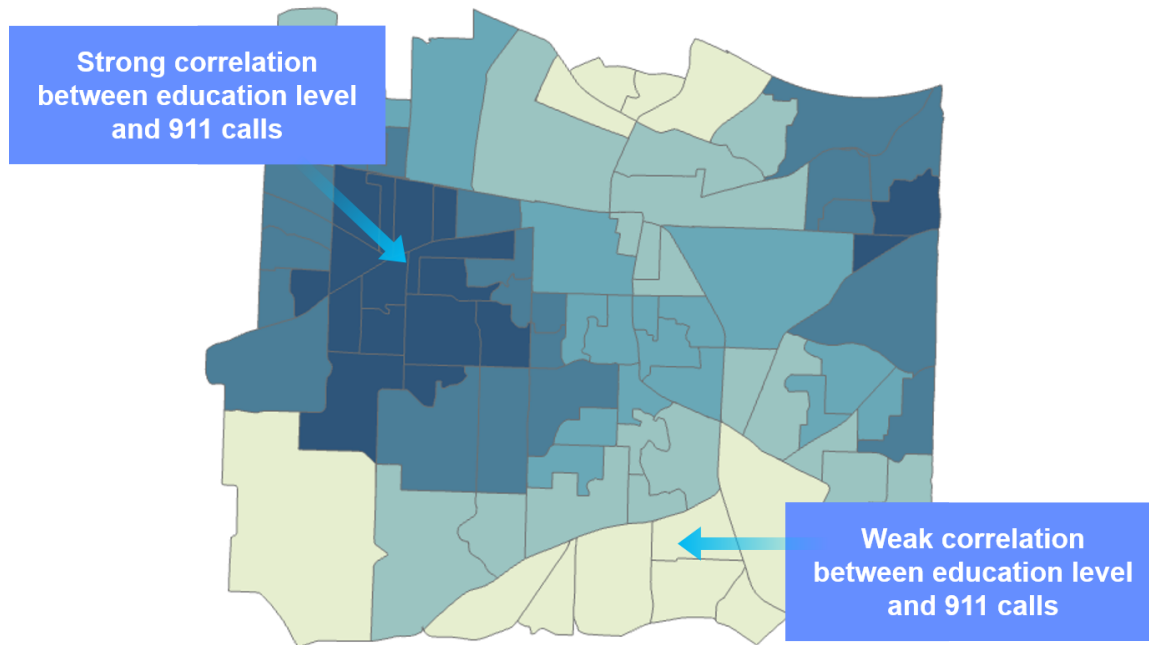


Figure 11.3. GWR results.

## GWR in action (continued)

---

### GWR tips

- Use GWR after you find a properly specified model using OLS and the same independent variables.
- Use GWR if the Koenker test statistic has an asterisk (\*) next to it after running OLS. The asterisk (\*) next to the Koenker statistic indicates statistically significant nonstationarity.
- Use GWR when you are predicting future values based on current and estimated values.
- An optional output of GWR is creating coefficient surfaces so that you can visualize the relationships between each independent variable and dependent variable to see where the relationships are stronger.
- GWR prints R-squared as a measure of goodness of fit for the model. Its value varies from 0.0 to 1.0, with higher values being preferable. It may be interpreted as the proportion of dependent variable variance accounted for by the regression model. The denominator for the R-squared computation is the sum of squared dependent variable values. Adding an extra explanatory variable to the model does not alter the denominator but does alter the numerator, giving the impression of improvement in model fit that may not be real.

## GWR in action (continued)

---

- Because of the previously described problem for the R-squared value, calculations for the adjusted R-squared value normalize the numerator and denominator by their degrees of freedom. This has the effect of compensating for the number of variables in a model, and consequently, the adjusted R-squared value is almost always smaller than the R-squared value. However, in making this adjustment, you lose the interpretation of the value as a proportion of the variance explained. In GWR, the effective number of degrees of freedom is a function of the bandwidth, so the adjustment may be quite marked in comparison to a global model like OLS. For this reason, the AICc is preferred as a means of comparing models.



[ArcGIS Pro Help: Interpreting GWR results](#)



## Perform GWR

In this exercise, you will use the same variables from a properly specified model to get a better result by allowing for spatial variation in the variable relationships. You will also use GWR to predict Medicare costs related to reducing dehydration rates.

In this exercise, you will perform the following tasks:

- Run GWR using a properly specified OLS model.
- Map coefficients to see variation over space.
- Predict using GWR.

## Step 1: Run GWR using a properly specified OLS model

You will resume with the previous map and example of determining factors that influence Medicare spending.

- a** Restore the ArcGIS Pro project, and verify that you are viewing the Regression Analysis map and the Generalized Linear Regression report window.



If you closed the GLR report window, browse to **C:\EsriTraining\SNAP\Results\Ex10**, and then open the GLR Report file.

- b** In the GLR report's GLR Diagnostics section, locate the Koenker statistic.

```
----- GLR Diagnostics -----
-----
Input Features:          Study Area   Dependent Variable:          TOTAL_COSTS_2010
Number of Observations:      110      Akaike's Information Criterion (AICc) [d]:      1672.966945
Multiple R-Squared [d]:      0.870841   Adjusted R-Squared [d]:      0.864631
Joint F-Statistic [e]:      140.241872 Prob(>F), (5,104) degrees of freedom:      0.000000*
Joint Wald Statistic [e]:    556.069919 Prob(>chi-squared), (5) degrees of freedom:      0.000000*
Koenker (BP) Statistic [f]:  27.470483 Prob(>chi-squared), (5) degrees of freedom:      0.000046*
Jarque-Bera Statistic [g]:   1.591597   Prob(>chi-squared), (2) degrees of freedom:      0.451221
```

### 1. What does a statistically significant Koenker value indicate?

---



---

Running GWR does not take more effort to find the correct variables, as you have done that work using OLS. You will use the same variables from the properly specified OLS model from the previous exercise when you run GWR.

- c** Close the Generalized Linear Regression report window.
- d** Search for and open the **Geographically Weighted Regression (GWR)** tool, and set the following parameters:

- Input Features: Study Area
- Dependent Variable: Total Costs 2010
- Model Type: Continuous (Gaussian)
- Explanatory Variables: HospBedsD, EvAndManD, ImagingD, HoustonD, PQI10D
- Output Features: **GWR**
- Neighborhood Type: Distance Band
- Neighborhood Selection Method: Golden Search

- e** Click Run.

- f In the Geoprocessing pane, click View Details to view the diagnostics.
- g Expand the Geographically Weighted Regression (GWR) window, scroll to the bottom, and find Model Diagnostics.

```

----- Model Diagnostics -----
R2:                                0.9026
AdjR2:                              0.8806
AICc:                               1667.7464
Sigma-Squared:                      189692.8124
Sigma-Squared MLE:                  155074.4410
Effective Degrees of Freedom:       89.9253

```

There are fewer diagnostics in the GWR report than in the OLS report, but there are adjusted R-squared and AIC values. The OLS model from the previous exercise had an adjusted R-squared of 0.86 and an AIC of 1672.



Find a properly specified model using OLS first, and then use the same variables in GWR.

2. What can you say about the adjusted R-squared and AIC values in the GWR model compared to the values in the OLS model?

---



---



---



---

By taking into account nonstationarity, GWR provides an improved model of Medicare spending.

- h Close the Geographically Weighted Regression (GWR) window.

## Step 2: Map coefficients to see variation over space

Like OLS, GWR creates a spatial output. You can symbolize the data using the coefficients to see the strength of the correlations that the variables have over space.

- a In the Contents pane, select the GWR layer, if necessary.
- b Click the Appearance tab.
- c Click the Symbology down arrow, and then choose Unique Values and then Graduated Colors.
- d In the Symbology pane, for Field, choose Coefficient (IMAGINGD).

- e Choose 7 classes and the Blues (Continuous) color scheme.

**Primary symbology**

Graduated Colors

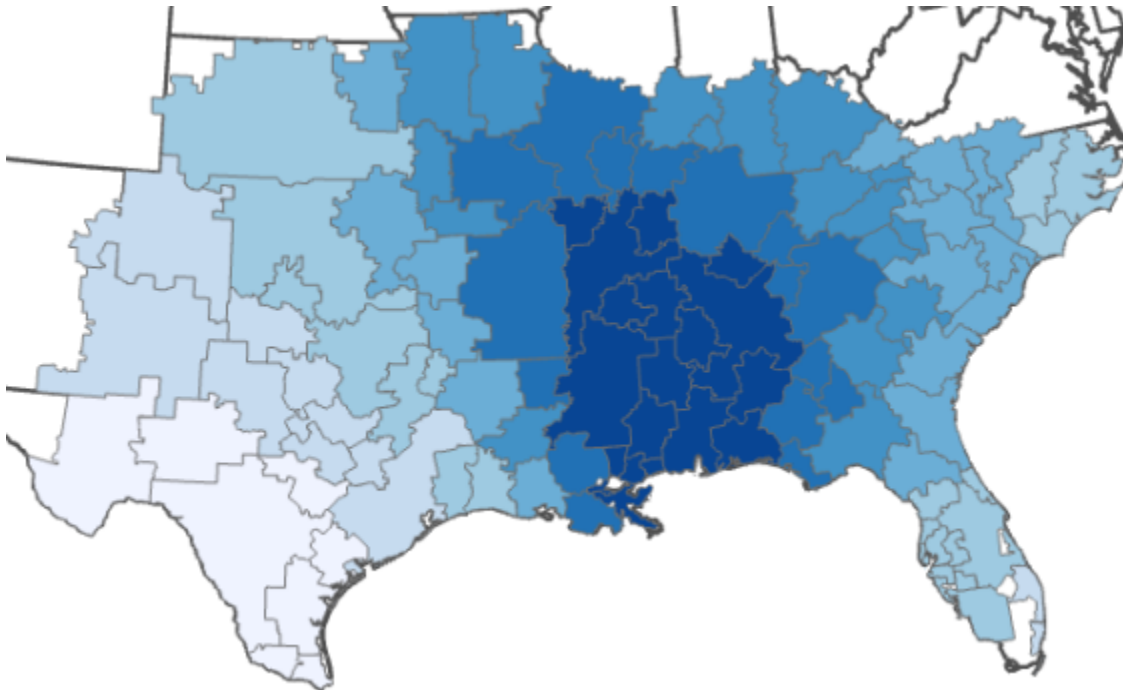
Field

Normalization

Method

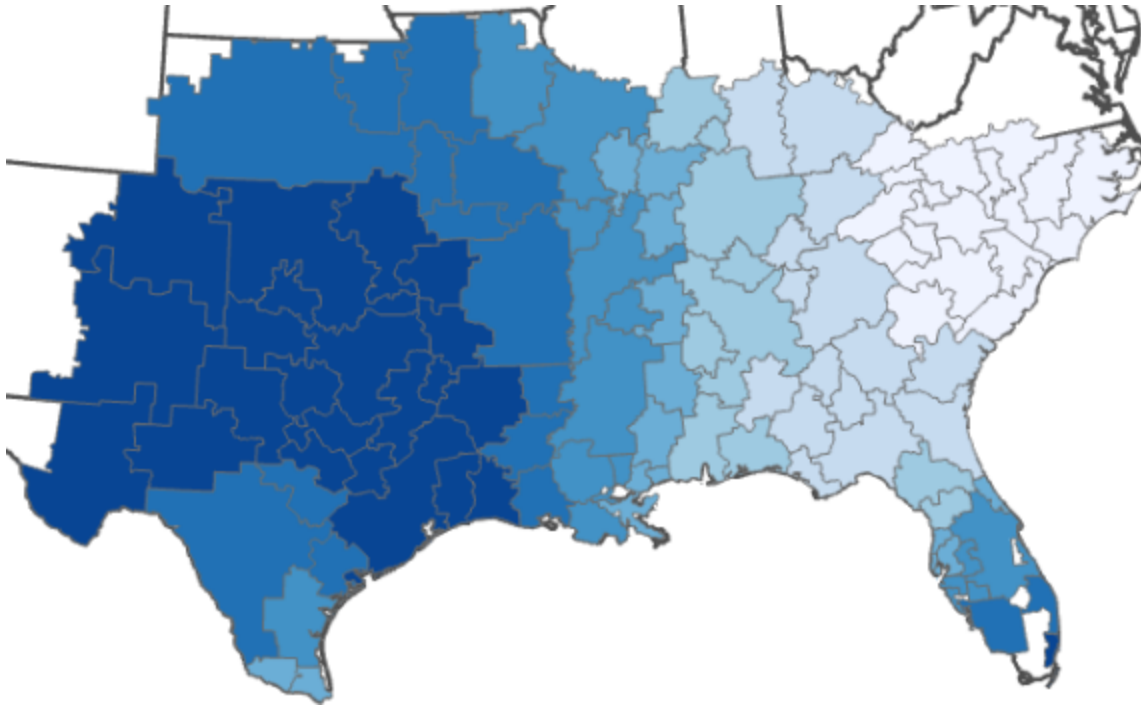
Classes

Color scheme



The darker areas show locations where the relationship between number of imaging events and Medicare spending is the strongest. Knowing where relationships are strongest can help you focus any remediation efforts or decide how to address the problem. Next, you will symbolize the GWR layer by another coefficient.

- f In the Symbology pane, change the Field to Coefficient (PQI10D), which is the dehydration rate, and keep the other settings as they are.



Viewing the coefficient map for dehydration rates indicates an underlying spatial process. In the western part of the study area (Texas), dehydration rates have more of an impact on Medicare spending than in the southeastern states, such as the Carolinas and Georgia. The map does not indicate that there is more dehydration in dark areas or that there is more spending; rather, it shows that the relationship between dehydration rates and Medicare spending is strongest.



Alternatively, you can set an optional parameter for GWR to create coefficient surfaces for each independent variable by specifying an output workspace. The coefficient surfaces are similar to how you symbolized the GWR layer based on the coefficient. The surface will show spatial correlation between variables, but it will be a raster dataset.

Based on the results, you can target outreach programs to help educate people on staying properly hydrated. You cannot assign resources everywhere, so GWR helps narrow down problem areas so that you can target efforts to resolve the issue.

### Step 3: Predict using GWR

You just ran GWR again using some additional parameters for prediction this time. You will use most of the same variables, but instead of the real dehydration admissions data, you will use a reduced dehydration variable previously created, which is the original dehydration admissions reduced by 50 percent.

- a In the Catalog pane, click the History tab, if necessary.


- b Open the previous run of the GWR tool, and leave all parameters that you set previously as they were.
- c Expand Prediction Options, and set the following parameters:
  - Prediction Locations: Study Area
  - Explanatory Variables To Match: Only change PQI10D under Field From Prediction Locations to ReducedDehydration
  - Output Predicted Features: **PredictReducedDehy**

Explanatory Variables to Match	
Field from Prediction Locations	Field from Input Features
HospBedsD	HospBedsD
EvAndManD	EvAndManD
ImagingD	ImagingD
HoustonD	HoustonD
ReducedDehydration	PQI10D

Output Predicted Features  
PredictReducedDehy

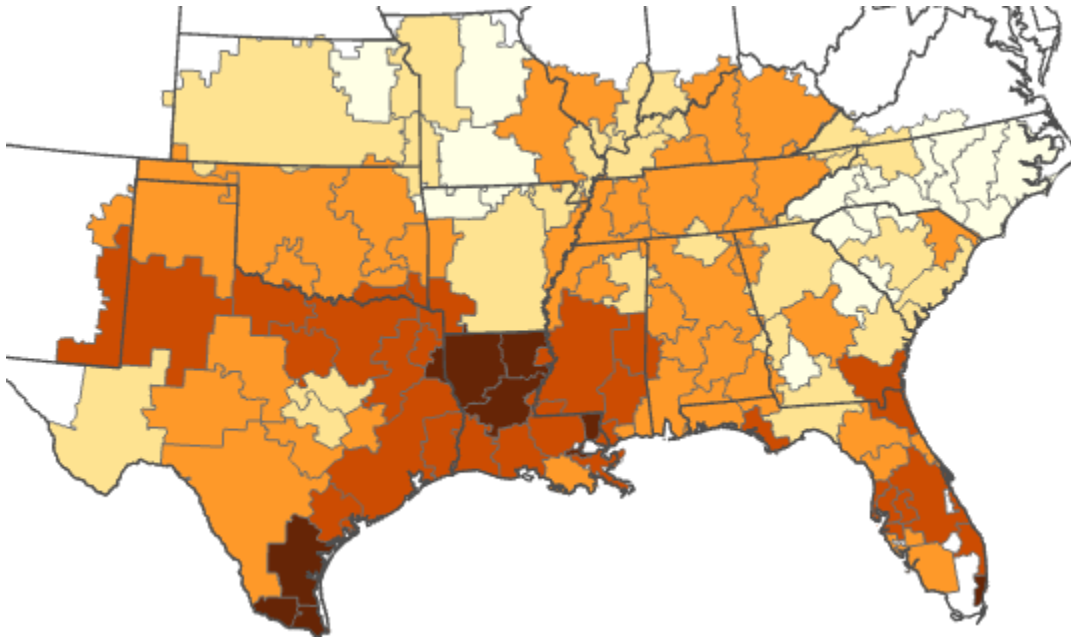
- d Click Run.

GWR starts by calibrating the model using the original data, so you will notice that the diagnostics in the report are the same. The difference is that after it completes the original analysis, GWR then predicts the impact based on any new variables provided; in this case, it was dehydration. You will update the symbology of the output prediction layer to match the original map of Medicare spending, using that predicted value for cost.

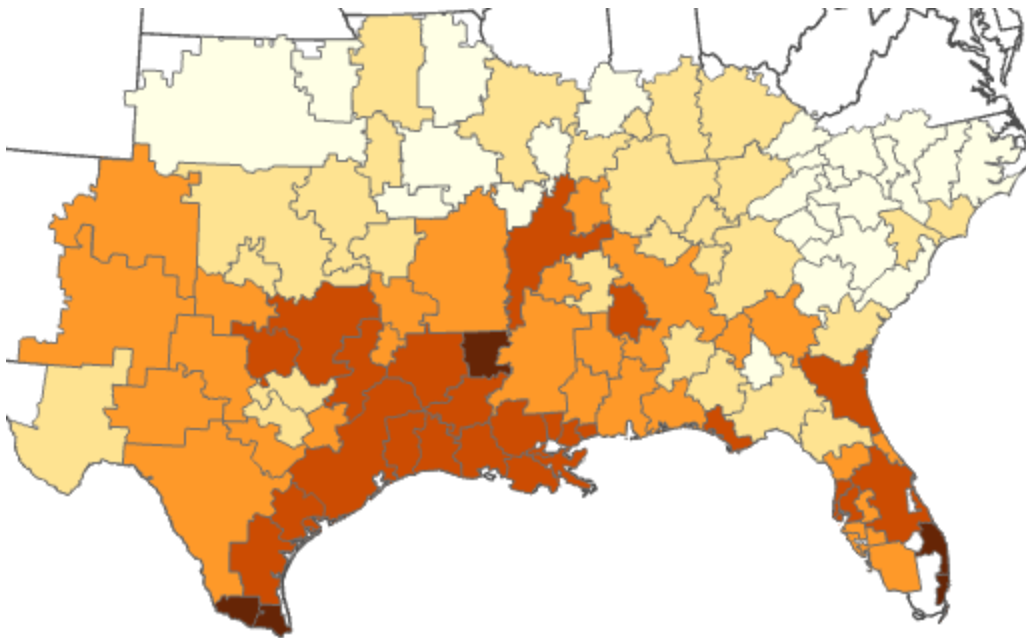
- e In the Contents pane, ensure that PredictReducedDehy is selected.
- f In the Symbology pane, ensure that the Primary Symbology is set to Graduated Colors.
- g In the upper-right corner of the Symbology pane, click the Options button  and choose Import Symbology.
- h In the Geoprocessing pane, for Symbology Layer, choose Study Area, and then click Run.
- i Return to the Symbology pane.
- j For Field, choose Predicted (TOTAL\_COSTS\_2010).

- k In the Contents pane, make PredictReducedDehy, States, and Study Area the only visible layers.
- l Swipe the prediction layer, and compare it to the original values.

Initial costs:



Predicted based on reducing dehydration by 50 percent:



You can see what the impact of reducing dehydration by 50 percent would be and the areas where it would be the most effective.

- m** Save the project, and leave ArcGIS Pro open.



## Lesson review

---

1. Explain the differences between OLS and GWR.

---

---

---

---

---

---

2. When should you run GWR?

---

---

---

---

---

# Answers to Lesson 11 questions

## How relationships change over space (page 11-2)

### Scenario 1: Housing price prediction

1. Is OLS, a global regression model, an appropriate choice to make these predictions? Why?  
**A global regression model may not be the best choice because the one regression equation applied to all features would not account for spatial variation. Home prices often vary with location. A strong positive relationship may also exist between housing values and housing age in one part of a city, but that relationship may be weaker in another part. For example, a city's historic district may contain many older, more valuable homes because of their historical importance and the costs associated with building renovation. But other parts of a city may have many older homes that have not been renovated or maintained—they may be dilapidated.**

### Exercise 11: Perform GWR (page 11-9)

1. What does a statistically significant Koenker value indicate?  
**It indicates that your variables exhibit nonstationarity or heteroscedasticity, or vary over geographic space and data space. Because of the varying relationships, running GWR may benefit your model.**
2. What can you say about the adjusted R-squared and AIC values in the GWR model compared to the values in the OLS model?  
**Adjusted R-squared is 0.88, indicating that you are telling more of the Medicare spending story in this area by allowing the relationships to vary across space. Additionally, the lower AIC value for GWR indicates that it provides a better fit than the OLS model, given that the models use the same dependent variable.**

# 12

## Geostatistical interpolation

You have used deterministic interpolation to create continuous surfaces from sample points using deterministic methods. You also learned that there is another method of interpolation called geostatistical interpolation. In this lesson, you will learn the basics of geostatistical interpolation and create a prediction surface.

### Topics covered

---

Deterministic interpolation recap

Geostatistical interpolation

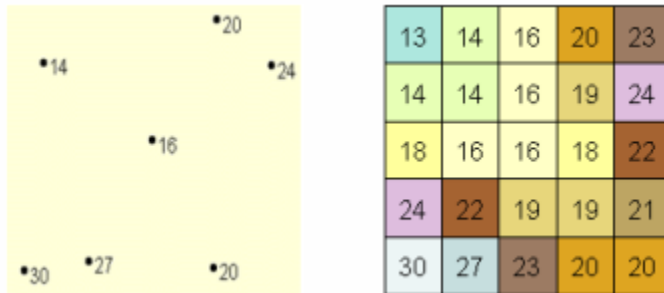
Kriging

Geostatistical workflow

Empirical Bayesian kriging

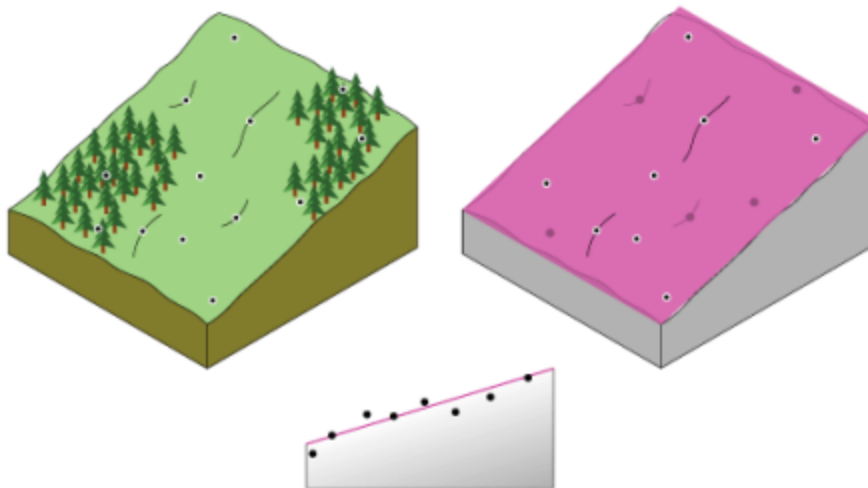
## Deterministic interpolation

All spatial interpolators attempt to predict the value of an attribute at unknown locations using attribute values from known sampled locations.



**Figure 12.1.** Interpolation estimates unknown values using known values.

Earlier in the course, you used deterministic interpolation methods to create surfaces from known point locations, predicting unknown values from known values. Deterministic interpolators use a mathematical formula to calculate this predicted value based on the degree of smoothing or similarity in relation to neighboring points.



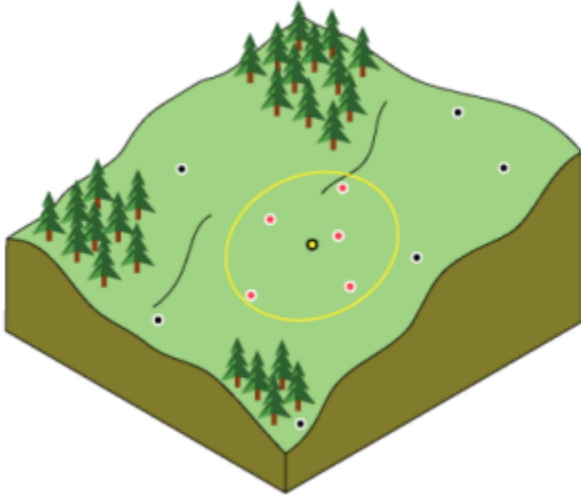
**Figure 12.2.** Some interpolators attempt to fit a mathematical function to a distribution of points, much like bending a piece of paper to fit the distribution.

When you use deterministic interpolators, the output is fully determined by the user-specification of the parameter values and the data. Changing these parameter values may produce different

## Deterministic interpolation (continued)

---

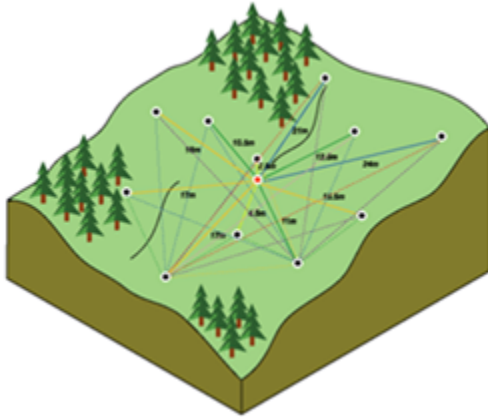
results, but they do not consider the statistical properties and underlying spatial structure of the data.



**Figure 12.3.** In IDW, a deterministic interpolation method, the user can specify a search neighborhood. Only values within the search neighborhood determine the unknown values.

## Geostatistical interpolation

Geostatistical interpolation uses the statistical properties of your measured points to predict the unknown locations across the surface. The process of modeling the statistical correlation between all pairs of points in a dataset allows the spatial dependence to be inferred based on the underlying spatial structure of the dataset.



**Figure 12.4.** Geostatistical interpolation models the statistical correlation between all pairs of known points based on both their distance apart and their values.

Geostatistics is based on the regionalized variable theory, which says the variation in a surface can be decomposed into three main components: a deterministic trend component, an autocorrelated error component, and a random error component.

Geostatistical example:



**Figure 12.5.** All these blue points represent a sample location that includes a temperature value. Thus, you can calculate a mean, which is a statistic. Deterministic methods ignore even basic statistics, such as the mean.

## Kriging

---

One of the most popular geostatistical interpolation methods is kriging. Kriging assumes that at least some of the spatial variation observed in natural phenomena can be modeled by random processes with spatial autocorrelation and require that the spatial autocorrelation be explicitly modeled.

### Kriging assumptions

Assumption	Description
<b>Spatial continuity</b>	Every location in the area has a value, but not all values are available to you.
<b>Spatial autocorrelation</b>	Locations closer in distance are assumed to be closer in value than locations farther apart.
<b>Stationarity</b>	The relationship between two points and their values depends only on the distance between them, not their exact location.
<b>Normally distributed</b>	Use histograms and other charts to verify normal distribution and apply transformation if needed.
<b>No global trends</b>	There is a constant average in your data values across the surface. You can remove trends.
<b>Spatial clustering</b>	The data is evenly distributed, not spatially clustered, as clusters will not appropriately represent the study area.

You can use kriging techniques to describe and model spatial patterns, predict values at unmeasured locations, and assess the uncertainty associated with a predicted value at the unmeasured locations.



[ArcGIS Pro Help: How kriging works](#)

[ArcGIS Pro Help: Kriging in Geostatistical Analyst](#)

[ArcGIS Pro Help: Understanding how to create surfaces using geostatistical techniques](#)

## Geostatistical workflow

The workflow for performing geostatistical interpolation is more complex than performing deterministic interpolation. Knowledge of the data and of the various geostatistical properties and options in the tools is vital to creating a valid prediction surface.

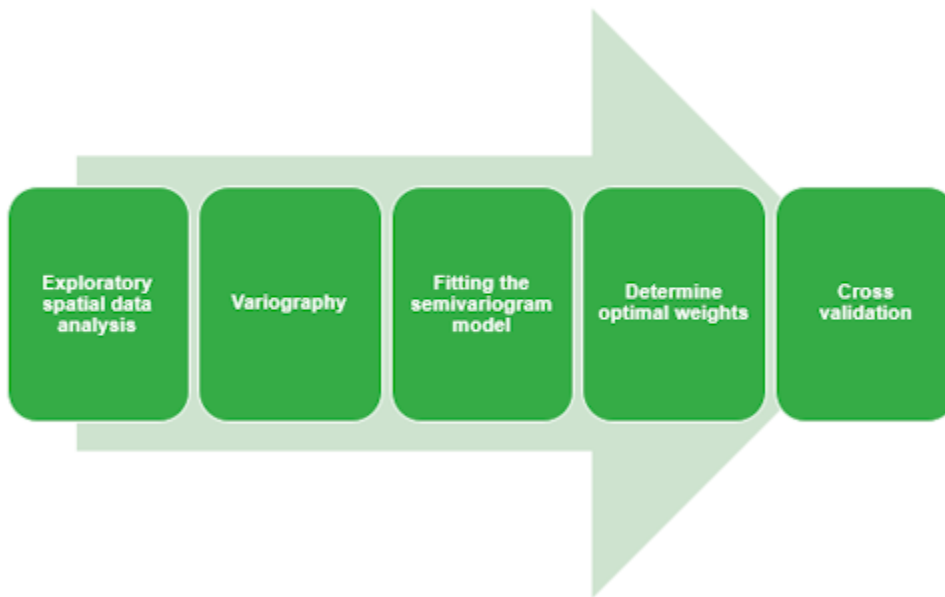


Figure 12.6. Geostatistical workflow.

### Exploratory spatial data analysis

Preparing for kriging is the first step. Discover if your data meets the kriging assumptions.

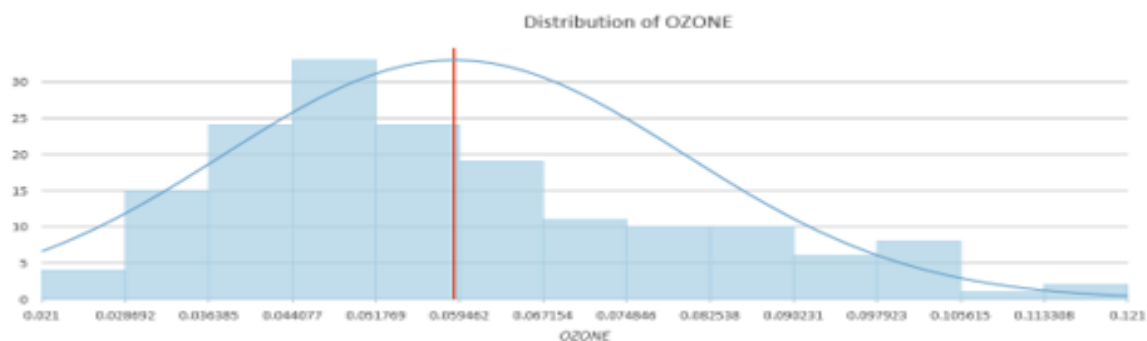


Figure 12.7. Checking the distribution of your data is a commonly used operation for data exploration.



Esri Training course: [Exploring Spatial Patterns in Your Data Using ArcGIS](#)

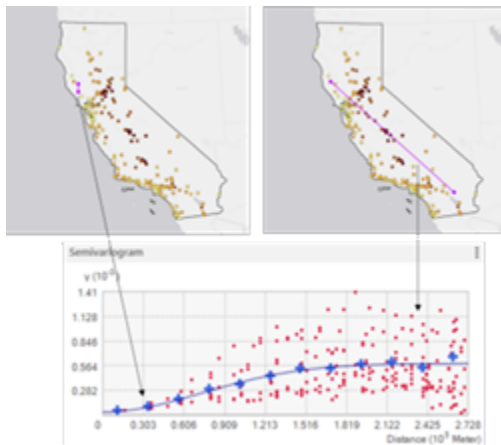


## Geostatistical workflow (continued)

### Describing the spatial structure of the data

Describing the spatial structure of data involves understanding the local spatial variation (spatial autocorrelation) in the sample points.

Semivariance is a measure of the difference in value between a pair of points. The semivariogram shows you that there are similarities between each pair of points at a short distance, but features become less similar at a certain distance.



**Figure 12.8.** Each red dot represents a pair of point locations. The x-axis represents the distance between paired locations, whereas the y-axis represents semivariance. A red dot in the lower left represents a pair of points that are close together in distance and are also similar in value. A red dot in the top-right corner represents a point pair that is far away in distance and also dissimilar in value.

### Fitting the semivariogram model

After you have created a description of the spatial structure (autocorrelation) in the data through the empirical semivariogram, you would then attempt to model this spatial structure using a mathematical equation that best fits the blue points, which represent the average semivariogram values for each bin (lag).

### Determining optimal weights

Measure distance from all sample points within a search neighborhood of the unknown point of interest. Then use the value modeled on the semivariogram for that distance to estimate weights used to predict the unknown value at that location.

## Geostatistical workflow (continued)

---

### Cross-validation

Cross-validation is a procedure for testing how well the model predicts values at unknown locations. In cross-validation, a piece of data whose value is known independently is removed from the dataset, and the rest of the data is used to predict its value. This estimate is then compared to the actual sample value to calculate the model error.



**ArcGIS Pro Help:** *The geostatistical workflow*

**ArcGIS Pro Help:** *Essential vocabulary for Geostatistical Analyst*

**ArcGIS Pro Help:** *Understanding the semivariogram: The range, sill, and nugget*

**ArcGIS Pro Help:** *Modeling a semivariogram*

## Use the Geostatistical Wizard to perform kriging

You have used deterministic interpolation methods to create surfaces from sample points. Now, you will use geostatistical techniques to create a prediction surface. You will analyze the same ozone sample points from an earlier exercise, but you will use kriging this time.

In this exercise, you will perform the following tasks:

- Explore data using charts.
- Perform kriging using the Geostatistical Wizard.



## Step 1: Set up the ArcGIS Pro project

In this step, you will prepare the project.

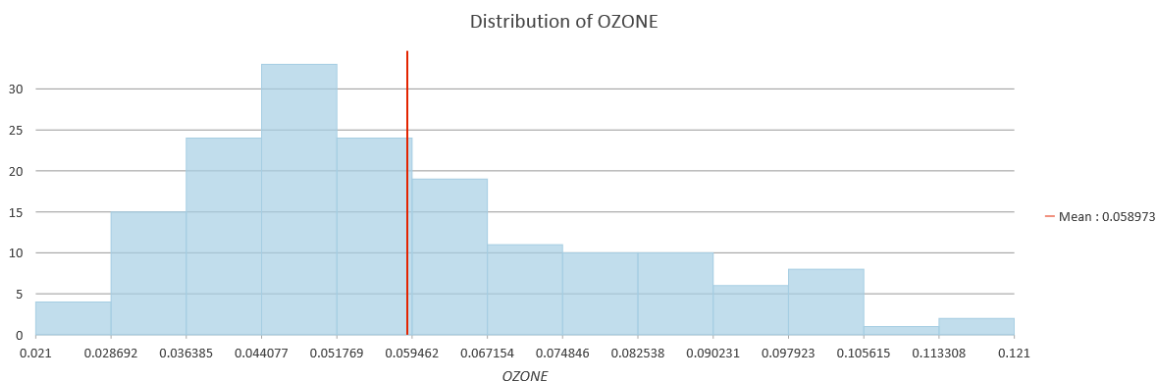
- a If necessary, start ArcGIS Pro and open the SNAPCourse project.
- b Open the Interpolation map.
 

*Hint:* Catalog pane > Project tab > Maps
- c Turn off all layers in the Interpolation map except Samples and StateBoundary.
- d If necessary, right-click StateBoundary and choose Zoom To Layer.
- e From the Analysis tab, click Environments.
- f Set the Extent to Same As Layer - StateBoundary, and then click OK.

## Step 2: Explore the data distribution

Exploring data is important for all analysis, but it is vital to kriging because of the assumptions that kriging makes about your data.

- a In the Contents pane, select the Samples layer.
- b From the Data tab, click Create Chart and choose Histogram.
- c In the Chart Properties pane, for Number, choose OZONE.

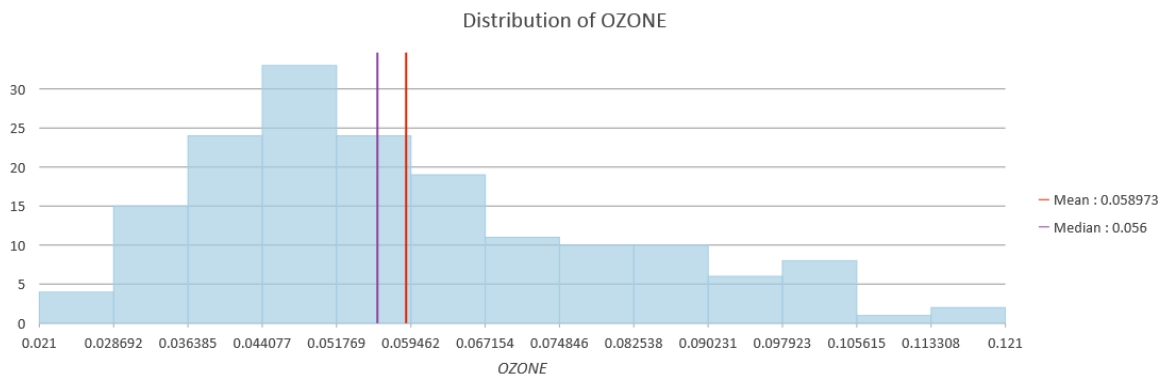


The mean of the data is displayed with the red vertical line in the chart.

The height of each bar represents the frequency of data within each bin. Generally, the important features of the distribution are its central value (for example, mean and median), spread, and

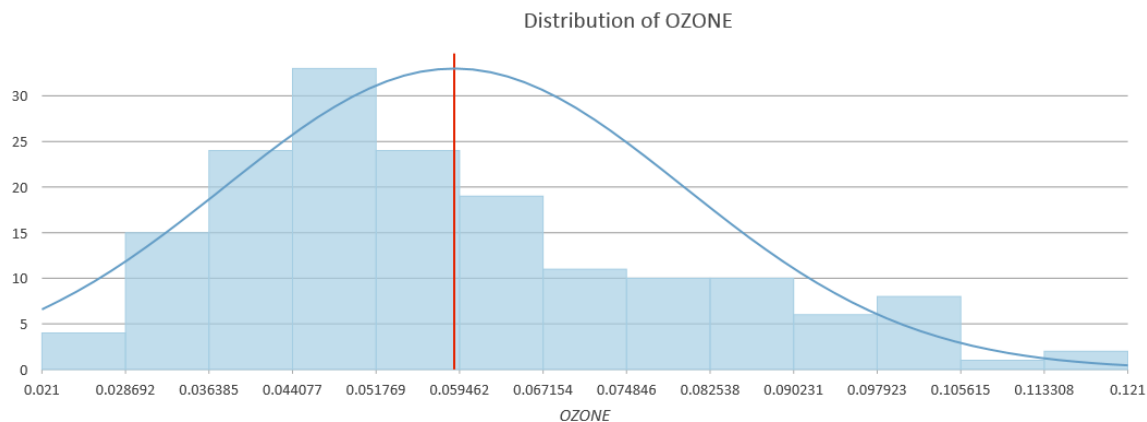
symmetry. The ozone data histogram indicates that the data is unimodal (one hump) and nearly symmetric. The right tail of the distribution indicates the presence of a relatively small number of sample points with large ozone concentration values. As a quick check, if the mean and the median are approximately the same value, you have one piece of evidence that the data may be approximately normally distributed.

- d** In the Chart Properties pane, check the Median box.



The mean and median values are 0.058 and 0.056, respectively, so they are very close.

- e** In the Chart Properties pane, check the Show Normal Distribution box.



Although the ozone measurements do not fit perfectly into the bell-shaped curve, other indicators (such as mean, median, and kurtosis) suggest a normal distribution. Kriging assumes that the data is normally distributed, so if your data is not normal, you should apply a transformation. You can visualize what a transformation will do to the data in the chart. Charts are connected to the features in the map, so you can select a column in the histogram and see the associated features selected in the map.

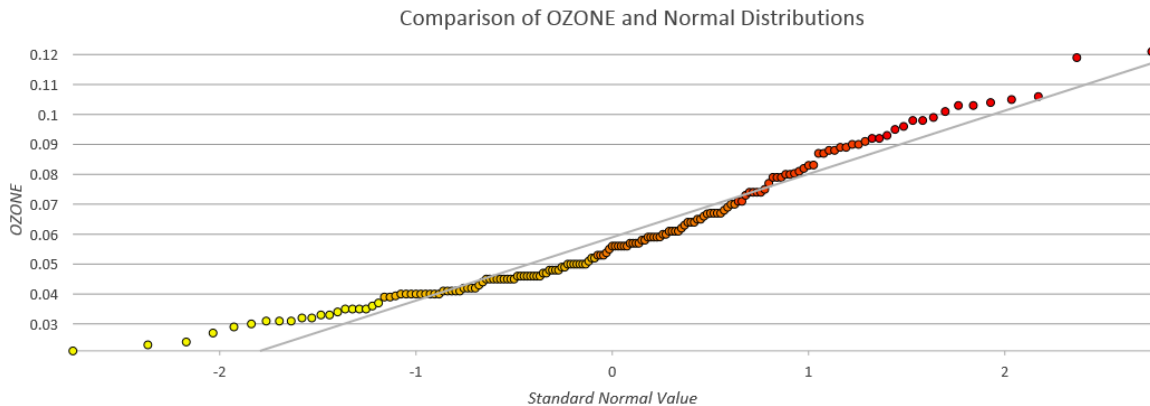
- f** In the histogram, draw a box around the four bars on the right to select them.



The features highlighted in the map belong to one of the four bins that you selected in the chart.

Next, you will continue your exploratory analysis using a QQ plot.

- g** Close the chart.
- h** On the Data tab, in the Selection group, click Clear to clear your selection.
- i** From the Data tab, click Create Chart and choose QQ Plot.
- j** In the Chart Properties pane, for Compare The Distribution Of, select OZONE.



This QQ plot is normal, as it is plotting the quantiles of a numeric variable (OZONE) against the quantiles of a normal distribution. If the distributions of the compared quantiles are identical, then the plotted points will form an approximate straight line. The farther the plotted points deviate from a straight line, the less similar the compared distributions are. In this case, the ozone values closely follow a normal distribution.

- k** Close the charts and the Chart Properties pane.

### Step 3: Perform kriging using the Geostatistical Wizard

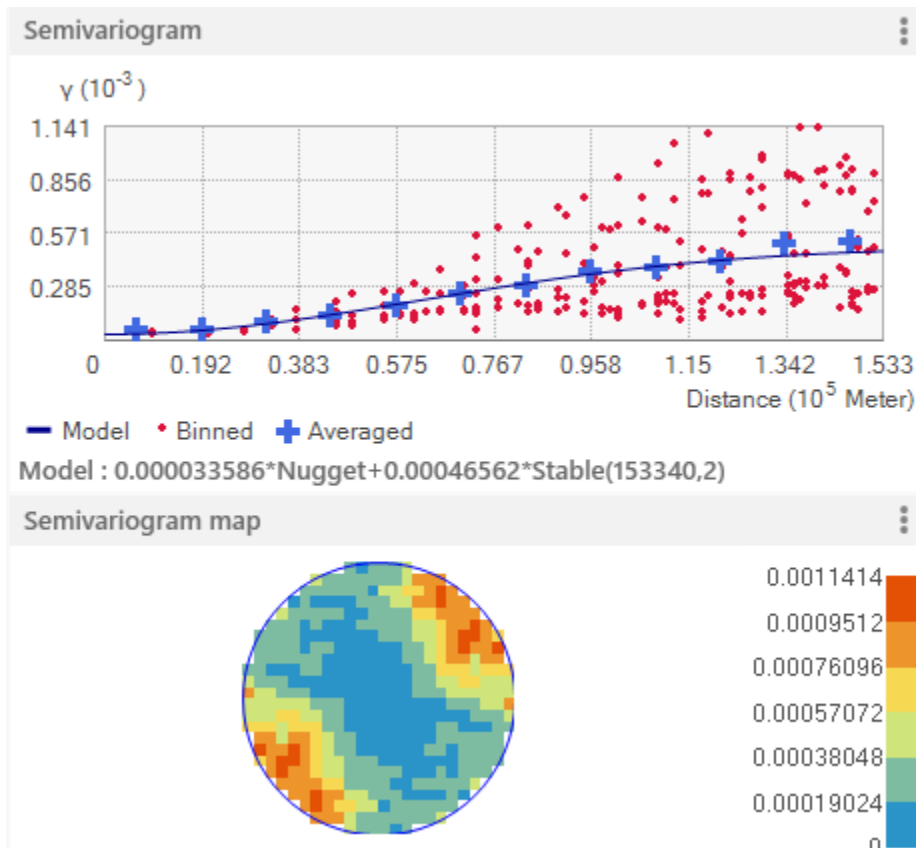
In this step, you will use the Geostatistical Wizard to perform kriging.

- a** From the Analysis tab, in the Tools group, click Geostatistical Wizard.
- b** For Geostatistical Methods, choose Kriging/CoKriging.
- c** Verify that the Input Dataset 1 Source Dataset is set to Samples.
- d** For Data Field, choose OZONE.

## Geostatistical Wizard - Kriging / CoKriging

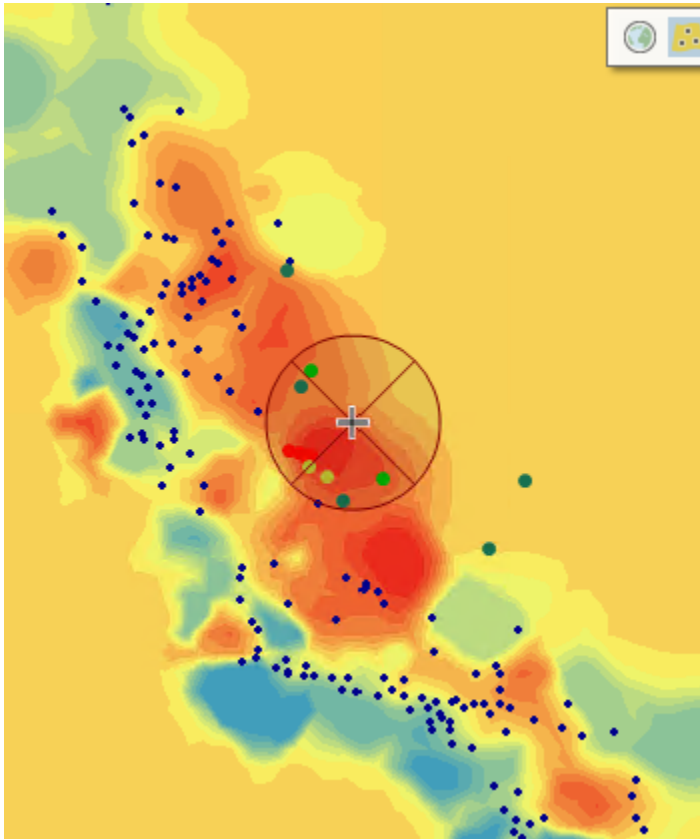
Geostatistical methods	Input Dataset 1
<input type="radio"/> Empirical Bayesian Kriging	Source Dataset <input type="text" value="Samples"/>
<input type="radio"/> EBK Regression Prediction	Data Field <input type="text" value="OZONE"/>
<input checked="" type="radio"/> <b>Kriging / CoKriging</b>	<b>Input Dataset 2</b>

- e Click Next.
- f For Simple Kriging, choose or verify that Prediction is selected.
- g For Transformation Type, choose None.
- h Click Next.
- i In the General Properties, for Function Type, choose Semivariogram.



The semivariogram model is displayed, which allows you to examine spatial relationships between measured points. Now you would like to fit the semivariogram model to capture the spatial relationships in the data and use it in the prediction model. The goals are to achieve the best fit and incorporate your knowledge of the phenomenon in the model. You can change parameters to get the best fit, or you can let ArcGIS Pro optimize the model.

**j** Click Next.



You can assume that as locations get farther from the prediction location, the measured values have less spatial autocorrelation with the prediction location. As these points have little or no effect on the predicted value, they can be eliminated from the calculation of that particular prediction point by defining a search neighborhood. You can control the size and shape of the search neighborhood and other properties.

**k** Click Next.

The final panel of the wizard is for cross validation. You learned about manual ways in which to validate surfaces earlier in the course. With the Geostatistical Wizard, validation is built in. Validation removes one data location and predicts the associated data using the data at the rest of the locations. The primary use for this tool is to compare the predicted value to the observed value to obtain useful information about some of your model parameters.

**l** Click Finish.

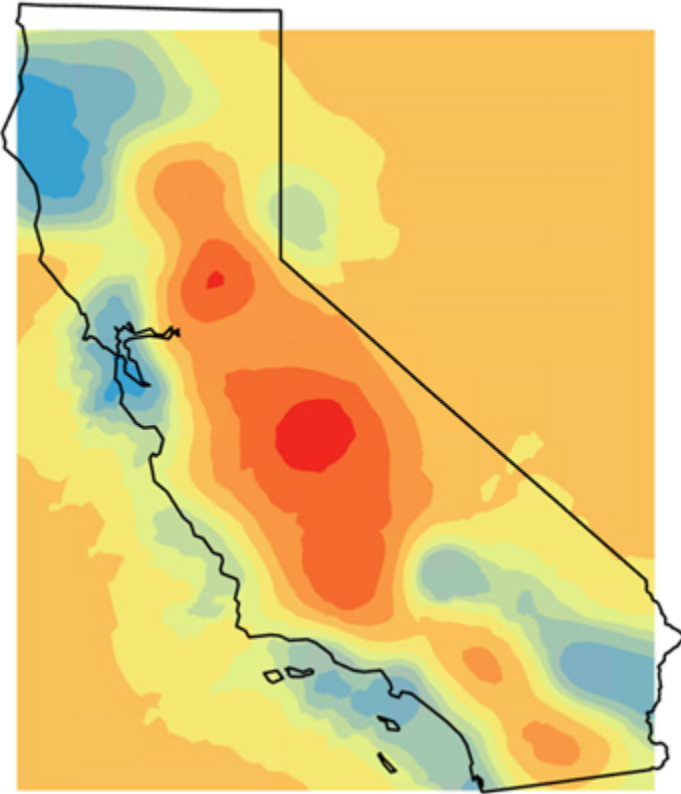


The Method Report window summarizes information on the method and its associated parameters that will be used to create the output surface.

**m** Click OK.



- n If necessary, zoom out in the map, and then turn off the Samples layer.



The surface created is a layer in the project and not a raster dataset in a geodatabase. To save the layer to disk, you would right-click it, point to Export Layer, and choose To Rasters.

- o Turn on the Samples layer.
- p Investigate the sample points by exploring the map.
- q Visually judge how well the default Kriging layer represents the measured ozone values.

In general, do high ozone predictions occur in the same areas where high ozone concentrations were measured?

You have created a geostatistical surface using kriging. Although you did not alter any parameters, you see that there are many important data considerations when you perform kriging.

- r Save the project.

## Step 4: Evaluate predicted value and error

One advantage of geostatistical interpolation is the ability to evaluate the error in the predicted values. You will use the kriging result as the surface to analyze a layer of California cities as point

locations where predictions and validations will be performed. The result is a point layer containing each city, its original attributes, a prediction value, and a standard error value.

You will first create a point layer from the Kriging layer.

- a In the Contents pane, right-click the Kriging layer, point to Export Layer, and choose To Points to open the GA Layer To Points tool.
- b In the Geoprocessing pane, set the following parameters:
  - Input Geostatistical Layer: Kriging
  - Input Point Observation Locations: **C:\EsriTraining\SNAP\Interpolation\CaliOzone.gdb\ca\_cities**
  - Output Statistics At Point Locations: **CityPredict**
- c Run the tool.
- d In the Contents pane, open the CityPredict attribute table.

Each city now has a predicted ozone value, as well as a standard error value that indicates the level of uncertainty associated with the ozone prediction for each city.

- e Close the table.
- f Save the project, and leave ArcGIS Pro open.

## Empirical Bayesian kriging (EBK)

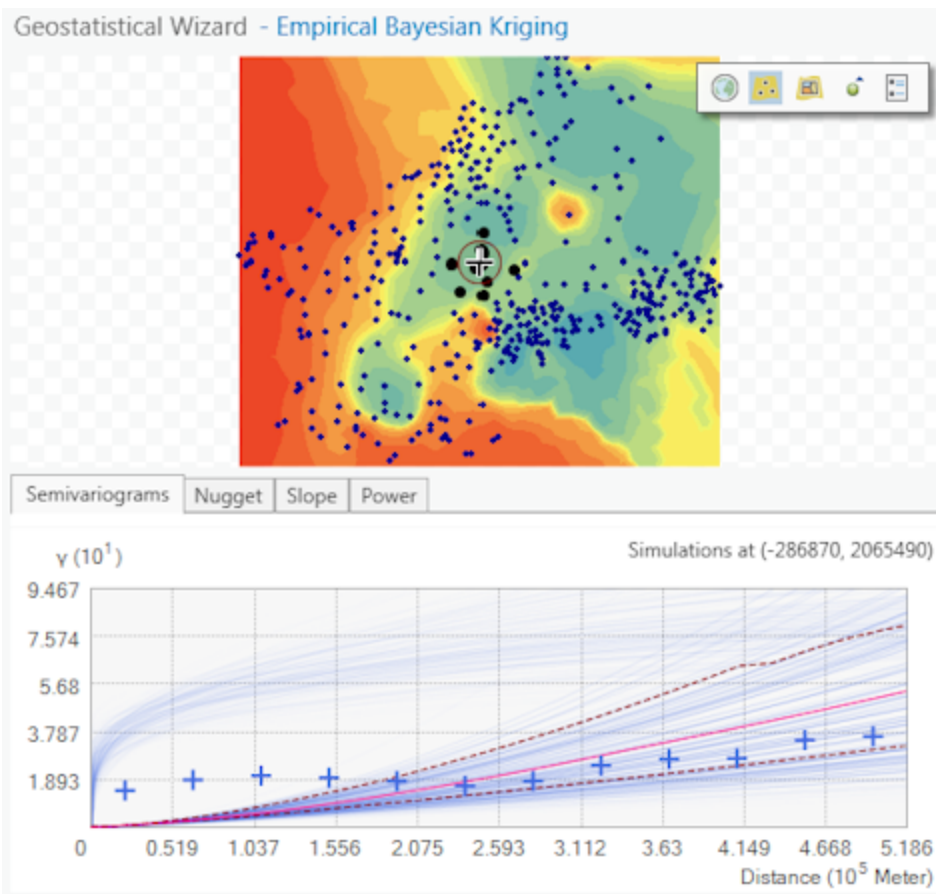
---

Empirical Bayesian kriging (EBK) is a geostatistical interpolation method that automates the most difficult aspects of building a valid kriging model. Other kriging methods in Geostatistical Analyst require you to manually adjust parameters to receive accurate results, but EBK automatically calculates these parameters through a process of subsetting and simulations. EBK offers a data-driven approach to interpolation.

### Advantages

Requires minimal interactive modeling.
Standard errors of prediction are more accurate than other kriging methods.
Allows accurate predictions of moderately nonstationary data.
More accurate than other kriging methods for small datasets.

## Empirical Bayesian kriging (EBK) (continued)



**Figure 12.9.** Simulated semivariograms from EBK.

There is also a tool called Empirical Bayesian Kriging 3D (EBK3D) that allows you to interpolate points in 3D to account for both horizontal and vertical changes in data values. Imagine that you have points that represent greenhouse gas samples throughout the atmosphere at varying altitudes. You could use EBK3D to interpolate values where no samples were recorded.

## Empirical Bayesian kriging (EBK) (continued)

---



**ArcGIS Pro Help:** *What is Empirical Bayesian kriging?*

## Lesson review

---

1. How is geostatistical interpolation different from deterministic interpolation?

---

---

---

---

---

2. What are the assumptions of kriging?

---

---

---

---

Throughout the course, you have witnessed and applied many analysis tools to examine the spatial relationships in data. Imagine being able to see the effects that mountains, valleys, buildings, and other 3D objects have on these relationships. Using 3D GIS, you can detect trends and patterns that are not as apparent in 2D. Further, many analyses questions can only be answered using 3D tools and visualization.

In this lesson, you will learn techniques for analyzing both surface and 3D feature data in ArcGIS Pro to identify patterns not apparent in 2D. You will perform line-of-sight analysis, buffer 3D features, and use 3D overlay tools. 3D capability is included in ArcGIS Pro, so it is unnecessary to have separate apps to handle the 3D visualization and analysis. However, 3D Analyst is required for 3D analysis tools.

## Topics covered

---

When to use 3D analysis

3D analysis examples and tools

## When to use 3D analysis

---

You have performed many types of analyses in the course, but all have been in 2D. You will learn about 3D analysis and situations when it would enhance or make your analysis possible.

**How could your analysis benefit from incorporating 3D, and what are some potential 3D analysis examples?**



## 3D analysis examples

---

There are many uses for 3D analysis, such as analyzing underground resources, determining visibility or line of sight, shadow-volume analysis, and volumetric and area analysis. In ArcGIS Pro, you can view and edit 3D data out of the box. If you want to perform 3D analysis using geoprocessing tools, you must have the 3D Analyst extension.

### Multipatch features

A multipatch feature is a GIS object that stores a collection of patches to represent the boundary of a 3D object as a single row in a database. Patches store texture, color, transparency, and geometric information representing parts of a feature. All multipatches store z-values as part of the coordinates used to construct patches. When you create a feature class, you can specify multipatch as its type, rather than point, line, or polygon.



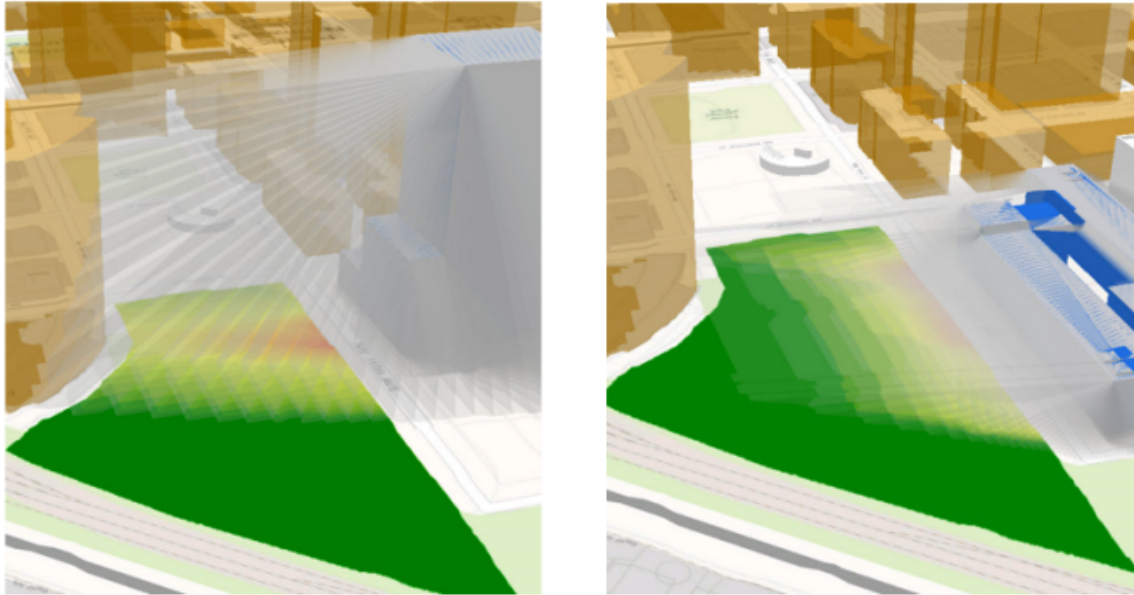
**Figure 13.1.** A common use of multipatch features is to represent buildings.

### Sun-shadow volume

Sun-shadow volume creates volumes that model shadows cast by each feature using sunlight for a given date and time. You can use sun-shadow volume analysis to visualize the effects of a new building on surrounding areas, such as a park.

## 3D analysis examples (continued)

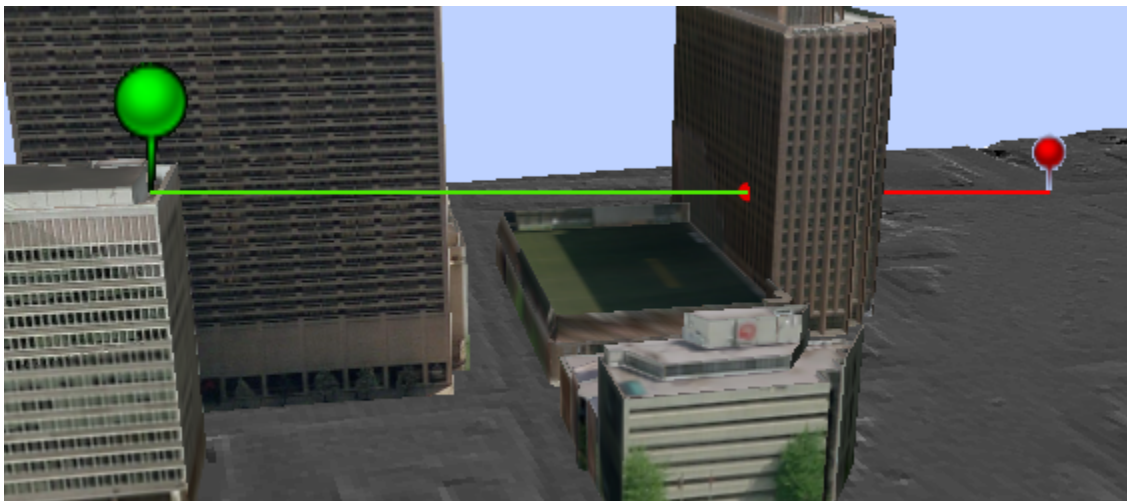
---



**Figure 13.2.** Shadow-impact analysis at different times of the day, or year.

## Line-of-sight analysis

Line-of-sight analysis determines the visibility of sight lines over obstructions consisting of a surface and an optional multipatch dataset.



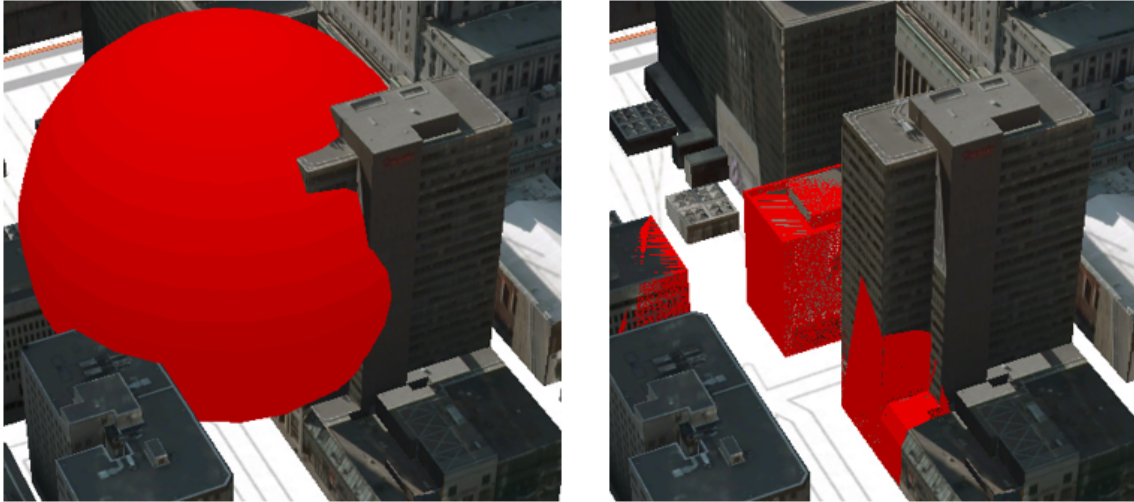
**Figure 13.3.** Line-of-sight analysis between two points.

## 3D analysis examples (continued)

---

### 3D feature analysis

The 3D Features toolset provides a collection of tools for constructing features and assessing geometric properties in three-dimensional space. You can buffer, intersect, and apply union to 3D features as you do with 2D features.



**Figure 13.4.** A 3D buffer intersected with 3D buildings.

- Buffer 3D creates a three-dimensional buffer around points or lines to produce spherical or cylindrical multipatch features.
- Intersect 3D computes the intersection of multipatch features to produce closed multipatches encompassing the overlapping volumes, open multipatch features from the common surface areas, or lines from the intersecting edges.

## Interactive 3D analysis

---

Exploratory analysis in 3D is a way of performing various forms of quick investigation by interactively creating graphics and editing analysis parameters on the fly. The interactive tools create analytical objects by clicking in the scene or using input source layers. You can manipulate analysis parameters and receive real-time visual feedback in the scene.



**Figure 13.5.** Interactive 3D analysis tools: Viewshed (on the left) and Line Of Sight (on the right).

## Interactive 3D analysis (continued)

---

There are four exploratory tools:

- **Line Of Sight** creates sight lines to determine if one or more targets are visible from a given observer location.
- **View Dome** determines the parts of a sphere that are visible from an observer located at the center.
- **Viewshed** determines the visible surface area from a given observer location through a defined viewing angle.
- **Slice** temporarily suppresses part of a scene's display to reveal hidden content. It can be applied to any content in the scene, making it possible to see inside buildings, explore stacked volumes, and push through subsurface geology.

Each tool uses a different method to achieve visibility analysis and has customizable creation methods and parameter values. The analysis feedback in the view is color-coded to distinguish what is obstructed, unobstructed, and out of range.



[ArcGIS Pro Help: Exploratory analysis tools](#)

## Perform 3D analysis

You will use 3D buildings and other features in Montreal, Quebec, to perform visibility analysis using 3D Analyst tools. You will also use 3D features tools to buffer and intersect 3D features.

In this exercise, you will perform the following tasks:

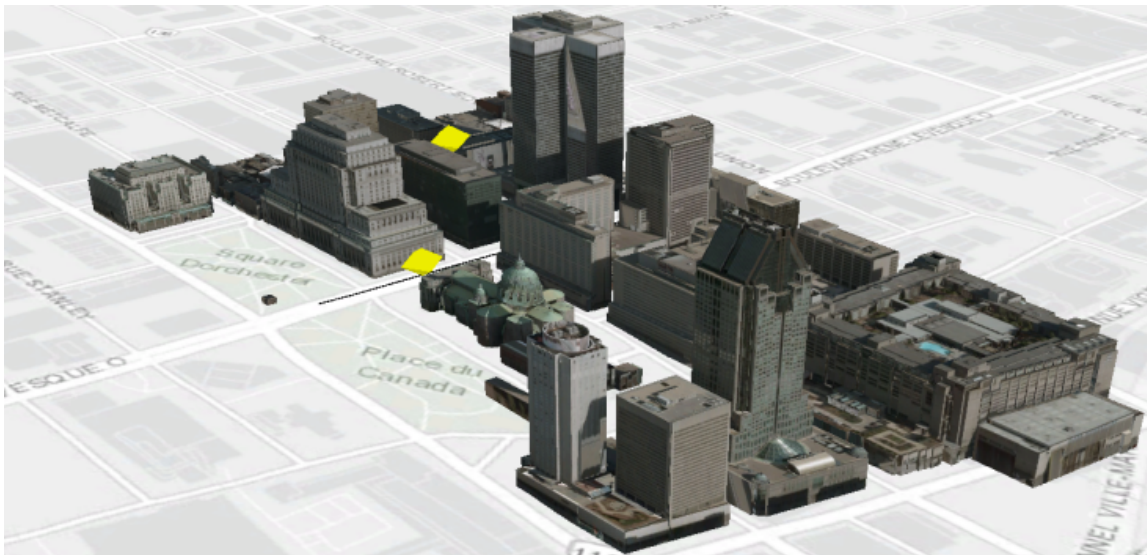
- Perform line-of-sight analysis.
- Buffer and intersect 3D features.



## Step 1: Set up the project

Before you work with 3D analysis tools, you will import a scene and set some environments.

- a If necessary, start or restore ArcGIS Pro and view the SNAPCourse project.
- b From the Insert tab, click Import Map.
- c Browse to **C:\EsriTraining\SNAP\3D Analysis** and import Montreal.mapx.



The map file contains the definition for a 3D scene that has buildings, a route, and observer points for Montreal, Quebec. Next, you will set analysis environments.

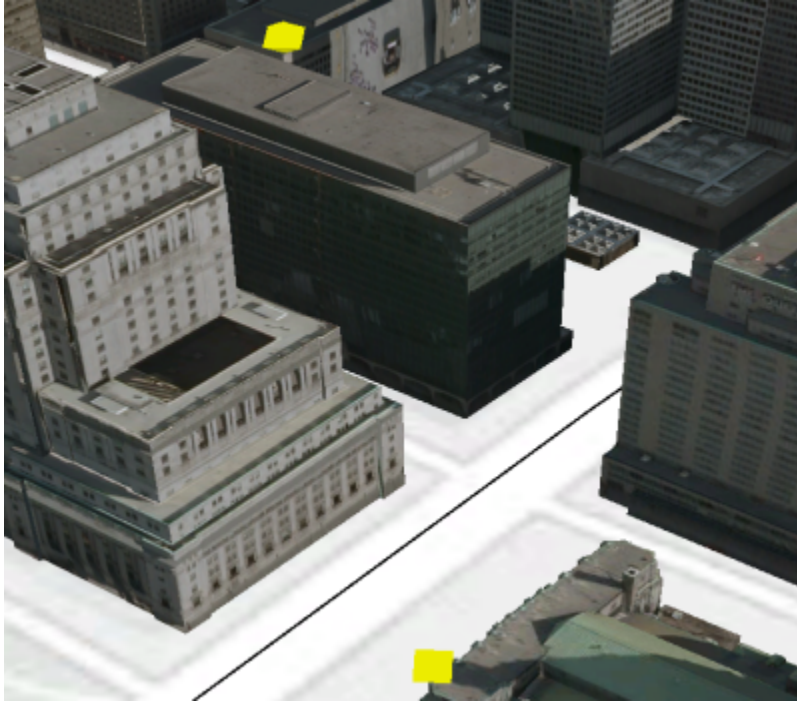
- d From the Analysis tab, click Environments.
- e For Output Coordinate System, choose Buildings.
- f For Extent, choose Same As Layer - Buildings.
- g For Mask, clear the mask if one is set.
- h Click OK.

## Step 2: Create sight lines

Special events—sporting events, concerts, or parades—can attract thousands of people, presenting security concerns for security planners and law enforcement. A key part of planning security at such events involves observing crowds. Surveillance can be achieved by placing surveillance cameras in strategic locations. Many law enforcement agencies may position officers

on rooftops or other high vantage points to observe crowd behavior during special events. In this exercise, you will perform a line-of-sight analysis for an event in Montreal.

- a In the Contents pane, right-click the Observers layer and choose Zoom To Layer.
- b Tilt and navigate the scene to see the observers and the route.



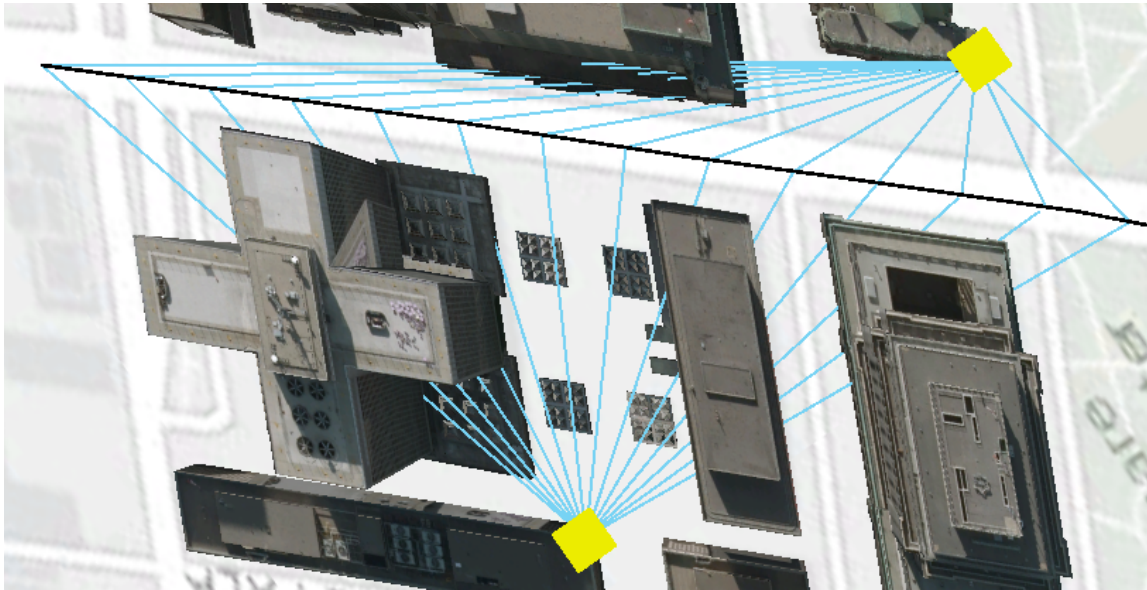
Your view may be different.

The two yellow points represent observers, either a camera or a person. You will use 3D functionality to determine what each potential observer can see along the route. First, you will create lines, called sight lines, between each of your observer points and the route. Sight lines are a required parameter in the Line Of Sight tool. You will space these lines 30 feet apart along the route.

- c In the Geoprocessing pane, view the toolsets in the 3D Analyst Tools toolbox.
- d Expand Visibility.
- e Open the Construct Sight Lines tool, and set the following parameters:
  - Observer Points: Observers
  - Target Features: Route
  - Output: **SightLines**
  - Sampling Distance: **30**



**f** Click Run.



Sight lines are added from each observer to the route. Some of these sight lines are obstructed by buildings, while others are not.

**g** In the Contents pane, turn off the SightLines layer.

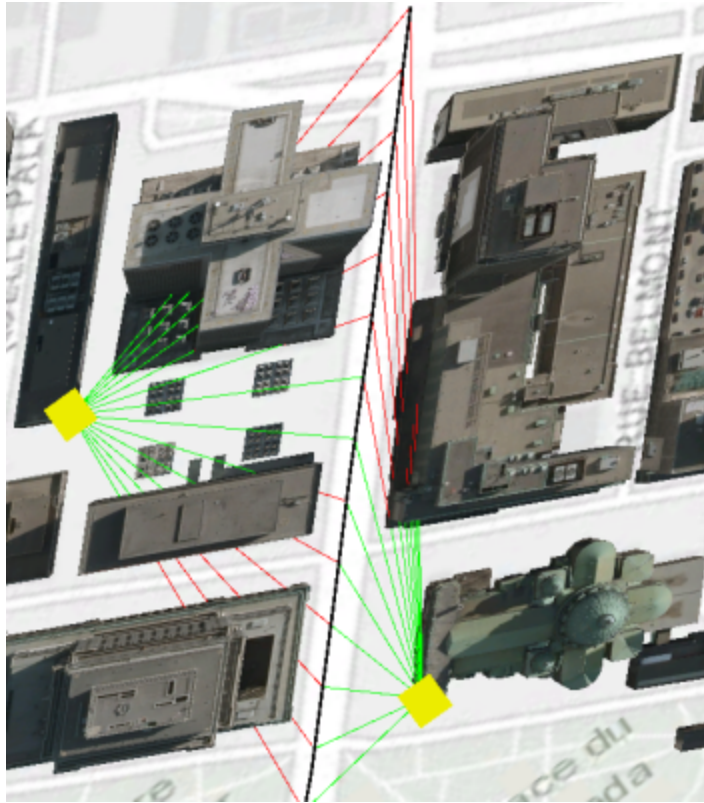
### Step 3: Perform line-of-sight analysis

Next, you will use the sight lines as an input for line-of-sight analysis.

**a** In the Geoprocessing pane, open the **Line Of Sight** tool, and then set the following parameters:

- Input Surface: MontrealDEM
- Input Line Features: SightLines
- Input Features: Buildings
- Output Feature Class: **LOS**

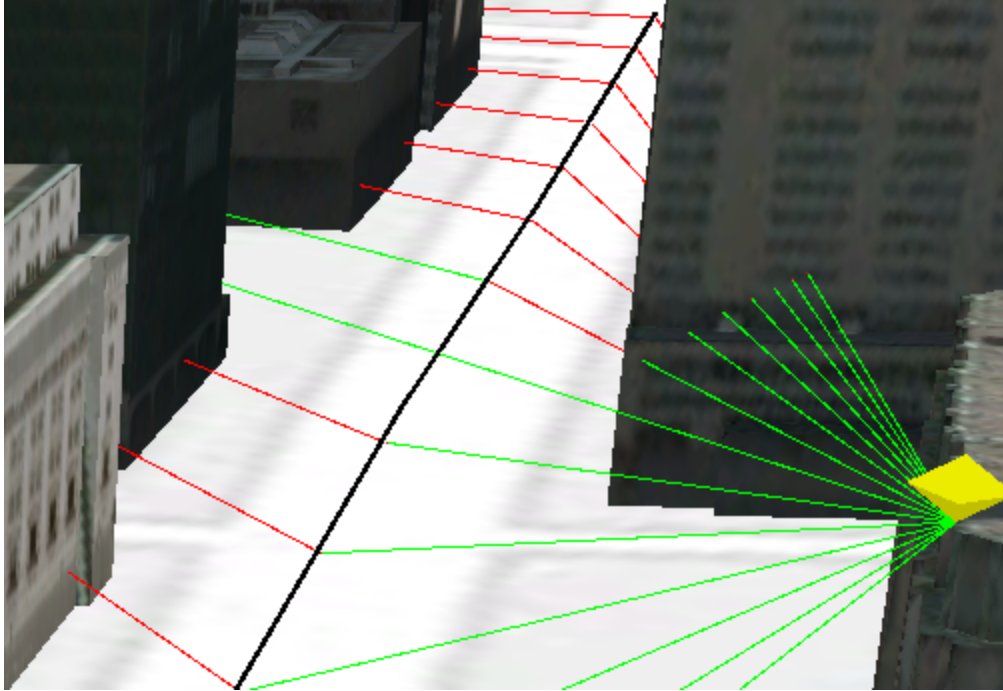
**b** Click Run.



LOS  
VisCode  
— Visible  
— Not visible

In the LOS layer, the green lines indicate visible lines of sight. The red lines indicate sight lines that are blocked (that is, the buildings block the observers' visibility of portions of the route).

- c Navigate around the scene to view the sight lines and how they are affected by the buildings.



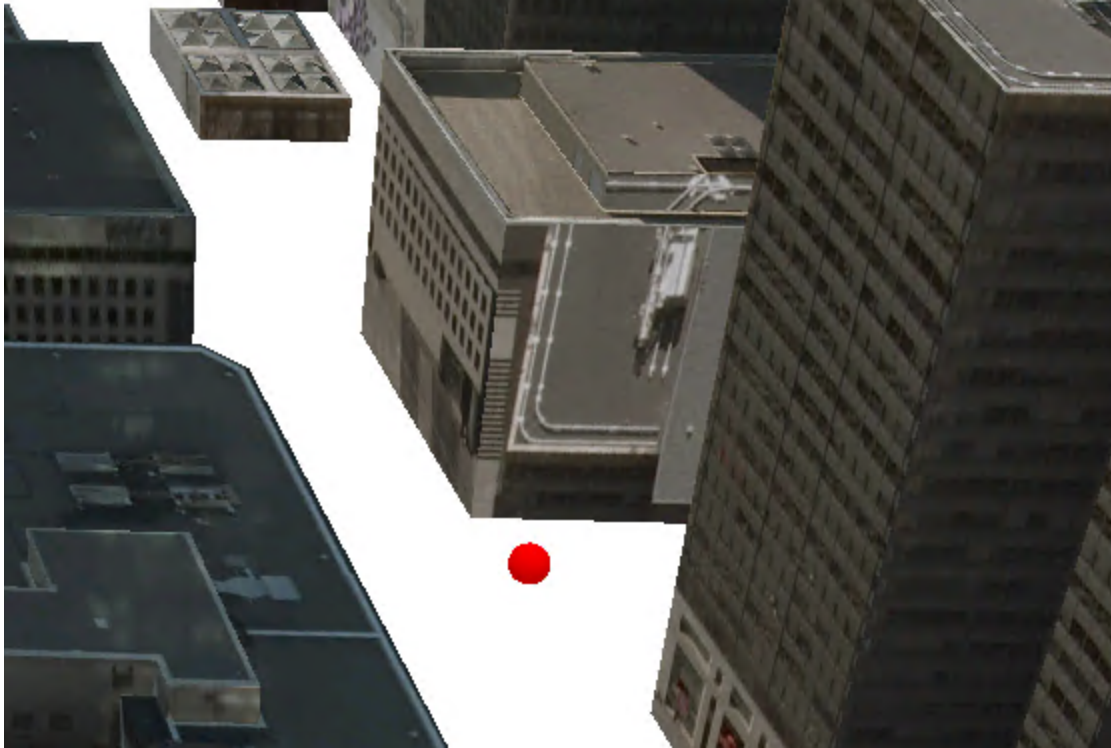
Sometimes, a line of sight can start as visible but become red (not visible) when it is obstructed.

- d Save the project.

#### Step 4: Create a 3D buffer

The next scenario involves a fictitious steam pipe explosion in a city. The pipe is potentially wrapped in asbestos, a known carcinogen, so the site must be declared an "asbestos containment area." The blast could have cracked windows and window seals, allowing contamination to enter buildings. The question that emerged was which buildings in the area would need to be sampled based on a set distance from the blast. You will use a 3D buffer and overlay tool to help answer the question about which buildings should be sampled.

- a In the Contents pane, make PipeExplosion and Buildings the only visible layers.
- b From the Map tab, click Bookmarks, and then choose Pipe Explosion.



The red symbol indicates where the steam pipe burst. You have used the Buffer tool to create buffers for 2D features, but there is another buffer tool specifically for 3D features that you will use for this analysis.

- c In the Geoprocessing pane, return to the 3D Analyst Tools toolbox and expand the 3D Features toolset.



The 3D Features toolset contains comparable tools for performing buffer, intersect, union, and near operations on 3D features.

- d Open the Buffer 3D tool, and set the following parameters:

- Input Features: PipeExplosion
- Output Feature Class: **Pipe3DBuffer**
- Distance: **50 Meters**

- e Click Run.

- f In the Contents pane, change the color of the Pipe3DBuffer layer to a bright red.

*Hint:* Right-click the symbol.



A planar buffer would be flat on a 2D surface, but because you are working in a 3D scene and using 3D analysis tools, the buffer is a multipatch feature that contains z-values.

**g** Save the project.

## Step 5: Intersect 3D features

Next, you will use a 3D intersect tool to find only the intersecting parts of the buildings. As with the intersect tools for 2D features, a 3D intersect combines attributes of intersecting features. In this example, you could determine building names, addresses, and possible resident information for the affected areas.

**a** In the Geoprocessing pane, return to the 3D Analyst Tools toolbox, open the Intersect 3D tool, and set the following parameters:

- Input Multipatch Features: Pipe3DBuffer
- Input Multipatch Features: Buildings
- Output Feature Class: **AffectedAreas**

**b** Click Run.

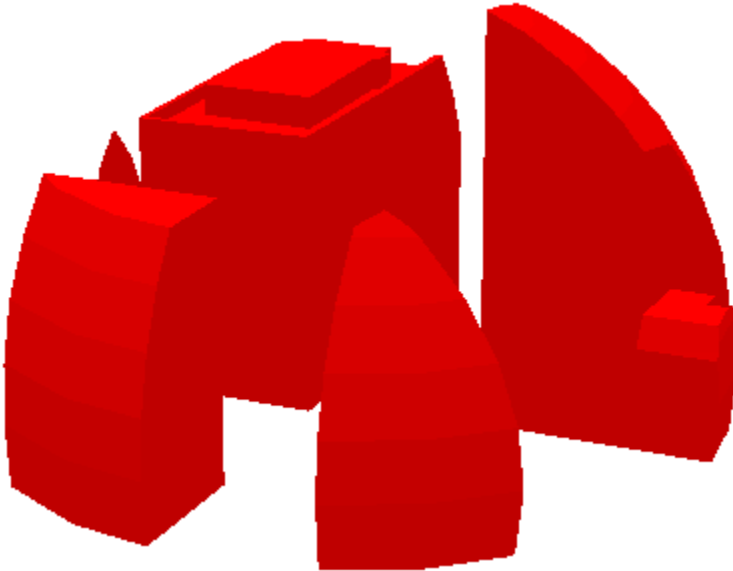


Disregard the warning. It appears because some multipatch features in the Buildings feature class are not fully enclosed, but this issue will not affect your results.

- c In the Contents pane, turn off the Pipe3DBuffer layer.
- d Tilt the scene to view the intersected features from various angles.

The texture of the buildings can interfere with the display of the overlapping features, so you can turn off the Buildings layer to better see the intersect result.

- e Make AffectedAreas the only visible layer.



The areas where the buildings are red is where the blast radius intersected with the buildings and should be tested for contamination.

With the Buildings layer off, you can clearly see the areas where the buildings and buffer intersected. If you have the appropriate data, you can also use the result of the intersection to select building interior features, such as rooms. This action will quickly give you a list of locations to check for breached windows for possible contamination.

- f Open the AffectedAreas attribute table.

BUFF_DIST	BUFF_QUAL	SourceID_2	BuildingShell Feature	Name	Description	Attribution/Source
50	20	3	BuildingShell/7/1			VM08_2009/GroundS...
50	20	5	BuildingShell/15/1			VM08_2009/GroundS...
50	20	10	BuildingShell/27/1			VM08_2009/GroundS...
50	20	31	BuildingShell/119/1			VM08_2009/GroundS...
50	20	33	BuildingShell/127/1			VM08_2009/GroundS...

The attributes for the buffer and the buildings are in the AffectedAreas layer. If the attribute table had more information, such as address, occupant name, and so on, it would also be in the table.

- g Close the table.

- h Save the project, and then exit ArcGIS Pro.

## Lesson review

---

1. **The core functionality of ArcGIS Pro includes 3D visualization. Does the core functionality also include 3D analysis geoprocessing tools?**

---

---

---

---

2. **What are some benefits of 3D analysis?**

---

---

---



# Answers to Lesson 13 questions

## When to use 3D analysis (page 13-2)

How could your analysis benefit from incorporating 3D, and what are some potential 3D analysis examples?

**Possible responses include the following:**

**Benefits:** Gain insight into the landscape and how real-world features relate to one another, for enhanced visualization and analysis, to solve spatial problems for which 3D is essential.

**Examples:** Underground or subsurface analysis (subways, mines, geologic), visibility analysis, obstruction analysis, and sun shadow volume analysis.

# Appendix A

## Esri data license agreement

ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, INC. ("ESRI"), IS WILLING TO LICENSE THE ENCLOSED ELECTRONIC VERSION OF THE TRAINING MATERIALS TO THE STUDENT ("YOU") ONLY UPON THE CONDITION THAT YOU ACCEPT ALL TERMS AND CONDITIONS CONTAINED IN THIS ESRI DATA LICENSE AGREEMENT ("AGREEMENT"). PLEASE READ THE TERMS AND CONDITIONS CAREFULLY. BY CLICKING, "I ACCEPT", YOU ARE INDICATING YOUR ACCEPTANCE OF THE ESRI DATA LICENSE AGREEMENT. IF YOU DO NOT AGREE TO THE TERMS AND CONDITIONS AS STATED, ESRI IS UNWILLING TO LICENSE THE TRAINING MATERIALS TO YOU.

**Training Materials Reservation of Ownership.** This Agreement gives You certain limited rights to use electronic and tangible versions of the digital or printed content required to complete a course, which may include, but are not limited to, workbooks, data, concepts, exercises, and exams ("Training Materials"). Esri and its licensor(s) retain exclusive rights, title, and ownership to the copy of Training Materials, software, data, and documentation licensed under this Agreement. Training Materials are protected by United States copyright laws and applicable international copyright treaties and/or conventions. All rights not specifically granted in this Agreement are reserved to Esri and its licensor(s).

**Grant of License.** Esri grants to You a personal, nonexclusive, nontransferable license to use Training Materials for Your own training purposes. You may run and install one (1) copy of Training Materials and reproduce one (1) copy of Training Materials. You may make one (1) additional copy of the original Training Materials for archive purposes only, unless Esri grants in writing the right to make additional copies.

Training Materials are intended solely for the use of the training of the individual who registered and attended a specific training course. You may not (i) separate the component parts of the Training Materials for use on multiple systems or in the cloud, use in conjunction with any other software package, and/or merge and compile into a separate database(s) or documents for other analytical uses; (ii) make any attempt to circumvent the technological measure(s) (e.g., software or hardware key) that effectively controls access to Training Materials; (iii) remove or obscure any copyright, trademark, and/or proprietary rights notices of Esri or its licensor(s); or (iv) use audio and/or video recording equipment during a training course.

**Term.** The license granted by this Agreement will commence upon Your receipt of the Training Materials and continue until such time that (1) You elect to discontinue use of the Training Materials or (2) Esri terminates this Agreement for Your material breach of this Agreement. This Agreement will be terminated automatically without notice if You fail to comply with any provision of this Agreement. Upon termination of this Agreement in either instance, You will return to Esri or destroy all copies of the Training Materials, including any whole or partial copies in any form, and

# Appendix A

## Esri data license agreement (continued)

deliver evidence of such destruction to Esri, and which evidence will be in a form acceptable to Esri in its sole discretion. The parties hereby agree that all provisions that operate to protect the rights of Esri and its licensor(s) will remain in force should breach occur.

**Limited Warranty.** Esri warrants that the media on which Training Materials is provided will be free from defects in materials and workmanship under normal use and service for a period of ninety (90) days from the date of receipt.

**Disclaimer of Warranties.** EXCEPT FOR THE LIMITED WARRANTY SET FORTH ABOVE, THE TRAINING AND TRAINING MATERIALS CONTAINED THEREIN ARE PROVIDED "AS IS," WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT. ESRI DOES NOT WARRANT THAT THE TRAINING OR TRAINING MATERIALS WILL MEET YOUR NEEDS OR EXPECTATIONS; THAT THE USE OF TRAINING MATERIALS WILL BE UNINTERRUPTED; OR THAT ALL NONCONFORMITIES, DEFECTS, OR ERRORS CAN OR WILL BE CORRECTED. THE TRAINING DATABASE HAS BEEN OBTAINED FROM SOURCES BELIEVED TO BE RELIABLE, BUT ITS ACCURACY AND COMPLETENESS, AND THE OPINIONS BASED THEREON, ARE NOT GUARANTEED. THE TRAINING DATABASE MAY CONTAIN SOME NONCONFORMITIES, DEFECTS, ERRORS, AND/OR OMISSIONS. ESRI AND ITS LICENSOR(S) DO NOT WARRANT THAT THE TRAINING DATABASE WILL MEET YOUR NEEDS OR EXPECTATIONS, THAT THE USE OF THE TRAINING DATABASE WILL BE UNINTERRUPTED, OR THAT ALL NONCONFORMITIES CAN OR WILL BE CORRECTED. ESRI AND ITS LICENSOR(S) ARE NOT INVITING RELIANCE ON THIS TRAINING DATABASE, AND YOU SHOULD ALWAYS VERIFY ACTUAL DATA, SUCH AS MAP, SPATIAL, RASTER, OR TABULAR INFORMATION. THE DATA CONTAINED IN THIS PACKAGE IS SUBJECT TO CHANGE WITHOUT NOTICE. IN ADDITION TO AND WITHOUT LIMITING THE PRECEDING PARAGRAPH, ESRI DOES NOT WARRANT IN ANY WAY TRAINING DATA. TRAINING DATA MAY NOT BE FREE OF NONCONFORMITIES, DEFECTS, ERRORS, OR OMISSIONS; BE AVAILABLE WITHOUT INTERRUPTION; BE CORRECTED IF ERRORS ARE DISCOVERED; OR MEET YOUR NEEDS OR EXPECTATIONS. YOU SHOULD NOT RELY ON ANY TRAINING DATA UNLESS YOU HAVE VERIFIED TRAINING DATA AGAINST ACTUAL DATA FROM DOCUMENTS OF RECORD, FIELD MEASUREMENT, OR OBSERVATION.

**Exclusive Remedy.** Your exclusive remedy and Esri's entire liability for breach of the limited warranties set forth above will be limited, at Esri's sole discretion, to (i) replacement of any defective Training Materials; (ii) repair, correction, or a workaround for Training Materials; or (iii) return of the fees paid by You for Training Material that do not meet Esri's limited warranty, provided that You uninstall, remove, and destroy all copies of the Training Materials and execute and deliver evidence of such actions to Esri.

# Appendix A

## Esri data license agreement (continued)

IN NO EVENT WILL ESRI BE LIABLE TO YOU FOR COSTS OF PROCUREMENT OF SUBSTITUTE GOODS OR TRAINING; LOST PROFITS; LOST SALES; BUSINESS EXPENDITURES; INVESTMENTS; BUSINESS COMMITMENTS; LOSS OF ANY GOODWILL; OR ANY INDIRECT, SPECIAL, EXEMPLARY, CONSEQUENTIAL, OR INCIDENTAL DAMAGES ARISING OUT OF OR RELATED TO THIS AGREEMENT, HOWEVER CAUSED OR UNDER ANY THEORY OF LIABILITY, EVEN IF ESRI HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. ESRI'S TOTAL CUMULATIVE LIABILITY HEREUNDER, FROM ALL CAUSES OF ACTION OF ANY KIND, WILL IN NO EVENT EXCEED THE AMOUNT ACTUALLY PAID BY YOU FOR THE PORTION OF THE TRAINING UNDER THIS AGREEMENT. THESE LIMITATIONS WILL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY.

**Export Regulation.** You must comply with all applicable laws and regulations of the United States including, without limitation, its export control laws. You expressly acknowledge and agree not to export, reexport, transfer, or release Esri-provided Training Materials, in whole or in part, to (i) any US embargoed country (including to a resident of any US embargoed country); (ii) any person or entity on the US Treasury Department Specially Designated Nationals List; (iii) any person or entity on the US Commerce Department Lists of Parties of Concern; or (iv) any person or entity where such export, reexport, or provision violates any US export control laws or regulations including, but not limited to, the terms of any export license or licensing provision and any amendments and supplemental additions to US export laws.

**Governing Law.** This Agreement is governed by and construed in accordance with the laws of the state in which training is being held or, in the case of training provided over the Internet, the laws of the State of California, without reference to its conflict of laws principles.

# Appendix B

## Answers to lesson review questions

### Answers to lesson 1 review questions

1. What are the six types of spatial analysis tools?  
**The six types are Temporal, Proximity, Overlay, Statistical, Network, and 3D.**
2. What helps you choose the appropriate datasets for your analysis?  
**Analysis criteria helps you choose.**

### Answers to lesson 2 review questions

1. What are some things to consider when preparing data for analysis?  
**Confirm that these data properties are suitable for your organizational standards:**
  - Data format (tabular or spatial)
  - Data source
  - Quality
  - Currency
  - Extent
  - Spatial reference
  - Scale
  - Raster resolution
  - Attributes**Enhance data to prepare it for analysis:**
  - Add attributes
  - Calculate values
  - Join fields
  - Edit features and attributes
  - Modify spatial reference
  - Display XY data
  - Extract features of interest

# Appendix B

## Answers to lesson review questions (continued)

2. How can environment settings help streamline your analysis workflows?

**Environment settings create standardized parameters for certain tools. For example, you can set an extent environment so that all results have the same extent.**

3. What should you consider when selecting an output cell size?

**Consider the cell size of the input rasters. The output cell size should not be any smaller than the cell size of the input rasters.**

### Answers to lesson 3 review questions

1. Explain the three ways that ArcGIS Pro measures proximity.

**Euclidean (straight-line distance), geodesic (considers earth's curvature), and cost (best route, driving times)**

2. Explain the difference between using a straight-line distance and using cost.

**Straight-line is an as-the-crow-flies distance. But for some applications, such as routing, a straight line does not accurately reflect distance. You can use cost, such as time, to create driving time data that accurately reflects traffic, population, and other factors.**

### Answers to lesson 4 review questions

1. What is overlay analysis?

**Overlay analysis is the geometric intersection of multiple datasets to combine, erase, modify, or update features in a new output dataset. Overlay analysis answers the question: What is on top of what?**

# Appendix B

## Answers to lesson review questions (continued)

2. If you use the Intersect tool with streams and watersheds as the inputs, what would the resulting feature class contain?

**Overlay tools output the simpler of the geometries from the inputs, so the output would contain the streams that fell within the watersheds. Further, the attribute table would have both stream and watershed attributes. Users could query streams and determine which watersheds that they fall within, or the other way around.**

### Answers to lesson 5 review questions

1. What are the methods for automating processes in ArcGIS Pro?

**The methods are batch geoprocessing, ModelBuilder, Python, and tasks.**

2. Why would you set model parameters for your model elements and variables?

**You set model parameters to share your model with other users who want to run it with their own data.**

### Answers to lesson 6 review questions

1. Describe interpolation.

**Interpolation is the process of estimating unknown values from a sample of known values. Typically, the input samples are points that contain a discretely captured value that you want to represent as a continuous surface.**

2. What is deterministic interpolation?

**Deterministic methods use mathematical functions (nonstatistical) for creating surfaces from measured points based either on extent of similarity or on the degree of smoothing.**

3. What are some ways in which you can validate surfaces created using interpolation?

**Use the Explore tool to click sample points and the surface at the same location to compare estimated values. You can also interpolate on a subset of your sample points and then use the withheld points to see how well the interpolator estimated values.**

# Appendix B

## Answers to lesson review questions (continued)

### Answers to lesson 7 review questions

1. What is the difference between binary and weighted overlay?  
**Binary determines whether a cell is either suitable or unsuitable while weighted ranks cells from low to high, based on suitability.**
2. Explain the difference between the Reclassify tool and the Rescale By Function tool.  
**Reclassify is when the user manually sets class breaks and which values go into each class. Rescale By Function is when the software determines the classes based on a function and your input data. Reclassify is best for discrete data and Rescale By Function is best for continuous data.**

### Answers to lesson 8 review questions

1. Explain how spatial statistics remove subjectivity from your data.  
**Displaying data using classification methods is highly subjective because mapmakers can show exactly what they want to portray to an audience. Spatial statistics remove that subjectivity by quantifying spatial patterns and relationships and by providing statistics to give you confidence in your decisions.**
2. Briefly describe descriptive and inferential statistics.  
**Descriptive tools summarize the data that you provide and return a value or a chart that describes a trend or central tendency. Inferential statistics tools test a sample and provide statistical values that represent the entire population and indicate whether you can reject that null hypothesis.**

### Answers to lesson 9 review questions

1. Explain how adding time can enhance your analysis results.  
**You can get a clearer picture of what is happening currently with your data, rather than having incidents from years ago influence the results.**



# Appendix B

## Answers to lesson review questions (continued)

2. Explain spatial and temporal variance.

**Spatial variance is how values differ at different geographic locations. Temporal variance is how a value at the same location varies over time.**

3. Differentiate between analyzing time snapshots of data and true space-time analysis.

**Analyzing by time snapshots groups features into arbitrary bins based on a day, week, month, year, and so on. Time snapshots may break up related data and do not give you the whole story. Space-time analysis assesses each feature separately so that other features within a specified time period will be analyzed together regardless of whether they spill over two months.**

### Answers to lesson 10 review questions

1. Explain the key components of the regression equation.

- **Dependent variable:** What you are trying to explain or predict.
- **Independent variable:** The factors that you believe contribute to or influence the variation in the dependent variable.
- **Coefficients:** A value associated with each independent variable in a regression equation, representing the strength and type of relationship that the independent variable has to the dependent variable.
- **Residuals:** The over- and under-predictions in the model, or the differences between actual observed values and predicted values.

2. Explain what the AIC value represents and how to use it.

**AIC is useful for comparing models using the same dependent variable, and you want this value to be lower. AIC is only comparable between models using the same dependent variable.**

# Appendix B

## Answers to lesson review questions (continued)

3. You want to perform OLS regression analysis but do not have key independent variables in the attribute table. What can you do to get the required information in the table?

**You could manually add attributes or join fields from other data sources. Further, if you do not have other data sources to get information, use the Enrich Layer tool to add attributes from ArcGIS Online.**

### Answers to lesson 11 review questions

1. Explain the differences between OLS and GWR.

**OLS is a global regression model and GWR is local. Global means one equation for all features and local means one equation for each feature. OLS does not account for spatial variation and GWR does. You can use GWR to make predictions based on estimated data. GWR does not provide the diagnostics to perform the six statistical checks, whereas OLS does.**

2. When should you run GWR?

**After you find a properly specified OLS model, you can then use the same variables in GWR. You want to predict alternative or future values using a model calibrated with existing values or the OLS diagnostics indicate statistically significant nonstationarity.**

### Answers to lesson 12 review questions

1. How is geostatistical interpolation different from deterministic interpolation?

**Deterministic interpolation fits a surface through your sample points using math functions to predict unknown values, whereas geostatistical interpolation uses spatial autocorrelation and statistics to predict unknown values.**

2. What are the assumptions of kriging?

**Assumptions include spatial continuity, spatial autocorrelation, stationarity, normally distributed, global trends, and spatial clustering.**

# Appendix B

## Answers to lesson review questions (continued)

### Answers to lesson 13 review questions

1. The core functionality of ArcGIS Pro includes 3D visualization. Does the core functionality also include 3D analysis geoprocessing tools?

**No, the 3D Analyst extension is required for 3D geoprocessing tools.**

2. What are some benefits of 3D analysis?

- **Enhanced visualization**
- **Ability to perform analyses not possible in 2D**
- **Opportunity to gain another perspective about your data**

# Appendix C

## Additional resources

Lesson 3	Resources
Choosing the best distance measure	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>How proximity tools calculate distance</i></li></ul>
Measuring cost	<ul style="list-style-type: none"><li>• Esri Training courses: <a href="#">Creating Optimized Routes Using ArcGIS Pro</a>, <a href="#">Creating an Origin-Destination Cost Matrix in ArcGIS Pro</a>, and <a href="#">Finding the Closest Facilities Using ArcGIS Pro</a></li></ul>

Lesson 5	Resources
Automation methods in ArcGIS Pro	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Create a new task</i></li><li>• Esri Training course: <a href="#">Automating Workflows Using ArcGIS Pro Tasks</a></li></ul>

Lesson 6	Resources
Interpolation methods	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Deterministic methods for spatial interpolation</i></li><li>• ArcGIS Pro Help: <i>What are geostatistical interpolation techniques?</i></li></ul>
Interpolation tools	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Classification trees of the interpolation methods offered in Geostatistical Analyst</i></li></ul>
Deterministic interpolation	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Subset Features</i></li><li>• ArcGIS Pro Help: <i>GA Layer To Points</i></li><li>• ArcGIS Pro Help: <i>Performing cross-validation and validation</i></li></ul>

# Appendix C

## Additional resources (continued)

Lesson 7	Resources
Types of raster overlay	<ul style="list-style-type: none"><li>• Esri Training course: <a href="#">Using Raster Data for Site Selection</a></li></ul>

Lesson 8	Resources
Types of spatial statistics	<ul style="list-style-type: none"><li>• Esri Press: <i>The Esri Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics</i></li></ul>
Interpreting inferential statistics	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>What is a z-score? What is a p-value?</i></li></ul>

Lesson 9	Resources
Space-time analysis	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Why hexagons?</i></li></ul>

Lesson 10	Resources
Six OLS checks	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Regression analysis basics</i></li><li>• ArcGIS Pro Help: <i>What they don't tell you about regression analysis</i></li></ul>
Exploratory regression	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>How Exploratory Regression works</i></li></ul>

Lesson 11	Resources
GWR in action	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Interpreting GWR results</i></li></ul>

# Appendix C

## Additional resources (continued)

Lesson 12	Resources
Kriging	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>How kriging works</i></li><li>• ArcGIS Pro Help: <i>Kriging in Geostatistical Analyst</i></li><li>• ArcGIS Pro Help: <i>Understanding how to create surfaces using geostatistical techniques</i></li></ul>
Geostatistical workflow	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>The geostatistical workflow</i></li><li>• ArcGIS Pro Help: <i>Essential vocabulary for Geostatistical Analyst</i></li><li>• ArcGIS Pro Help: <i>Understanding the semivariogram: The range, sill, and nugget</i></li><li>• ArcGIS Pro Help: <i>Modeling a semivariogram</i></li><li>• Esri Training course: <a href="#">Exploring Spatial Patterns in Your Data Using ArcGIS</a></li></ul>
Use the Geostatistical Wizard to perform kriging	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Search neighborhoods</i></li></ul>
Empirical Bayesian kriging (EBK)	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>What is Empirical Bayesian kriging?</i></li></ul>

Lesson 13	Resources
Interactive 3D analysis	<ul style="list-style-type: none"><li>• ArcGIS Pro Help: <i>Exploratory analysis tools</i></li></ul>

**Contact Esri Training Services**

T 909 793 2853 x1585

F 909 793 5953

[esri.com/training](http://esri.com/training)

380 New York Street

Redlands, California 92373-8100 USA



**esri**

**THE  
SCIENCE  
OF  
WHERE®**