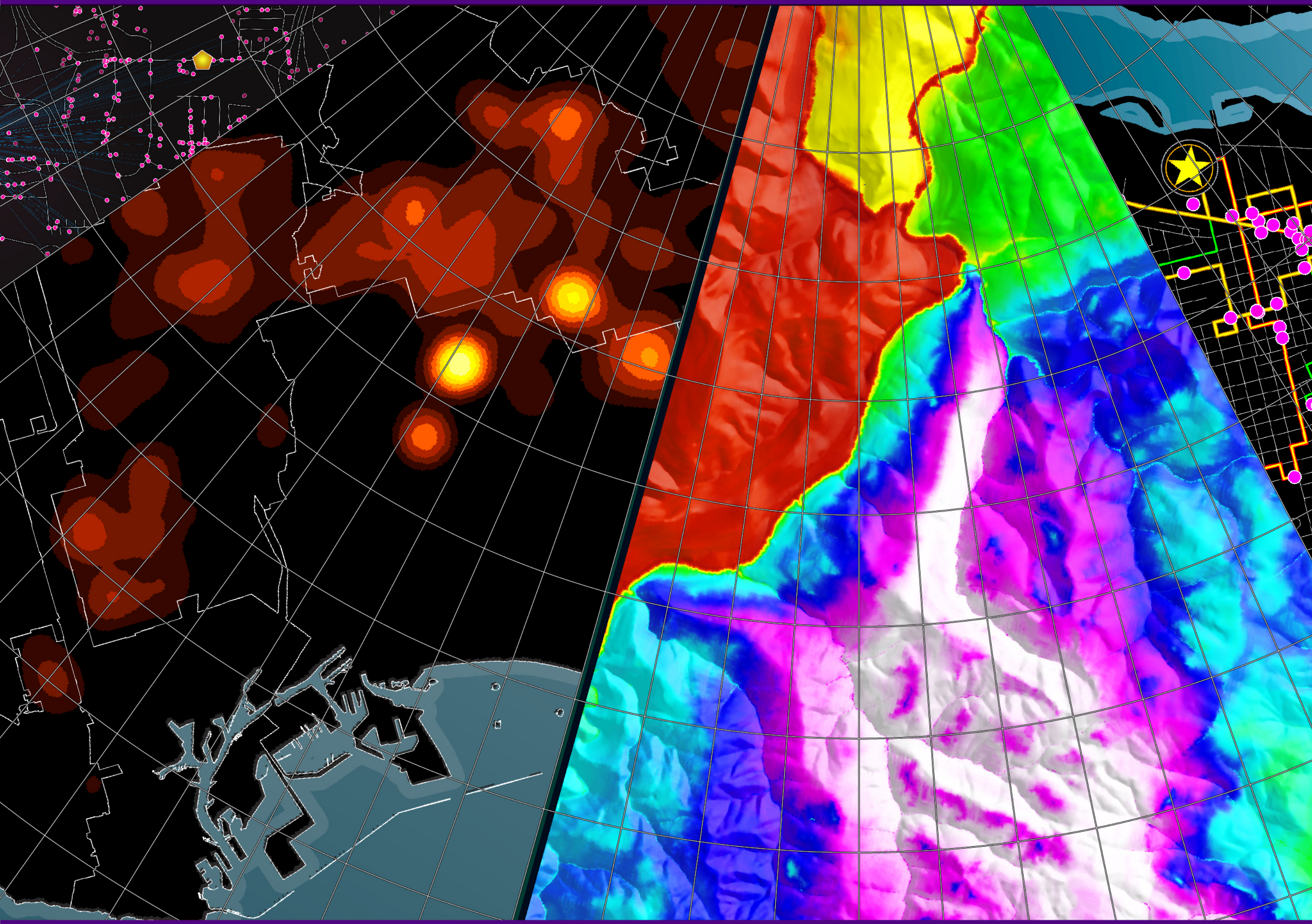
2020

Essential Modeling Techniques for Geospatial Analysis Using ArcGIS: An Intermediate-Level GIS Workbook

Nicolas R. Malloy
Humboldt State University

Essential Modeling Techniques for Geospatial Analysis Using ArcGIS

An Intermediate-Level GIS Workbook, First Edition



Nicolas R. Malloy

GEOSPATIAL INSTITUTE

Essential Modeling Techniques for Geospatial Analysis Using ArcGIS

An Intermediate-Level GIS Workbook, First Edition

NICOLAS R. MALLOY

PUBLISHED BY

GEOSPATIAL INSTITUTE

EUREKA, CA

Essential Modeling Techniques for Geospatial Analysis Using ArcGIS
An Intermediate-Level GIS Workbook, First Edition

Nicolas R. Malloy

Published by Geospatial Institute.

www.geospatial.institute

Notice of Rights

Notice of Liability

Trademarks

Contents

Preface

In the spring of 2010, the Humboldt State University formed the Geospatial Task Force to improve the geospatial curriculum. Assigned to develop a practical series of Geospatial courses that would serve students across multiple programs, two primary areas of assessment were considered. First, the existing curriculum was evaluated for redundancy and overlap. Second, professional requirements were *identifi*ed to eliminate obsolete content and replace it with relevant job skills.

As a member of the Geospatial Task Force, I conducted interviews with both alumni and students to gain first-hand insight into our assessment goals. The consensus from those who had experience with geospatial courses at HSU was that the Intermediate Geographic Information Systems course was outdated and lacked relevancy in terms of job skills and modern analytical methods. This assessment was confirmed when course content was evaluated based on standards defined in the U.S. Department of Labor Geospatial Technology Competency Model.

This book is the result of the work and development that followed over the years following the Geospatial Task Force recommendation. Here, readers will find an introduction to several geospatial modeling techniques. Though some tutorials presented here cover similar concepts, each represents a complete and independent exercise. The modeling techniques shown here only scratch the surface of what is possible for each. The intent is to introduce readers to a varied array of geospatial modeling techniques and to prepare students for more advanced work.

I sincerely hope that by working through these tutorials, you will develop the skills you need to be successful in the workplace.

—Nicolas R. Malloy

Print Book or eBook?

This book is available as both an eBook as wells as a print book. The eBook format offers several advantages over a traditional printed textbook related to cost, dissemination, accessibility, interactivity, and sustainability. One of the initial goals at the start of this endeavor was to create free or low-cost educational materials for geospatial science. Publishing a book in an electronic format provides me with a wide array of pricing options. For example, some distribution outlets, such as Amazon Kindle Direct Publishing™ (KDP), allow authors to participate in low-cost subscription programs like Kindle Unlimited™. Readers can pay a low monthly subscription fee to gain access to millions of titles. Authors can also choose a limited number of days when the cost of their book is $0.00. When these days coincide with the start of a college term, students may have access to the book for free.

Because an eBook is delivered electronically, there is no need to drive to a physical bookstore. Readers can access the book from anywhere, using any device that supports eBooks. Printed books are not accessible. This book uses the ePub 3.0 specification, which is natively accessible for readers that need to use screen readers. Additionally, eBooks offer many interactive options, such as links to external resources, files, and videos, that do not translate to the printed book. Finally, eBooks are environmentally friendly. In the words of William McDonough and Michael Braungart, "this book is not a tree."

However, despite the benefits of an electronic book, I understand that some readers still prefer a hardcopy book. Many readers benefit from the tangible aspect of a physical book and find hardcopy text easier to read. Whenever interactive elements are introduced, such as links to websites, videos, and data sources, hardcopy readers can find the URL to these resources in the footnotes.

Navigating the eBook

I designed this textbook as a reflowable eBook. As a result, there are differences in how to read and browse the text when compared to traditional hardback books. Specific navigation options will depend on the device and the settings used. Generally, readers should swipe left and right or use navigation buttons to turn pages. The internal table of contents links to different sections within the book. The Navigation table of contents, which is specific to the reader's device, mirrors the internal table of contents. This book also contains an interactive index that links to specific words and pages within the text. Though the index displays page numbers following the terms of interest, there are no real page numbers in a reflowable eBook. The specific number of pages changes depending on the reader's preferences, such as typeface and font size. At the time of this

publication, there is no means to change the index page numbers to reflect these preferences dynamically. Therefore, the page numbers in the index do not refer to actual pages but instead provide a means to link to the correct location within the book. It may help to think of them as relative positions within the textbook rather than specific pages.

The eBook displays hyperlinks in blue italic font. Hyperlinks may navigate to external resources, and a web browser may open when they are activated. The specific browser will depend on the reader's default settings. Typically, these external resources are YouTube videos, external websites, or data files. Readers should tap or click hyperlinks to activate them.

The eBook contains many images to assist in learning GIS software. Each image includes ALT text for screen reader accessibility. The pictures change size depending on the size of the screen on the device. Readers can tap twice or double-click an image to enlarge it. Some mobile devices allow readers to tap twice a second time to zoom into the picture further. Experiment with the reading device to see what interactions are available.

What the Reader Should Know

The tutorials presented here are for readers with an intermediate level experience using GIS. Readers should be capable of solving analytical problems using both vector and raster operations. Students are expected to have a solid foundation of both geospatial concepts and experience using GIS software, including the following:

» Projections and Transformations
» Datums
» Geographic and Projected Spatial Reference Systems
» Table Joins
» Adding XY data
» Geoprocessing Operations (buffer, clip, dissolve, erase, intersect, etc.)
» Attribute Queries
» Spatial Queries

Readers that need additional help should review the tutorials covered in the book ***Geospatial Concepts: The Fundamentals of Geospatial Science***[1]

Learning Outcomes

The tutorials focus on the academic, workplace, and industry sector competencies outlined in the U.S. Department of Labor Geospatial Technology Competency Model, including:

» Reading
» Writing
» Core Geospatial Abilities and Knowledge
» Working with Tools & Technology
» Planning and Organizing
» Critical & Analytical Thinking
» Creative Thinking
» Problem Solving
» Positioning & Data Acquisition
» Critical & Analytical Thinking

1 URL: https://www.amazon.com/dp/B07SKZJYQ5

Required Hardware and Software

This book assumes that the reader has access to ArcGIS Desktop™ Advanced 10.6.1 or higher. The authors assume that readers have a computer that runs the Microsoft Windows™ 10 operating system and with enough power and memory to meet the ArcGIS Software requirements.

Other software requirements include the following:

» Google Chrome

» 7-Zip (7-Zip is a free, open-source file compression/decompression utility)

» Microsoft Office

» Adobe Acrobat Reader

Additionally, some tutorials recommend access to a GPS receiver that can connect to a computer for data download. Any hand-held GPS receiver that can connect to a Windows PC to download data should work fine. The base model for the Garmin GPSMAP 64 GPS receiver is the recommended model. However, readers are not required to choose this model. Readers are encouraged to look for used or refurbished models that fit their budget, such as the Garmin eTrex 10. I provide sample data for readers that do not have access to a GPS receiver.

Internet Connection

For some content referenced in this book, access to a reliable high-speed internet connection is required.

Public Domain Data

Whenever possible, this book uses public domain data acquired from various websites belonging to governmental and non-profit organizations. The goal is to provide readers with data acquisition knowledge and experience. Due to the nature of obtaining data from public sources, there may be times when the data is temporarily unavailable. Additionally, some datasets may change over time, causing the results of a given tutorial to vary from the outcomes described within this book. Whenever possible, I provide back-up copies of the datasets used in the tutorials via links provided in the appendices at the end of the book.

About the Chapters

This book contains four chapters and six intermediate-level tutorials. The tutorials represent approximately two-weeks of work for a three-credit 16-week semester course. Tutorials may take between four to six hours to complete, depending on their complexity. When possible, I provide an estimated time to complete tutorials. Additional references, such as video content and external websites, may also be mentioned throughout the text.

Chapter 1: Review and Self-Assessment

When this book talks about geospatial modeling, it refers to methods that take the GIS user beyond basic mapping, attribute queries, spatial queries, and overlay operations. Geospatial modeling involves a high degree of subjectivity and allows for the modification of user-defined criteria. Typically, changes are made quickly and used to explore different scenarios and outcomes. This chapter focuses on the approaches to consider when introducing a geospatial component to model a problem, behavior, or phenomenon.

Additionally, this chapter provides a tutorial designed to help readers test areas of knowledge learned previously, or for readers that need additional review

Chapter 2: Modeling Network Paths

Network paths include the phenomenon that must move through a fixed system or structure. These paths may consist of stream networks, transportation networks, and utility networks. This chapter covers the elements that make up a network path model and provides tutorials that offer practical applications of these concepts.

Chapter 3: Modeling Overland Paths

Overland paths model movement or routes through places with no infrastructure or systematic constraints. Modeling overland paths is often based on modeling which path has the least cost. For example, determining the path for a new fire road in a national park might consider cost factors such as slope, vegetation, and proximity to sensitive areas such as streams or wildlife habitat. When modeling the large-scale movement of wildlife over diverse topography, cost factors might include elevation, slope, vegetation, proximity to water sources, and natural barriers to movement. Cost factors for modeling the spread of a wildfire from the point of origin might include slope, wind direction, and fuel sources. This chapter covers the elements that make up an overland path model.

Chapter 4: Modeling Surfaces

Surface modeling is a representation of a geographic feature or phenomenon that can be measured continuously across some part of the earth's surface. This chapter covers two surface modeling techniques used to create surfaces, density, and interpolation.

Acknowledgments

This book is the result of years of development, fact-checking, testing, and revision. I would like to thank all of the people that made this book possible, including both fellow faculty members and my students. I also thank the reviewers who provided feedback and advice for this publication.

Chapter 1: Review and Self Assessment

A Review of the Geospatial Modeling Process

When this book talks about **geospatial modeling**, it refers to methods that take the GIS user beyond basic mapping, attribute queries, spatial queries, and overlay operations. Geospatial modeling involves a high degree of subjectivity and allows for the modification of user-defined criteria. Typically, changes are made quickly and used to explore different scenarios and outcomes.

Generally, there is a broad range of methods associated with different disciplines and specializations. For example, an urban planner may use a very specialized set of tools, software, and techniques to solve problems, answer questions, and manage resources. An environmental planner will have a different set of problems and resources issues to address. They utilize a very different set of tools and methods. In many cases, these experts will use mathematical or statistical software outside of a geographic information system. They may even go back and forth, providing input to a GIS while also using the output.

The specific goals, methods, and criteria will vary among disciplines. Discussing these external methods and tools is beyond the scope of this book. Instead, this book focuses on the approaches to consider when introducing a geospatial component. Some general steps in the geospatial modeling process apply to a wide range of applications.

The process for creating a geospatial model usually contains the following steps:

- Stating the goal
- Defining the criteria
- Gathering the data
- Running the model
- Verifying the results
- Documentation and communication

Stating the Goal

Usually, one declares the purpose of the analysis in terms of solving a problem or answering a question. When indicating the goal of the analysis, one should try to be as specific as possible. The more specific one is, the easier it will be to develop and outline one's criteria.

The following are some examples of problems with a geospatial component:

- Where is the best place to install nest boxes for barn owls near a vineyard in Sonoma County?
- How often does the bathymetry of the Eureka Slough change over time?
- What factors influence crime in Los Angeles?
- What places in our State and National Parks are most affected by illegal marijuana grows?

While these questions are an excellent place to start, it would be more useful to state the goals and objectives in a way that includes additional details. For example, consider the problem with placing barn owl nest boxes. Rodents such as gophers and voles can cause significant damage to vineyards as they chew on roots and bark. In Napa Valley, California, some farmers are turning to barn owls as an environmentally friendly means to control the rodent population. To attract barn owls, farmers install empty nest boxes around their vineyards. But the question is, where to install them? If one states the problem in a general way, one might come up with something like this:

The objective of this analysis is to determine the best location for installing barn owl nest boxes.

This statement is a good starting point. However, framing the problem or objective in a more specific way will help to determine what one's criteria will be and what data one might need to acquire. Instead, one might use the following wording:

The objective of this analysis is to determine the environmental factors influencing barn owl selection of nest boxes, including land cover, vegetation types, hydrology, and proximity to urban features, such as roads.

By framing the problem with a detailed statement, one begins to solidify what the model might look like and what information one will need to build it.

Defining the criteria

The criteria one chooses and the weights assigned to them form the bedrock of the analysis. To start, one should create an essential list of criteria. This list determines the data one will need for the model. For example, a simple list of criteria for the barn owl nest box selection might include:

- land cover types
- roads
- waterways
- elevation
- slope
- field observations

The next thing to do is to decide on the relative importance or weight of each item. Defining the criteria and their influences are difficult to *quantify*, and the process may contain a high degree of subjectivity. As a result, this may be a portion of one's analysis that is the most difficult to defend. It is during this stage in the geospatial modeling process where background research becomes necessary. This research includes, but is not limited to, the following:

- Reviewing published literature on the topic
- Consulting with others knowledgeable in the field
- Gathering information through community outreach
- Using one's first-hand knowledge or expertise

Gathering the Data

The criteria that one outlines can serve as a list of map layers that one will need to gather. If one is lucky, much of the required data will already be in a geospatial format. However, most often, one must collect non-spatial data, such as Excel tables or tab-delimited text files, and convert them into a usable form. It may also be necessary to create the data oneself by collecting information out in the field using a GPS unit.

Running the Model

The methods that one inevitably uses is based upon the outline of criteria, and the data one has gathered. Most often, your model will consist of a series of functions, each corresponding to one or more steps in your criteria. Ideally, models should be flexible enough to modify the criteria with ease to explore different scenarios and outcomes.

Verifying the Results

After running one's model, the next step will be to determine if the results are valid. The easiest way to do this is to decide for oneself if the results make sense, given the stated goal or problem, the criteria, and one's knowledge of the study area. The results should relate to the criteria and the relative weights given. One may also want to share the results with colleagues and, if possible, have them run the model independently and compare the outcome. If time and resources allow, one may also go out into the field and ground-truth the results. One thing to keep in mind is that a geospatial model is only a generalized representation of reality and highly subjective. One should not expect the results to reflect reality in a specific way. After checking the results, it may be necessary to modify the criteria and rerun the model.

Documentation and Communication

Perhaps the most crucial step in the geospatial modeling process is documenting one's work and communicating the results. The most common form this takes in the geospatial industry is through written reports or papers, accompanied by graphic communication in the form of charts, graphs, and maps. Unfortunately, for many geospatial analysts, communication and documentation are an afterthought. The quality of your analysis will not matter if colleagues, clients, and decision-makers cannot understand the methods and results. Worse yet, if one's delivery is inferior as a result of poor writing skills and cartographic design, the only thing one might communicate is a lack of professionalism. If readers take the time to improve their writing skills and their cartographic design, they will be able to set themselves apart and make a good impression when it matters most.

Tutorial: Advanced Geospatial Analysis Skills

Review and Self-Assessment

This tutorial is designed to help readers test areas of knowledge learned previously, or for readers that need additional review. Readers will find very few step-by-step instructions during this activity. Instead, this tutorial gives readers a series of problems to solve. Readers that need additional help should review the tutorials covered in the book ***Geospatial Concepts: The Fundamentals of Geospatial Science***[1] (**Figure 1.1**).

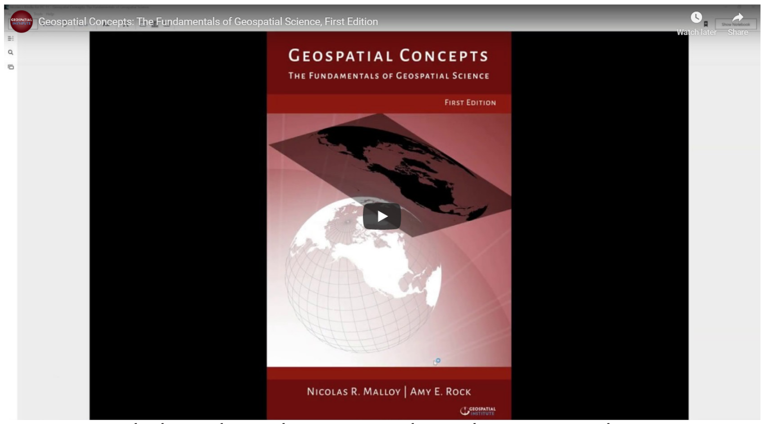


Figure 1.1: Watch this video to learn more about the Geospatial Concepts eBook.[2]

Estimated time to complete this tutorial: 4 hours

Learning Outcomes

Readers should be able to accomplish the following outcomes by the end of this tutorial:

- » Review how to acquire data from a public source
- » Review checking spatial references
- » Review spatial selections
- » Review attribute selections
- » Review creating layers from selected features
- » Review vector overlay operations
- » Review raster analysis tools such as slope and raster calculator
- » Review adding XY data
- » Review table joins
- » Review classification methods for data symbolization

Setting up Your Workspace

In a typical workflow, you work on geospatial data using a local hard drive. When done, you compress your data and back up your work to your cloud storage so that you can retrieve the files from anywhere. When referring to a **local hard drive**, it means you are working on data physically located on the computer in front of you.

In contrast, some computers also include networked drives. **Networked drives** link to cloud storage and save the data elsewhere. Examples include services like OneDrive or Google Drive. For this tutorial, use the **desktop** as your local hard drive location. You may also use an external USB drive if you plan to work in multiple places.

You must avoid using networked drives while you work.
They increase the processing time and can cause technical glitches.

1 URL: https://www.amazon.com/dp/B07SKZJYQ5
2 Video URL: https://youtu.be/_5l9na4DLyM

In this book, you use a particular folder structure. Start by creating your workspace folder on the local hard drive. A **workspace** is a folder or series of folders that contain all of your project files. The top-level folder in your workspace should indicate the activity or the project on which you are working. Organize all of your work within the workspace folder. On your **desktop**, create a new folder and give it a descriptive name, such as **GIS_Review**. Be sure there are no spaces. You may use underscores instead of spaces. Inside this folder, create the following three subfolders: *original*, *working*, and *final*. Having a standardized folder structure helps to keep a project organized, primarily when you are working with multiple partners. The folder structure you see here (**Figure 1.2**) is the standard used in each of the tutorials presented in this book.

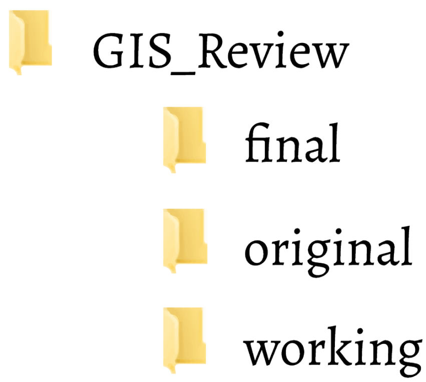


Figure 1.2: This diagram represents a basic folder structure used in this book.

As the name indicates, use the **original folder** for storing original, unaltered data. As you are working on a project, if, for some reason, your working version of the data gets lost or corrupted, you can go back to your *original* folder and find a fresh copy of the data. Use the **working folder** for data that you *create* or *alter* while working on your project. Use the **final folder** for storing any output you produce as a result of your work, such as images, maps, tables, or reports. Setting up a standard folder structure for a project is good practice and a habit you should develop.

Disable Background Geoprocessing

In the ArcGIS software, the *Background Geoprocessing* setting is often turned on by default. This setting allows users to continue to work while *a* tool is running in the background. However, sometimes this setting will stop tools from running or cause other unforeseen problems. To reduce that chances of the ArcGIS software crashing during this exercise, turn this setting off. After launching ArcMap, open the *Geoprocessing options* from the *Geoprocessing* menu. Under *Background Geoprocessing*, uncheck the box next to the word *Enable* (**Figure 1.3**).

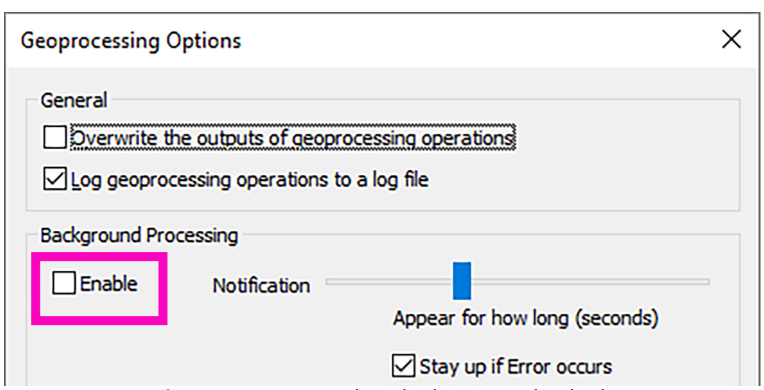


Figure 1.3: Be sure that the box is unchecked.

Spatial Analyst Extension

The steps in this activity involve using the *Spatial Analyst* extension. After launching ArcMap, make sure this extension is activated.

Projections, Datums, and Spatial Reference Systems

The following *Skill Drills* use the knowledge you have learned previously; thus, the instructions do not contain many step-by-step details. You should refer to previous coursework or texts if needed.

Skill Drill: Acquire data from the Los Angeles GIS Data Portal

The ***Los Angeles County GIS Data Portal***[1], maintained by the County of Los Angeles, is a website where you can download many different types of GIS datasets. The website includes instructional material on how to use the site, including text and videos. If you need a refresher, I recommend watching the ***instructional videos***[2] before you begin. Navigate to the Los Angeles County GIS Data Portal and use the keyword *city boundaries* to search for city polygons on the website (**Figure 1.4**). Click on the heading that says *City Boundaries and Annexations / Detachments*.

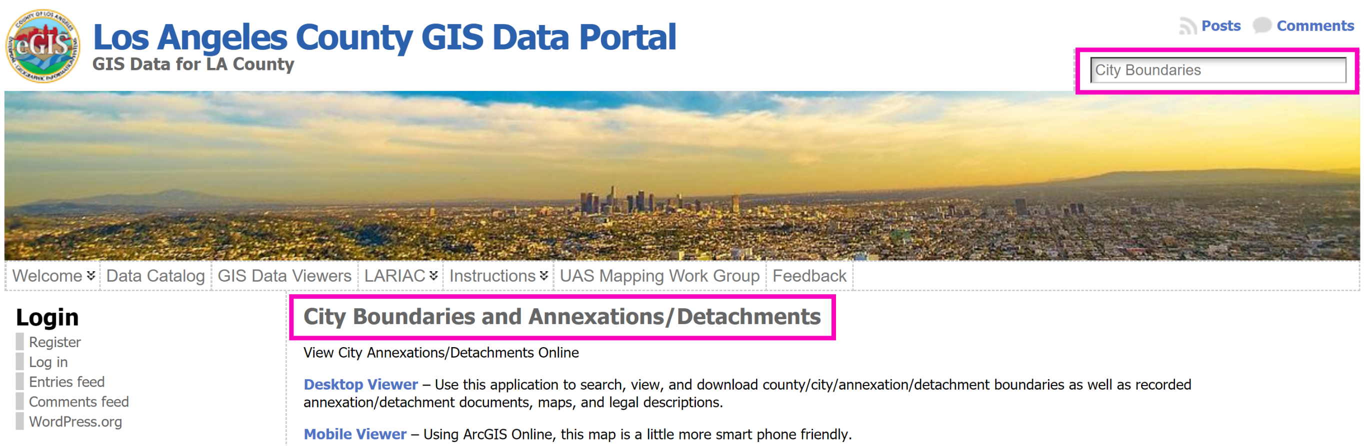


Figure 1.4: The link to the city boundaries data looks like a section heading. It serves both purposes but can be misleading.

Scroll down and click the link that says *City Boundaries Polygons* (**Figure 1.5**)and download the file. Decompress the file in your *original* folder.

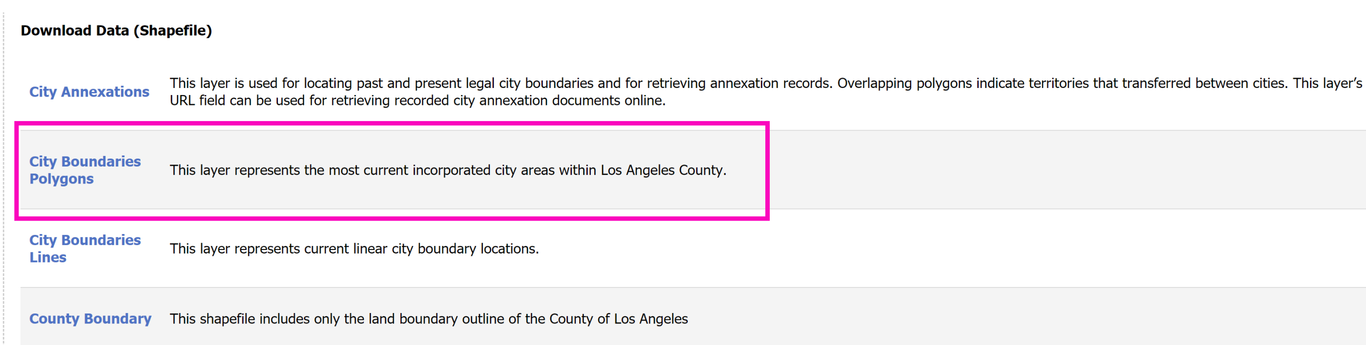
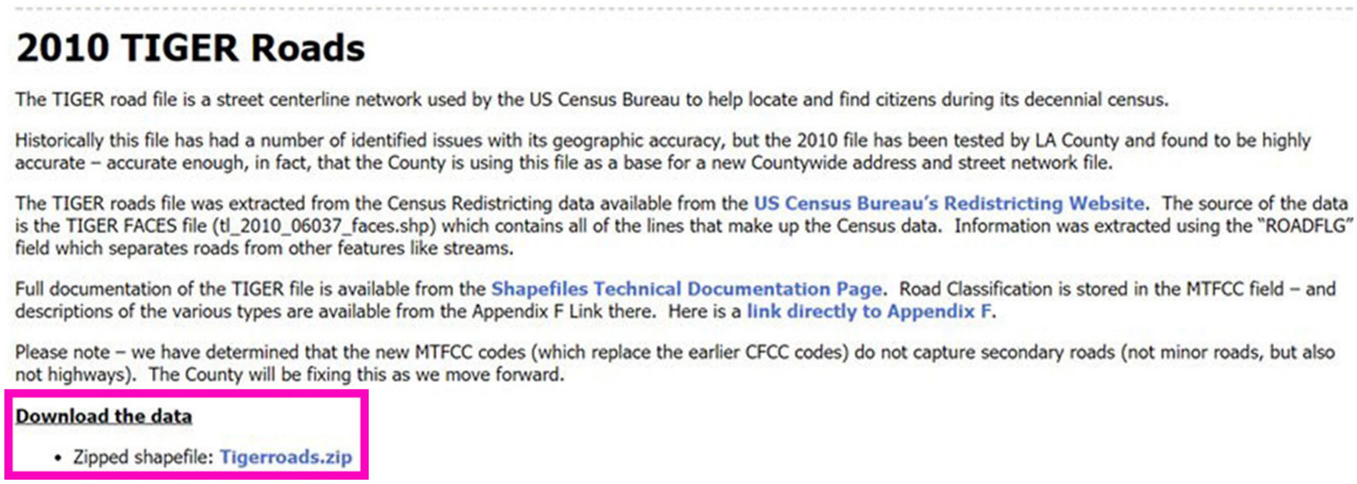


Figure 1.5: Right-click on the link and select Save link as, then navigate to your *original* folder.

If, for some reason, the website is down or the data is no longer available,
you can download a backup copy of the data for this tutorial using the link provided in Appendix A.

In ArcMap, **open a new blank map document** and add the layer to your *Table of Contents*. This layer will define the spatial reference for the *Data Frame* window if it is the first layer loaded into the *Table of Contents*. Repeat the steps for locating and downloading data from the Lost Angeles GIS Data Portal. This time use the keyword *roads* to find the *2010 TIGER roads layer* (**Figure 1.6**). Download the data and add it to your map document *Table of Contents*.

1 URL: https://egis3.lacounty.gov/dataportal/
2 URL: http://egis3.lacounty.gov/dataportal/about/videos/



2010 TIGER Roads

The TIGER road file is a street centerline network used by the US Census Bureau to help locate and find citizens during its decennial census.

Historically this file has had a number of identified issues with its geographic accuracy, but the 2010 file has been tested by LA County and found to be highly accurate – accurate enough, in fact, that the County is using this file as a base for a new Countywide address and street network file.

The TIGER roads file was extracted from the Census Redistricting data available from the US Census Bureau's Redistricting Website. The source of the data is the TIGER FACES file (tl_2010_06037_faces.shp) which contains all of the lines that make up the Census data. Information was extracted using the "ROADFLG" field which separates roads from other features like streams.

Full documentation of the TIGER file is available from the Shapefiles Technical Documentation Page. Road Classification is stored in the MTFCC field – and descriptions of the various types are available from the Appendix F Link there. Here is a link directly to Appendix F.

Please note – we have determined that the new MTFCC codes (which replace the earlier CFCC codes) do not capture secondary roads (not minor roads, but also not highways). The County will be fixing this as we move forward.

Download the data

- Zipped shapefile: Tigerroads.zip

Figure 1.6: The 2010 Tiger Roads link on the Los Angeles County GIS Data Portal.

In a separate document, write down the answers to the following questions:

Question 01: What is the projection of the dataframe in the current map document (.mxd)?

Question 02: What is the datum of the dataframe in the current map document (.mxd)?

Question 03: What is the spatial reference system of the dataframe in the current map document (.mxd)?

Question 04: What is the projection of the Tiger Roads layer?

Question 05: What is the datum of the Tiger Roads layer?

Question 06: What is the spatial reference system of the Tiger Roads layer?

Question 07: If you wanted to create a new Tiger Roads layer that used a different spatial reference system, datum, and projection, *which* tool in ArcMap would you need to use?

Vector Analysis

Use your vector analysis skills and the layers downloaded in the previous steps to answer the questions below. Remember, vector analysis includes attribute selections, spatial selections, and some overlay operations.

In a separate document, write down the answer to the following questions:

Question 08: How many incorporated cities share a border with the City of Pasadena? Be sure to exclude polygons whose city type is unincorporated.

Question 09: What is the total length (in kilometers) of all the roads completely within the incorporated city boundaries that share a border with the City of Pasadena? Be sure to exclude polygons whose city type is unincorporated.

Hint: You will need to create separate layers for the City of Pasadena and also for the bordering cities.

Raster Analysis

In this step, you will download data from the USGS National Map Viewer. Then you will use your raster analysis skills to answer the question below. Remember, raster analysis includes understanding the spatial resolution (cell size) of the data, using raster analysis tools, and performing mathematical computations on raster datasets.

Skill Drill: Acquire Elevation Data from the USGS National Map Viewer

Previously, you learned how to acquire data from public sources. Here you will download a digital elevation model from the *USGS National Map Viewer*. You will then use the digital elevation model to answer the question below. Navigate to the ***USGS National Map Viewer***[1]. On the right, zoom in and center the map to the approximate region of interest. For this activity, zoom

1 URL: https://viewer.nationalmap.gov/basic/

into **Mount Shasta, California**. On the left, under *Elevation Products*, check the box next to *1/3 arc-second DEM*. Then click. *Find Products* (**Figure 1.7**).

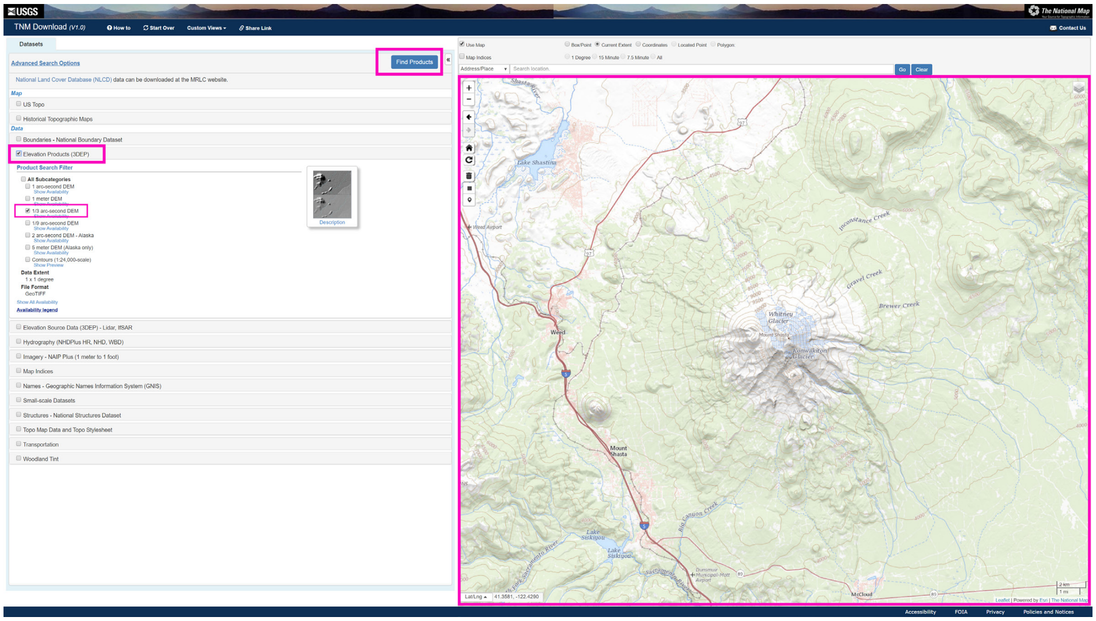


Figure 1.7: The USGS National Map Viewer.

On the *Products* tab, you will see a list of datasets available for download. To determine which digital elevation models cover the area of interest, you can click on the *Footprint* link to see the boundary of the dataset (**Figure 1.8**). You may need to zoom out a little to see the entire footprint. At this time, only one digital elevation model covers the area of interest. Click the *Download* link to download. Save the file in your *original* folder and decompress it.

If for some reason, the website is down or the data is no longer available, you can download a backup copy of the data for this tutorial using the link provided in Appendix A.

Check the spatial reference of the digital elevation model. Then, use the *Project Raster* tool to create a **copy** of the data with the NAD 83 UTM Zone 11N spatial reference. Save the *Output Raster Dataset* to your *working* folder. Under Output Coordinate System, choose NAD_1983_UTM_Zone_11N. When you are ready, *click OK*. In ArcMap, **open a new blank map document**, add the layer to your *Table of Contents*. If asked to create pyramids, say *Yes*.

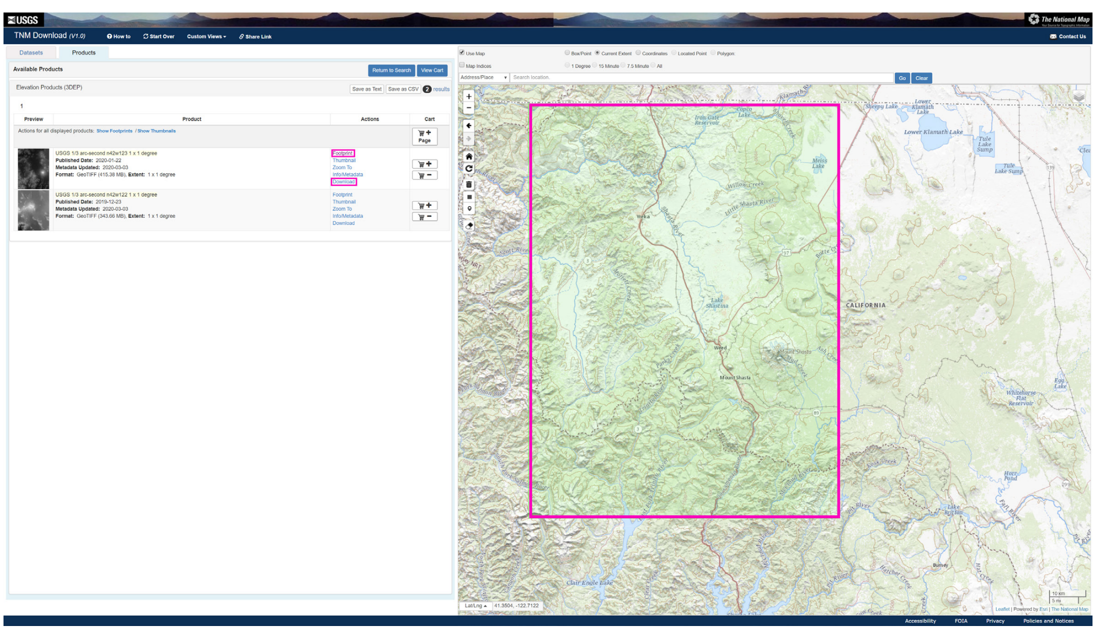


Figure 1.8: The USGS National Map Viewer *Products* page.

In a separate document, write down the answer to the following questions:

Question 10: What was the original Spatial Reference System before projecting the digital elevation model?

Question 11: What is the total area (in square meters) of all the regions having a slope of equal to or less than 40 degrees?

Tabular Data and Classification Methods

In this step, you will download data from the National Historical Geographic Information System (NHGIS). You will then use data classification methods to symbolize Census data and *identify* patterns. You will also create a CSV table and add it to ArcMap as XY data to answer the questions below.

Skill Drill: Downloading Data from the National Historical Geographic Information System

In previous courses, you learned how to acquire data from public sources. Here you will download Census data from the National Historical Geographic Information System (NHGIS). You will then use data to answer the questions below. Navigate to the NHGIS website at ***https://www.nhgis.org/data*** and click the login link (**Figure 1.9**).

Figure 1.9: Locate the login link on the upper right.

On the Data Finder page, apply the first data filter by clicking on the *Geographic Levels* button (**Figure 1.10**).

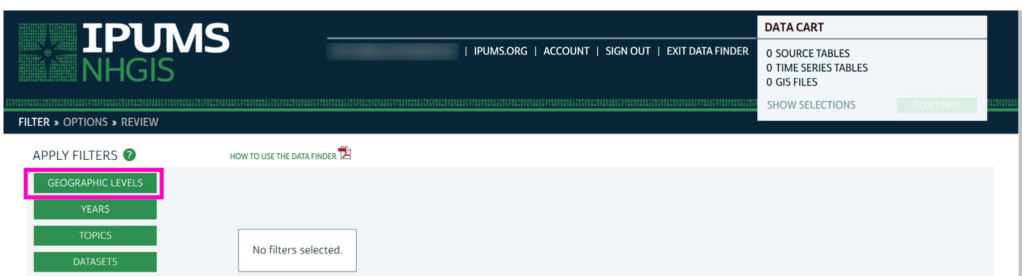


Figure 1.10: The geographic levels define the spatial units, such as states, counties, Census tracts, or Census blocks.

Click the plus sign next to *State* and click *Submit* (**Figure 1.11**).

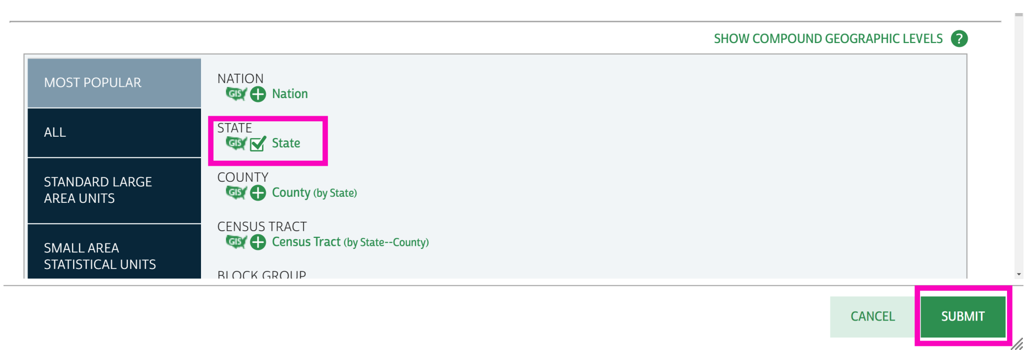


Figure 1.11: Geographic Levels Options.

When the data has finished loading, select the *Time Series Tables* tab. Scroll down and select *Median Household Income in Previous Year* (**Figure 1.12**).

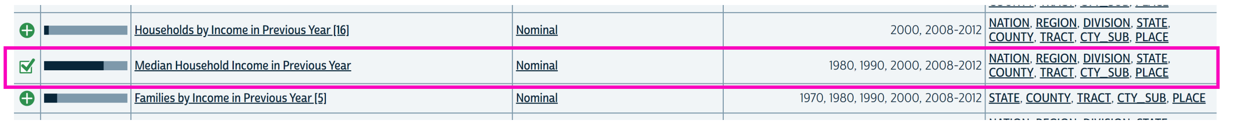


Figure 1.12: You may find the table easier to locate if you change the number of results on the page to 500. Then press CTRL F and type in the name "Median Household Income."

When you are ready, scroll up and click *Continue*. You will be taken to the data options page. You still need to select a GIS boundary file (shapefile) to go with the dataset. Open the *GIS Files* tab. Then, scroll down and locate the 2015 State TIGER/Line shapefile (**Figure 1.13**). Click the plus sign on the left.

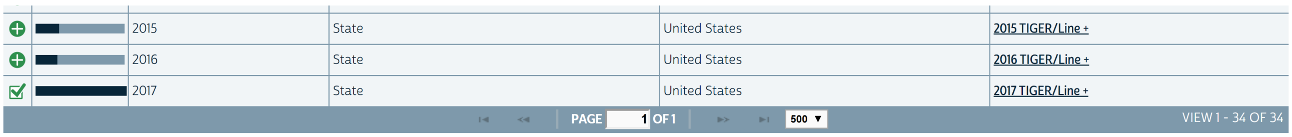


Figure 1.13: Locate the 2015 U.S. State shapefile near the bottom of the page.

Then, scroll up and click *Continue*. On the next page, click *Continue* again. Make sure the *Table File Structure* is set to *Comma Delimited* (**Figure 1.14**). Check the radio button next to *Time Varies by File* is checked as well. Enter a brief description describing the data. When done, click the *Submit* button. It may take five to ten minutes for the data to be ready. You should receive an email when it is complete. Follow the instructions in your email to download the data. Save it to your *original* folder and decompress it.

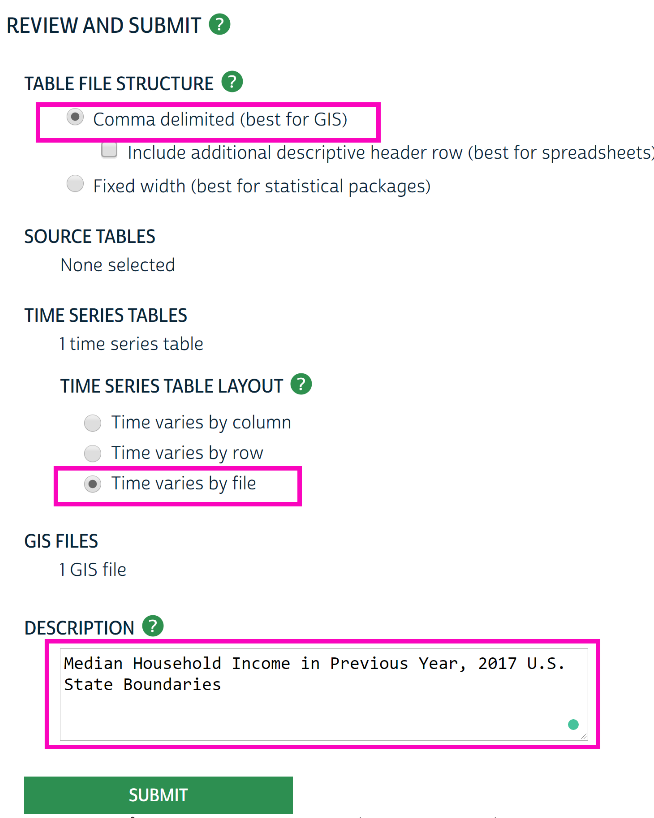


Figure 1.14: NHGIS Review page settings.

If for some reason, the website is down or the data is no longer available, you can download a backup copy of the data for this tutorial using the link provided in Appendix A.

Skill Drill: Table Join and Data Classification

In ArcMap, **open a new blank map document** and add the *U.S. State 2017* layer to your *Table of Contents*. Perform a table join on the *U.S. State 2017* layer and join the *20125 State* CSV data to the shapefile using the GISJOIN field. Read the metadata located in the text file (codebook) with a similar name as the 20125 State CSV to interpret the meaning of the field names. Determine which field name stands for **Median income in the previous year: Households**. Make a note of this as you will need to know which field to use to answer the question below.

In a separate document, write down the answer to the following question:

Question 12: Symbolize the United States using the field that represents *Median income in previous year: Households* and **normalize** it with the *Shape_Area* field. Use the **Quantile** method of classification and **five** classes. Which states fall into the HIGHEST category? You may count the District of Columbia and Puerto Rico as states for this question.

Hint: Perform these steps in the symbology tab of the layer properties. You will need to inspect the map visually to get the correct answer. It helps if you label the states with their names.

Skill Drill: Creating a CSV Table to Import As XY data

In this step, you will review how to create a CSV table and add it to a map document as XY data. Use the results to answer the question below. In Microsoft Excel, create a CSV table with the following field headers: name, latitude, longitude (**Figure 1.15**). Enter the name, latitude, and longitude values show above in the appropriate fields. *Save as* a CSV table and close Excel.

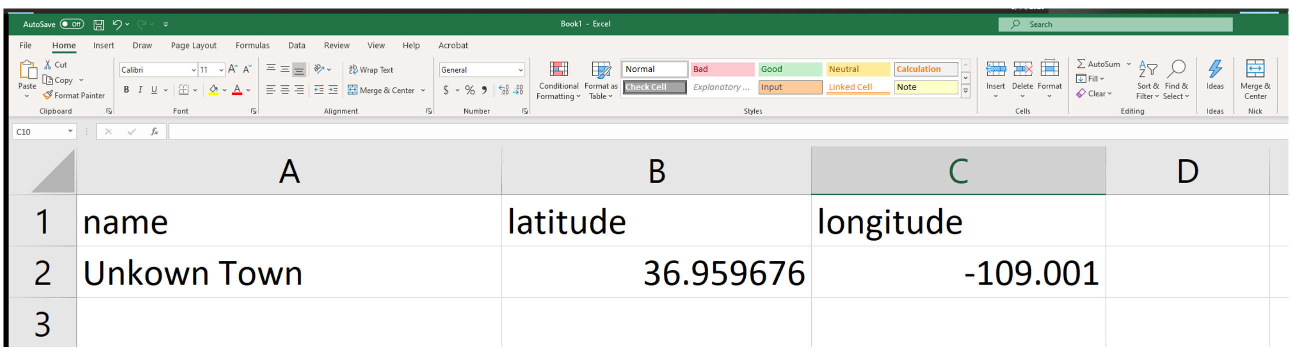


Figure 1.15: Be sure that your Excel table matches the table shown here.

To answer the question below, assume that a GPS receiver set to the geographic coordinate system **GCS WGS 1984**, and was used to obtain the geographic coordinates you see in the table above. The coordinates are in decimal degrees.

Hint: It's important to remember that each geodetic datum as a unique set of latitude and longitude values. For example, the latitude and longitude values for your home using the North American Datum of 1983 will be different than the latitude and longitude values using the North American Datum of 1927. Likewise, the World Geodetic Datum of 1984 (WGS 1984) will use yet another set latitude and longitude values to define the location of your home.

In a separate document, write down the answer to the following question:

Question 13: Add the CSV file you created to the map document used in the previous step and display it as a point feature. In which state is the unknown town located?

Chapter 2: Modeling Network Paths

Network paths include the phenomenon that must move through a fixed system or structure. These paths may consist of stream networks, transportation networks, and utility networks. This chapter covers the elements that make up a network path model and provides tutorials that offer practical applications of these concepts.

Geospatial analysts use network paths to solve many problems, including the following:

- The allocation of resources
- Determining the best location based on demand
- Delineating a service area based on travel time
- Finding the optimal route

Components of a Network

The Network Dataset

To run a network analysis, one first needs to create or obtain a network dataset. A **network dataset** is a vector data model built from a collection of topologically connected network elements. It represents a linear network and models the movement within. **Network elements**, such as edges, junctions, and turns, are derived from **network sources**, feature classes that represent the network, such as roads, intersections, pipes, and valves. A network dataset stores the connectivity between the source feature classes. Usually, the source features, such as lines and points, are *topologically unaware*[1] of each other. If two linear features intersect in a typical feature class, such as two street segments in a road layer, neither line is aware of the other. A network dataset is different because it keeps track of which source features are coincident. It also has a connectivity policy, which one can modify to define further which coincident features are genuinely connected.

For example, suppose one wanted to model movement across a freeway overpass in Arcata, California (**Figure 2.1**). The line segments representing the overpass and the freeway below are coincident, meaning, the portions that intersect seem to occupy the same place geographically. Typically, these linear segments lack connectivity and topological awareness. In a network dataset, the linear features, called edges, and junctions, the places where lines connect, are topologically aware. When one performs a network analysis, the GIS knows which paths along the network are feasible. This quality makes it possible to model overpasses and underpasses and the movement through them.

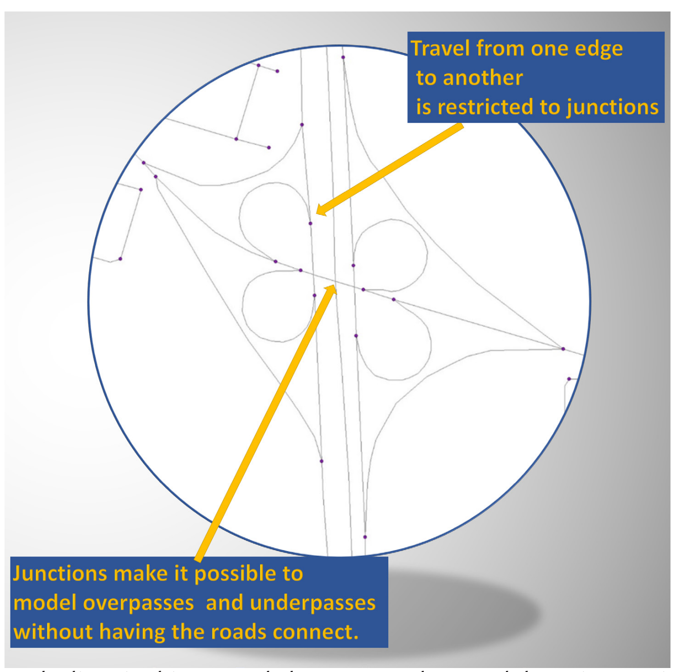


Figure 2.1: The lines in this network dataset are edges, and the points are junctions.

1 The term topology refers to the spatial relationships between connecting or adjacent features.

Network Elements

The following network elements make up a network dataset:

- Edges
- Junctions
- Turns

Edges

The edges are the segments over which travel takes place and use linear feature classes as sources. Examples of feature classes that serve as sources include stream segments, road segments, utility lines, and rail lines. One may assign any number of attributes edges, including those that may affect travel through the network, such as travel time and direction.

Junctions

Junctions connect edges and facilitate navigation from one edge to another. For example, in a street network, each intersection becomes a junction. If a junction source feature is not provided, the GIS will create a junction feature class automatically during the network dataset creation process. Examples of junctions include street intersections, dead ends, on-ramps, off-ramps, pipe valves, and stream network confluences. One may also assign attribute information to junctions, such as the time it takes to stop at a stoplight.

Turns

Turns are optional network elements that store additional connectivity rules that can affect movement between two or more edges. For example, a turn element may include a road intersection where left turns are prohibited.

Network Cost Variables

A GIS can usually create a network dataset if provided with a linear feature. However, it may take additional time to optimize a network dataset with the **cost variables** needed to generate a robust model.

Cost may be measured using the following:

- Distance
- Time
- Money

Geospatial analysts often base network analysis on modeling which path has the least cost. For example, the least-cost-path for an emergency responder would be the quickest path rather than the shortest one. The least-cost-path for a fleet of garbage trucks would probably factor in travel time, time at stops, labor costs, fuel costs, vehicle maintenance, and service time when returning to unload. The cost variables may also include more subjective measurements, such as the difficulty of the movement of salmon upstream or across a fish ladder.

The Vehicle Routing Model

A vehicle routing model is a type of network analysis that will help choose optimal routes from a given depot, based on which orders need to be serviced by which route (**Figure 2.2**). The **orders** are a network analysis class, often a point feature in the model, that represent places where something is picked up or delivered. *Orders* could have attributes such as volume, weight, service time. *Orders* can also have attributes that describe restrictions, such as a specific type of person or vehicle allowed to service that order, or a particular time window when stops are permitted. The **routes** are often thought of as merely a specific path, among many possible paths. However, in a network routing model, routes are also a class that defines the vehicle, the driver, and the linear features on which they travel. A **depot** is usually the route starting point and ending point. It can also represent a place where the route renews itself by unloading or reloading cargo before continuing to the next order. *Depot*s can have attributes that determine when the route is allowed to begin, when it must end, how long it takes to unload and reload cargo, and how long it takes to service the vehicle between orders or at the end of a shift.

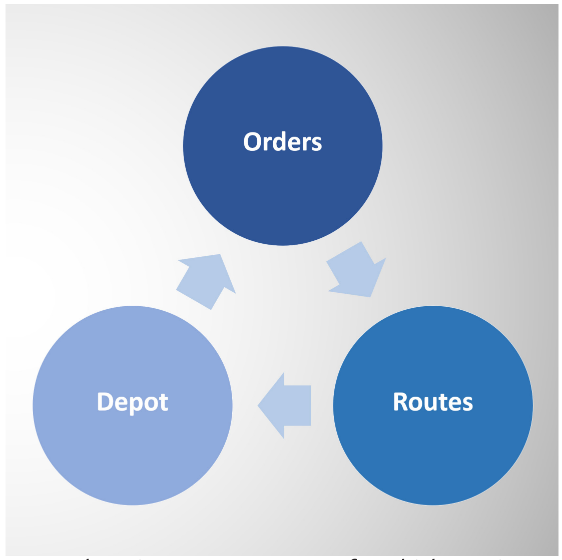


Figure 2.2: The primary components of a vehicle routing model.

To help one understand these concepts, let's assume one must address the following scenario: A local waste management authority would like to begin an organic waste diversion program for the collection of compostable organic waste. This waste stream represents a significant portion of waste generated in Eureka, CA. The diversion of organic waste has the potential to reduce hauling costs, offset greenhouse gas emissions, and generate energy. Organic waste diversion represents the newest frontier in the waste management industry. The waste management authority will begin a pilot program that includes a small number of participating businesses and restaurants. In this scenario, one is tasked with optimizing the organic waste diversion by finding the best routes, a combination of trucks, drivers, and paths through a street network.

In this example, the depot would be the location of the garbage dump. Attributes for this depot might include:

- The name and address of the depot
- The starting time for the routes
- The ending time for the routes
- The curb approach or direction the garbage trucks must approach the depot for route renewals.
- The attributes for the orders in this scenario include characteristics based on the restaurants serviced by the garbage trucks. One might include attributes such as the following:
- The name and address of the restaurant or business
- The average organic waste volume for each location
- One or more time-windows when pick-ups are allowed
- A maximum violation time if pick-ups fall outside the time-window
- The service time required at that location
- And any other restrictions, such as a specific vehicle type, or a curb approach

The attributes for the routes in this scenario include information related to the garbage trucks, the drivers, and the street network such as the following:

- The name and description of the garbage truck, the driver, and the route
- The name of the starting depot
- The name of the ending depot
- The service time required at the start and end of a route
- The capacity of the truck
- The cost per unit of work time, which may factor in the cost of the driver's wages
- The cost per unit of distance, which may factor in the cost of vehicle fuel and maintenance
- The time in which overtime pay begins
- The cost of overtime pay
- The maximum number of orders, the maximum total time, and the maximum total distance of the route

The depots, orders, and route attributes are just some of the essential elements that make up a vehicle routing model (**Figure 2.3**). Additional complexity can be introduced by including elements such as:

- *Route* renewal information
- *Route* zones that specify a work territory
- *Route* seed points used to weight a cluster of order locations
- Breaks that may be required by the drivers, such as meal-times and rest periods

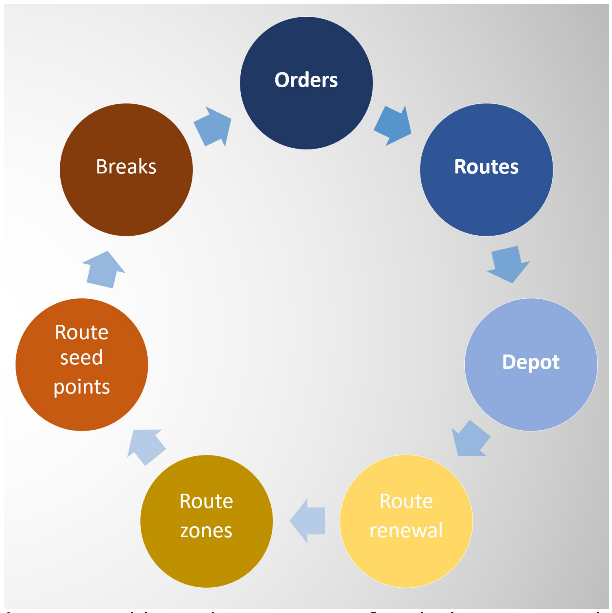


Figure 2.3: Additional components of a vehicle routing model

The Network Allocation Model

Geospatial analysts use a **network allocation model** to represent the interactions between facilities and demand points. It uses the premise that a specific location provides services or resources and that there is a demand for that resource nearby. This model can be used to predict the amount of demand at each facility to make planning decisions. It can also help to determine which areas are not served and provide information on the optimal location for the placement of a new facility.

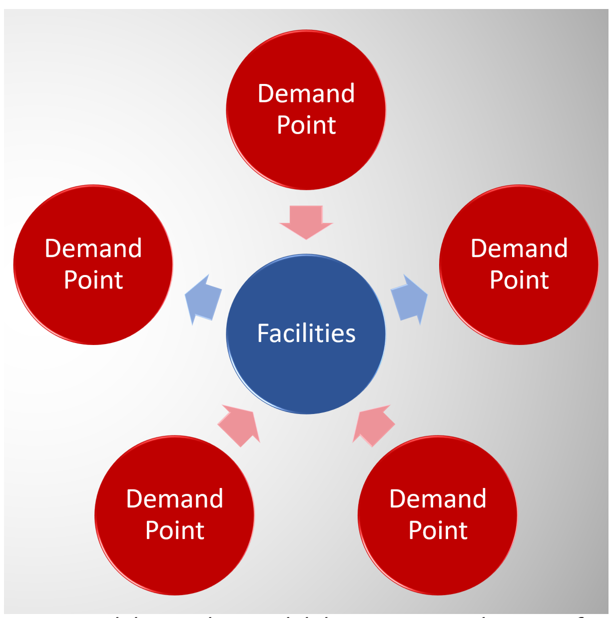


Figure 2.4: A network allocation model is used to model the interactions between facilities and demand points.

Demand Points

The **demand points** in a network allocation model will depend on the type of service provided. For example, demand points for a grocery store could be based on overall population data such as census blocks. Demand points for a daycare center might be based on a subset of the census data based on age group, family size, or marital status. Demand points for a fire station might be based on either population or the history of incident responses. One can also assign weights to a set of demand points based on frequency, density, distance from a facility, or any other factor that could be *justified*.

Facilities

Facilities represent the locations where services originate. They can be places where people travel, such as a health clinic. They may also be places from which services travel, such as a delivery of goods from a warehouse to a set of retail stores. In a network allocation model, one can have multiple facilities competing with one another. One can also designate one or more facilities as required locations. In contrast, others in the model could be designated as potential locations that must compete with other potential facilities.

Allocation Cost Factors

The interaction between facilities and demand points are based on a set of **cost variables**. Previously, I described a few cost variables, such as time, distance, and money. However, the cost can also be based on less tangible factors, such as the likelihood of a specific population group to visit a particular facility. In a GIS, the cost is sometimes referred to as an **impedance**. When modeling network allocation, it is typical to set a maximum impedance value or **impedance cutoff**. This cutoff is a way to limit the allocation of demand points to a facility once one meets a certain cost threshold.

For example, a fire station may need to respond to emergency calls within a specific time limit. If one is allocating response zones for a fire station, one might assign an impedance cutoff to a maximum of five minutes (**Figure 2.5**). That way, any demand points farther than five minutes away will not be allocated to the station. Or for example, a retail store may find that customers will not travel more than twenty minutes to shop at a specific location. In the model, one could apply an impedance cutoff of twenty minutes so that any census blocks farther than this distance will not count as part of the demand for any particular location

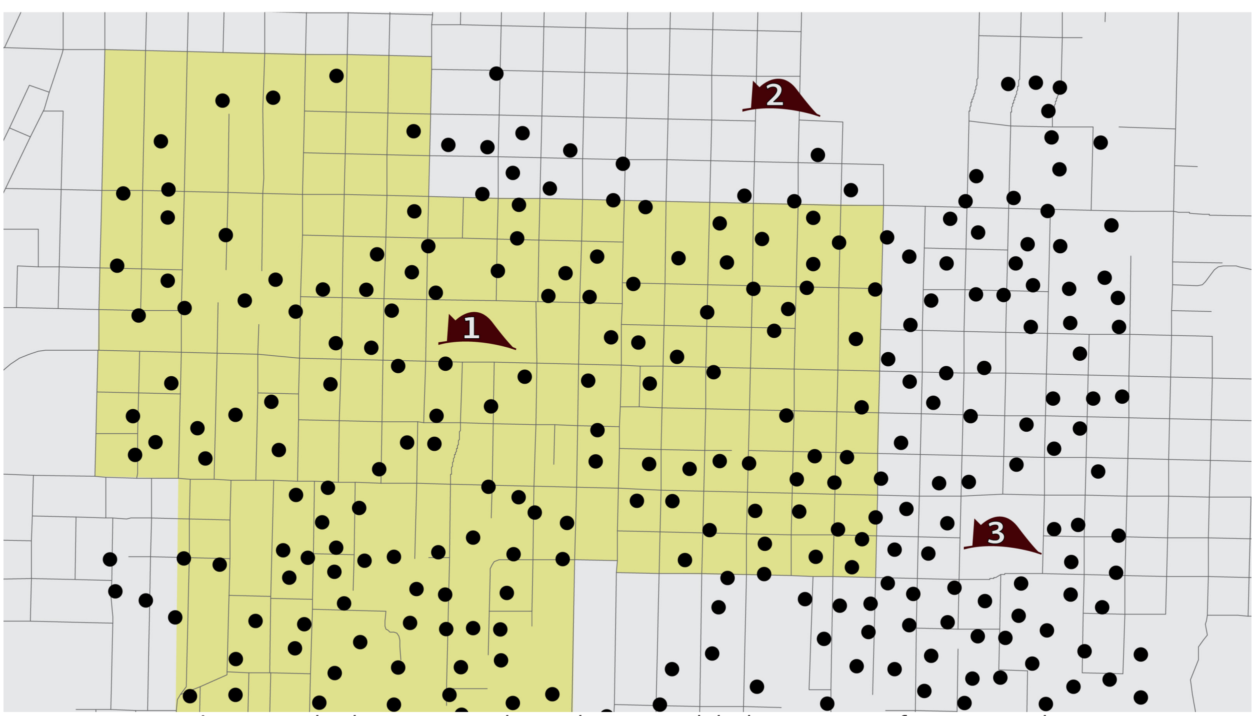


Figure 2.5: The dots represent demand points, and the hats represent fire stations. The yellow area represents a five-minute travel time from fire station one.

Defining the Criteria

The criteria one chooses for the model come from the objectives of the analysis. To start, one should try to develop a detailed objective statement and use that to define a list of criteria. For example, let us assume one must address the following scenario: The Arcata Fire Protection District has decided to explore the possibility of updating their facilities and relocating the existing fire departments serving the cities of Arcata and McKinleyville. There are currently three fire stations in this district. The Arcata Fire Protection District has *identifi*ed many potential alternative sites for each of the existing stations. Still, they are not sure which locations optimize services between these two cities. An analysis of service calls indicates that most calls are not related to fires, but instead involve other categories, such as motor vehicle accidents and medical assistance calls. The Arcata Fire department has a detailed history of incident responses for the past seven years. They have a preferred response time of two minutes when determining a service area. Still, they can extend a service area that includes a five-minute response time. They would like the model to consider both a two-minute and five-minute response time.

The objective statement might consist of the following statement:

> *"The objective of this analysis is to evaluate the potential locations of new fire stations compared with the locations of the three existing fire stations, factoring in the volume of service calls within a response time-window of two to five minutes."*

Based on this objective statement, one might define our criteria as the following:

- A history of incident responses over seven years
- The frequency of calls at each location
- The location of alternative sites from Arcata to McKinleyville
- The location of existing fire stations
- A two-minute response time window which captures the highest demand based on incident history
- A five-minute response time window which captures the highest demand based on incident history
- A comparison between alternative sites and existing fire stations based on demand weight allocated to the facility.
- A limit of only three fire stations chosen by the model as the best candidates

The demand points for this model is the history of incident responses. The facilities are the existing and alternative fire stations. The allocation cost variables include the volume of incident responses and the distance from the facilities. Our impedance cutoff is an area defined, first by a two-minute response time window, then by a five-minute response window. The three final candidates are chosen based on the total allocated demand weight.

Allocated Demand Weight

Allocated demand weight is the allocated sum of the demand from all demand points assigned to the facility. In this scenario, the demand points are weighted by the frequency of calls at a specific location. A neighborhood with a history of many incident responses will have more weight than an area with only a few. Demand weight decays with distance so that as demand points land farther from a facility, less weight gets allocated from that point. For example, suppose *Neighborhood A* has a higher volume of calls than *Neighborhood B*, and one wanted to allocate demand from these two points to the Arcata Fire Station (**Figure 2.6**). If distance was not a factor, one could say that *Neighborhood A* allocates a higher demand weight to the Arcata Fire station than *Neighborhood B*.

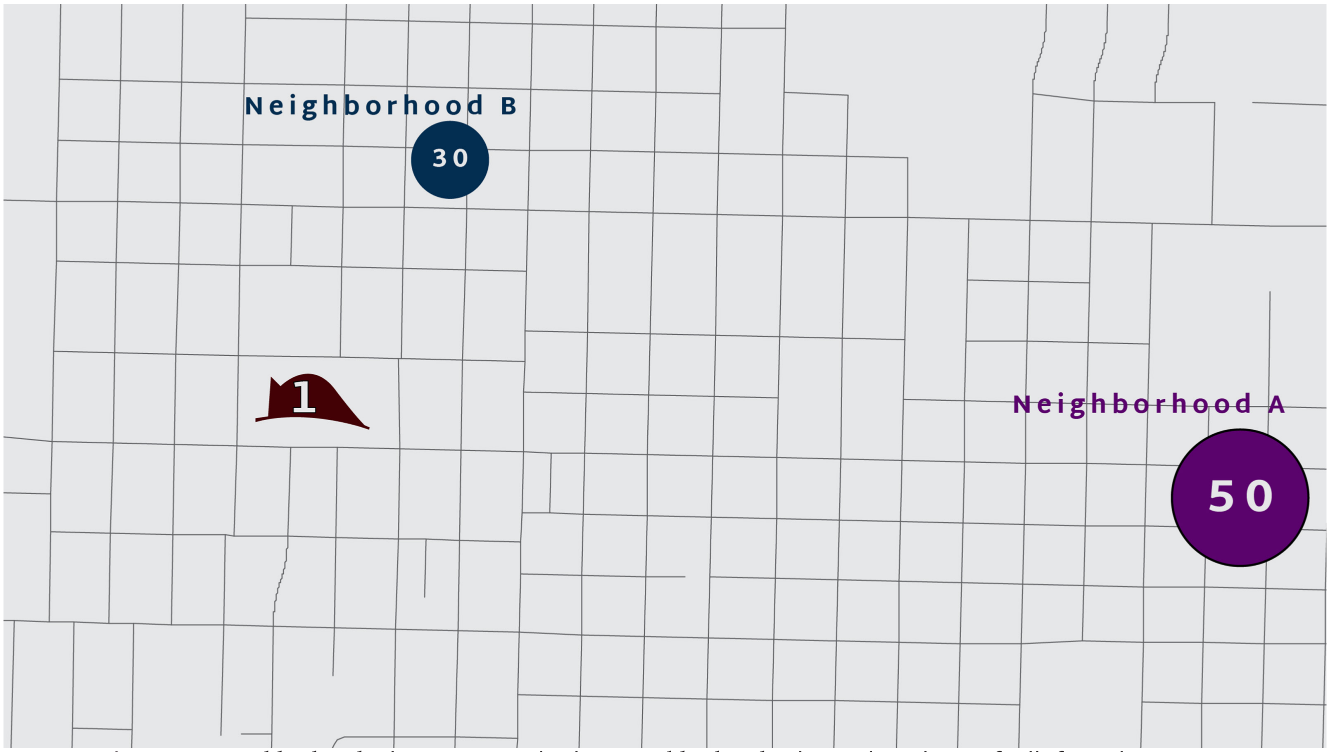


Figure 2.6: *Neighborhood A* has more weight than *Neighborhood B* due to the volume of calls from that region.

However, if one determined that the demand weight decayed with distance, *Neighborhood A* might allocate less demand weight than *Neighborhood B*, if it is farther away from the Arcata fire station (**Figure 2.7**). Further, if *Neighborhood A* is beyond the five-minute impedance cutoff window, it won't have any weight at all. The best location for a fire station would be the one with the most allocated demand weight.

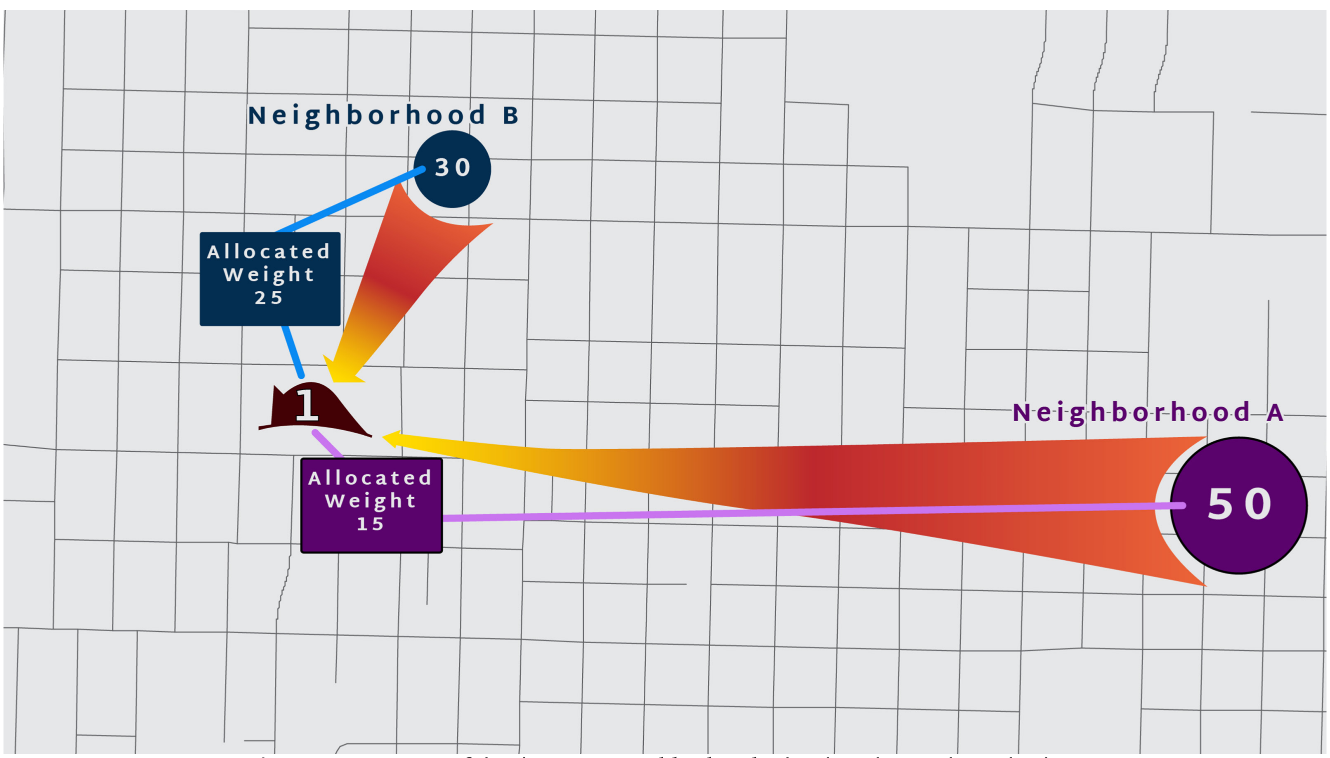


Figure 2.7: Because of the distance, *Neighborhood A* has less demand weight than *Neighborhood B*. This is the principle behind allocated demand weight.

Tutorial: Optimizing Organic Waste Diversion Using a Vehicle Routing Model

In this tutorial, you will explore modeling network paths using a **vehicle routing model (VRM)**. You will create a **network dataset** and modify route parameters to determine the least-cost paths down the length of a network.

Estimated time to complete this tutorial: 6 hours

Learning Outcomes

Readers should be able to accomplish the following outcomes by the end of this tutorial:

- » Review how to acquire data from a public source
- » Geocode an address
- » Review adding XY data
- » Review Data Management Tools: project, define projection
- » Create a network dataset
- » Model the vehicle routing based on the least-cost paths
- » Adjust model parameters to create and manage delivery routes based on a set of path constraints
- » Evaluate the results of network analysis based on cost

Scenario

In this fictional scenario, the Humboldt Waste Management Authority has begun an organic waste diversion pilot program for the collection of compostable organic waste. This waste stream represents a significant portion of the waste generated in Humboldt County. The diversion of organic waste has the potential to reduce hauling costs, offset greenhouse gas emissions, and produce energy. You will use the following criteria in your analysis:

- » The name and addresses of the restaurants and businesses
- » The average organic waste volume at each location
- » The time-windows when pick-ups are allowed
- » The service time required at each location
- » The service time required at the start and end of each route
- » The operating costs and capacity for a side-loading truck
- » The operating costs and capacity for a rear-loading truck
- » The cost of labor
- » The maximum total time for each route based on a 9-hour work shift.

Conduct this analysis using the **Universal Transverse Mercator (UTM) system** along with the **North American Datum of 1983 (NAD83)**. Humboldt County lies in **Zone 10** of the UTM system. All of your data must be in this spatial reference system at the **start** of your analysis. Create working copies of your data in this spatial reference system using the *Project* tool in ArcMap as needed.

Setting up Your Workspace

In a typical workflow, you work on geospatial data using a local hard drive. When done, you compress your data and back up your work to your cloud storage so that you can retrieve the files from anywhere. When referring to a **local hard drive**, it means you are working on data physically located on the computer in front of you.

In contrast, some computers also include networked drives. **Networked drives** link to cloud storage and save the data elsewhere. Examples include services like OneDrive or Google Drive. For this tutorial, use the **desktop** as your local hard drive location. You may also use an external USB drive if you plan to work in multiple places.

You must avoid using networked drives while you work.
They increase the processing time and can cause technical glitches.

In this book, you use a particular folder structure. Start by creating your workspace folder on the local hard drive. A **workspace** is a folder or series of folders that contain all of your project files. The top-level folder in your workspace should indicate the activity or the project on which you are working. Organize all of your work within the workspace folder. On your **desktop**, create a new folder and give it a descriptive name, such as **VehicleRoutingModel**. Be sure there are no spaces. You may use underscores instead of spaces. Inside this folder, create the following three subfolders: *original*, *working*, and *final*. Having a standardized folder structure helps to keep a project organized, primarily when you are working with multiple partners. The folder structure you see here (**Figure 2.8**) is the standard used in each of the tutorials presented in this book.

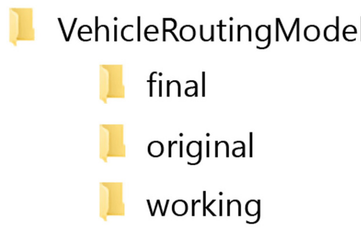


Figure 2.8: This diagram represents a basic folder structure used in this book.

As the name indicates, use the **original folder** for storing original, unaltered data. As you are working on a project, if, for some reason, your working version of the data gets lost or corrupted, you can go back to your *original* folder and find a fresh copy of the data. Use the **working folder** for data that you *create* or *alter* while working on your project. Use the **final folder** for storing any output you produce as a result of your work, such as images, maps, tables, or reports. Setting up a standard folder structure for a project is good practice and a habit you should develop.

Disable Background Geoprocessing

In the ArcGIS software, the *Background Geoprocessing* setting is often turned on by default. This setting allows users to continue to work while *a* tool is running in the background. However, sometimes this setting will stop tools from running or cause other unforeseen problems. To reduce that chances of the ArcGIS software crashing during this exercise, turn this setting off. In ArcMap, **open a new blank map document**. Open the *Geoprocessing options* from the *Geoprocessing* menu. Under *Background Geoprocessing*, uncheck the box next to the word *Enable* (**Figure 2.9**).

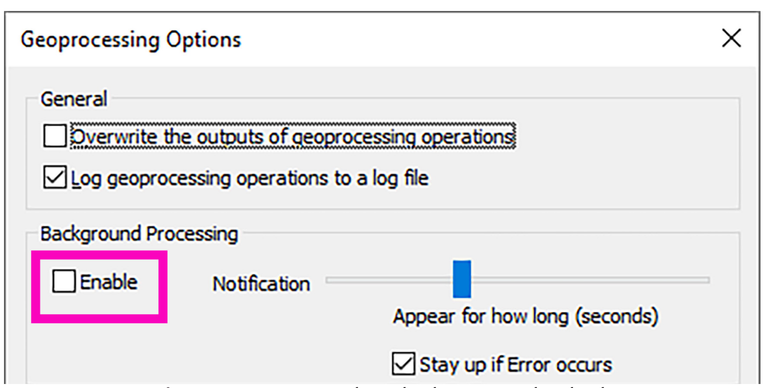


Figure 2.9: Be sure that the box is unchecked.

Network Analyst Extension

The steps in this activity involve using the *Network Analyst* extension. After launching ArcMap, make sure this extension is activated. You will also need to open the *Network Analyst Toolbar* and dock it near the top of your window for easy access. Click on the *Show/Hide Network Analyst Window* button to open the *Network Analyst Window*. Locate this button on the *Network Analyst Toolbar* (**Figure 2.10**).

Figure 2.10: Click the *Show/Hide Network Analyst Window* button

The *Network Analyst Window* is where you will enter or modify your vehicle routing model parameters. After running the model, the results load into a series of feature classes that appear in the *Table of Contents*. Novices tend to confuse the *Table of Contents* and the *Network Analyst Window* during this activity. For clarity, I recommend that you dock the *Network Analyst Window* on the right side of your screen while keeping the *Table of Contents* to the left (**Figure 2.11**).

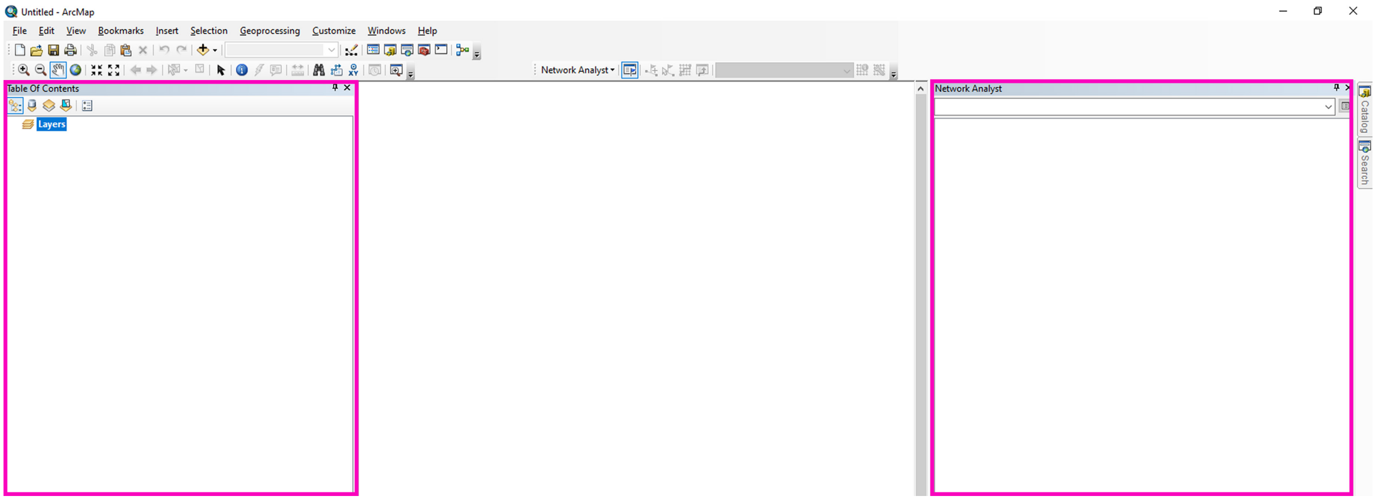


Figure 2.11: The *Table of Contents* docked on the right, and the *Network Analyst Window* docked on the left. Readers that separate the *Table of Contents* from the *Network Analyst Window* encounter less confusion.

Skill Drill: Acquire GIS data from the Humboldt County Website

To create a network dataset, you will need a road layer. In previous courses, you learned how to acquire data from public sources. Here you will download a shapefile containing the roads for Humboldt County. Navigate to the ***Humboldt County GIS Data Download page***[1]. Locate a shapefile containing the road data and download the file. Currently, there are multiple road layers on the website. You want the one that says *Humboldt County GIS Roadway Centerline* (**Figure 2.12**)

Figure 2.12: Locate the roadway centerline shapefile under the frequently requested data sets.

If, for some reason, the website is down or the data is no longer available, you can download a backup copy of the data for this tutorial using the link provided in Appendix A.

1 URL: https://humboldtgov.org/276/GIS-Data-Download

Decompress the file in your *original* folder and check the spatial reference system. If necessary, use the *Project* tool to create a copy of the data with the **NAD 83 UTM Zone 10** spatial reference system. Save the output to your *working* folder (**Figure 2.13**).

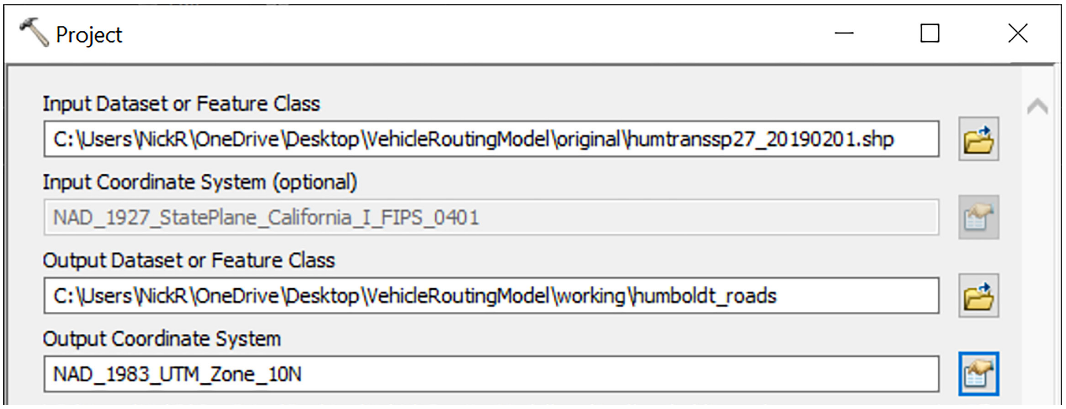


Figure 2.13: The *Project* tool never alters the original data. It creates a copy of the data with a new spatial reference system.

Add the new shapefile to your *Table of Contents* (**Figure 2.14**). This layer defines the spatial reference for the dataframe window because it is the first layer loaded into the *Table of Contents*.

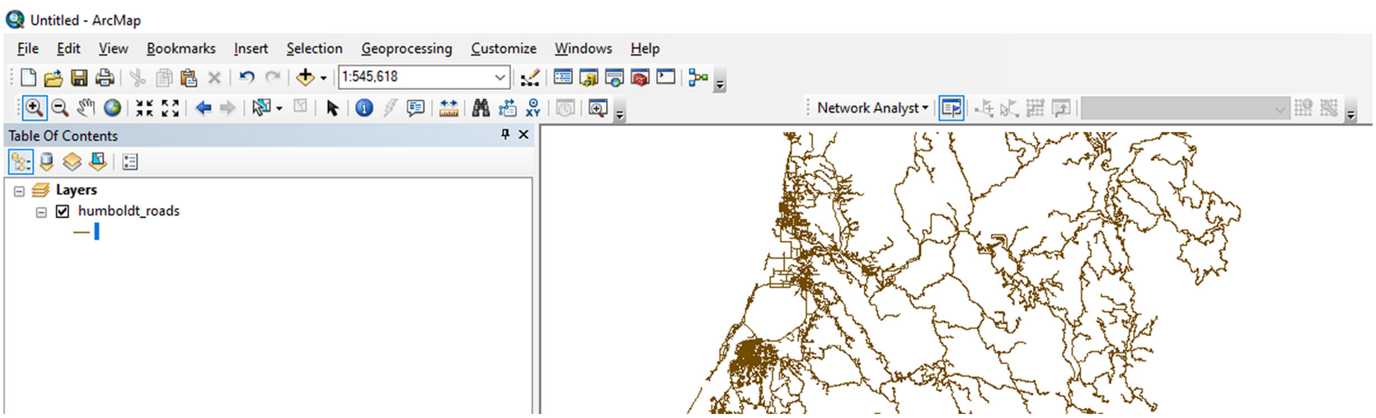


Figure 2.14: In ArcMap, the dataframe window adopts the spatial reference from the *first* layer loaded into the *Table of Contents*.

Return to the ***Humboldt County website***[1] and download the *County Boundary* shapefile (**Figure 2.15**).

Figure 2.15: Locate the County boundary shapefile under the frequently requested data sets.

1 URL: https://humboldtgov.org/276/GIS-Data-Download

Decompress the file into your *original* folder. Use the *Project* tool to create a copy of the data with the **NAD 83 UTM Zone 10** spatial reference system. Save the output to your *working* folder. Add the new shapefile to your *Table of Contents* (**Figure 2.16**)

In ArcMap, you may need to refresh your original folder in the Arc Catalog Window after decompressing the ZIP file.

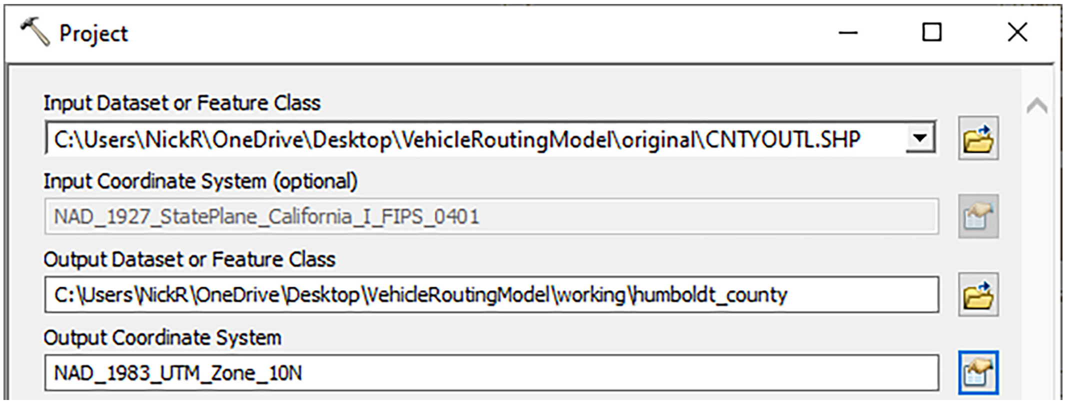


Figure 2.16: The *Project* tool never alters the original data. It creates a copy of the data with a new spatial reference system.

Repeat these steps for the *City Boundaries* shapefile. Use the *Project* tool to create a copy of the data with the **NAD 83 UTM Zone 10** spatial reference system. Save the output to your *working* folder. Add the new shapefile to your *Table of Contents* and arrange the layers so that each layer is visible in the *Data Frame* Window.

Optionally, you may change the Data Frame background color to blue to represent the Humboldt Bay.

Skill Drill: Clip the Road layer to the City of Eureka

Network analysis can place significant demands on a computer's processing power and memory. To save time during the analysis, you will need to constrain the road network to the area around the City of Eureka. Use an attribute query to create a layer based on the areas related to the City of Eureka and the surrounding regions. At this time, Humboldt County refers to these regions as the *City of Eureka* and the *City of Eureka Sphere* in the attribute table (**Figure 2.17**). Check for new alterations in the data in case the names in the attribute table have recently changed. Export the selected features as a new shapefile.

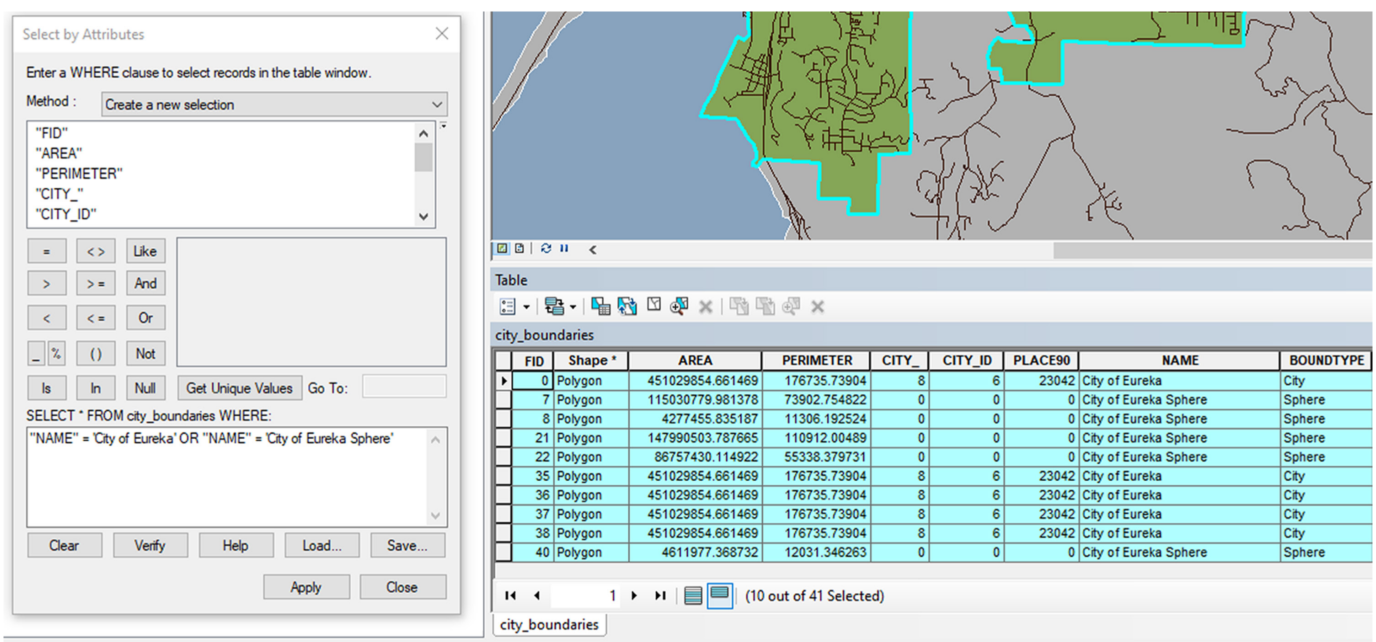


Figure 2.17: You can export the selected features to create a permanent shapefile.

Some roads may extend slightly outside of the City of Eureka and related regions. To prevent altering the geometry of the road segments needed in our analysis, create a one-mile buffer around the City of Eureka layer (**Figure 2.18**)

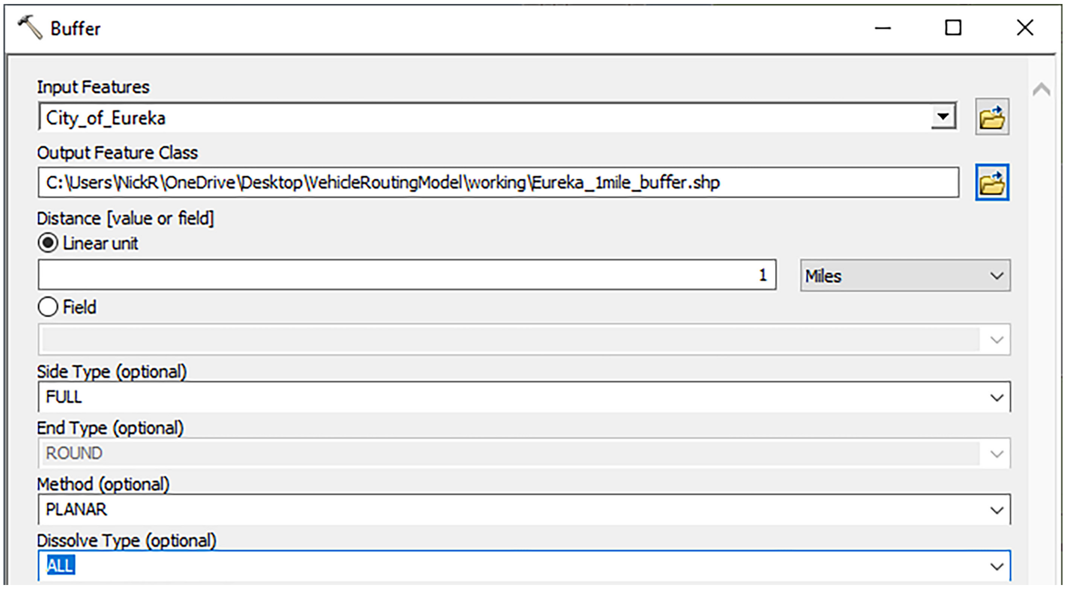


Figure 2.18: Create a one-mile buffer around the City of Eureka layer. Check to make sure your settings match.

Clip the Humboldt road layer and use the buffer as the *Clip Feature* (**Figure 2.19**). This clip operation ensures that your network dataset serves your needs while excluding the remaining roads in Humboldt County from our analysis.

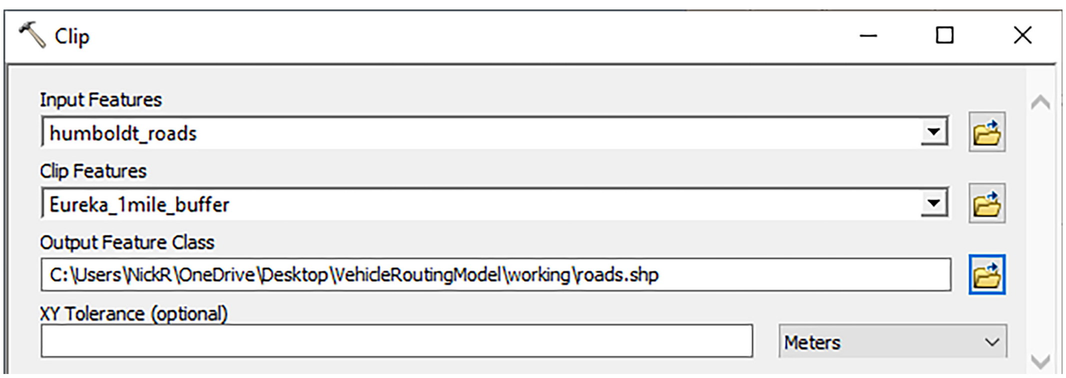


Figure 2.19: Use the *Clip* tool to constrain the roads to the Eureka area. Check to make sure your settings match.

Go to the ArcMap properties window and set the Pathnames to *Store relative pathnames to data sources* (**Figure 2.20**).

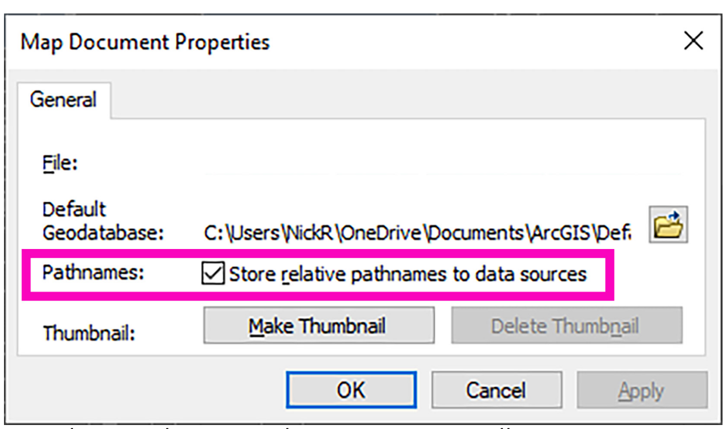


Figure 2.20: The store relative pathnames to data sources setting allows ArcMap to remember the location of the data sources relative to its current position. *This image has been modified for clarity.*

Take a moment to save your map document. Be sure to give the map document a meaningful name. Take care to save the map document in your workspace folder, *VehicleRoutingModel* (**Figure 2.21**).

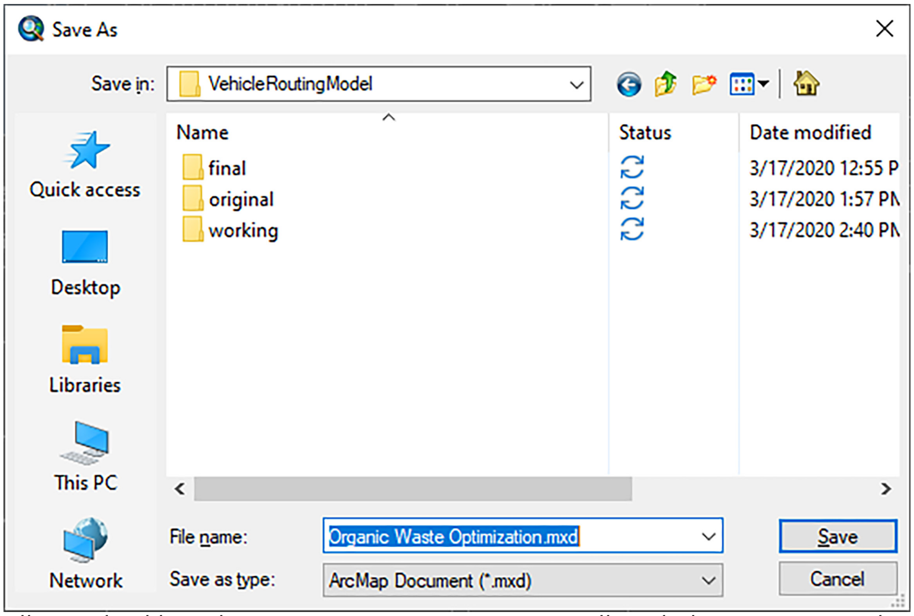


Figure 2.21: Generally, you should avoid spaces. However, empty spaces are allowed when naming map document files (.mxd)

Remove all unnecessary layers so that only the clipped roads and the Humboldt County boundary remain (**Figure 2.22**).

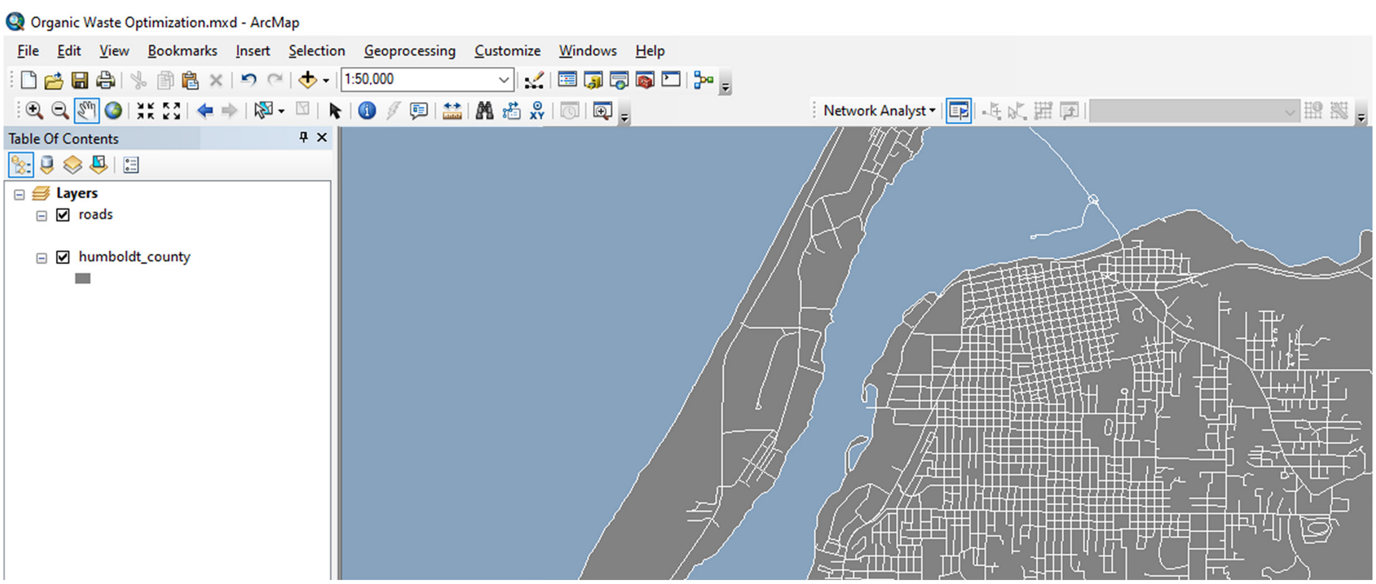


Figure 2.22: Maintaining a clean *Table of Contents* helps to prevent confusion and errors.

Skill Drill: Add a Time Cost Attribute to the Road layer

The vehicle routing model requires a time unit **cost variable** for the network dataset. Create a new field in the attribute table of the clipped-road layer (**Figure 2.23**). Name this field **Minutes**. This field will record the travel time, in minutes, for each road segment.

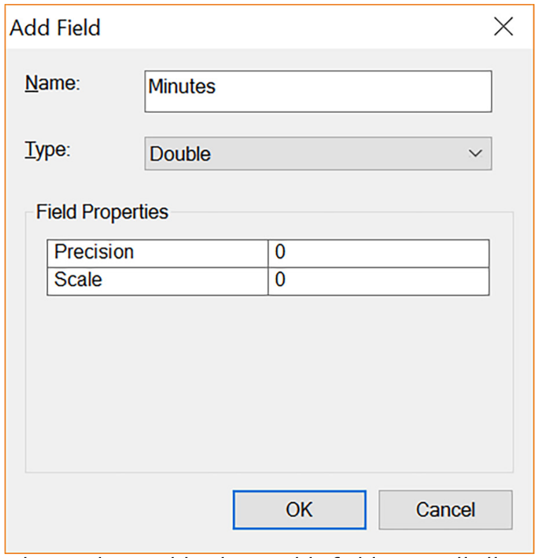


Figure 2.23: When adding a new field to the attribute table, the Double field type will allow for larger numbers with decimal places.

In most cases, the road segments will have values of less than a one minute travel time. Therefore, the field type should be either a float or a double. If you are ever unsure which to choose, I recommend that you always choose the *double* field type since it will allow for a larger number if necessary.

To calculate the travel time, in minutes, you will use the length of each road segment along with the speed limit. The road layer should have the length of each segment recorded in a *Length* field. However, since you clipped the road layer, these values will need to be updated to account for any road segments that may have been altered during the *Clip* operation. To update the length values, use the *Calculate Geometry* function in the attribute table (**Figure 2.24**). Recall that you can access Calculate Geometry by right-clicking on the field name representing the length of road segments. Make sure the values are saved in **meters**.

Note: If you can't find a field representing length, you can create your own by adding a new field in the attribute table. Set the field type to Double. Then use Calculate Geometry to populate the field.

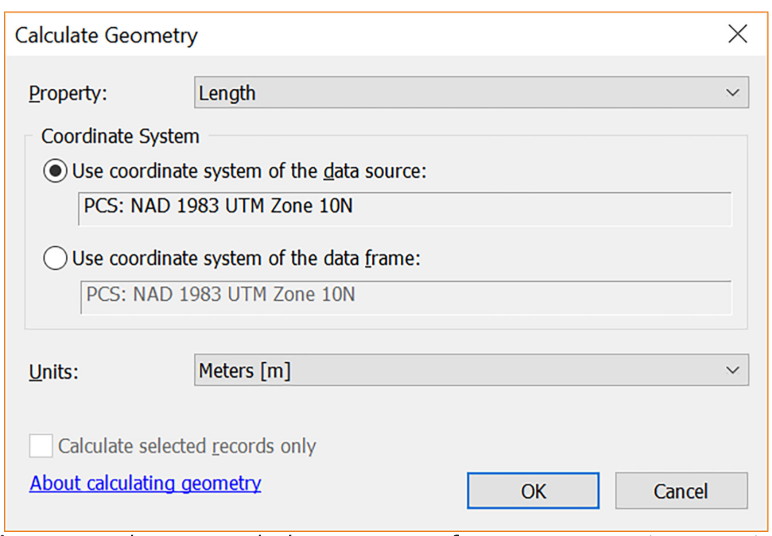


Figure 2.24: Always use Calculate Geometry after any geoprocessing operation that alters the feature geometry, such as a clip, or an intersect.

The road layer has two attributes that record the speed limit, AB_Speed, and BA_Speed. These attributes represent the speed limit, in miles per hour, for each direction. For consistency, use the AB_Speed in your calculation. Use the *Field Calculator* to compute the travel time and to populate the *Minutes* field. Your formula should factor in the **AB_Speed** (miles per hour) and the **Length** (meters) of each road segment to determine the travel time in minutes (**Figure 2.25**). Remember, you can access *Field Calculator* by right-clicking on the field name, *Minutes*.

Coming up with the right formula will involve converting miles per hour to meters per minute. Then it requires using the length of the roads to determine the travel time in minutes.

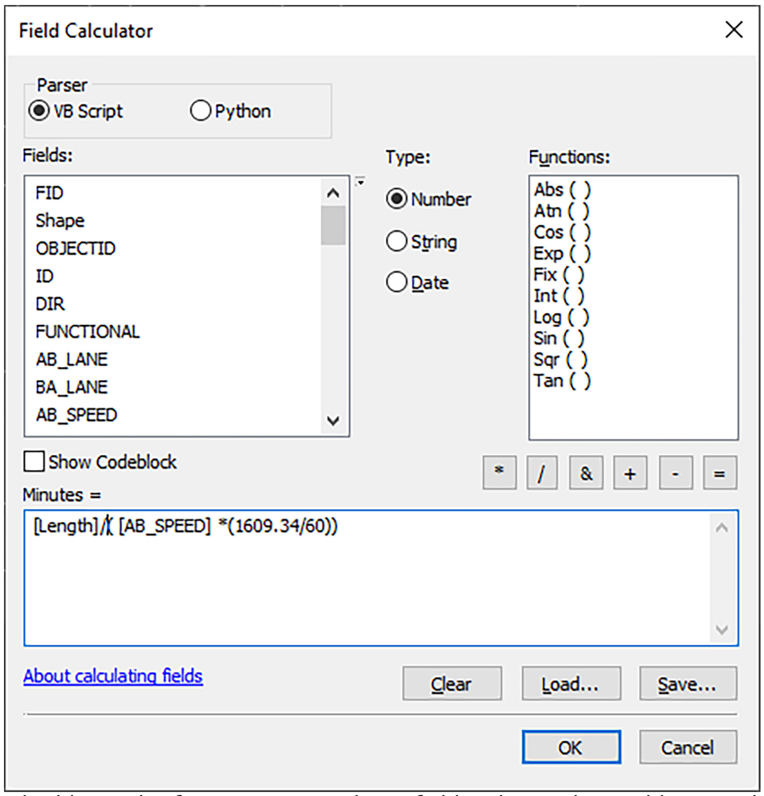


Figure 2.25: You can build complex functions to populate a field in the attribute table using the *Field Calculator*.

Once you have successfully calculated the minutes, you can verify the accuracy of your calculation by checking the travel time of the HWY 255 Bridge over Humboldt Bay. Just north of the exit for Woodley Island, the travel time of the road segment on HWY 255 is 1.268598 minutes (**Figure 2.26**).

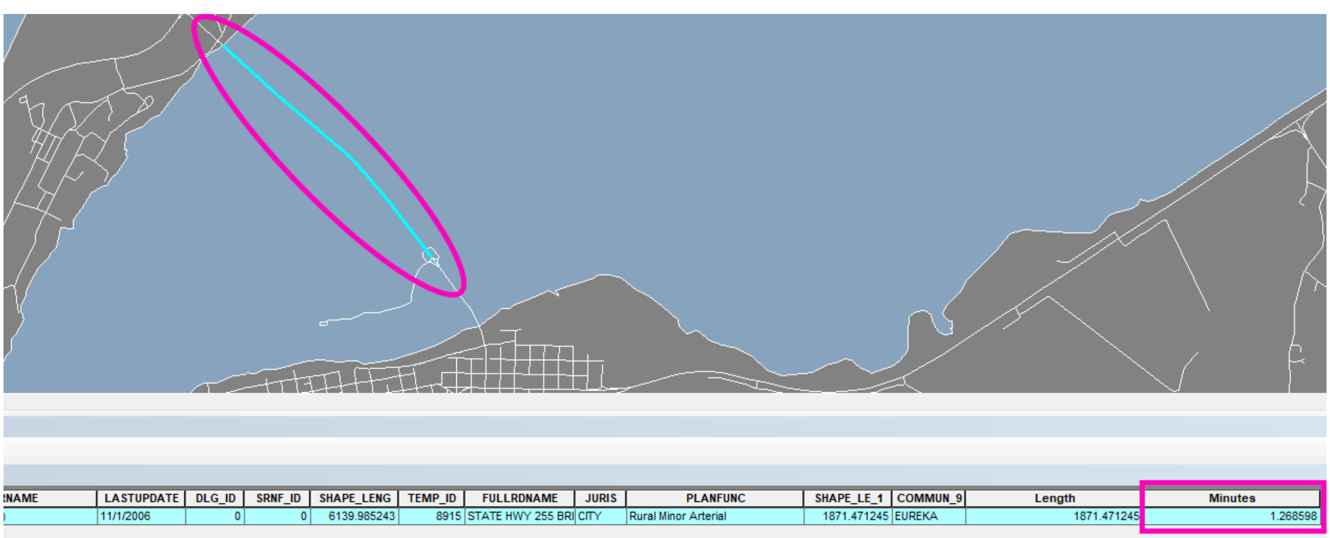
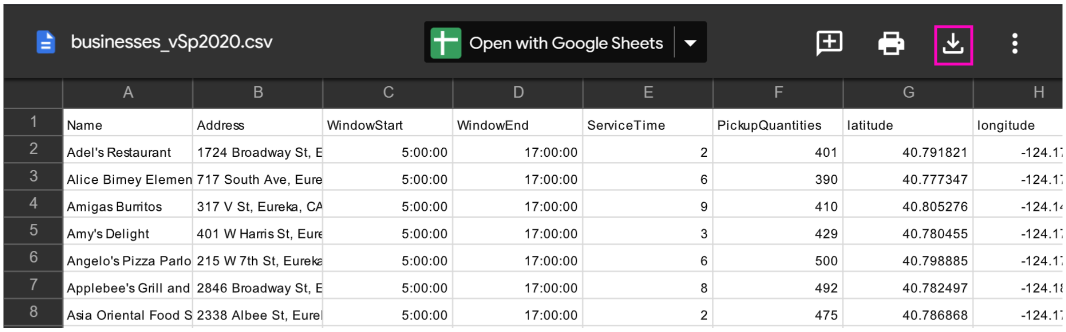


Figure 2.26: Use this road segment of HWY 255 to verify that your formula in the *Field Calculator* is correct.

Skill Drill: Adding XY Data

In this fictional scenario, a small number of local businesses have decided to participate in the pilot program. Information about these businesses, including location and the average weight of organic waste, is available in a ***comma-separated values (CSV) file***[1]. Download the file and save it to your *original* folder (**Figure 2.27**).



	A	B	C	D	E	F	G	H
1	Name	Address	WindowStart	WindowEnd	ServiceTime	PickupQuantities	latitude	longitude
2	Adel's Restaurant	1724 Broadway St, E	5:00:00	17:00:00	2	401	40.791821	-124.1
3	Alice Birney Elemen	717 South Ave, Eure	5:00:00	17:00:00	6	390	40.777347	-124.1
4	Amigas Burritos	317 V St, Eureka, CA	5:00:00	17:00:00	9	410	40.805276	-124.1
5	Amy's Delight	401 W Harris St, Eure	5:00:00	17:00:00	3	429	40.780455	-124.1
6	Angelo's Pizza Parlo	215 W 7th St, Eureka	5:00:00	17:00:00	6	500	40.798885	-124.1
7	Applebee's Grill and	2846 Broadway St, E	5:00:00	17:00:00	8	492	40.782497	-124.1
8	Asia Oriental Food S	2338 Albee St, Eurel	5:00:00	17:00:00	2	475	40.786868	-124.1

Figure 2.27: For the best results, use the Chrome browser to download the CSV file. Click the download button on the upper right.

Launch Microsoft Excel open the CSV file. Here you will see a table of businesses participating in the pilot program. Take a moment to read through available information about the businesses. Included in the table are a series of fields you will use in your vehicle routing model. Each business has a name and address. A time window, specified by the *Window Start* and *Window End* fields, indicates when organic waste pick-ups are allowed in the model. In this scenario, most businesses allow pick-ups throughout the day, yet, a few only allow pick-ups between 5 am and 8 am. The service time indicates the average time, in minutes, spent at each location during a pick-up. The pick-up quantities field indicates the average weight of the organic waste in pounds. This attribute will be used in the model in conjunction with the cargo capacity of the garbage truck. Together, these attributes determine when a truck is full and must return to the garbage station to unload before continuing. After you have looked over the CSV table, be sure to close Microsoft Excel before returning to the ArcGIS software.

It is important to remember that each geodetic datum as a unique set of latitude and longitude values. For example, the latitude and longitude values for your home using the North American Datum of 1983 will be different than the latitude and longitude values using the North American Datum of 1927. Likewise, the World Geodetic Datum of 1984 will use yet another set latitude and longitude values to define the location of your home. Like most latitude and longitude values, you obtain from the internet or from a GPS receiver, the decimal degrees in this CSV table are currently in the geographic spatial reference system **WGS 1984**. When adding XY data, the ArcGIS software only reads the decimal degrees. It is unaware to which spatial reference system these decimal degree values belong. Always specify the datum *source* when adding XY data.

In the *Catalog Window*, expand the *original* folder and right-click on the business CSV file. Select *Create Feature Class*, then choose *From XY Table* (**Figure 2.28**).

In ArcMap, you may need to refresh your original folder to see the CSV file.

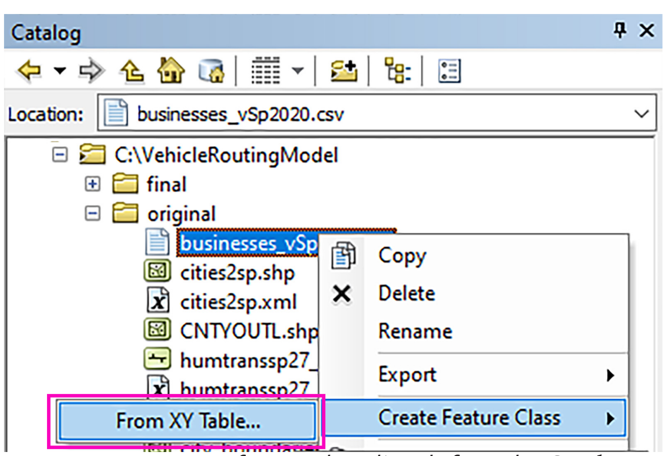


Figure 2.28: You can create a feature class directly from the *Catalog Window*.

1 URL: http://bit.ly/OrgWaste_business_data

When the *Create Feature Class from XY Table* window opens, click the button that says *Coordinate System of Input Coordinates*. Open the *Geographic* folder, then the *World* folder. Locate **GCS WGS 1984** and set that as the spatial reference.

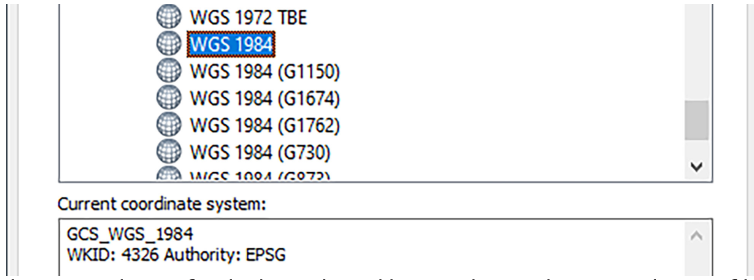


Figure 2.29: The source datum for the latitude and longitude coordinates in the CSV file is WGS 1984.

On the *Create Feature Class from XY Table* window, click the yellow file folder icon and save the output to the *working* folder. Name the file *businesses*. Next, click the *Advanced Geometry Options* button (**Figure 2.30**).

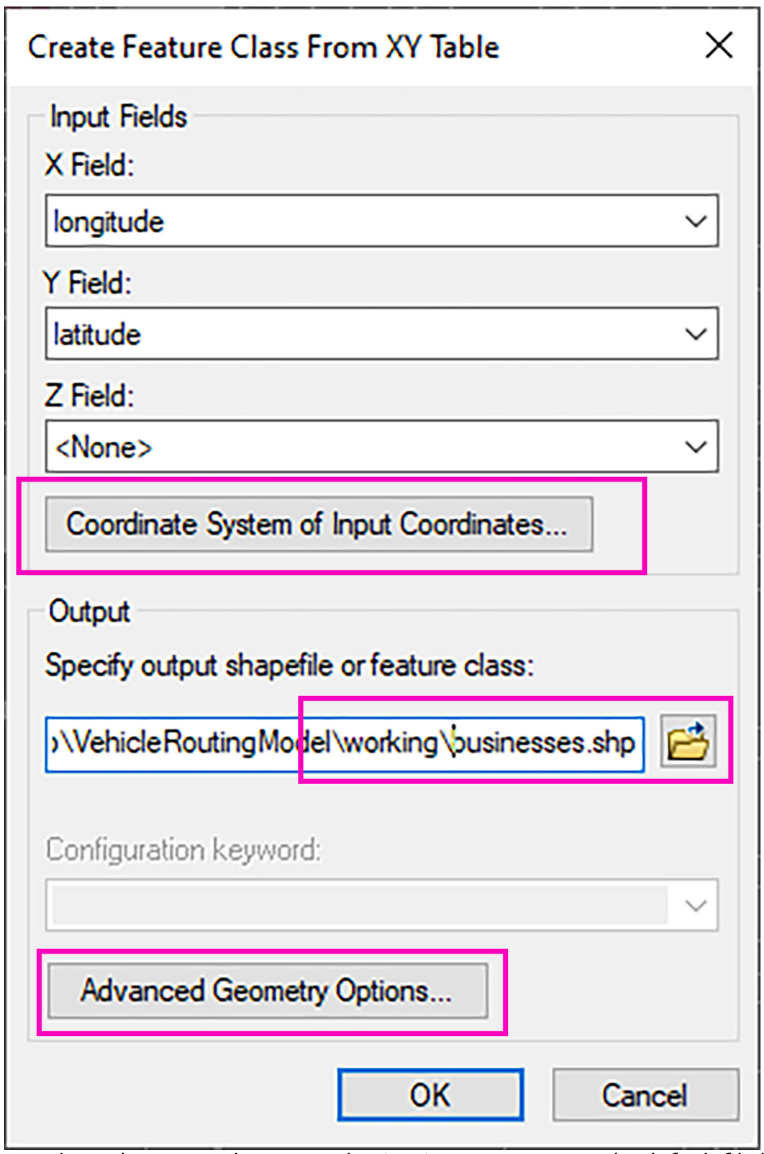


Figure 2.30: Always browse to the output destination. Never accept the default file location.

When the *Advanced Geometry Options* window appears, click the radio button next to *Use the map's spatial reference* (**Figure 2.31**). This option has the effect of transforming the output shapefile so that it uses **NAD 83 UTM Zone 10** as its spatial reference (**Figure 2.31**).

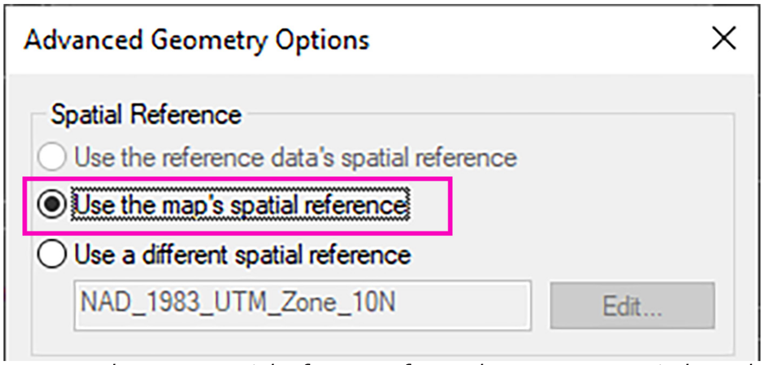


Figure 2.31: The map's spatial reference refers to the *Data Frame* window, which currently uses **NAD 83 UTM Zone 10** as the display projection.

Add the new businesses layer to the *Table of Contents* (**Figure 2.32**).

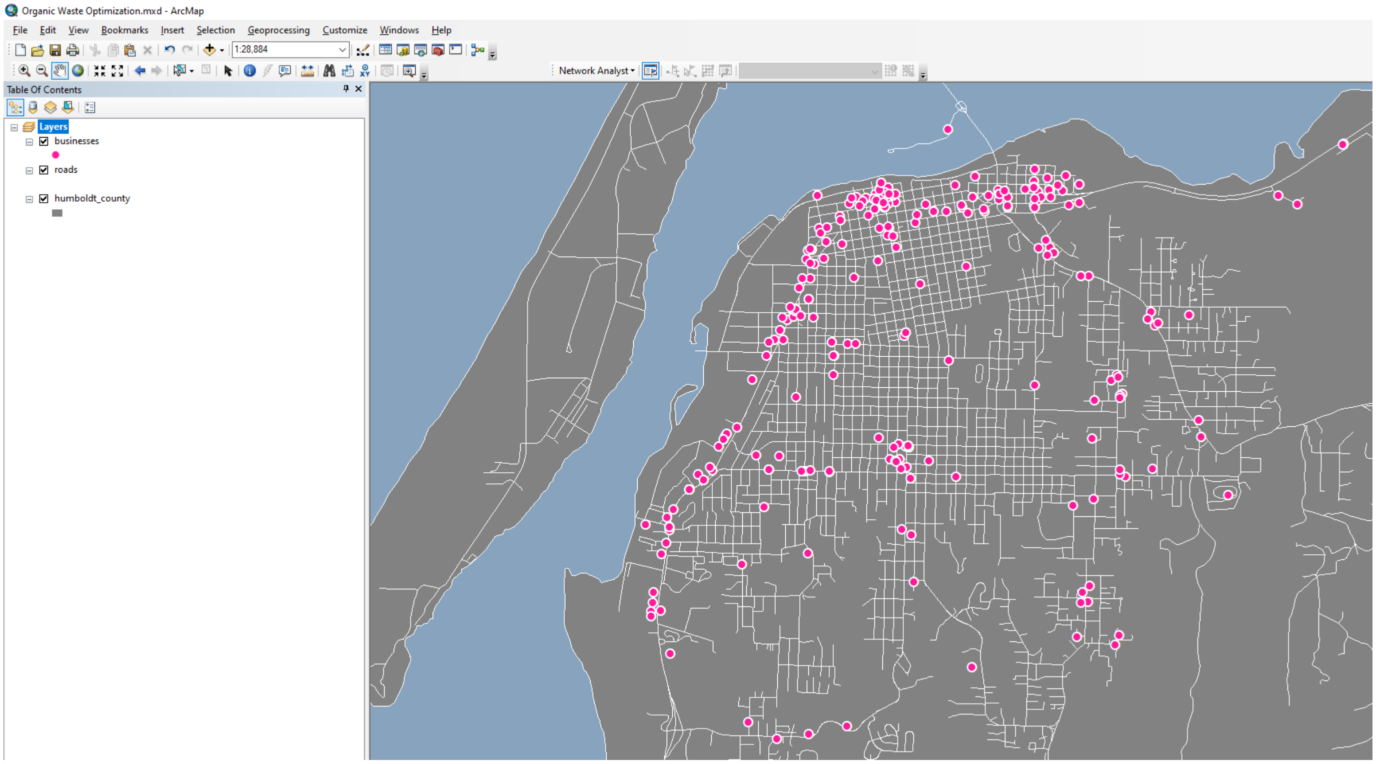


Figure 2.32: Only the business layer with the correct spatial reference should appear in the *Table of Contents*.

Skill Drill: Geocoding an Address and Creating a CSV Table to Import As XY Data

The Eureka Garbage Dump will be used in the vehicle routing model as a place where the route will begin and end. It will also serve as the location for route renewals. Without a route renewal, the garbage trucks will eventually fill to capacity. At this point, the route will end whether or not all of the locations have received service. By adding a route renewal location to the model, the garbage trucks will navigate back to the garbage dump, unload the cargo, and then continue to service the next business on the route. The Eureka Garbage Dump is located at the following address:

1059 W Hawthorne St, Eureka, CA 95501

Use **Google Maps** to geocode the address and obtain the latitude and longitude values for this address in decimal degrees. Google Maps uses **WGS 1984** as the datum for latitude and longitude values. In Microsoft Excel, create a CSV table with the following field headers: name, latitude, longitude. You may use *Eureka Garbage Dump* as the name attribute for the route renewal (**Figure 2.33**). Enter the latitude and longitude values in the appropriate fields. *Save as* a CSV table in your *original* folder. Be sure to *close* Excel when you are done saving the file.

In ArcMap, you may need to refresh your original folder to see the CSV file.

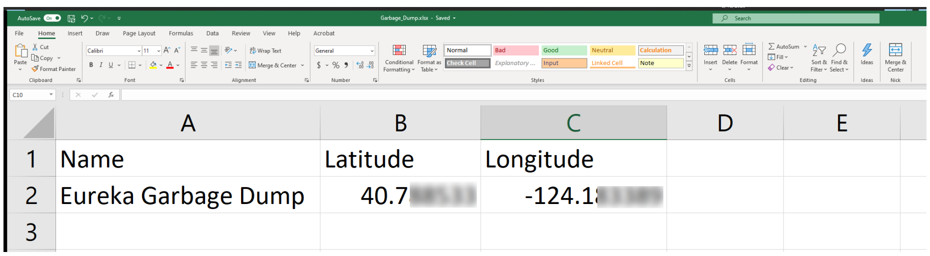


Figure 2.33: You will need to enter the latitude and longitude values acquired from Google Maps manually.

In the *Catalog Window*, expand the *original* folder and right-click on the garbage dump CSV file. Select *Create Feature Class*, then choose *From XY Table*. Follow the same workflow as for the businesses CSV to create a file that uses **NAD 83 UTM Zone 10**. Add the new file to your *Table of Contents* (**Figure 2.34**).

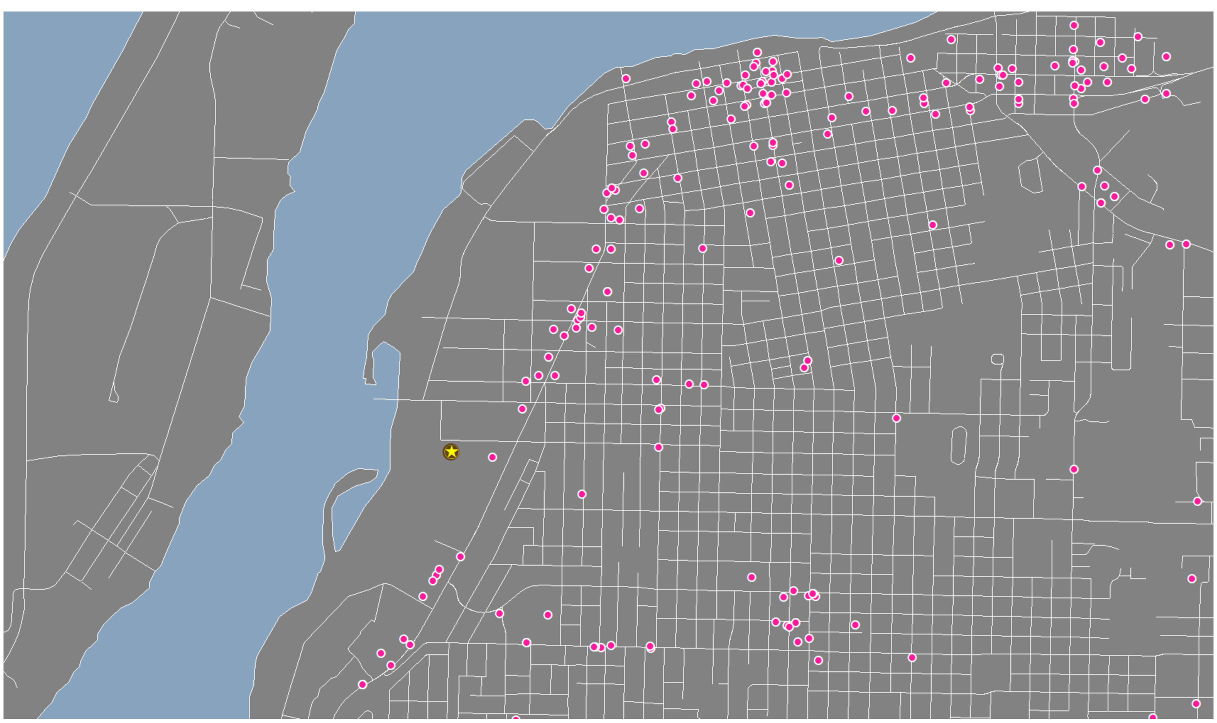


Figure 2.34: The Eureka Garbage Station is indicated here as a yellow star.

Creating A Network Dataset Using ArcCatalog

To conduct network analysis, you must first create a **network dataset** with roads (edges) and intersections (junctions). The ArcGIS software creates a network dataset from source features, such as lines and points. It then stores the connectivity between these source features. In this instance, the GIS uses the road layer as the source feature for the edges. It creates a junction feature class automatically for the road intersections.

Open the ArcCatalog window in ArcMap and navigate to your *working* folder. In the Arc*Catalog Window*, right-click on the shapefile of your roads and select *New Network Dataset* (**Figure 2.35**).

Note: If the Network Analyst extension is not currently activated, this option may appear grayed out. To correct this problem, activate the Network Analyst extension.

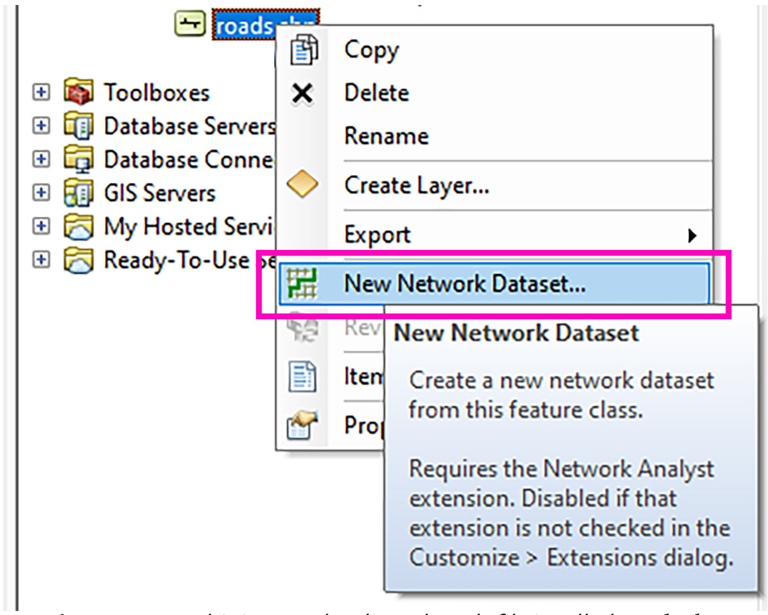


Figure 2.35: In this image, the clipped roads file is called *roads.shp*.

The New Network Dataset window opens up, and a default name is generated by the ArcGIS software ending in **_ND** (**Figure 2.36**). You may accept the default name if you wish and *click Next*.

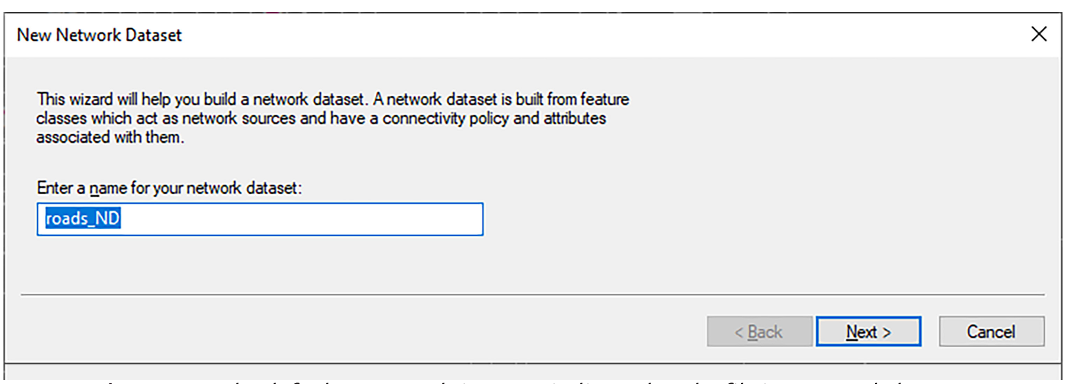


Figure 2.36: The default name ends in ND to indicate that the file is a network dataset.

Click Next to accept the default *Yes* for modeling turns in the network. On the next page, you have the option to change the default connectivity settings. In most cases, you will not need to change these settings. By default, road segments connect to each other at endpoints. To check these settings, click on the Connectivity button (**Figure 2.37**). Make sure the source feature, the road layer, is set to *End Point*. Then *click OK* and *Next*.

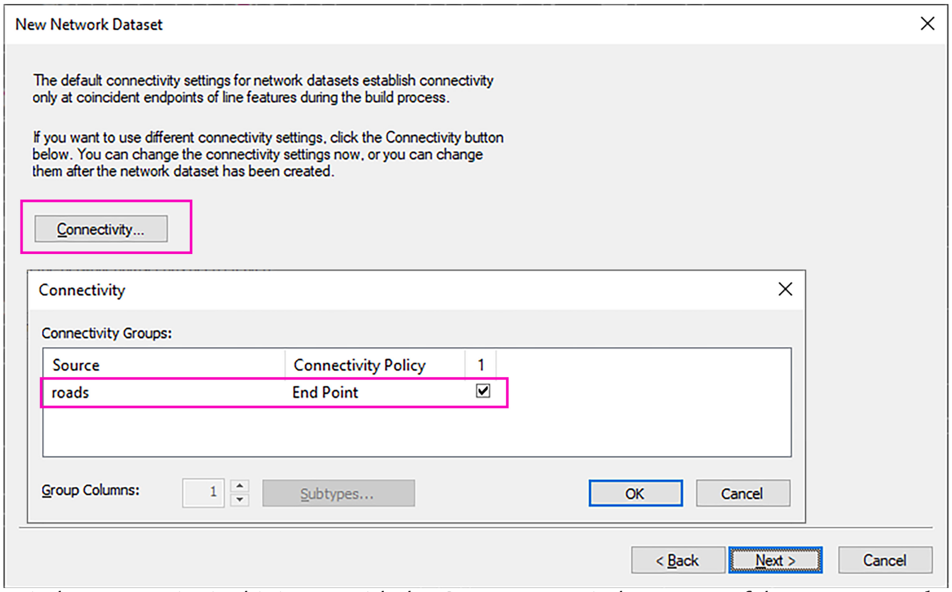


Figure 2.37: Two windows are active in this image with the *Connectivity* window on top of the *New Network Dataset* window.

When elevation fields are used, two edges that have the same endpoint locations, but different elevations remain disconnected. This is useful for modeling multilevel transportation infrastructure. This dataset does not contain a field with elevation information. *Click Next* to accept the default None for elevation settings.

The next window displays the attribute information for the network dataset (**Figure 2.38**). The ArcGIS software automatically tries to assign standard cost attribute fields, such as Length and Minutes, to the network dataset. Check the fields the ArcGIS software will use by clicking on the **Evaluators** button. This opens the *Evaluators* dialog window. From the drop-down menu next to Attributes, select *Length*. Under the *Source Values* tab, set the *Value* field to the Length field. If you will recall, this is the field updated via Calculate Geometry earlier. The *road* source feature is listed twice, one for each direction. Change the Value field to Length for both directions.

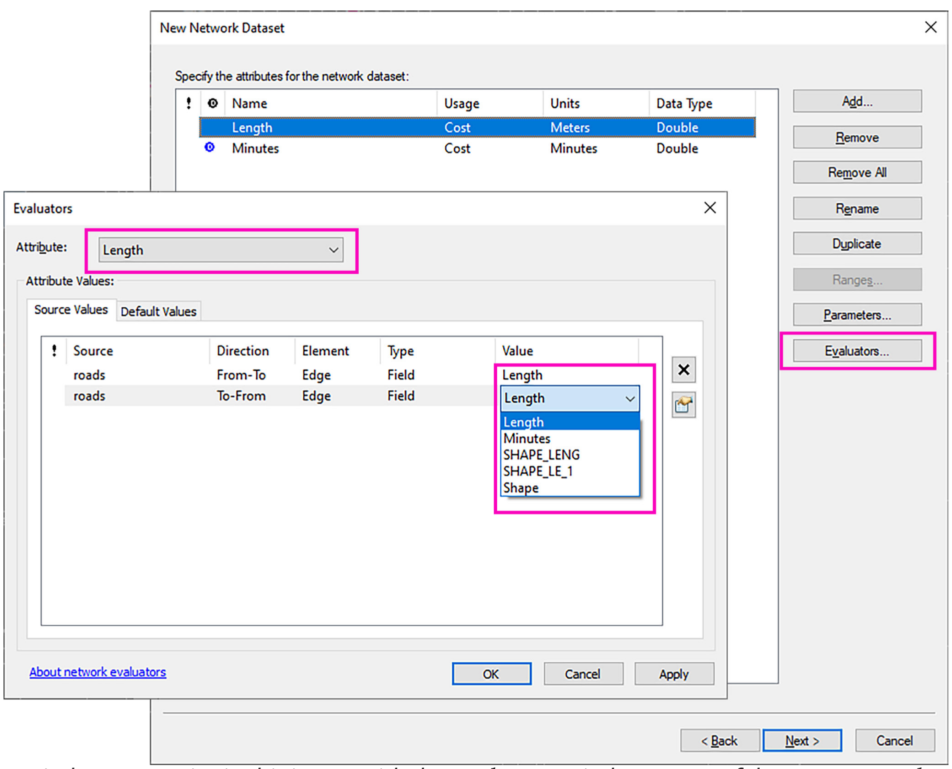


Figure 2.38: Two windows are active in this image with the *Evaluators* window on top of the *New Network Dataset* window

From the drop-down menu next to Attributes, select *Minutes*. Again, you will see the *road* source feature listed twice, one for each direction. Make sure the *Value* field for each direction is set to *Minutes* (**Figure 2.39**). If you will recall, this *Minutes* field was created earlier and records the time it takes to traverse the edges in the network dataset. Then *click OK* and *Next*.

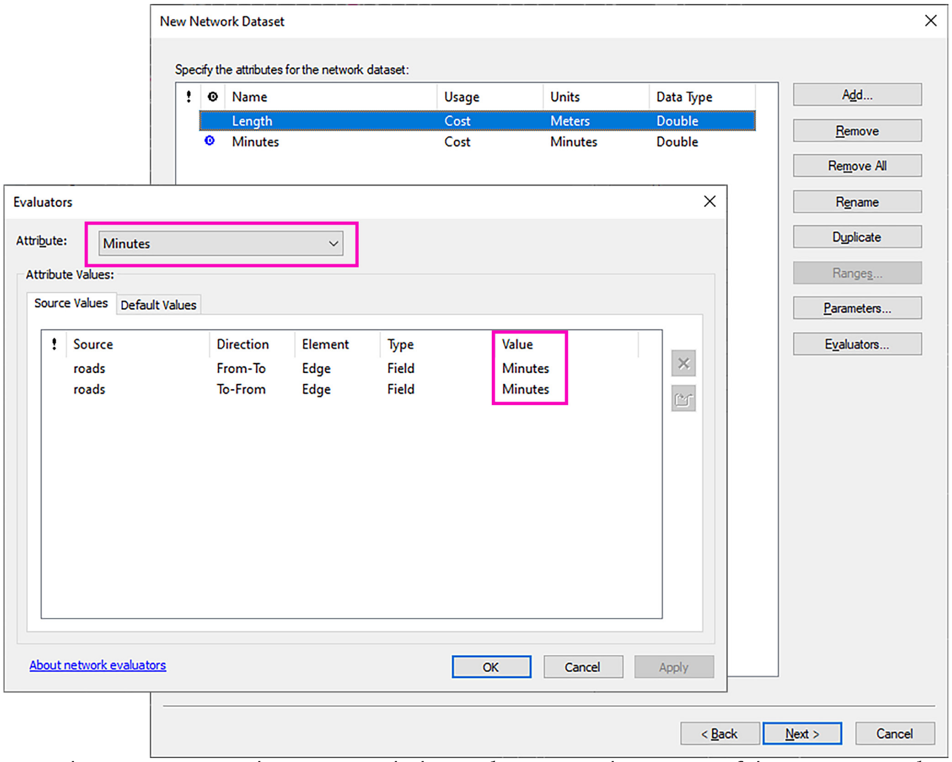


Figure 2.39: Two windows are active in this image with the *Evaluators* window on top of the *New Network Dataset* window

On the next window, *click Next* to accept the Default Travel mode. On the next window, *click Next* to accept the default Yes for driving directions settings. The last window will display a summary of the network dataset settings and attributes (**Figure 2.40**). Click *Finish* to create the network dataset.

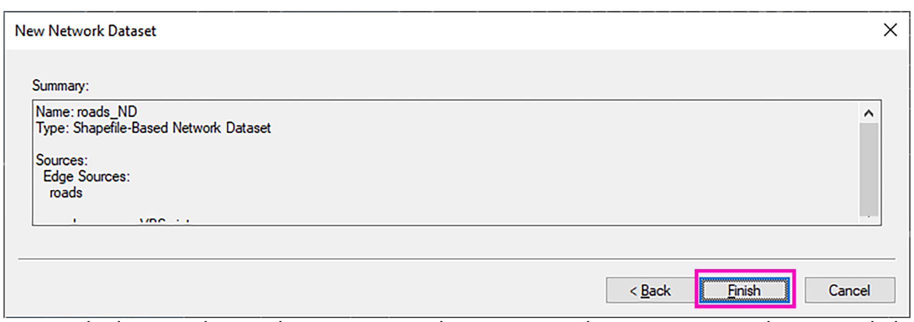


Figure 2.40: The last window in the New Network Dataset window summarizes the network dataset.

Click Yes to build the Network Dataset. Then *click No* to the dialog box that asks if you also want to add all feature classes that participate in the network dataset to the map (**Figure 2.41**).

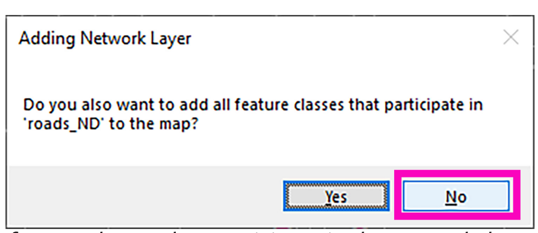


Figure 2.41: Adding all the feature classes that participate in the network dataset to the *Table of Contents* is not necessary. The network dataset references these feature classes from the *working* folder.

Be sure that only the garbage dump, the businesses, the network dataset, and the Humboldt County boundary are loaded in the *Table of Contents* (**Figure 2.42**). This prepares you to conduct network analysis with minimal clutter.

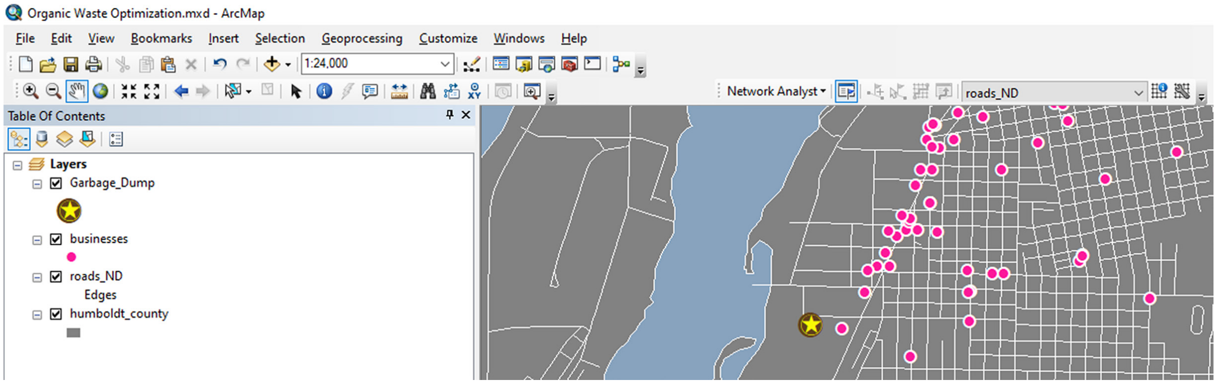


Figure 2.42: Maintaining a clean *Table of Contents* helps to prevent confusion and errors.

Setting up a Vehicle Routing Problem (VRP)

There are a significant number of variables that can be used while modeling a *Vehicle Routing Problem* (VRP). Often, they relate to routes, drivers, vehicles, pick-up locations, drop-off locations, and costs. This can help you to optimize routes for a wide variety of logistical operations. This scenario uses the operating expense of the trucks broken down by cost per mile. You will also factor in the cost of labor in terms of cost per minute. In the first model, you explore the results when using a route comprised of only one truck and one driver. In the *Network Analyst Toolbar*, choose *New Vehicle Routing Problem* (**Figure 2.43**).

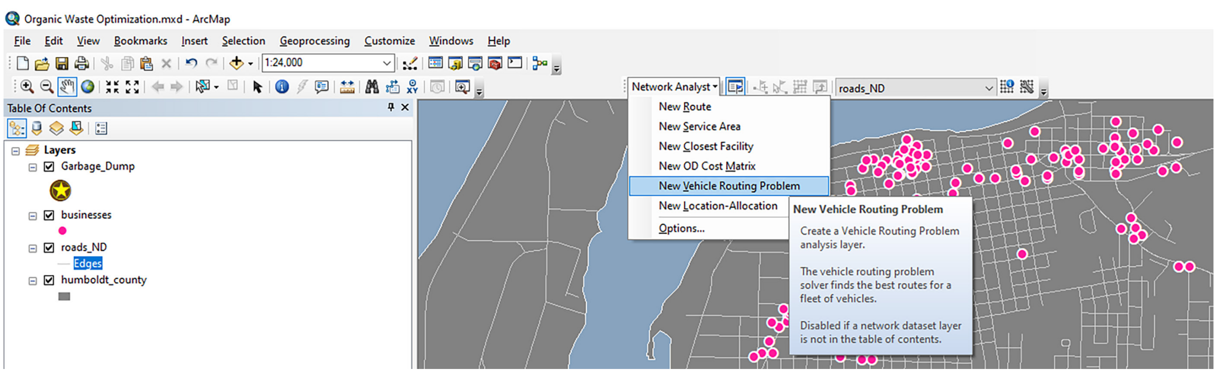


Figure 2.43: The New *Vehicle Routing Problem* option on the *Network Analyst* toolbar.

Note: It is possible to create multiple vehicle routing problems (VRPs) in one map document. Each represents a different model. In this exercise, you will only work with one. Be careful not to inadvertently create multiple VRPs.

In the *Table of Contents*, a new group layer appears. At the same time, in the *Network Analyst Window*, an empty *Vehicle Routing Problem* gets created (**Figure 2.44**).

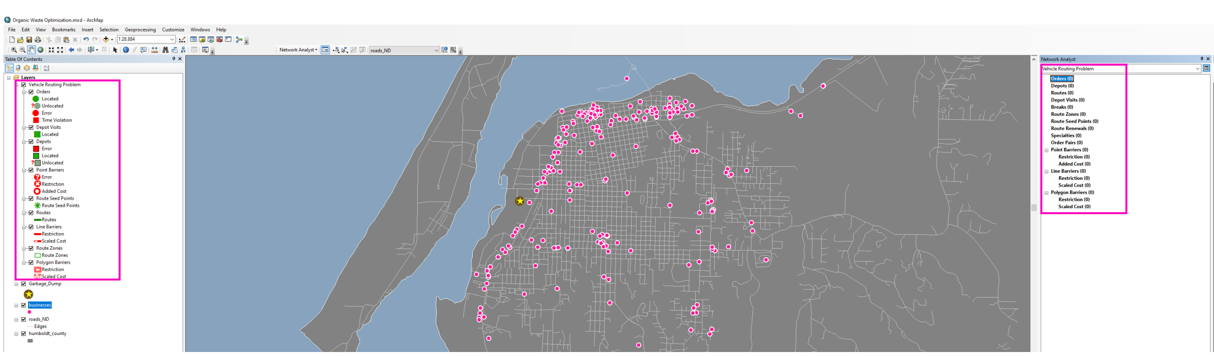


Figure 2.44: If you have not done so already, consider separating the *Table of Contents* from the *Network Analyst Window* by docking them across from each other on opposite sides of the ArcMap window.

The group layer in the *Table of Contents* is comprised of empty **memory feature classes** (**Figure 2.45**). These data types behave similarly to shapefiles but only exist *temporarily* in the computer's memory. When you close ArcMap, you will no longer have access to these layers. They must be exported if you want to make them permanent. However, they are temporary for a reason. By using memory feature classes, you can re-run the model multiple times with different sets of parameters. This workflow saves you from unnecessarily creating shapefiles that you may not need to keep in the long term. This saves on overall disk storage space and makes managing data easier as you conduct your analysis.

The *Vehicle Routing Problem* in the *Network Analyst Window* represents the model (**Figure 2.45**). It is *here* that you will enter the model parameters. Each time you adjust the model parameters, you will do so in this window as well.

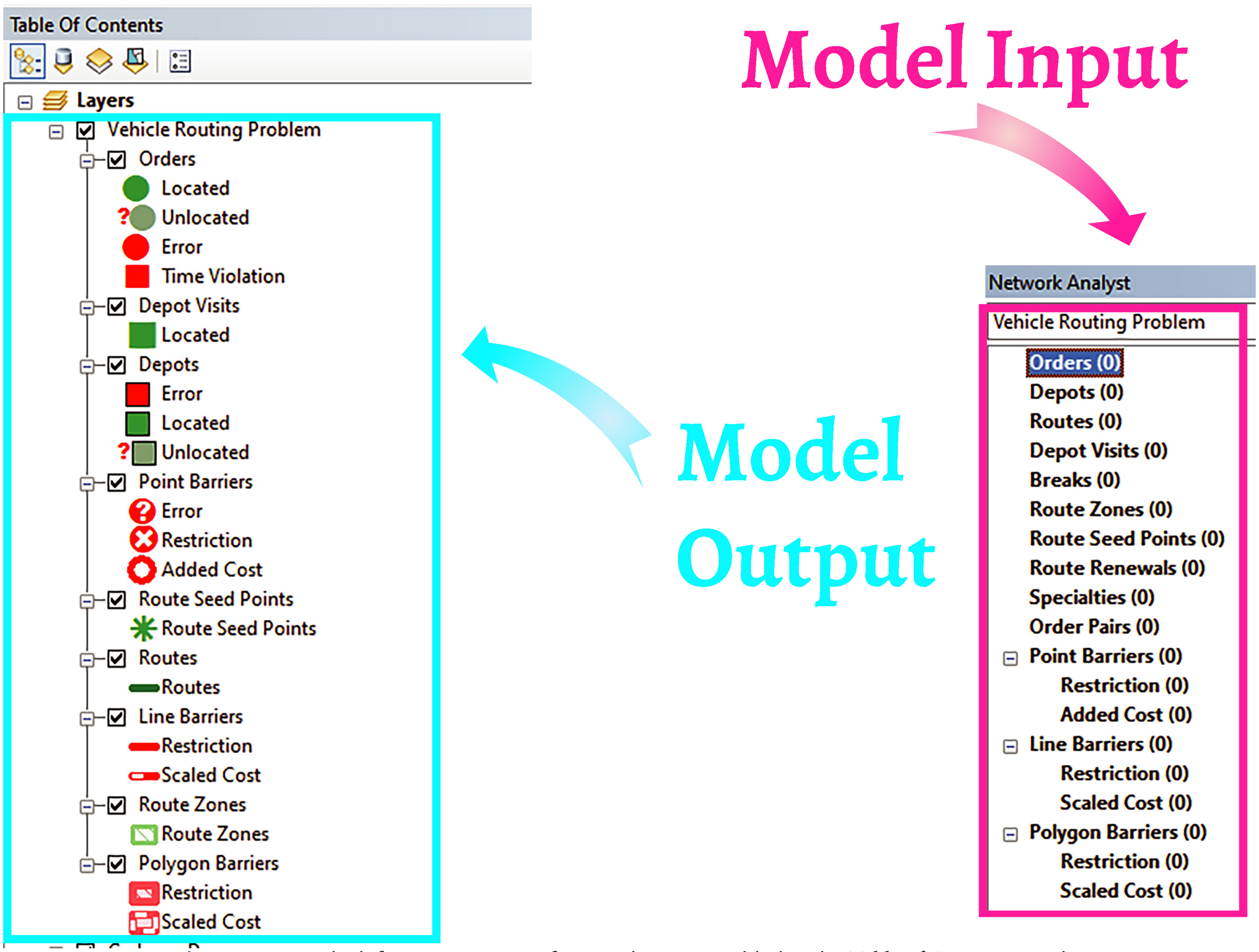


Figure 2.45: On the left, empty memory feature classes are added to the *Table of Contents*. On the right, an empty *Vehicle Routing Problem* is added to the *Network Analyst Window*.

Loading Orders into the Model

As you will recall from earlier in the chapter, the **orders** are a network class that represents places where something is **picked up or delivered**. In this scenario, organic waste will be picked up from participating businesses. The participating businesses represent our orders. *Orders* also have attributes that influence cost or place limitations on the model. In this model, the orders have attributes that include a **time window,** which indicates when organic waste pick-ups are allowed. The **service time** attribute indicates the average time, in minutes, spent at each location during a pick-up. The **pick-up quantities** attribute indicates the average weight of the organic waste in pounds. Finally, the **curb approach** attribute limits the route direction at the order location. Right-click on *Orders* in the *Network Analyst Window* and select *Load Locations* (**Figure 2.46**).

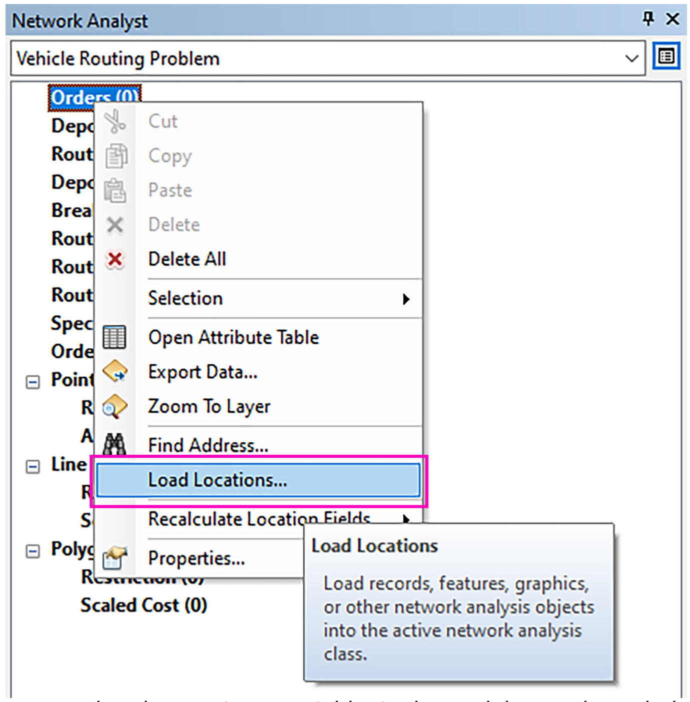


Figure 2.46: It is important to remember that any input variables in the model enter through the *Network Analyst Window*.

From the drop-down menu next to Load From, select the business layer (**Figure 2.47**). Order parameters are defined under the Location Analysis Properties pane using two attributes, Field, and Default Value. If the layer does not have a field that matches the property, a default value may be entered instead. Use the following values in the Location Analysis Properties. Keep the **default values** for any properties **not listed** here.

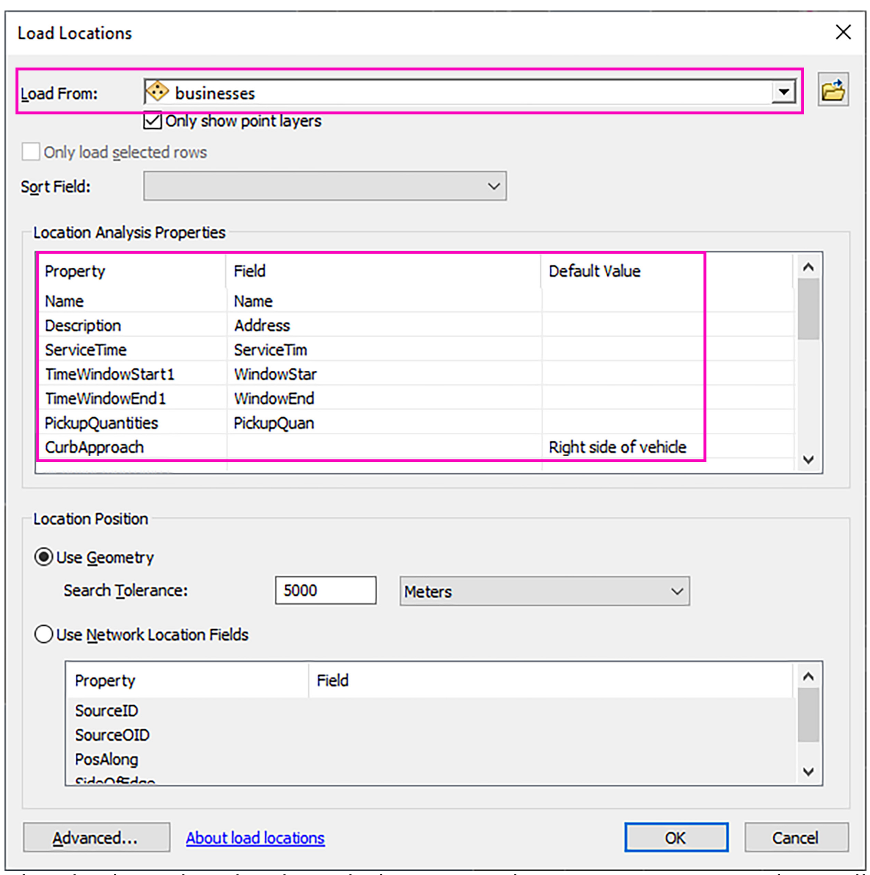


Figure 2.47: The image here has been altered so that only the custom values appear. You may need to scroll down to see them all.

Once the values above are entered, *click OK*. The *orders* should load in **two places**, the *Table of Contents* and the *Network Analyst Window* (**Figure 2.48**). The orders in the *Network Analyst Window* represent our model **input** variables. Any modification of the model parameters, or input to the model, are entered here. In the *Table of Contents*, the memory feature class, also named *Orders*, is now populated with the order attributes as well. **The additional output** will be written into the attribute tables of the memory feature class *after* running the model.

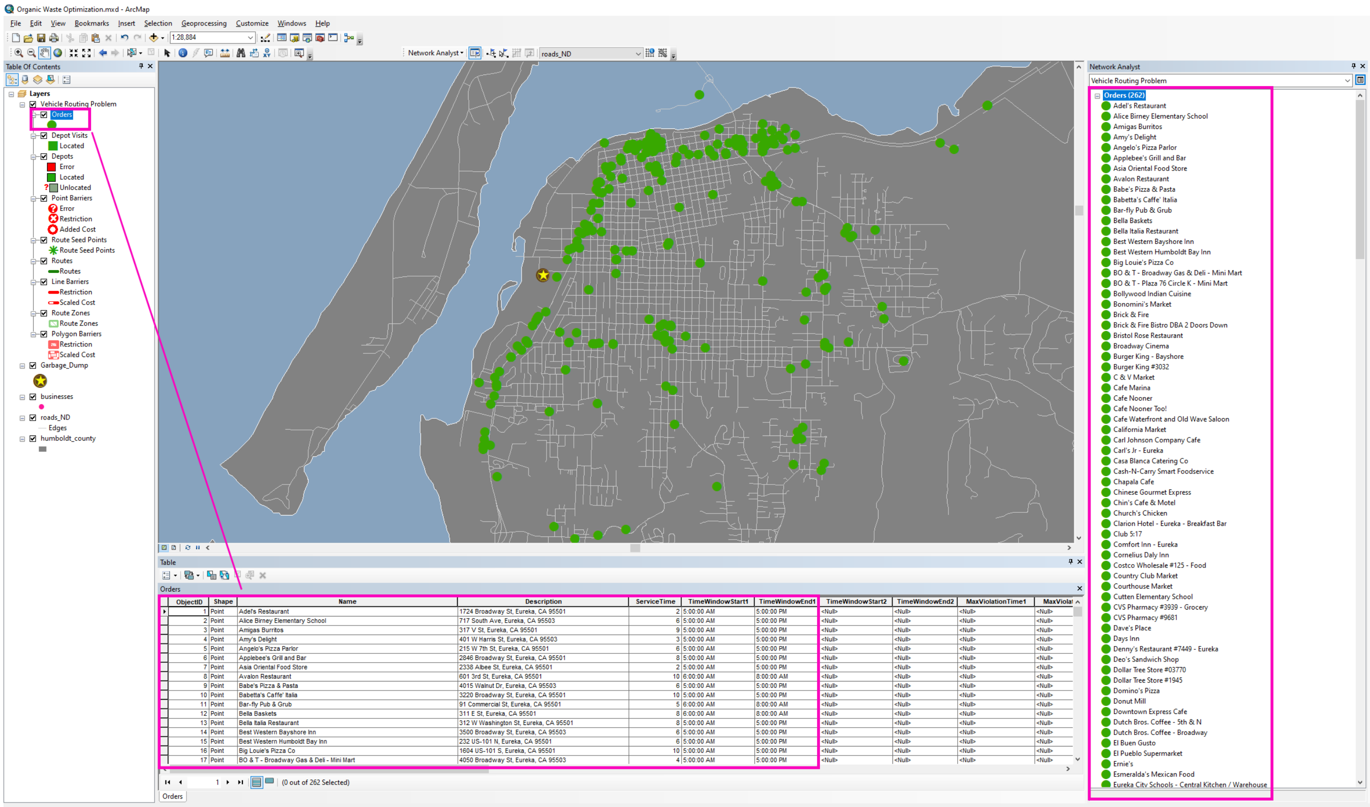


Figure 2.48: Some of the NULL values in the memory feature class attribute table will be replaced with data once the model is run.

Take a moment to open the attribute table for the *Orders* memory feature class in the *Table of Contents* and inspect the information (**Figure 2.48**). The attribute table should look very similar to the businesses shapefile, but with some additional attribute fields. In many ways, memory feature classes look and behave like shapefiles. However, it is essential to remember that the information in the attribute table of a memory feature class is **temporary**. Any changes to the model in the *Network Analyst Window* will also change the attributes in the memory feature class after re-running the model. You must export the memory feature class to make the information permanent. Close the attribute table for the *Orders* memory feature class. Next, you will load the *Depot* into the model.

Loading the Depot into the Model

As you will recall from earlier in the chapter, the **Depot** is a *Route* starting point and ending point. In this scenario, the *Depot* is the Eureka Garbage Station. The garbage trucks will begin and end their routes at this location. *Depot*s also have attributes that influence cost or place limitations on the model. In this model, the depots have attributes that include a **starting time** and an **ending time**. The model will be limited to calculating routes within this time window.

Right-click on *Depot*s in the *Network Analyst Window* and select Load Locations. From the drop-down menu next to Load From, select the garbage dump layer (**Figure 2.49**). *Depot* parameters are defined under the Location Analysis Properties pane using two attributes, Field, and Default Value. If the layer does not have a field that matches the property, a default value may be entered instead.

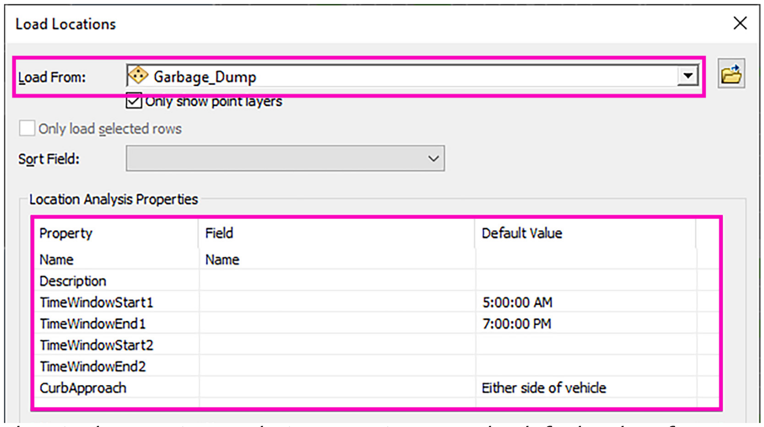


Figure 2.49: Use these values in the Location Analysis Properties. Keep the default values for any properties not listed here.

Note: Inadvertently switching the AM and PM is a mistake many readers often make when creating the depot. Enter your model variables carefully.

Adding Route Parameters into the Model

As you will recall from earlier in the chapter, the *Routes* are more than just a path. In a vehicle routing model, **Routes** are a network class that defines the **vehicle**, the **driver**, and the **linear features** on which they travel. *Routes* also have attributes that influence cost or place limitations on the model. In this model, the *Routes* have attributes that include the *Depot*, where the route will start and end. The *starting* and *ending service time* attributes indicate the time it takes to prepare and service the garbage trucks at the start and end of each day. The *earliest* and *latest start time* attributes indicate the time window when the route will first begin.

Routes also have attributes related to the **vehicle**. The **capacities** attribute indicates the carrying capacity of the garbage truck before it needs to be emptied. In this scenario, a side-loading garbage truck can carry twenty thousand pounds before it needs to be emptied. When the model runs, this attribute will be compared to the **pick-up quantities** attribute of the *Orders* to determine when the truck is full. The model assumes that both the *capacity* and the *pick-up quantities* use the same unit of measurement. The **cost per unit distance** attribute defines the operating cost of the vehicle per mile, including fuel and maintenance.

Finally, *Routes* have attributes related to the **driver**. These attributes include driver salaries as dollars per minute, under the cost per unit time value. The overtime start and cost per unit overtime tracks when any overtime pay accrues. The max total time attribute defines the maximum length of time a single driver can work. In this scenario, the total number of hours a driver can work is limited to twelve hours.

Right-click on *Routes* in the *Network Analyst Window* and select Add Item (**Figure 2.50**).

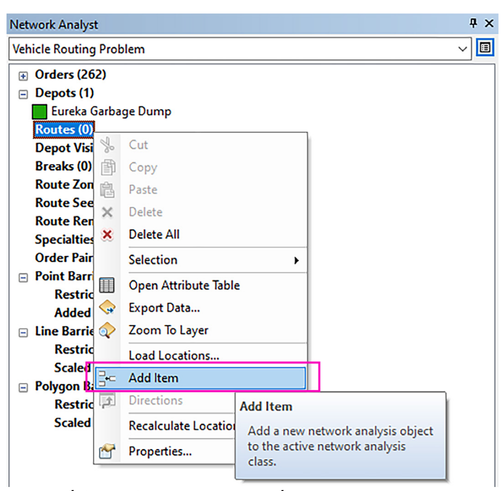


Figure 2.50: This action opens up the *Route Properties* window.

Under the Value column, enter the following values in the *Route Properties* window (**Figure 2.51**). Keep the default values for any properties not listed here. Once the values above are entered, *click OK*.

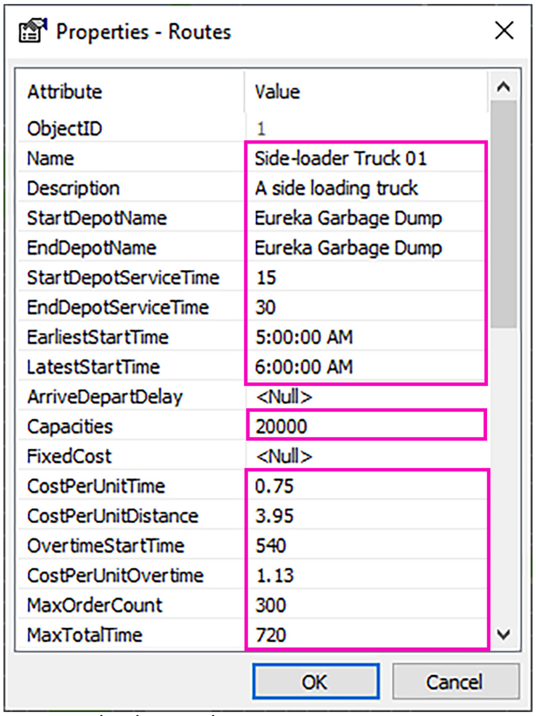


Figure 2.51: Check to make sure your route properties match.

The routes should now be loaded in two places (**Figure 2.52**). The routes in the *Network Analyst Window* represent our model parameters. Any modification of the model parameters, or input to the model, are entered here. In the *Table of Contents*, the memory feature class, also named *Routes*, is now populated with the route attributes, the output from the model. Unlike with the *Orders*, the *Routes* do not appear on the map in the data frame until after running the model. Next, you will add a *Route Renewal* into the model.

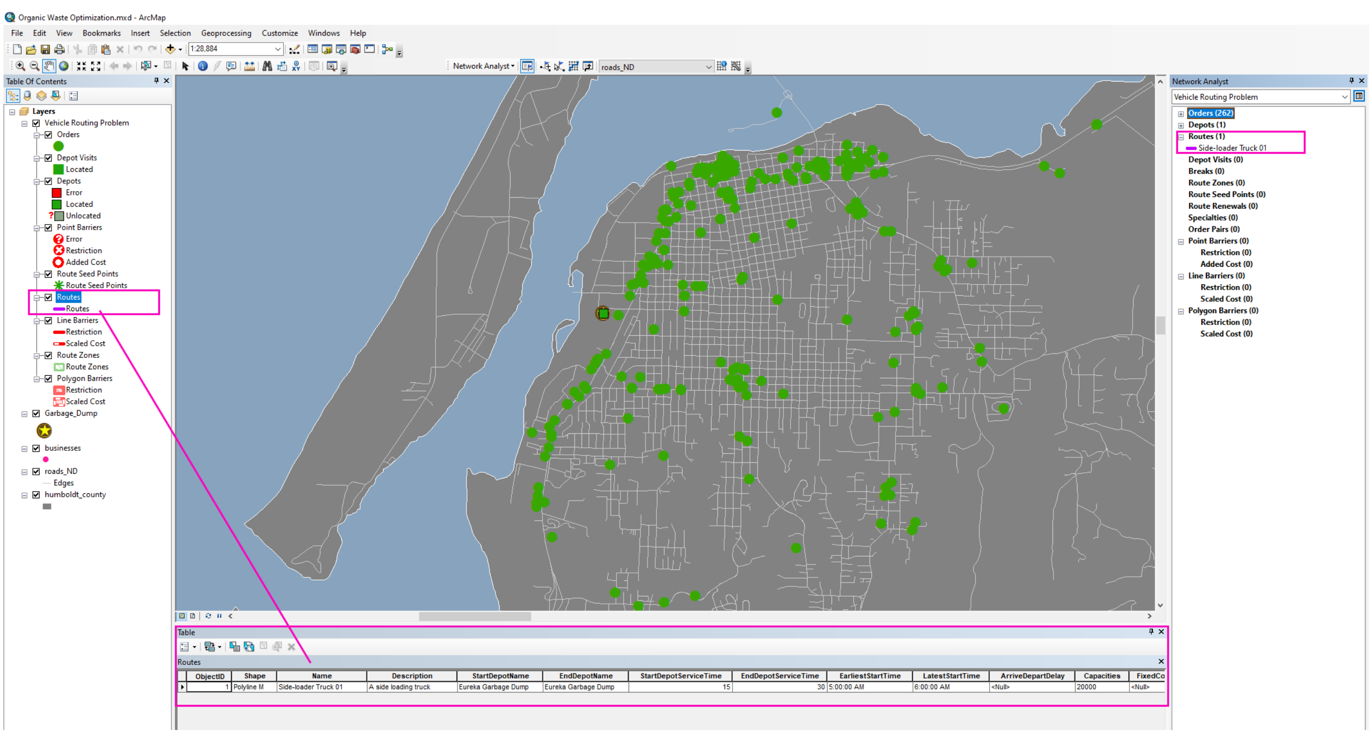


Figure 2.52: Some of the NULL values in the memory feature class attribute table will be replaced with data once the model is run.

Adding a Route Renewal into the Model

Recall from earlier in the chapter, the **Route Renewal** defines the location where a vehicle can unload or reload cargo. The *Depot* often serves as the route renewal location, though any additional locations may be used. In this scenario, the garbage station will serve as both the *Depot* and the *Route Renewal*. The garbage trucks will start out empty at the beginning of the route. As the vehicle stops at each order, the pick-up quantity is tracked and compared to the truck capacity. When the garbage truck is full, it must return to the route renewal location to unload the cargo before continuing on its route.

Note: If the route renewal is missing or defined incorrectly, the route will end prematurely, even if all of the orders have not yet been fulfilled.

One must also consider that the truck may fill up before the route is finished. In this instance, it will need to return to the garbage station to empty its load before continuing. Right-click on *Route Renewals* in the *Network Analyst Window* and select *Add Item* (**Figure 2.53**).

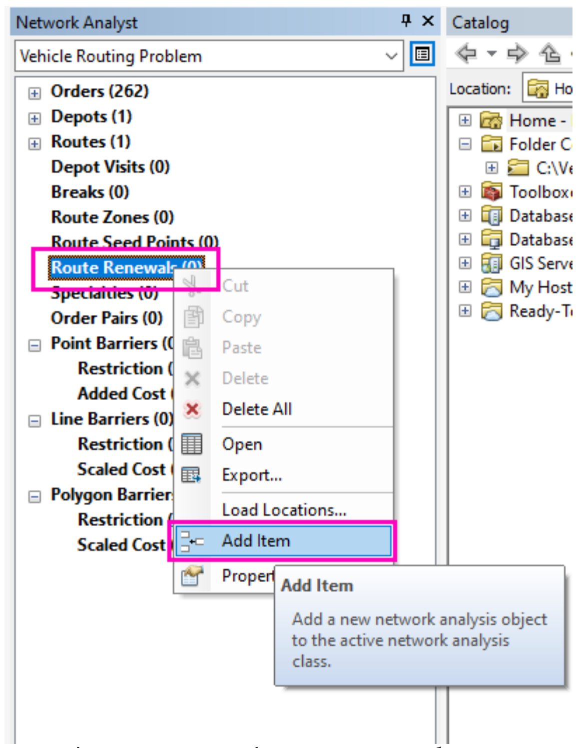


Figure 2.53: This action opens the *Route Renewal* properties window.

This opens up the *Route Renewal* Properties window. Next to *Depot Name*, select **Eureka Garbage Dump** from the drop-down menu. Next to *Route*Name select **Side-loader truck 01** from the drop-down menu (**Figure 2.54**). Next to *ServiceTime*, enter **10**. The service time attribute represents the time it takes the truck to unload the cargo before continuing on. Once the values are entered, *click OK*.

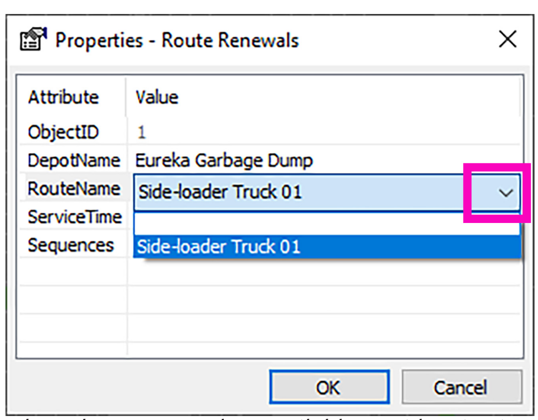


Figure 2.54: Always use the drop-down menu when available. Avoid typing in model variable parameters.

Adjusting the Analysis Setting of Model

The vehicle routing problem also has global properties that must be set. So far, you have been working with attributes for individual model parameters such as orders, routes, and depots. Here you will adjust the analysis settings of the entire vehicle routing problem. These settings will influence all of the model parameters. Click on the *Vehicle Routing Problem* Properties button in the *Network Analyst Window* (**Figure 2.55**).

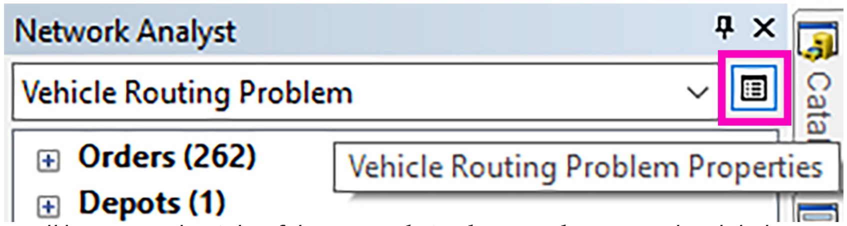


Figure 2.55: The small button on the right of the *Network Analyst Window* opens the global properties for the VRM.

Navigate to the *Analysis Settings* tabs (**Figure 2.56**). Verify that the *Time Attribute* is set to **Minutes (Minutes)**. Set the *Distance Attribute* to **Length (Meters)**. Verify that the *Distance Field Units* are set to **Miles**. This setting determines the units for other model attributes such as *cost per unit distance*, which in this scenario is **dollars per mile**. Set the *U-Turns at Junctions* to **Allowed Only at Dead Ends**. Leave all other settings as default and *click OK*.

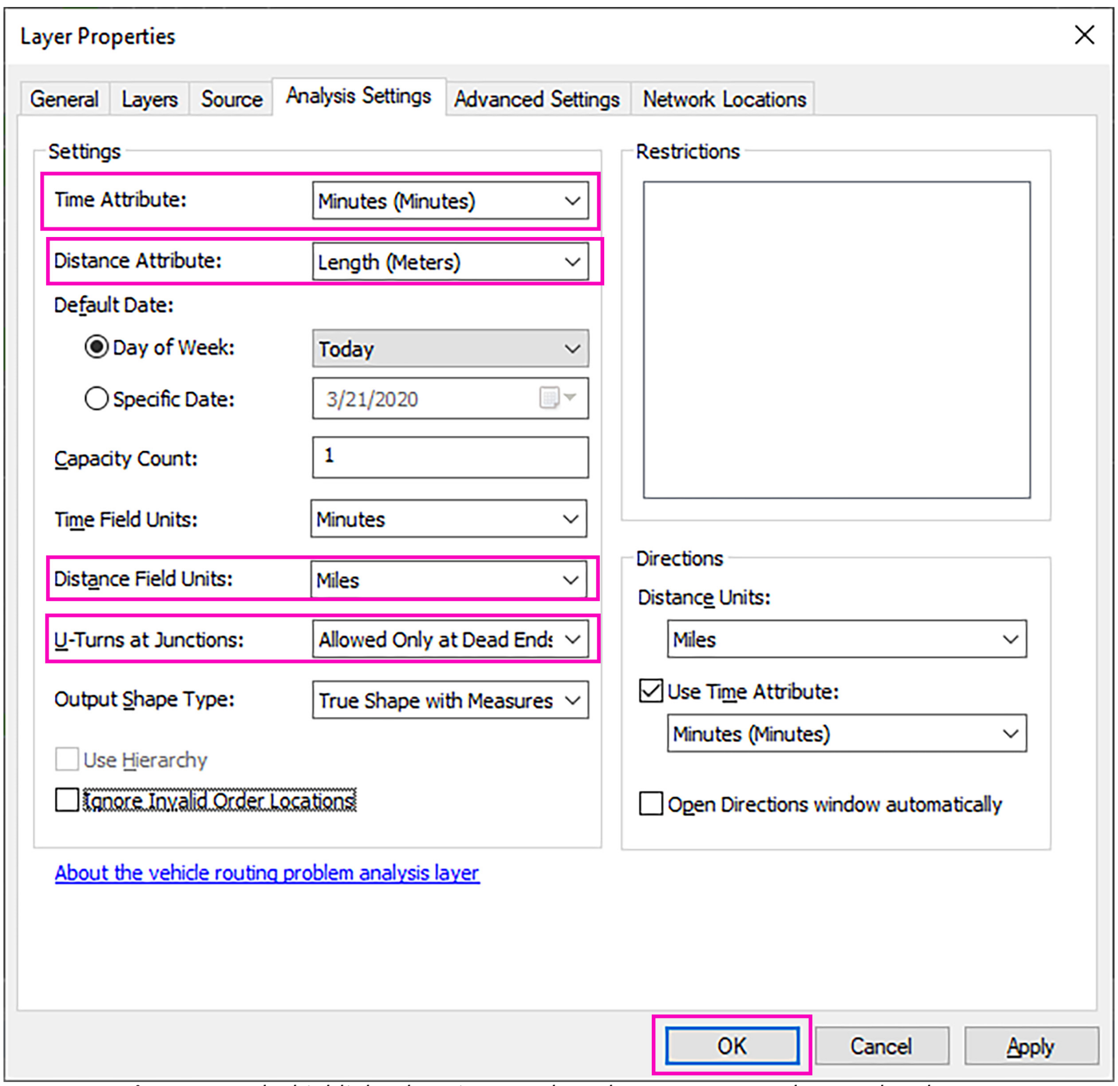


Figure 2.56: The highlighted sections are the only parameters readers need to change.

Running the Vehicle Routing Model

One of the advantages of creating a vehicle routing model via the *Network Analyst* extension is the ability to run the model multiple times while making small adjustments in between each run. In this activity, you run several models, adjusting the parameters each time. For the first run of the model, all of the parameters were set in the previous steps. To run the model currently loaded in the *Network Analyst Window*, click on the *Solve* button in the *Network Analyst Toolbar* (**Figure 2.57**).

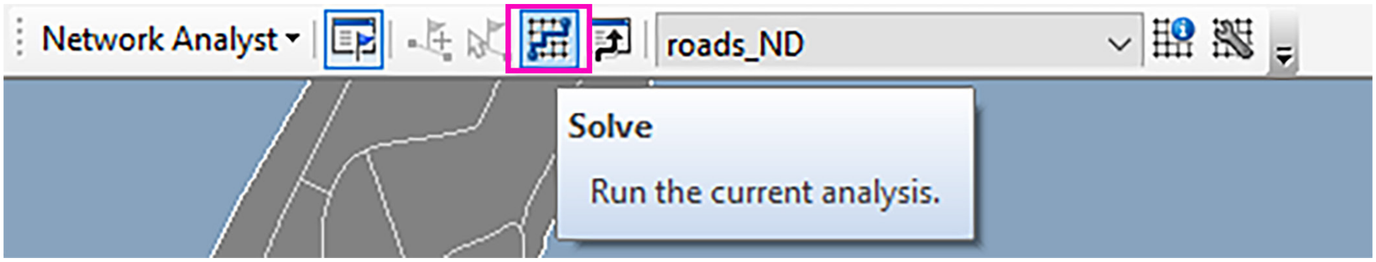


Figure 2.57: Every time one wishes to run or re-run a network model, one should press the *Solve* button.

Note: Time to go grab a coffee!
Depending on the speed of your computer, the model may take a while to run. Expect to wait up five or ten minutes.
Take a moment to stand, walk around, or stretch. Do not try to work in the ArcGIS software while the model is running.

You may or may not get a warning message (**Figure 2.58**). Warning messages like the one below appear when some of the orders do not receive service or if the service was late.

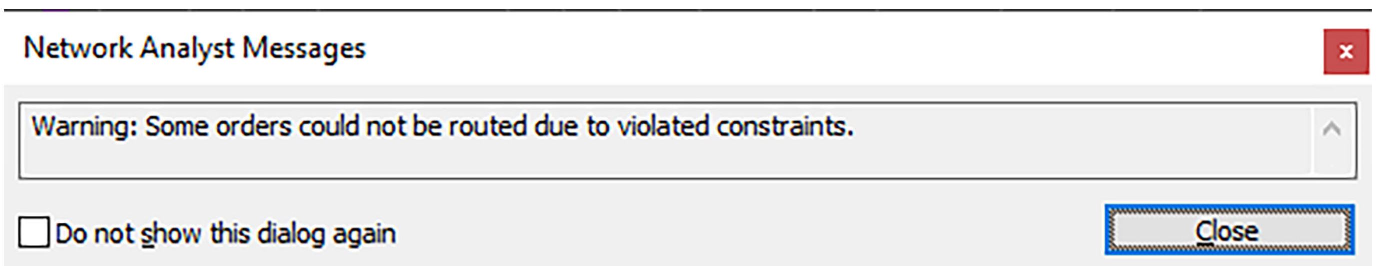


Figure 2.58: When a warning message appears, one must investigate why.

When the *Network Analyst* completes the first run, it writes the results into the temporary memory layers in the *Table of Contents*. It adds the *Route* to the map (**Figure 2.59**). Additional information also appears in the *Network Analyst Window*.

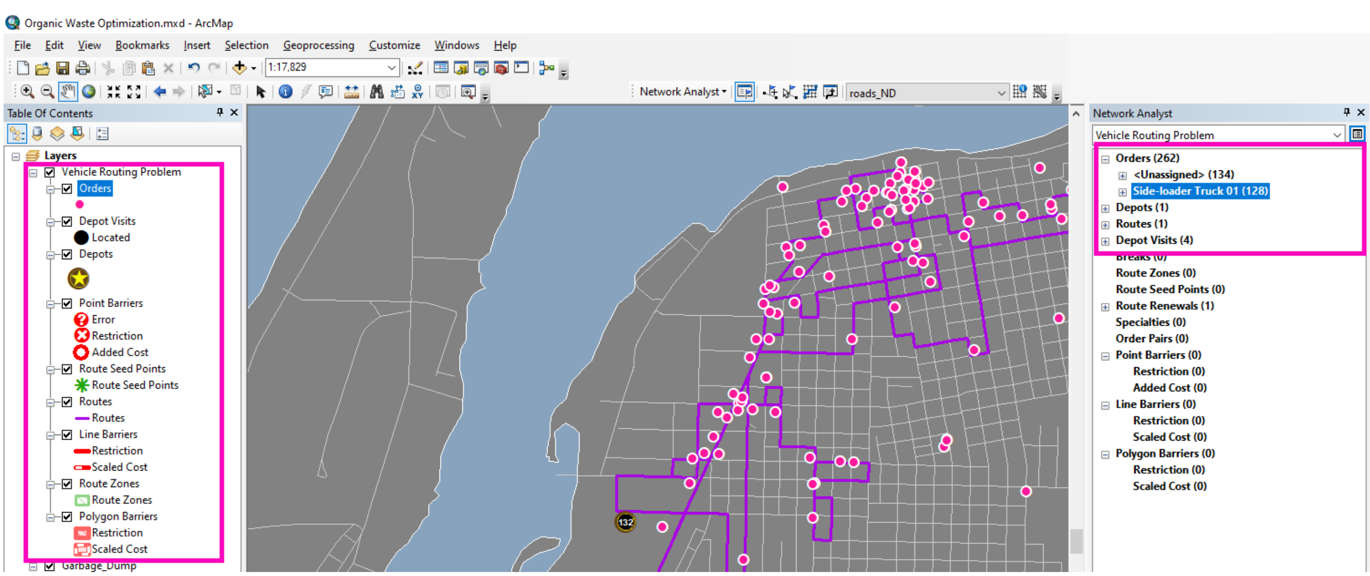


Figure 2.59: Your **results may vary** from the one shown here.

Interpreting the Results

Take a moment to explore the results. In this model, not all of the restaurants received services from the garbage company. The error message flags this result but does not provide much information. One must investigate the results and check the model parameters for errors to determine why not all orders where reached. To start, take a closer look at the *Network Analyst Window* (**Figure 2.60**).

A quick glance confirms that the route named *Side-loader Truck 01* reached **128** businesses and that **134** businesses were not reached. There were a total of **four** visits to the *Depot* (Eureka Garbage Dump). Recall that when the garbage truck fills up to capacity with organic waste, it returns to the *Depot* to empty. Then the truck returns to service orders. From this information, one may extrapolate that the route was operating correctly, but ran out of time. The question to ask is, why?

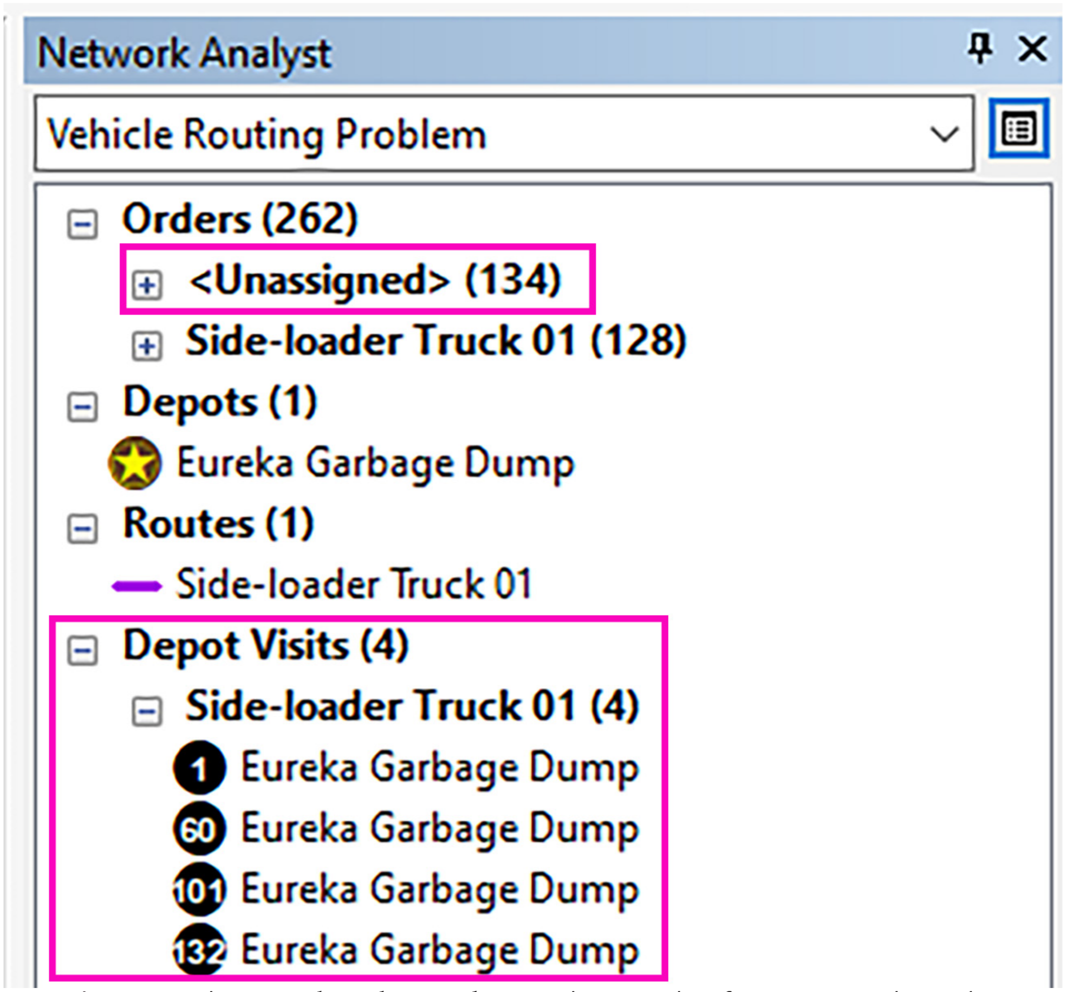


Figure 2.60: The *Network Analyst Window* provides many clues for interpreting the results. Actua**l results may vary** as the instructor changes the data over time.

There could be several reasons for running out of time before reaching all of the *Orders*:

» The *Depot* Time Window Start and Time Window End values may have the incorrect AM or PM time causing the *Depot* to open late or close early.

» The service times, or start and end times, for the *Route* may have been entered incorrectly.

» The truck capacities in the route parameters may have been entered incorrectly so that the truck filled up too quickly.

» The max total time for the *Route* may have been entered incorrectly.

» All parameters were entered correctly, but the *Route* reached the maximum total time allowed for a single driver, including overtime.

Investigate each of these possibilities and model constraints before moving on to the next step.

Documenting the Results.

Open the attribute table for the *Routes* memory feature class located in the *Table of Contents* (**Figure 2.61**).

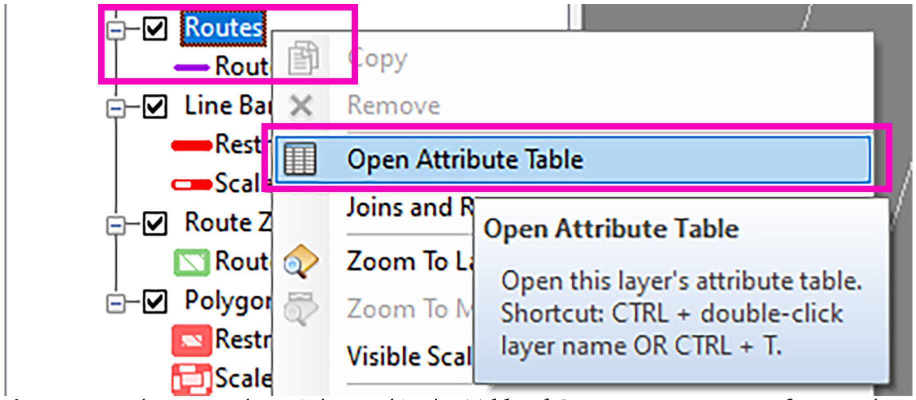
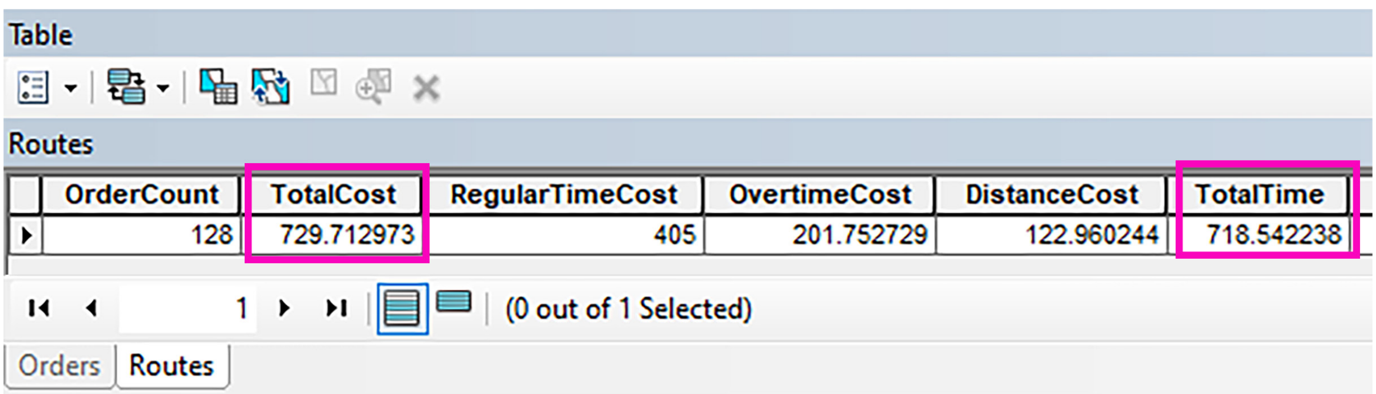


Figure 2.61: The *Routes* layer is located in the *Table of Contents* as a memory feature class.

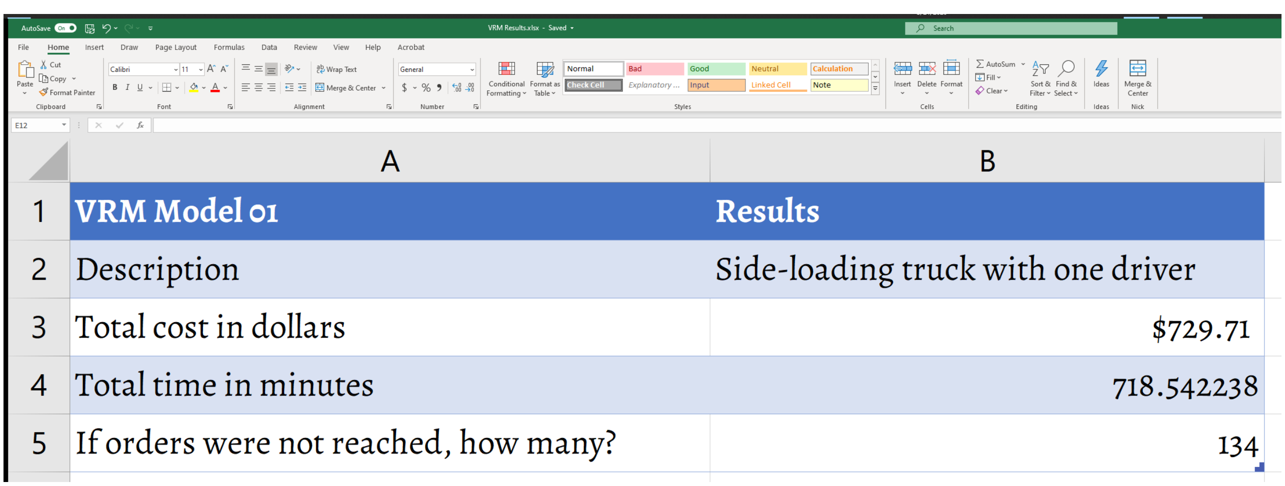
Here one may find the results of the model (**Figure 2.62**). You already know that in this model, **134** businesses were not reached. The *Routes* attribute table also tells you how much the *Route* cost and how much time was spent. In this example, the total time is 718.542238 minutes. The maximum time allowed in this model is 720 minutes, a twelve-hour shift. These results tell you that the driver worked nearly twelve hours before stopping. Another stop would have caused the driver to exceed twelve hours. This situation seems like a reasonable explanation of why 134 orders were not reached.



OrderCount	TotalCost	RegularTimeCost	OvertimeCost	DistanceCost	TotalTime
128	729.712973	405	201.752729	122.960244	718.542238

Figure 2.62: Actual **results may vary** as the instructor changes the data over time.

Open a blank Excel workbook and create a table recording the following results from your first model. Save the file to your *final* folder for future reference.



	A	B
1	VRM Model 01	Results
2	Description	Side-loading truck with one driver
3	Total cost in dollars	$729.71
4	Total time in minutes	718.542238
5	If orders were not reached, how many?	134

Figure 2.63: Actual **results may vary** as the instructor changes the data over time.

Skill Drill: Adding a Second Garbage Truck to the Model

In the first model, you ran it with only **one** *Route*, representing one driver and one garbage truck. In this step, you add a **second** route before running the model. To add a second route, expand the *Routes* in the *Network Analyst Window*. Right-click on Side-loader Truck 01 and select *Copy*. Then, right-click again and choose *Paste* (**Figure 2.64**).

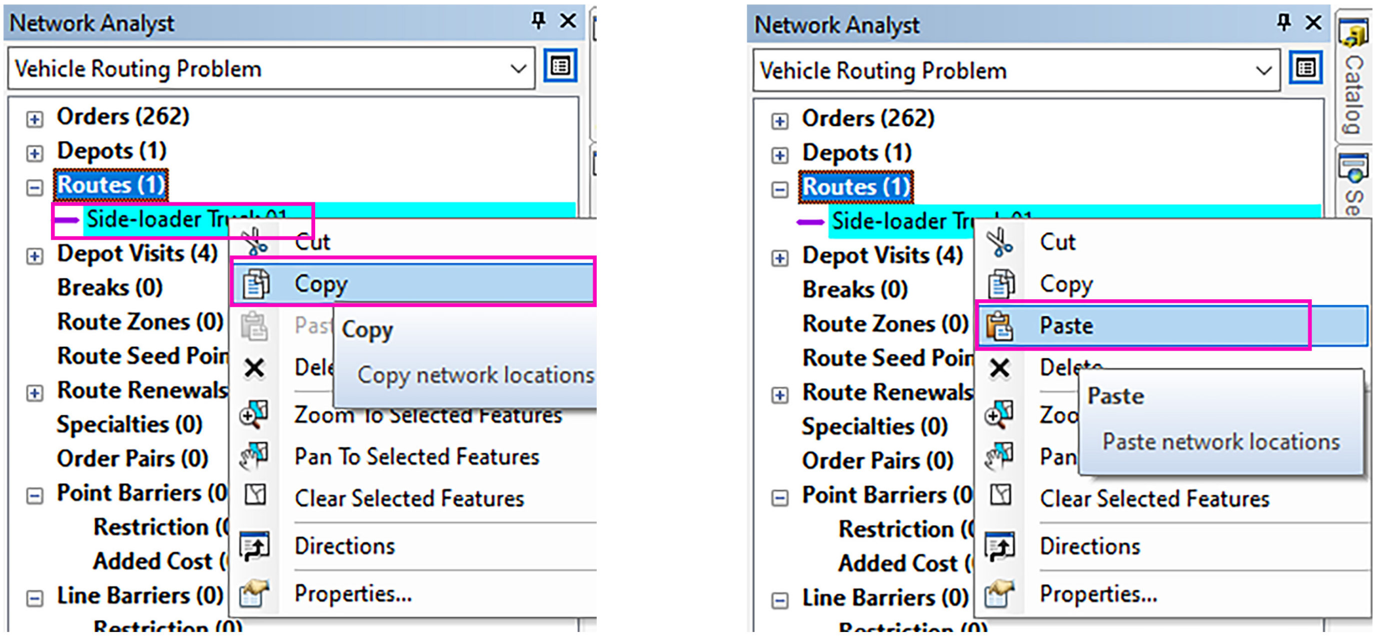


Figure 2.64: Copying the route duplicates the route properties, saving time, and reducing the chances of introducing error.

The second route has precisely the same attributes as the first route. However, you must change the route name. Call the second route, Side-Loader Truck 02 (**Figure 2.65**).

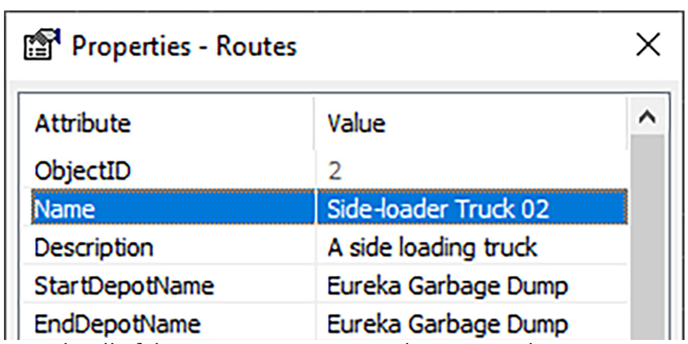


Figure 2.65: In this example, all of the *Route* properties need to remain the same except for the *Route* name.

Next, add a *Route Renewal* for *Side-loader Truck 02*. To add a second route renewal, follow the same steps you used to create the first route renewal in the section *Adding a Route Renewal into the Model*. Be sure the *Route Name* attribute is set to the new route, *Side-loader Truck 02*. Also, be sure that the *Eureka Garbage Dump* is the *Depot* for both *Route Renewals* (**Figure 2.66**).

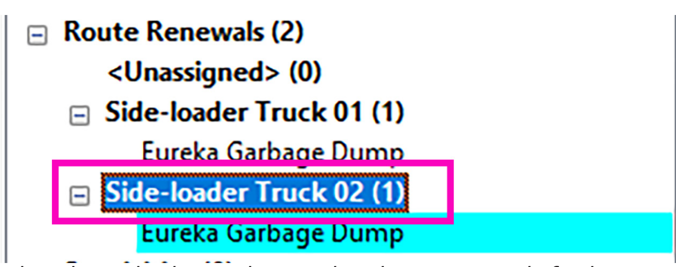


Figure 2.66: When done, check to make sure that there are a total of only **two** route renewals.

Once the second route and route renewal are added, run the model by clicking on the *Solve* button in the *Network Analyst Toolbar*. When you re-run the model, the new results overwrite the previous results in the memory feature classes in the *Table of Contents*.

In the *Table of Contents*, open the layer properties for the *Routes* and navigate to the *Symbology* tab (**Figure 2.67**). On the left, choose *Unique Values* under *Categories*. For the *Value Field*, choose *Name*. Uncheck the box next to *<all other values>*. Click the line symbol for each route and chose a contrasting color for each. When ready, *click OK*.

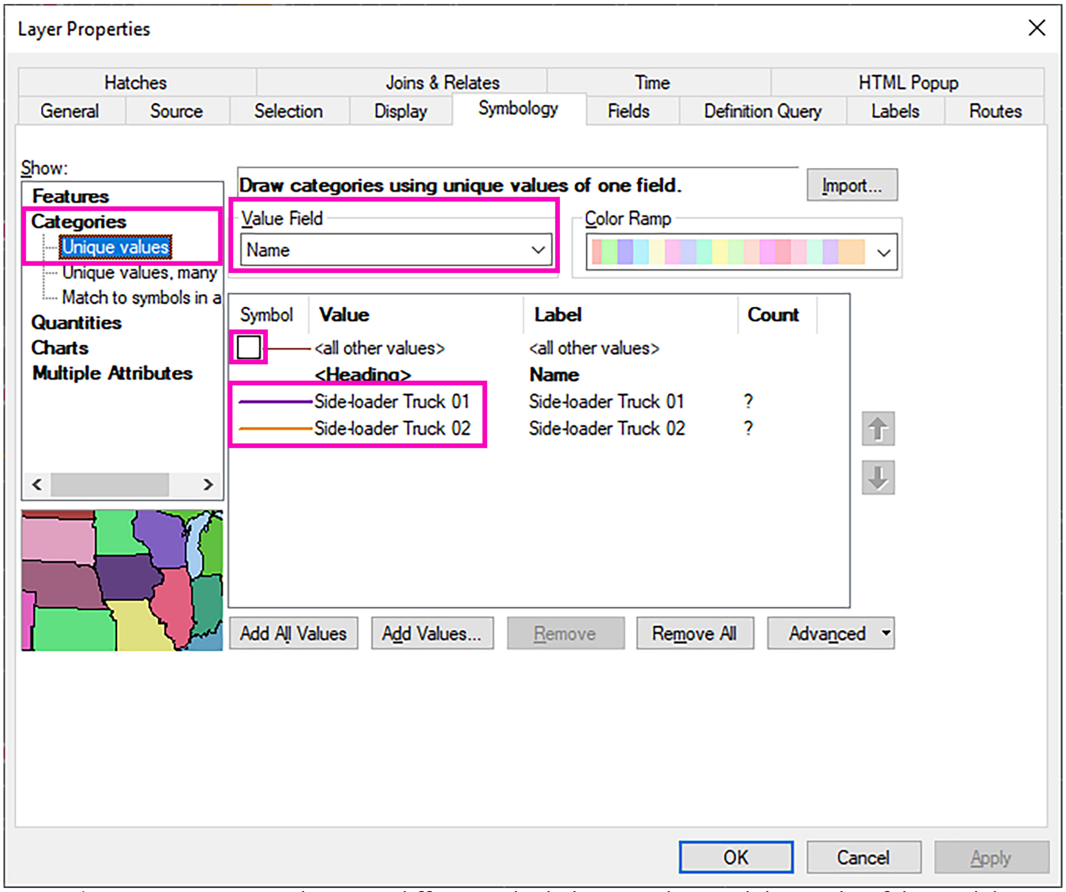


Figure 2.67: Giving each route a different color helps to understand the results of the model

The map should update with each route visible as a separate color (**Figure 2.68**).

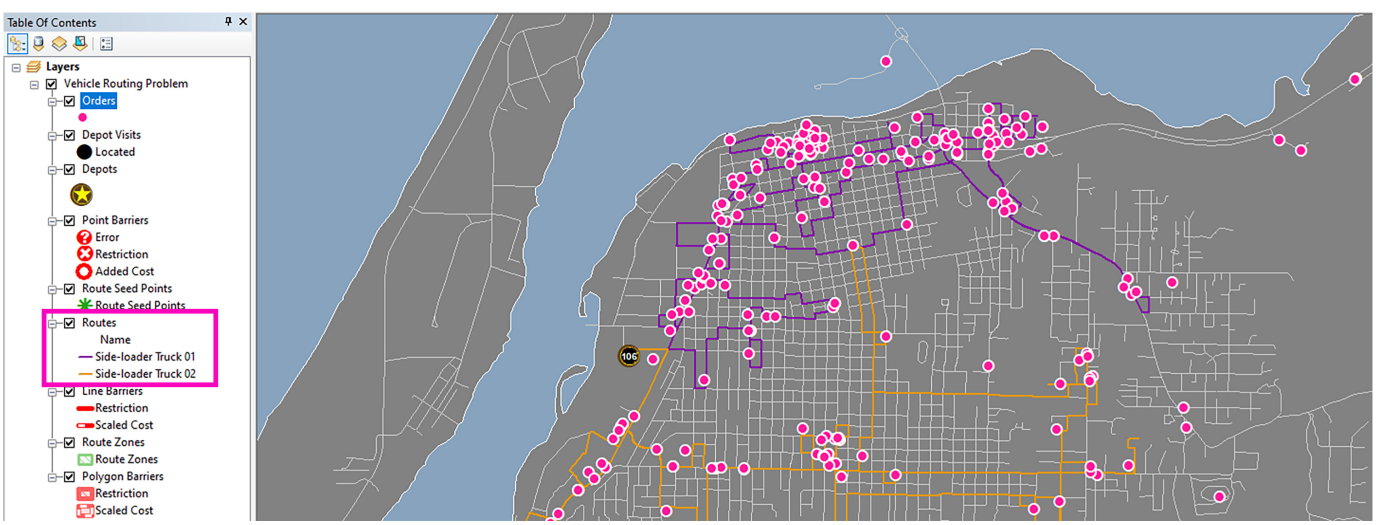


Figure 2.68: The path each route took to try to reach all orders is easily distinguished by the different colors.

Open the *Routes* attribute table and record the results of the second run on a separate table in Microsoft Excel (**Figure 2.69**).

	A	B	C	D	E
1	VRM Model 01	Results		VRM Model 02	Results
2	Description	Side-loading truck with one driver		Description	Two Side-loading trucks with one driver each
3	Total cost in dollars	$729.71		Side-Loader 01 time in minutes	714.424956
4	Total time in minutes	718.542238		Side-Loader 01 cost in dollars	$700.57
5	If orders were not reached, how many?	134		Side-Loader 02 time in minutes	715.754850
6				Side-Loader 02 cost in dollars	$771.16
7				Total time in minutes	1430.179806
8				Total cost in dollars	$1,471.73
9				If orders were not reached, how many?	45

Figure 2.69: Actual **results may vary** as the instructor changes the data over time.

In this instance, two side-loading trucks were not enough to service all of the businesses in one day. The VRM needs additional modifications to accomplish the goal of reaching all orders.

Skill Drill: Adjusting the Model to include Rear-Loading Trucks

After running the first two models, it is clear that *at least* two routes are needed to complete all of the orders. However, additional options need to be considered, such as using a rear-loading garbage truck instead of a side-loading truck. The rear-loading trucks have a higher cargo capacity. Still, they require two people to operate, one to drive, and another to help load heavy garbage cans into the back of the truck. As a result, the operating costs per mile are higher due to employee wages.

In the *Network Analyst Window*, open the properties for the Side-loader Truck 1 and make the following changes:

» **Name**: Rear-Loader 01
» **Description**: Rear-loading truck with two employees
» **Capacities**: 38400
» **CostPerUnitTime**: 1.5
» **CostPerUnitDistance**: 2.75
» **CostPerUnitOvertime**: 2.25

Repeat the changes for the second route, but give it the name Rear-Loader 02 (**Figure 2.70**).

Properties - Routes

Attribute	Value
ObjectID	1
Name	Rear-Loader 01
Description	Rear-loading truck with two employees
StartDepotName	Eureka Garbage Dump
EndDepotName	Eureka Garbage Dump
StartDepotServiceTime	15
EndDepotServiceTime	30
EarliestStartTime	5:00:00 AM
LatestStartTime	6:00:00 AM
ArriveDepartDelay	<Null>
Capacities	38400
FixedCost	<Null>
CostPerUnitTime	1.5
CostPerUnitDistance	2.75
OvertimeStartTime	540
CostPerUnitOvertime	2.25
MaxOrderCount	300
MaxTotalTime	720

Properties - Routes

Attribute	Value
ObjectID	3
Name	Rear-Loader 02
Description	Rear-loading truck with two employees
StartDepotName	Eureka Garbage Dump
EndDepotName	Eureka Garbage Dump
StartDepotServiceTime	15
EndDepotServiceTime	30
EarliestStartTime	5:00:00 AM
LatestStartTime	6:00:00 AM
ArriveDepartDelay	<Null>
Capacities	38400
FixedCost	<Null>
CostPerUnitTime	1.5
CostPerUnitDistance	2.75
OvertimeStartTime	540
CostPerUnitOvertime	2.25
MaxOrderCount	300
MaxTotalTime	720

Figure 2.70: The rear-loaders have a higher capacity and a lower maintenance cost, but need at least two people to operate.

Be sure that there are still only two routes total in the VRM. Many readers make mistakes here.

Next, update the *Route Renewals* so that they reference the correct route names (**Figure 2.71**). Failure to do so will cause an error.

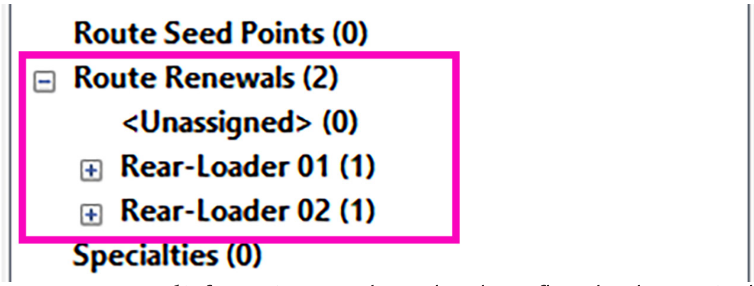


Figure 2.71: The *Route Renewal* information must be updated to reflect the changes in the *Route* name.

Leave all of the other attributes unchanged. When you are ready, run the model by clicking on the *Solve* button in the *Network Analyst Toolbar*. The new model will run, and the new results will overwrite the previous results in the memory feature classes in the *Table of Contents*. Record the results on a third table in Microsoft Excel in a similar fashion as the first two tables (**Figure 2.72**).

VRM Model 01	Results
Description	Side-loading truck with one driver
Total cost in dollars	$729.71
Total time in minutes	718.542238
If orders were not reached, how many?	134
VRM Model 02	**Results**
Description	Two Side-loading trucks with one driver each
Side-Loader 01 time in minutes	714.424956
Side-Loader 01 cost in dollars	$700.57
Side-Loader 02 time in minutes	715.754850
Side-Loader 02 cost in dollars	$771.16
Total time in minutes	1430.179806
Total cost in dollars	$1,471.73
If orders were not reached, how many?	45
VRM Model 03	**Results**
Description	Two Rear-loading trucks with Two drivers each
Side-Loader 01 time in minutes	717.151181
Side-Loader 01 cost in dollars	$1,295.30
Side-Loader 02 time in minutes	719.469102
Side-Loader 02 cost in dollars	$1,316.07
Total time in minutes	1436.620283
Total cost in dollars	$2,611.37
If orders were not reached, how many?	38

Sheet1

Figure 2.72: Your **results may vary** as the instructor changes the data over time.

In this instance, two rear-loading trucks could not fulfill all of the Orders despite the increased capacity.

Skill Drill: Adding a Third Garbage Truck to the Model

Using the skills you have acquired, add a third garbage truck to the model. Experiment with different combinations of rear-loading and side-loading trucks. Determine which combination incurs the **least cost** to the garbage company, but can reach **all of the orders**. Once you have determined the optimal route parameters, save the **final** results on a fourth table in Microsoft Excel. Be sure to save the Excel file to your *final* folder for future reference.

Skill Drill: Creating a Map of the Results

You should be familiar with the steps needed to create a map layout of your results. Design a map for use as a figure in a report or summary. Ideally, the map should be designed at a size of approximately **6 by 7 inches**. Include a north arrow, a scale bar, and a legend. The map should include the **garbage dump**, the **businesses**, and the three routes with a separate color for each.

Note: You can change the colors of the routes in the layer properties symbology tab.

When your map layout is complete, export the map as a PNG file with a resolution of 300dpi (**Figure 2.73**). Save the file in your *final* folder.

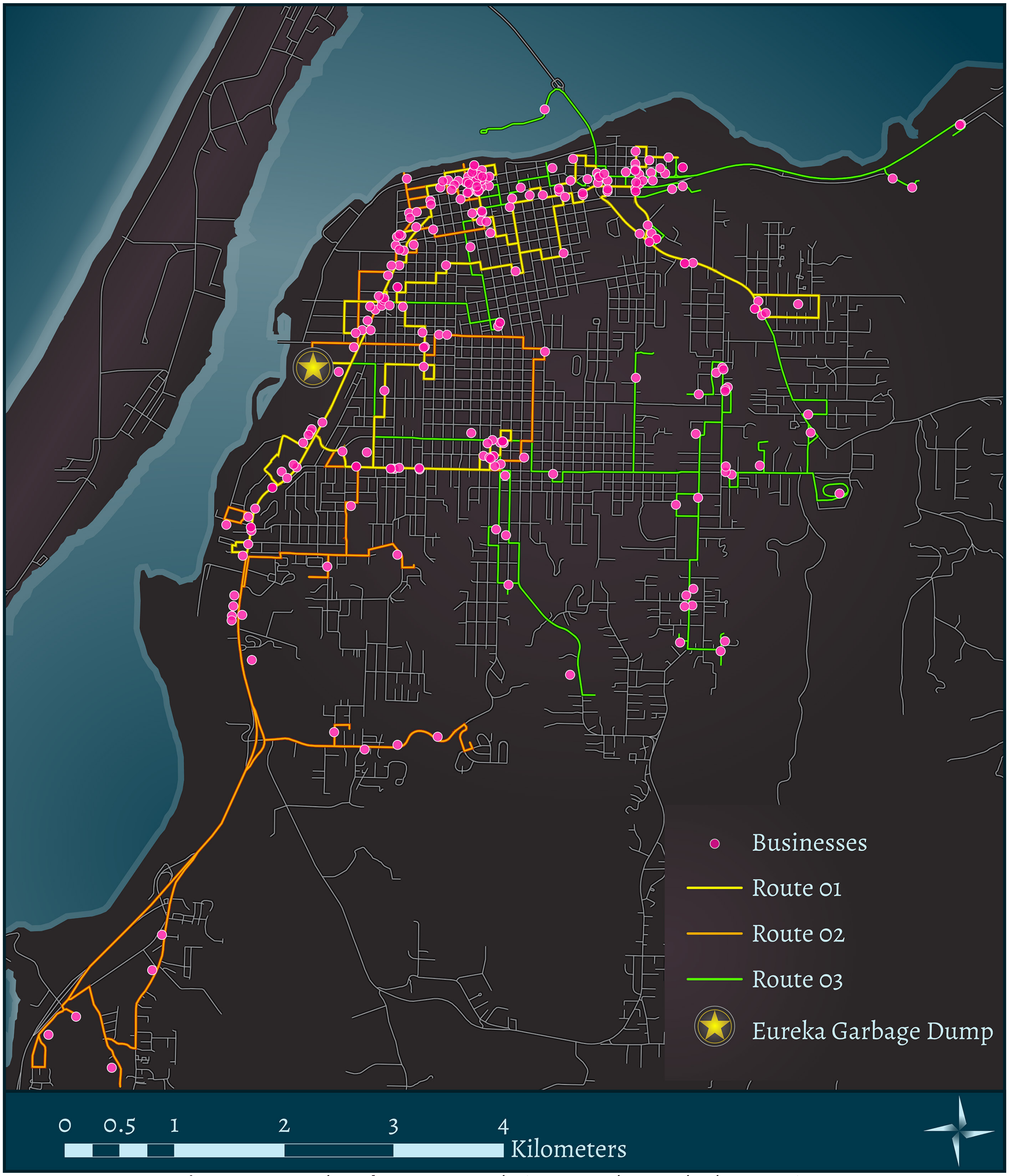


Figure 2.73: Actual **results may vary** as the instructor changes the data over time.

Tutorial: Demand-Based Site Selection for Fire Stations Using a Network Allocation Model

In this tutorial, readers explore modeling network paths using a network allocation model. Readers will create a network dataset, which will allow them to allocate resources based on demand weight and create the least-cost paths along the length of a network.

Estimated time to complete this tutorial: 6 hours

Learning Outcomes

Readers should be able to accomplish the following outcomes by the end of this tutorial:

» Review how to acquire data from a public source
» Geocode an address
» Review adding XY data
» Review Data Management Tools: project, define projection
» Create a network dataset
» Model network allocation based on the least-cost paths
» Adjust model parameters to allocate resources based on demand weight
» Evaluate the results of a network analysis based on demand

Scenario

In this scenario, the Arcata Fire Protection District has decided to explore the possibility of updating their facilities and relocating the three existing fire departments serving the cities of Arcata and McKinleyville. Some potential sites have been *identifi*ed, but the city planners are not sure which location will optimize services within the district. You will attempt to answer this question using a location-allocation solver available through the ArcGIS *Network Analyst* extension.

You will use the following criteria in your analysis:

» A history of incident responses
» A two minute response time window which captures the highest demand based on incident history
» A five minute response time window which captures the highest demand based on incident history
» A comparison of several proposed sites within the Arcata Fire Protection District and related areas
» A comparison between proposed sites and existing fire stations

Conduct this analysis using the **Universal Transverse Mercator (UTM) system** along with the **North American Datum of 1983 (NAD83)**. Humboldt County lies in **Zone 10** of the UTM system. All of your data must be in this spatial reference system at the start of your analysis. Create working copies of your data in this spatial reference system using the *Project* tool in ArcMap as needed.

Setting up Your Workspace

In a typical workflow, you work on geospatial data using a local hard drive. When done, you compress your data and back up your work to your cloud storage so that you can retrieve the files from anywhere. When referring to a **local hard drive**, it means you are working on data physically located on the computer in front of you.

In contrast, some computers also include networked drives. **Networked drives** link to cloud storage and save the data elsewhere. Examples include services like OneDrive or Google Drive. For this tutorial, use the **desktop** as your local hard drive location. You may also use an external USB drive if you plan to work in multiple places.

You must avoid using networked drives while you work.
They increase the processing time and can cause technical glitches.

In this book, you use a particular folder structure. Start by creating your workspace folder on the local hard drive. A **workspace** is a folder or series of folders that contain all of your project files. The top-level folder in your workspace should indicate the activity or the project on which you are working. Organize all of your work within the workspace folder. On your **desktop**, create a new folder and give it a descriptive name, such as **NetworkAllocationModel**. Be sure there are no spaces. You may use underscores instead of spaces. Inside this folder, create the following three subfolders: *original*, *working*, and *final*. Having a standardized folder structure helps to keep a project organized, primarily when you are working with multiple partners. The folder structure you see here (**Figure 2.74**) is the standard used in each of the tutorials presented in this book.

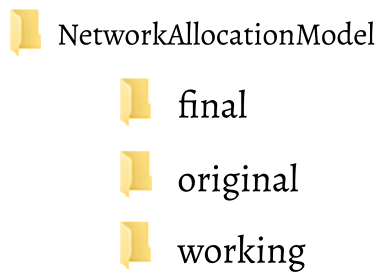


Figure 2.74: This diagram represents a basic folder structure used in this book.

As the name indicates, use the **original folder** for storing original, unaltered data. As you are working on a project, if, for some reason, your working version of the data gets lost or corrupted, you can go back to your *original* folder and find a fresh copy of the data. Use the **working folder** for data that you *create* or *alter* while working on your project. Use the **final folder** for storing any output you produce as a result of your work, such as images, maps, tables, or reports. Setting up a standard folder structure for a project is good practice and a habit you should develop.

Disable Background Geoprocessing

In the ArcGIS software, the *Background Geoprocessing* setting is often turned on by default. This setting allows users to continue to work while *a* tool is running in the background. However, sometimes this setting will stop tools from running or cause other unforeseen problems. To reduce that chances of the ArcGIS software crashing during this exercise, turn this setting off. In ArcMap, **open a new blank map document**. Open the *Geoprocessing options* from the *Geoprocessing* menu. Under *Background Geoprocessing*, uncheck the box next to the word *Enable* (**Figure 2.75**).

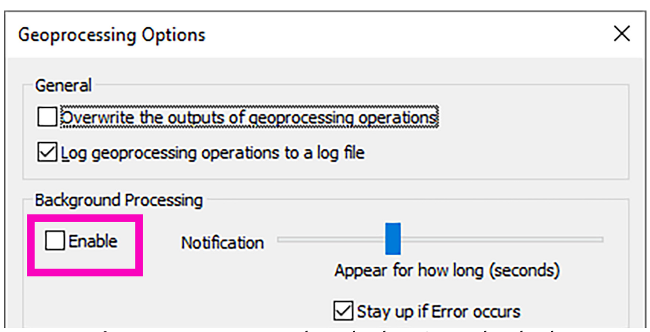


Figure 2.75: Be sure that the box is unchecked.

Network Analyst Extension

The steps in this activity involve using the *Network Analyst* extension. After launching ArcMap, make sure this extension is activated. You will also need to open the *Network Analyst Toolbar* and dock it near the top of your window for easy access. Click on the *Show/Hide Network Analyst Window* button to open the *Network Analyst Window*. Locate this button on the *Network Analyst Toolbar* (**Figure 2.76**).

Figure 2.76: Click the *Show/Hide Network Analyst Window* button

The *Network Analyst Window* is where you will enter or modify your vehicle routing model parameters. After running the model, the results load into a series of feature classes that appear in the *Table of Contents*. Novices tend to confuse the *Table of Contents* and the *Network Analyst Window* during this activity. For clarity, I recommend that you dock the *Network Analyst Window* on the right side of your screen while keeping the *Table of Contents* to the left (**Figure 2.77**).

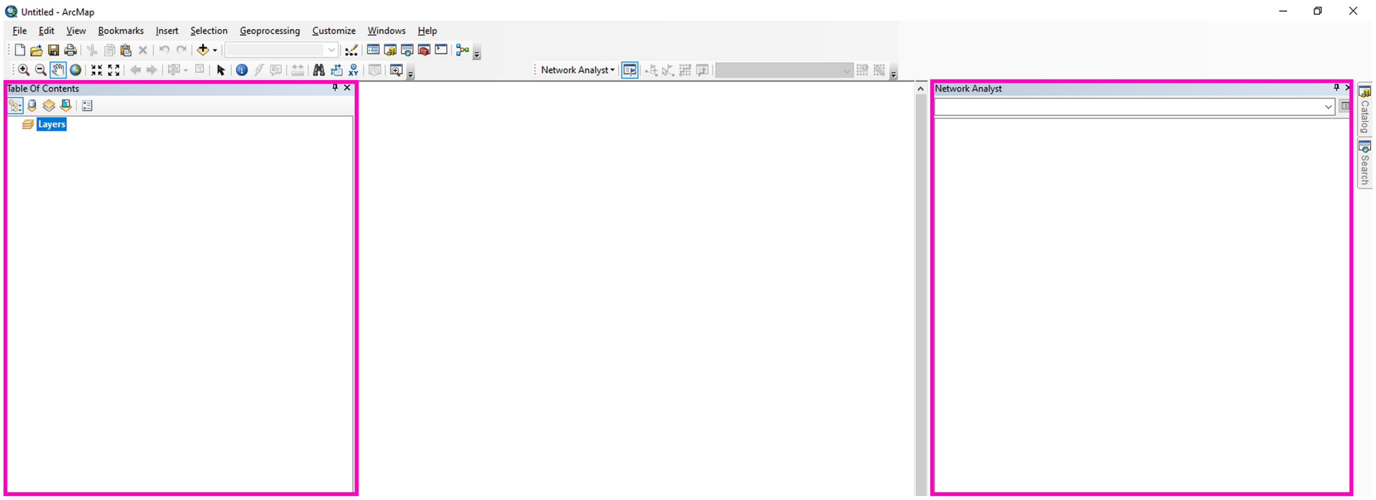


Figure 2.77: The *Table of Contents* docked on the right, and the *Network Analyst Window* docked on the left. Readers that separate the *Table of Contents* from the *Network Analyst Window* encounter less confusion.

Skill Drill: Acquire GIS data from the Humboldt County Website

To create a network dataset, you will need a road layer. In previous courses, you learned how to acquire data from public sources. Here you will download a shapefile containing the roads for Humboldt County. Navigate to the ***Humboldt County GIS Data Download page***[1]. Locate a shapefile containing the road data and download the file. Currently, there are multiple road layers on the website. You want the one that says *Humboldt County GIS Roadway Centerline* (**Figure 2.78**)

Figure 2.78: Locate the roadway centerline shapefile under the frequently requested data sets.

If, for some reason, the website is down or the data is no longer available, you can download a backup copy of the data for this tutorial using the link provided in Appendix A.

1 URL: https://humboldtgov.org/276/GIS-Data-Download

Decompress the file in your *original* folder and check the spatial reference system. If necessary, use the *Project* tool to create a copy of the data with the **NAD 83 UTM Zone 10** spatial reference system. Save the output to your *working* folder (**Figure 2.79**).

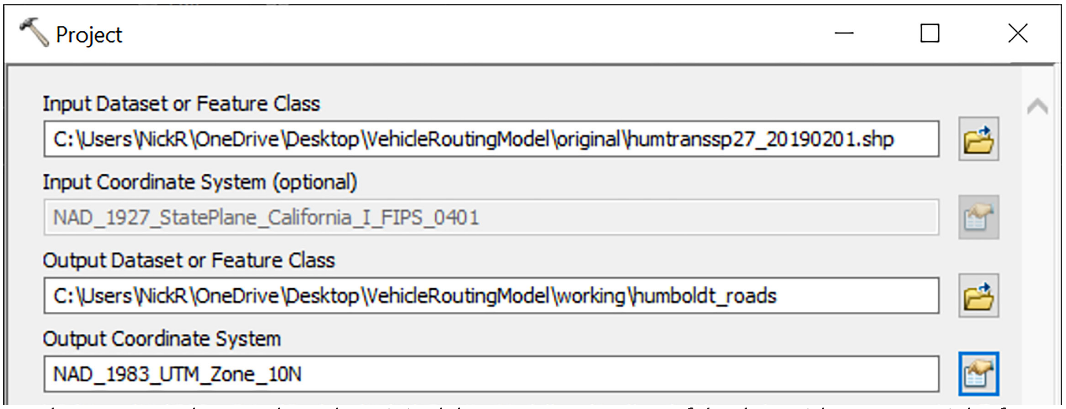


Figure 2.79: The *Project* tool never alters the original data. It creates a copy of the data with a new spatial reference system.

Add the new shapefile to your *Table of Contents* (**Figure 2.80**). This layer defines the spatial reference for the dataframe window because it is the first layer loaded into the *Table of Contents*.

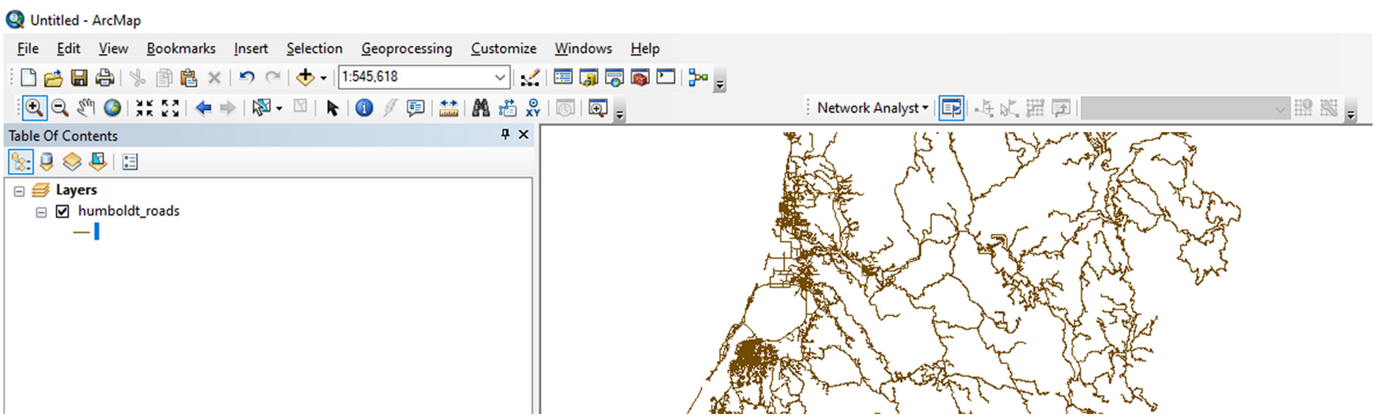


Figure 2.80: In ArcMap, the dataframe window adopts the spatial reference from the *first* layer loaded into the *Table of Contents*.

Return to the ***Humboldt County website***[1] and download the *County Boundary* shapefile (**Figure 2.81**).

Figure 2.81: Locate the County boundary shapefile under the frequently requested data sets.

Decompress the file into your *original* folder. Use the *Project* tool to create a copy of the data with the **NAD 83 UTM Zone 10** spatial reference system. Save the output to your *working* folder. Add the new shapefile to your *Table of Contents* (**Figure 2.82**)

In ArcMap, you may need to refresh your original folder in the Arc Catalog Window after decompressing the ZIP file.

1 URL: https://humboldtgov.org/276/GIS-Data-Download

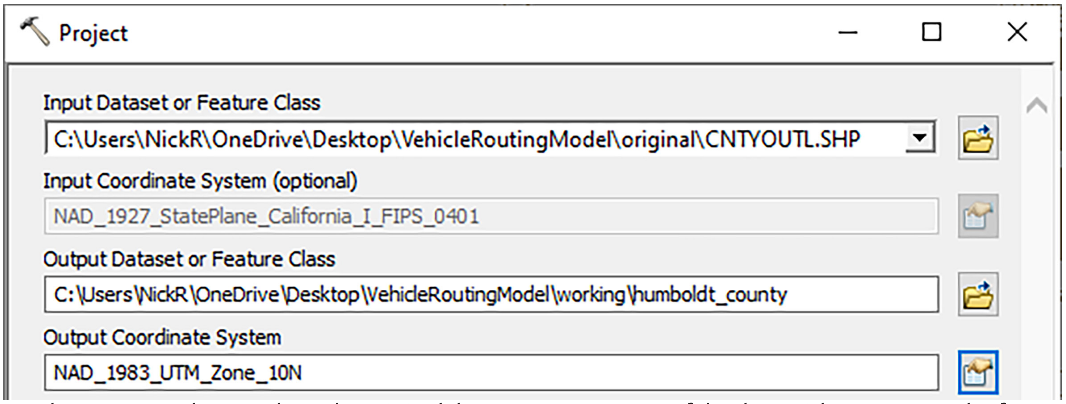


Figure 2.82: The *Project* tool never alters the original data. It creates a copy of the data with a new spatial reference system.

Repeat these steps for the *Fire Districts and Response Areas* shapefile. Use the *Project* tool to create a copy of the data with the **NAD 83 UTM Zone 10** spatial reference system. Save the output to your *working* folder. Add the new shapefile to your *Table of Contents* and arrange the layers in the *Table of Contents* so that each layer is visible in the *Data Frame* Window.

Optionally, you may change the Data Frame background color to blue to represent the Humboldt Bay.

Skill Drill: Clip the Road layer to the City of Arcata and McKinleyville

Network analysis can place significant demands on a computer's processing power and memory. To save time during the analysis, you will need to constrain the road network to the areas serviced by the **Arcata Fire Protection District**. Using an attribute query, create a layer based on the areas related to the Arcata Fire Protection District. At this time, Humboldt County designates these areas as *Arcata Fire* under the *Agency* field in the attribute table (**Figure 2.83**). Check for new alterations in the data in case the names in the attribute table have recently changed. Export the selected features as a new shapefile. Add the new layer to the *Table of Contents*.

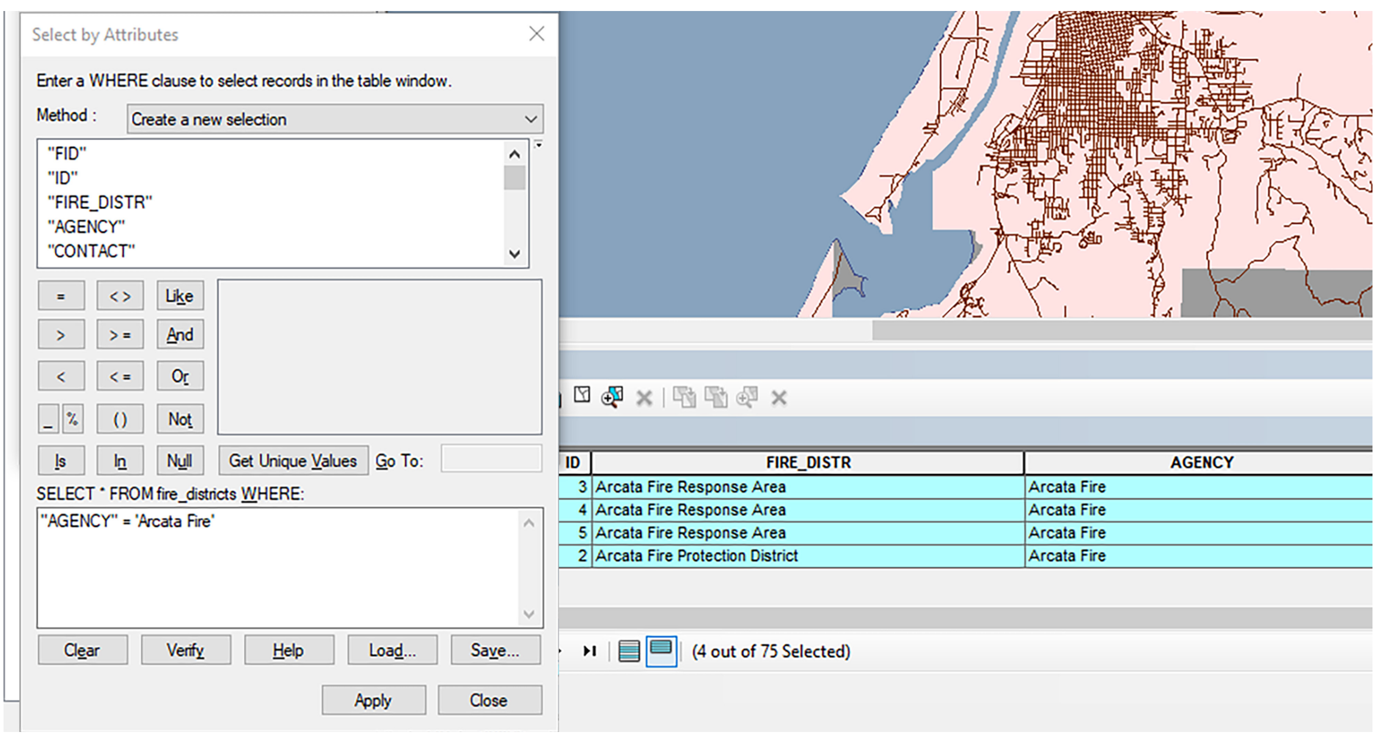


Figure 2.83: You can export the selected features to create a permanent shapefile.

Some roads may extend just outside of the Arcata Fire Protection District. To prevent altering the geometry of the road segments needed in the analysis, create a one-mile buffer around the Arcata Fire Protection District (**Figure 2.84**)

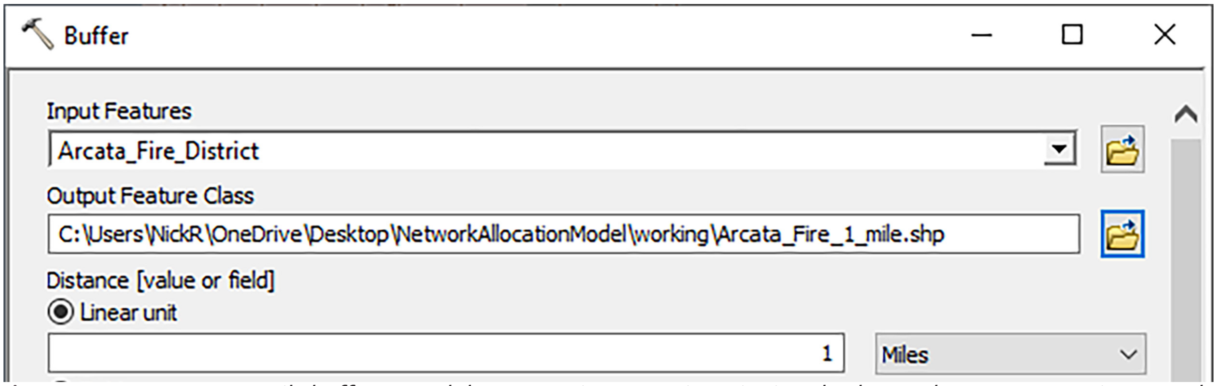


Figure 2.84: Create a one-mile buffer around the Arcata Fire Protection District. Check to make sure your settings match.

Clip the Humboldt road layer and use the buffer as the *Clip Feature* (**Figure 2.85**). This clip operation ensures that your network dataset serves your needs while excluding the remaining roads in Humboldt County from our analysis.

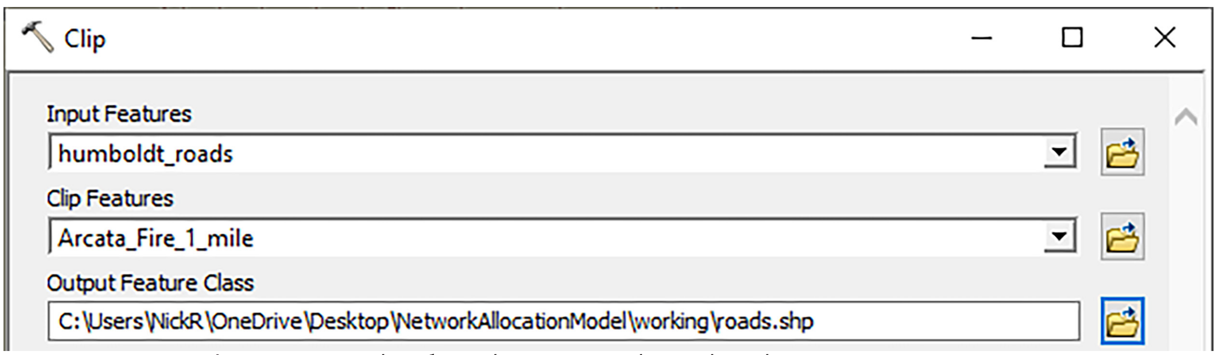


Figure 2.85: Use the *Clip* tool to constrain the roads to the Arcata Fire Protection District buffer. Check to make sure your settings match.

Go to the ArcMap properties window and set the Pathnames to *Store relative pathnames to data sources* (**Figure 2.86**).

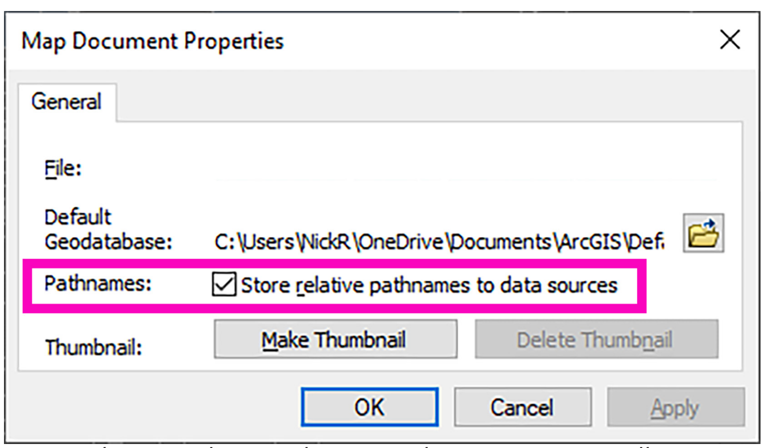


Figure 2.86: The store relative pathnames to data sources setting allows ArcMap to remember the location of the data sources relative to its current position.

Take a moment to save your map document. Be sure to give the map document a meaningful name. Take care to save the map document in your workspace folder, *VehicleRoutingModel* (**Figure 2.87**).

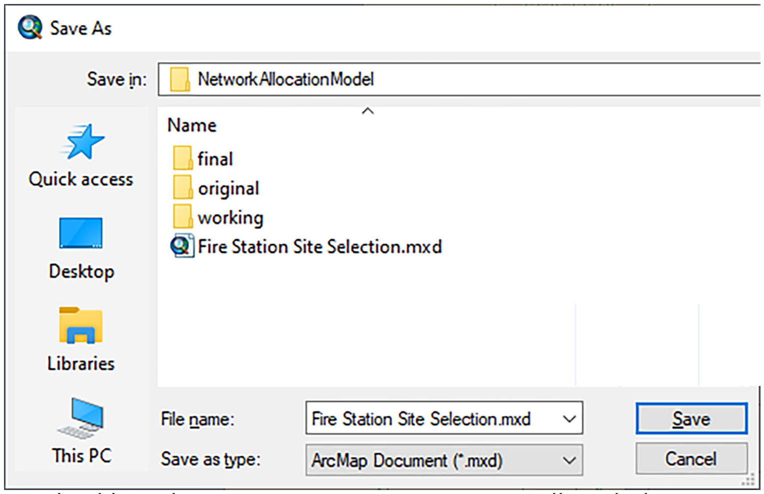


Figure 2.87: 1.84: Generally, you should avoid spaces. However, empty spaces are allowed when naming map document files (.mxd)

Remove all unnecessary layers so that only the clipped roads and the Humboldt County boundary remain (**Figure 2.88**).

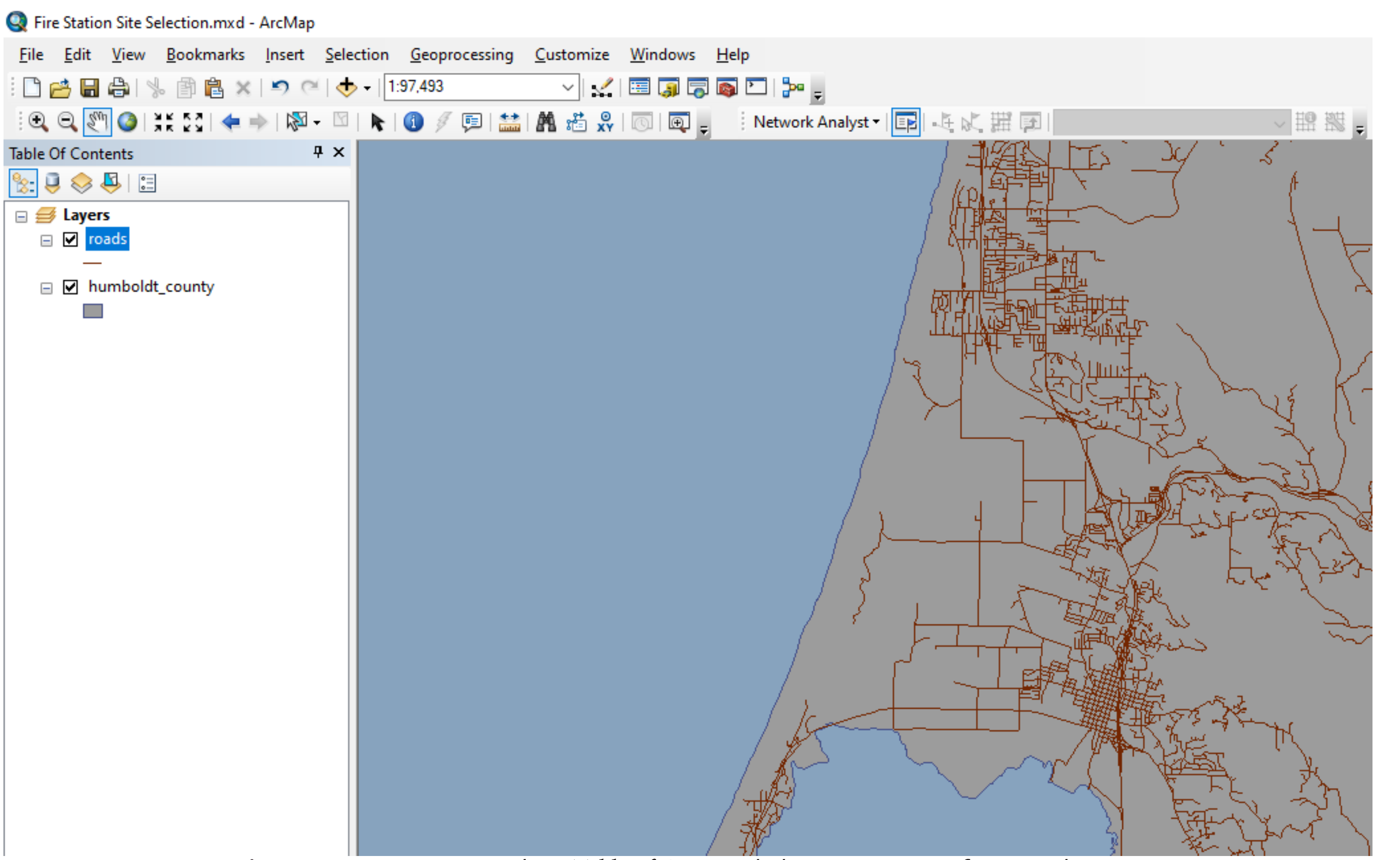


Figure 2.88: Maintaining a clean *Table of Contents* helps to prevent confusion and errors.

Skill Drill: Add a Time Cost Attribute to the Road layer

The vehicle routing model requires a time unit **cost variable** for the network dataset. Create a new field in the attribute table of the clipped-road layer (**Figure 2.89**). Name this field **Minutes**. This field will record the travel time, in minutes, for each road segment.

In most cases, the road segments will have values of less than a one minute travel time. Therefore, the field type should be either a float or a double. If you are ever unsure which to choose, I recommend that you always choose the *double* field type since it will allow for a larger number if necessary.

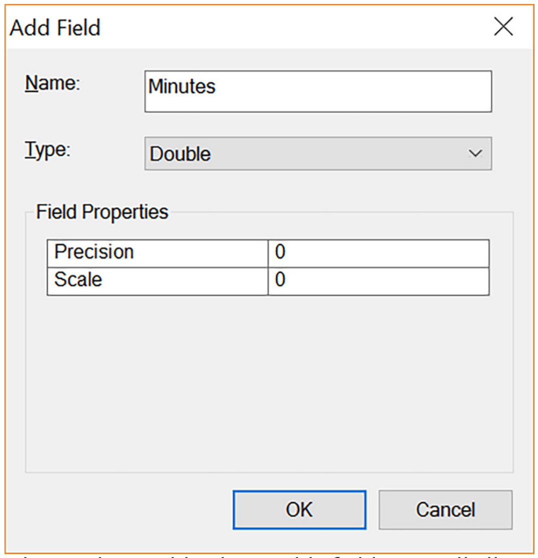


Figure 2.89: When adding a new field to the attribute table, the Double field type will allow for larger numbers with decimal places.

To calculate the travel time, in minutes, you will use the length of each road segment along with the speed limit. The road layer should have the length of each segment recorded in a *Length* field. However, since you clipped the road layer, these values will need to be updated to account for any road segments that may have been altered during the *Clip* operation. To update the length values, use the *Calculate Geometry* function in the attribute table (**Figure 2.90**). Remember, you can access Calculate Geometry by right-clicking on the field name representing the length of road segments. Make sure the values are saved in **meters**.

Note: If you can't find a field representing length, you can create your own by adding a new field in the attribute table. Set the field type to Double. Then use Calculate Geometry to populate the field.

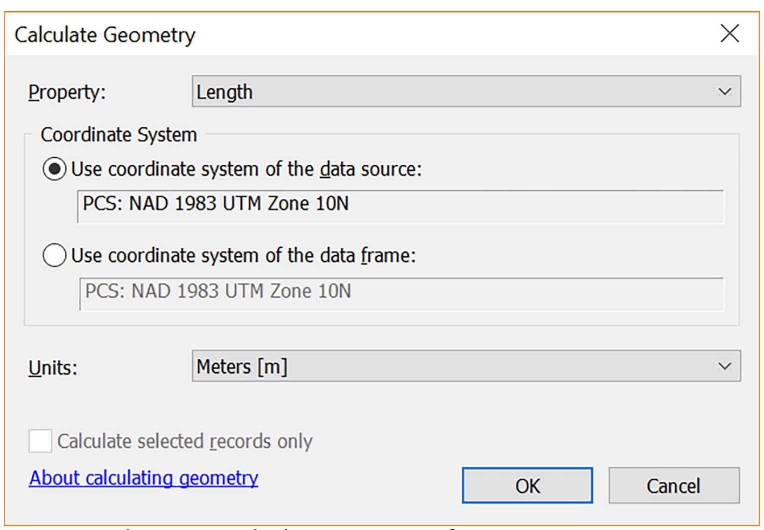


Figure 2.90: Always use Calculate Geometry after any geoprocessing operation that alters the feature geometry, such as a clip, or an intersect.

The road layer has two attributes that record the speed limit, AB_Speed and BA_Speed. These attributes represent the speed limit, in miles per hour, for each direction. For consistency, use the AB_Speed in your calculation. Use the *Field Calculator* to compute the travel time and to populate the *Minutes* field. Your formula should factor in the **AB_Speed** (miles per hour) and the **Length** (meters) of each road segment to determine the travel time in minutes (**Figure 2.91**). Remember, you can access *Field Calculator* by right-clicking on the field name, *Minutes*.

Coming up with the right formula will involve converting miles per hour to meters per minute. Then it requires using the length of the roads to determine the travel time in minutes.

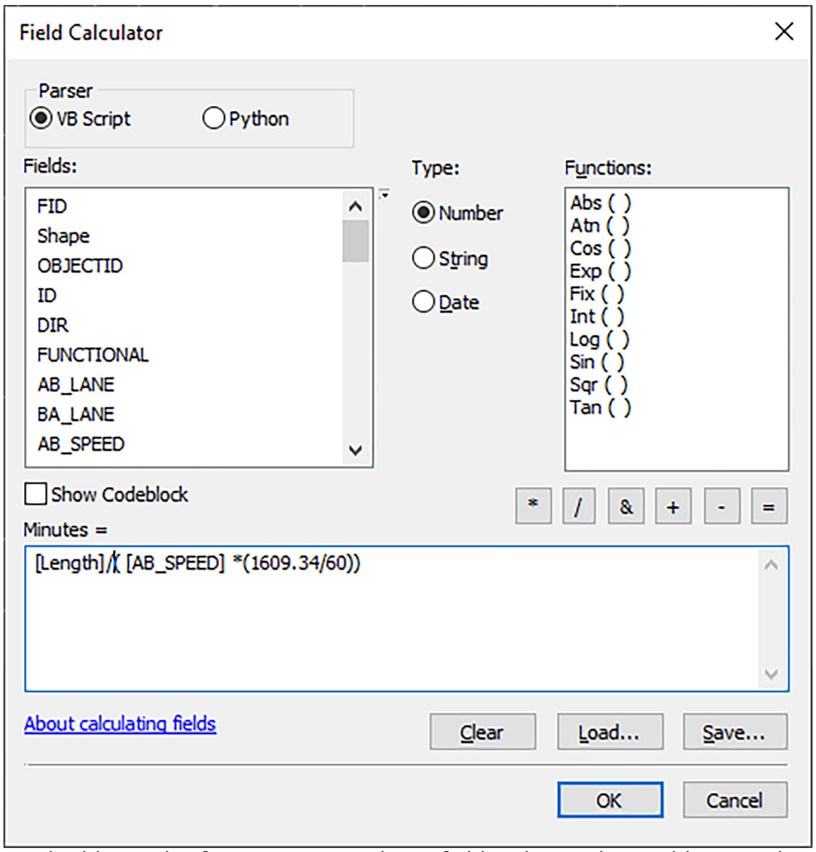


Figure 2.91: You can build complex functions to populate a field in the attribute table using the *Field Calculator*.

Once you have successfully calculated the minutes, you can verify the accuracy of your calculation by checking the travel time of a portion of G Street in Arcata, just before the on-ramp to HWY 101. Just west of the on-ramp, the travel time of the road segment on G Street is 1.059253 minutes (**Figure 2.92**).

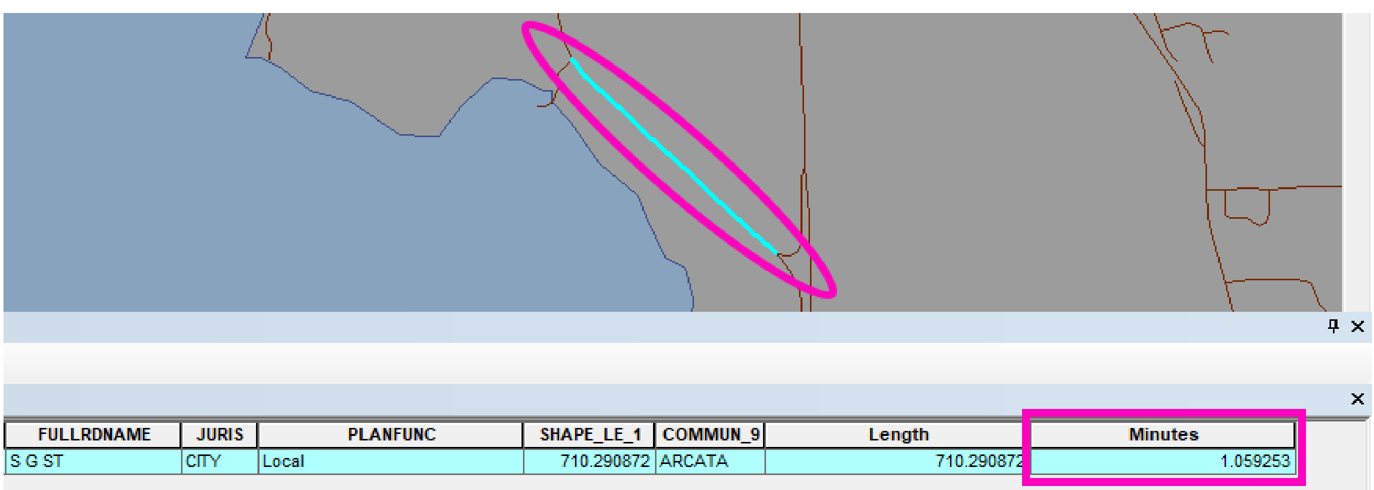


Figure 2.92: Use this road segment of HWY 255 to verify that your formula in the *Field Calculator* is correct.

Skill Drill: Potential Fire Stations as XY Data

In this fictional scenario, several sites were proposed as potential replacement locations for the new fire stations. These locations must be evaluated to determine the top three choices for the Arcata Fire Protection District and related regions. Information about these proposed fire station locations is available in a ***comma-separated values (CSV) file***[1]. Download the file and save it to your *original* folder (**Figure 2.93**).

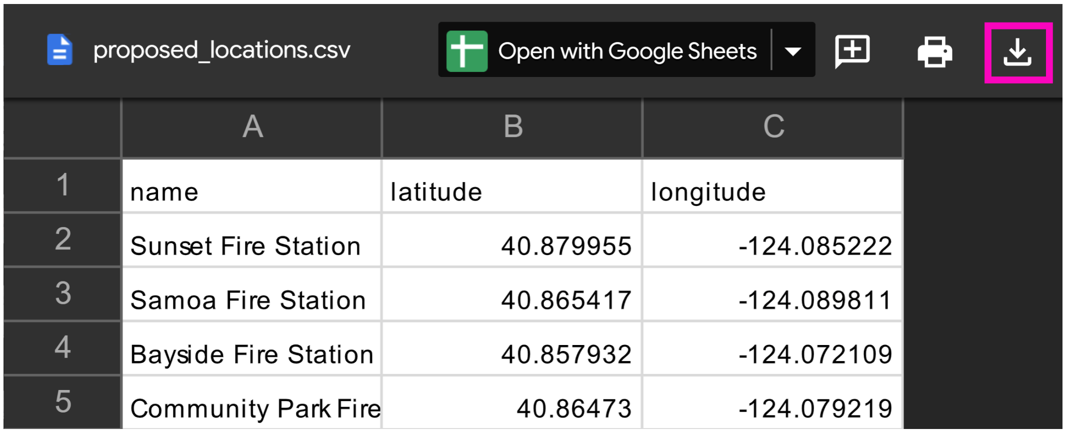


Figure 2.93: For the best results, use the Chrome browser to download the CSV file. Click the download button on the upper right.

It is important to remember that each geodetic datum as a unique set of latitude and longitude values. For example, the latitude and longitude values for your home using the North American Datum of 1983 will be different than the latitude and longitude values using the North American Datum of 1927. Likewise, the World Geodetic Datum of 1984 will use yet another set latitude and longitude values to define the location of your home. Like most latitude and longitude values, you obtain from the internet or from a GPS receiver, the decimal degrees in this CSV table are currently in the geographic spatial reference system **WGS 1984**. When adding XY data, the ArcGIS software only reads the decimal degrees. It is unaware to which spatial reference system these decimal degree values belong. Always specify the datum *source* when adding XY data.

In the *Catalog Window*, expand the *original* folder and right-click on the proposed locations CSV file. Select *Create Feature Class*, then choose *From XY Table* (**Figure 2.94**).

In ArcMap, you may need to refresh your original folder to see the CSV file.

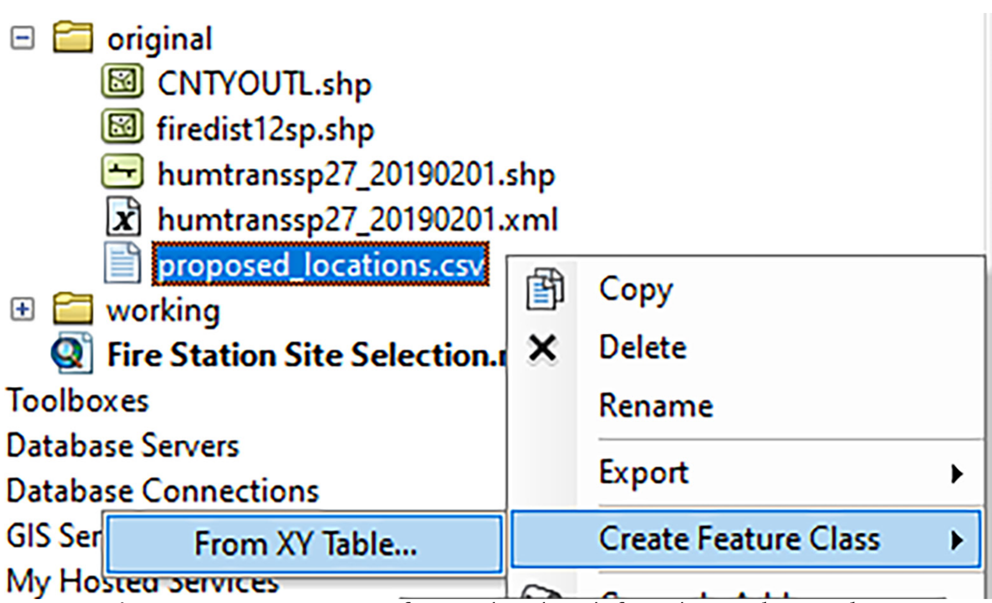


Figure 2.94: You can create a feature class directly from the *Catalog Window*.

1 URL: https://bit.ly/Proposed_Fire_Stations

When the *Create Feature Class from XY Table* window opens, click the button that says *Coordinate System of Input Coordinates*. Open the *Geographic* folder, then the *World* folder. Locate **GCS WGS 1984** and set that as the spatial reference (**Figure 2.95**).

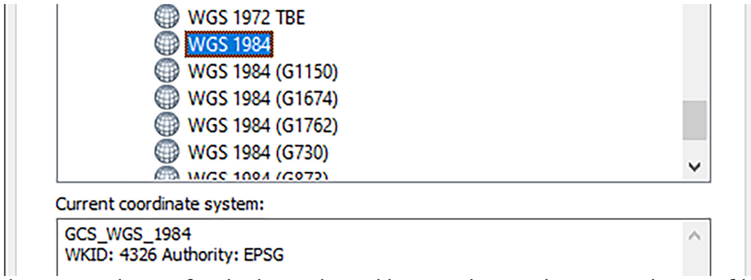


Figure 2.95: The source datum for the latitude and longitude coordinates in the CSV file is WGS 1984.

On the *Create Feature Class from XY Table* window, click the yellow file folder icon and save the output to the *working* folder. Name the file *proposed_locations*. Next, click the *Advanced Geometry Options* button (**Figure 2.96**).

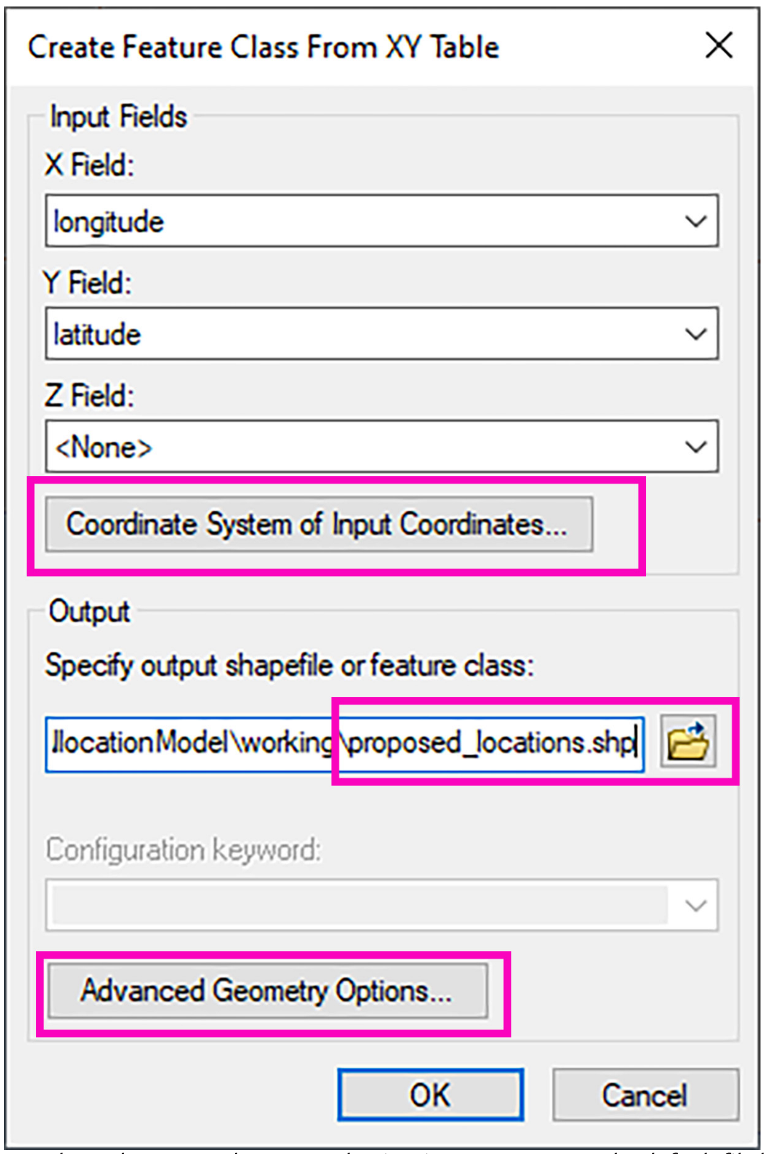


Figure 2.96: Always browse to the output destination. Never accept the default file location.

When the *Advanced Geometry Options* window appears, click the radio button next to *Use the map's spatial reference* (**Figure 2.97**). This option has the effect of transforming the output shapefile so that it uses **NAD 83 UTM Zone 10** as its spatial reference.

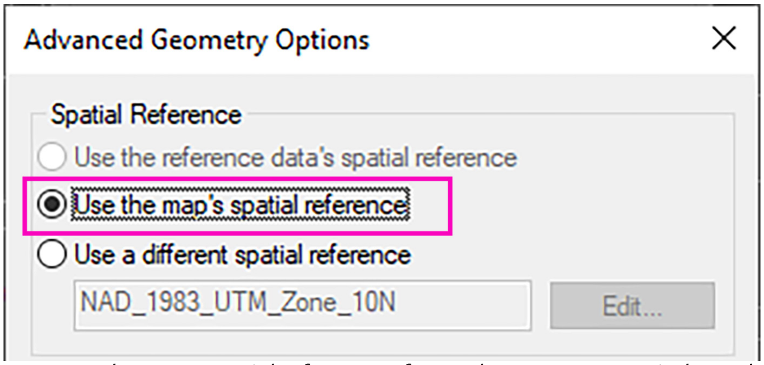


Figure 2.97: The map's spatial reference refers to the *Data Frame* window, which currently uses **NAD 83 UTM Zone 10** as the display projection.

Add the new businesses layer to the *Table of Contents* (**Figure 2.98**).

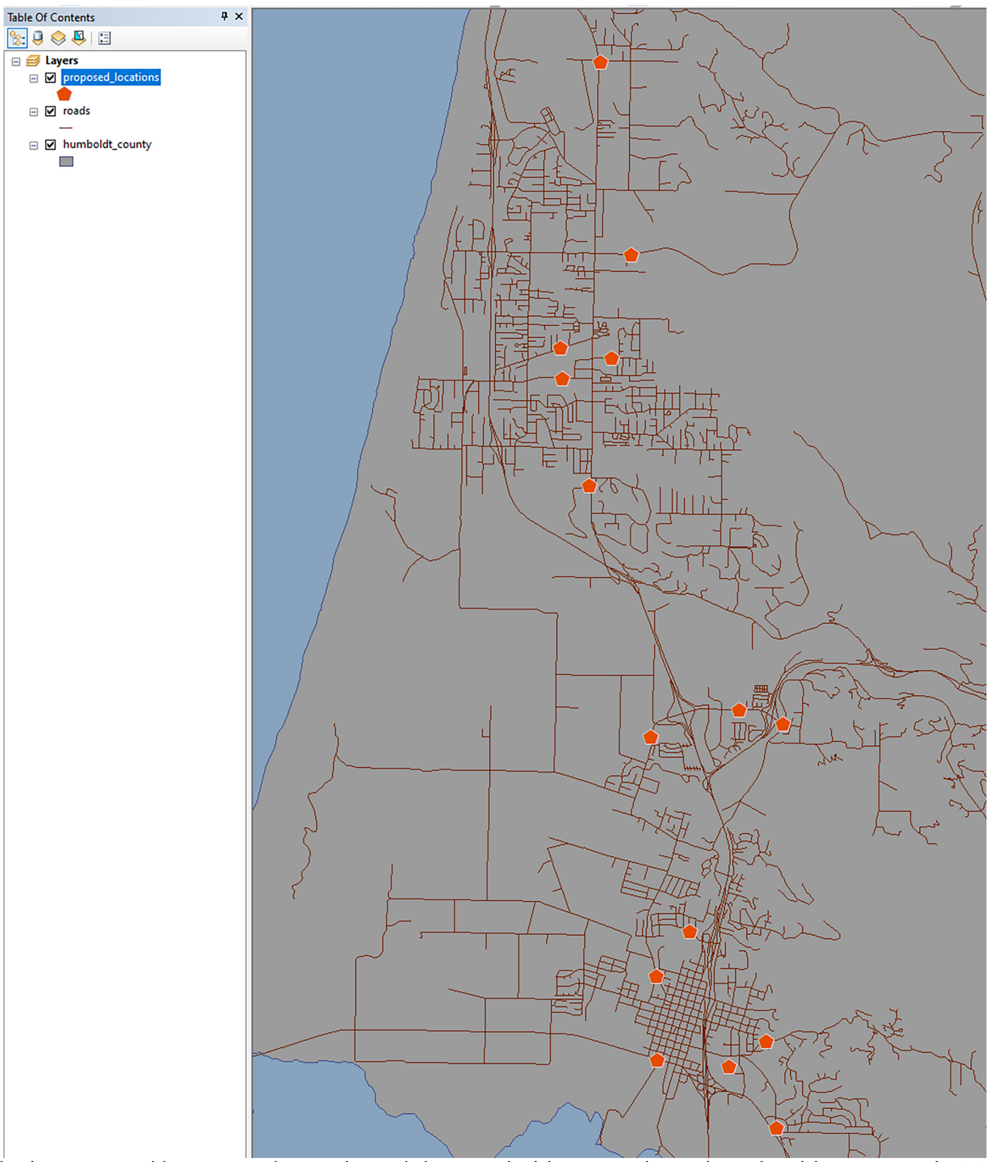


Figure 2.98: Only the proposed locations, the roads, and the Humboldt County boundary should appear in the *Table of Contents*.

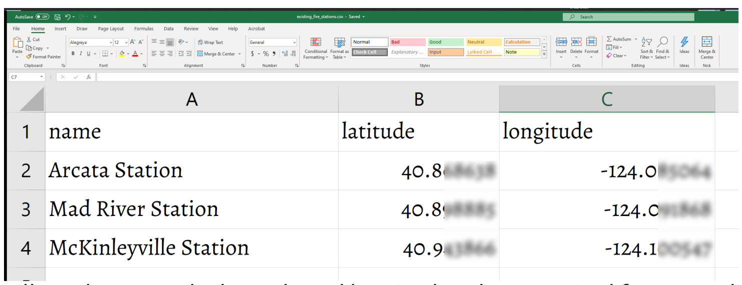
Skill Drill: Geocoding an Address and Creating a CSV Table to Import As XY Data

Eventually, you will compare the top three proposed locations with the three existing fire stations. The existing fire stations are located at the following addresses:

Arcata Station
631 9th Street, Arcata CA 95521
Mad River Station
3235 Janes Road, Arcata CA 95521
McKinleyville Station
2149 Central Avenue, McKinleyville CA 95519

Use **Google Maps** to geocode the address and obtain the latitude and longitude values for this address in decimal degrees. Google Maps uses **WGS 1984** as the datum for latitude and longitude values. In Microsoft Excel, create a CSV table with the following field headers: name, latitude, longitude. You may use the names provided above as the name attribute for the existing stations. Enter the latitude and longitude values in the appropriate fields. *Save as* a CSV table and close Excel.

In ArcMap, you may need to refresh your original folder to see the CSV file.



	A	B	C
1	name	latitude	longitude
2	Arcata Station	40.8	-124.0
3	Mad River Station	40.8	-124.0
4	McKinleyville Station	40.9	-124.1

Figure 2.99: You will need to enter the latitude and longitude values acquired from Google Maps manually.

In the *Catalog Window*, expand the *original* folder and right-click on the existing fire stations CSV file. Select *Create Feature Class*, then choose *From XY Table*. Follow the same workflow as for the proposed locations CSV to create a file that uses **NAD 83 UTM Zone 10**. Add the new file to your *Table of Contents* (**Figure 2.100**).

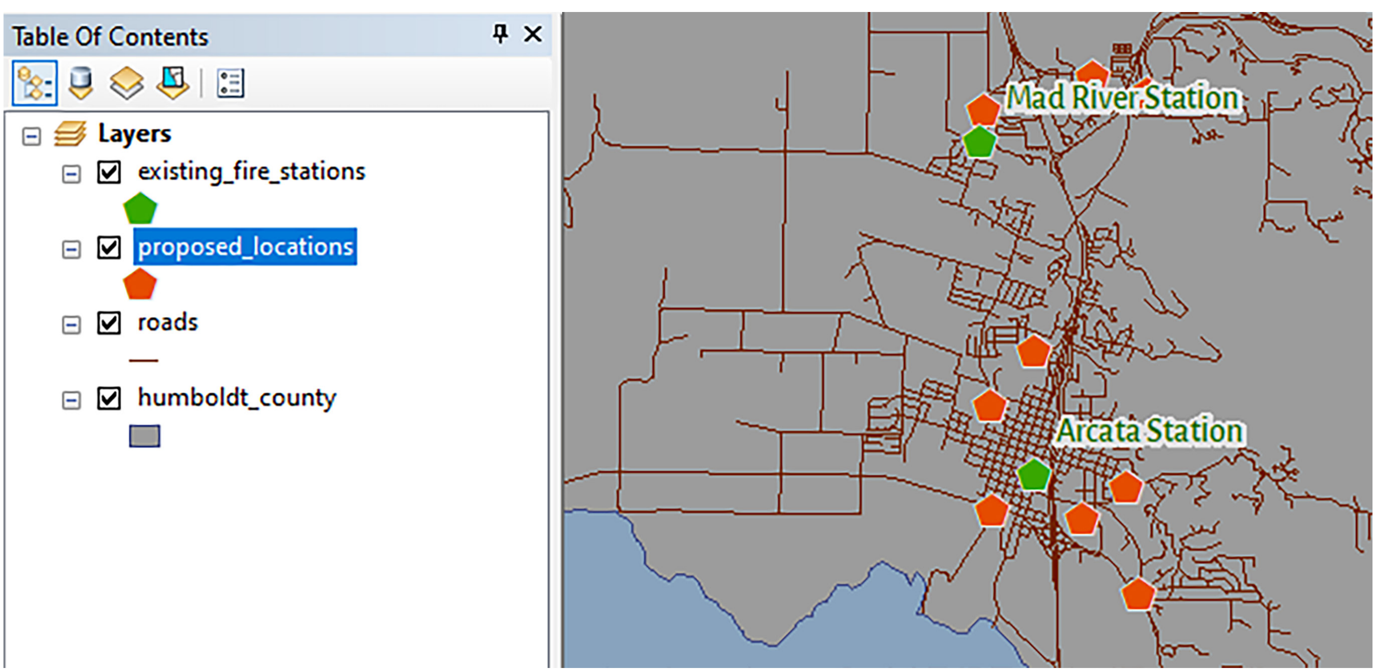


Figure 2.100: The existing fire stations are shown here in green. The proposed locations are shown on the map in orange. You can use the *Label Features* option to display the names of the existing fire stations.

Skill Drill: Adding the Fire Incident History as XY Data

By now, you should be familiar with creating a feature class from XY data. Using the skills you learned previously, download the history of service calls to the ***Arcata Fire District from a CSV***[1] table saved on Google Drive. As before, the latitude and longitude values come from the GCS WGS 1984. Like you did earlier, use **NAD 83 UTM Zone 10** for the output spatial reference and save the file to your *working* folder. Add the new shapefile to the map.

Creating A Network Dataset Using ArcCatalog

To conduct network analysis, you must first create a **network dataset** with roads (edges) and intersections (junctions). The ArcGIS software creates a network dataset from source features, such as lines and points. It then stores the connectivity between these source features. In this instance, the GIS uses the road layer as the source feature for the edges. It creates a junction feature class automatically for the road intersections.

Open the ArcCatalog window in ArcMap and navigate to your *working* folder. In the Arc*Catalog Window*, right-click on the shapefile of your roads and select *New Network Dataset* (**Figure 2.101**).

Note: If the Network Analyst extension is not currently activated, this option may appear grayed out. To correct this problem, activate the Network Analyst extension.

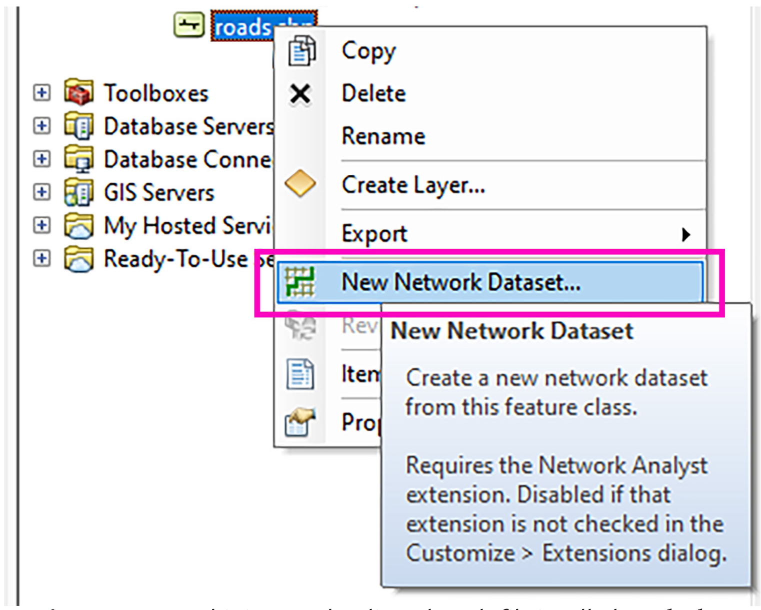


Figure 2.101: In this image, the clipped roads file is called *roads.shp*.

The New Network Dataset window opens up, and a default name is generated by the ArcGIS software ending in **_ND** (**Figure 2.102**). You may accept the default name if you wish and *click Next*.

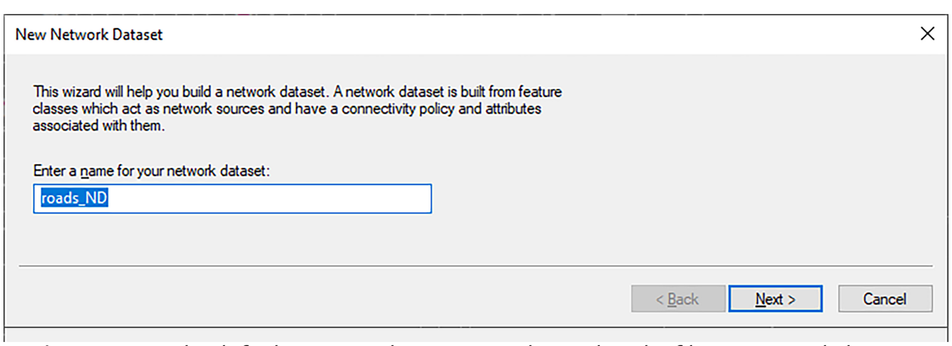


Figure 2.102: The default name ends in ND to indicate that the file is a network dataset.

1 URL: https://bit.ly/ArcataFireData

Click Next to accept the default *Yes* for modeling turns in the network. On the next page, you have the option to change the default connectivity settings. In most cases, you will not need to change these settings. By default, road segments connect to each other at endpoints. To check these settings, click on the Connectivity button (**Figure 2.103**). Make sure the source feature, the road layer, is set to *End Point*. Then *click OK* and *Next*.

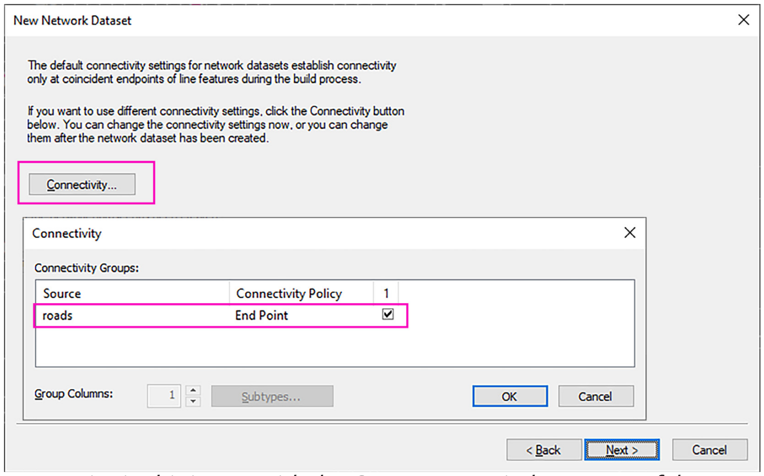


Figure 2.103: Two windows are active in this image with the *Connectivity* window on top of the *New Network Dataset* window.

When elevation fields are used, two edges that have the same endpoint locations, but different elevations remain disconnected. This is useful for modeling multilevel transportation infrastructure. This dataset does not contain a field with elevation information. *Click Next* to accept the default None for elevation settings.

The next window displays the attribute information for the network dataset (**Figure 2.104**). The ArcGIS software automatically tries to assign standard cost attribute fields, such as Length and Minutes, to the network dataset. Check the fields the ArcGIS software will use by clicking on the **Evaluators** button. This opens the *Evaluators* dialog window. From the drop-down menu next to Attributes, select *Length*. Under the *Source Values* tab, set the *Value* field to the Length field. If you will recall, this is the field updated via Calculate Geometry earlier. The *road* source feature is listed twice, one for each direction. Change the Value field to Length for both directions.

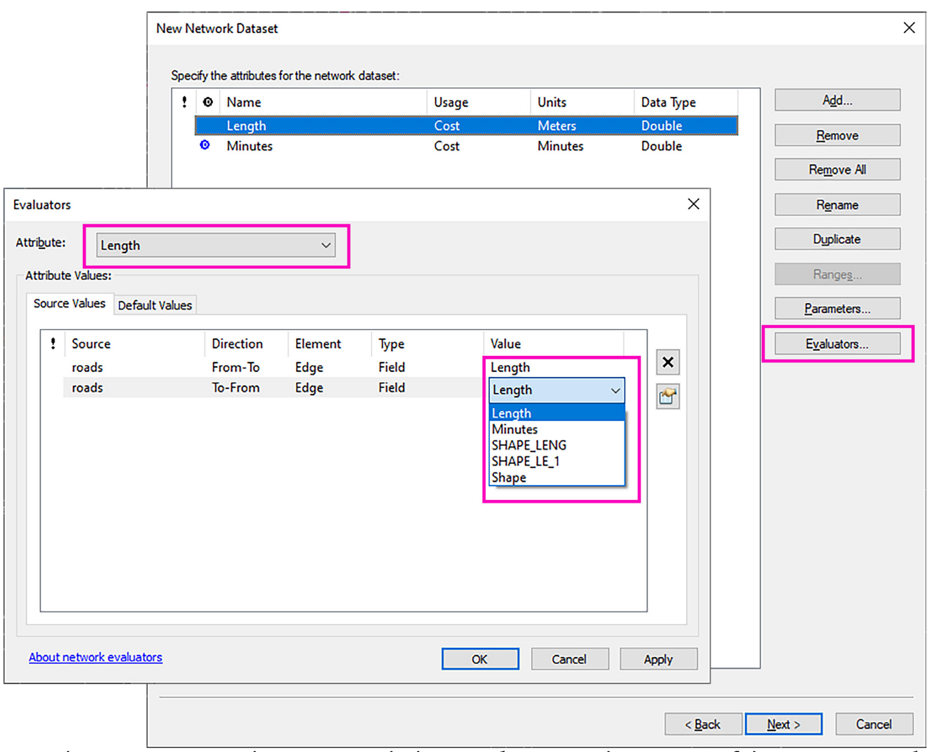


Figure 2.104: Two windows are active in this image with the *Evaulators* window on top of the *New Network Dataset* window

From the drop-down menu next to Attributes, select *Minutes*. Again, you will see the *roads* source feature listed twice, one for each direction. Make sure the *Value* field for each direction is set to *Minutes* (**Figure 2.105**). If you will recall, this *Minutes* field was created earlier and recorded the time it takes to traverse the edges in the network dataset. Then *click OK* and *Next*.

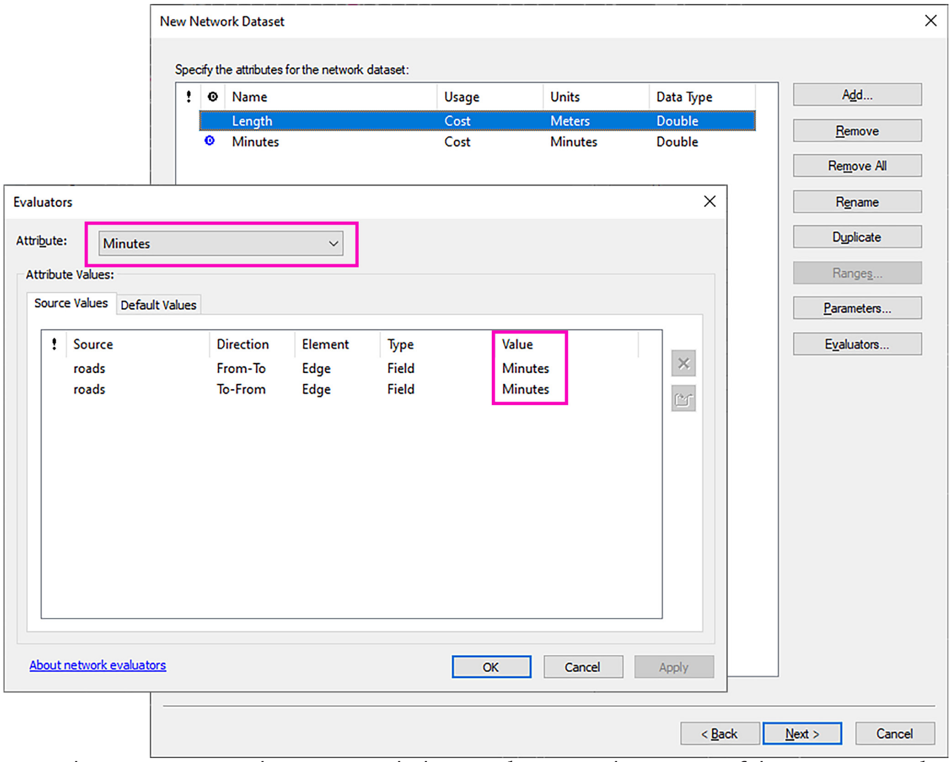


Figure 2.105 Two windows are active in this image with the *Evaulators* window on top of the *New Network Dataset* window

On the next window, *click Next* to accept the Default Travel mode. On the next window, *click Next* to accept the default Yes for driving directions settings. The last window will display a summary of the network dataset settings and attributes (**Figure 2.106**). Click *Finish* to create the network dataset.

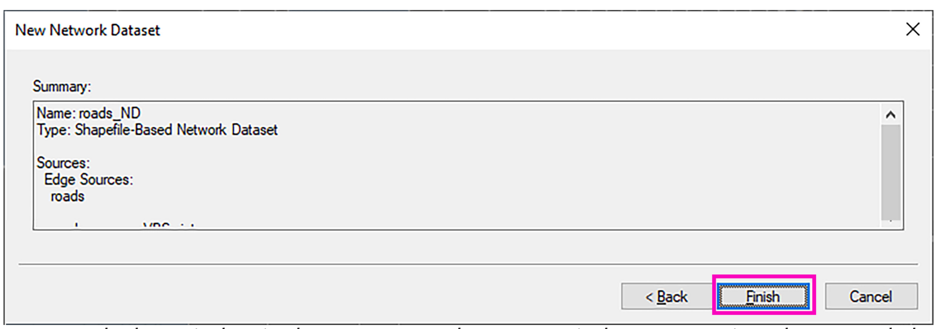


Figure 2.106: The last window in the New Network Dataset window summarizes the network dataset.

Click Yes to build the Network Dataset. Then *click No* to the dialog box that asks if you also want to add all feature classes that participate in the network dataset to the map (**Figure 2.107**).

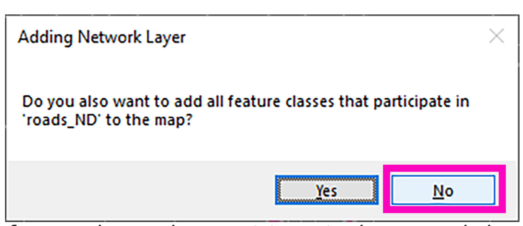


Figure 2.107: Adding all the feature classes that participate in the network dataset to the *Table of Contents* is not necessary. The network dataset references these feature classes from the *working* folder.

Be sure that only the incident history, the existing fire stations, the proposed locations, and the network dataset are loaded in the *Table of Contents* (**Figure 2.108**). This prepares you to conduct network analysis with minimal clutter.

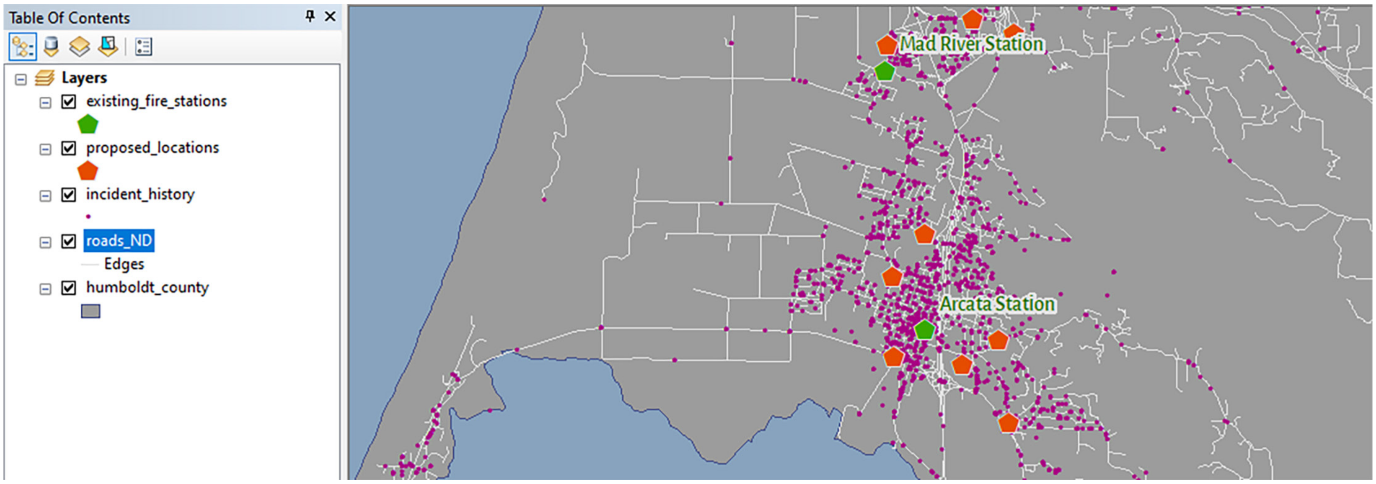


Figure 2.108: Maintaining a clean *Table of Contents* helps to prevent confusion and errors.

Setting up a Network Allocation Model

Network analysis involves calculating the **least-cost path** along the length of the network. **Cost** is also referred to as an **impedance**. Cost is usually measured in terms of distance, time, or money. For example, if time is the most relevant cost, the least-cost path from a fire station to a fire will be the quickest path, not the shortest one. In this step, you will use the *location-allocation solver* in *Network Analyst* to create a network allocation model. A **network allocation model** helps you choose which **facilities** from a set of facilities to choose from based on their potential interaction with **demand points**. In this scenario, the *facilities* will be the proposed fire station locations. The *demand points* will be the history of incidents within the Arcata Fire Protection District. The *impedance* is the travel time from facilities to demand points. This impedance will constrain the model to specific time windows that you define. The objective will be to chose **three facilities** among the candidates, maximizing the number of demand points that could be reached within a set amount of time.

In the *Network Analyst Toolbar*, choose *New Location-Allocation* (**Figure 2.109**).

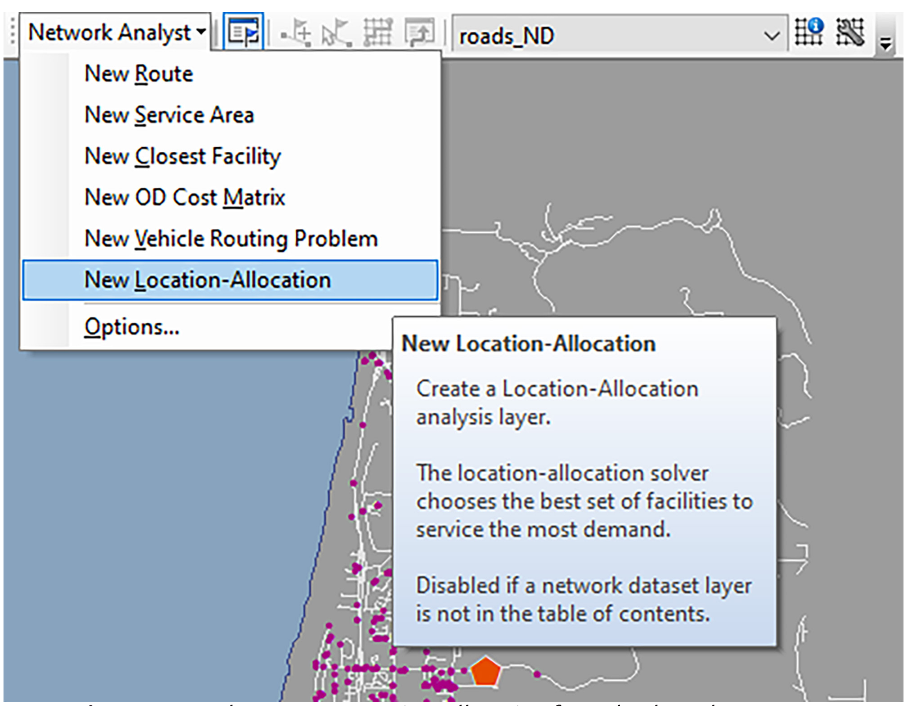


Figure 2.109: Choose New Location-Allocation from the drop-down menu.

In the *Table of Contents*, a new group layer is added (**Figure 2.110**). This group layer is comprised of empty **memory feature classes**. These data types behave similarly to shapefiles but only exist temporarily in the computer's memory. When you close ArcMap, you will no longer have access to these layers. They must be exported if you want to make them permanent. However, they are temporary for a reason. By using memory feature classes, you can re-run our model multiple times with different sets of parameters. This workflow keeps you from unnecessarily creating shapefiles that you may not need to keep in the long term. This saves on overall disk storage space and makes managing data easier as you conduct your analysis. In the *Network Analyst Window*, an empty network allocation model is added. It is here that you will enter your model parameters. Each time you adjust your model parameters, you will do so here as well.

Note: It is possible to create multiple network allocation models in one map document. Each represents a different model. In this exercise, you will only work with one. Be careful not to inadvertently create multiple models.

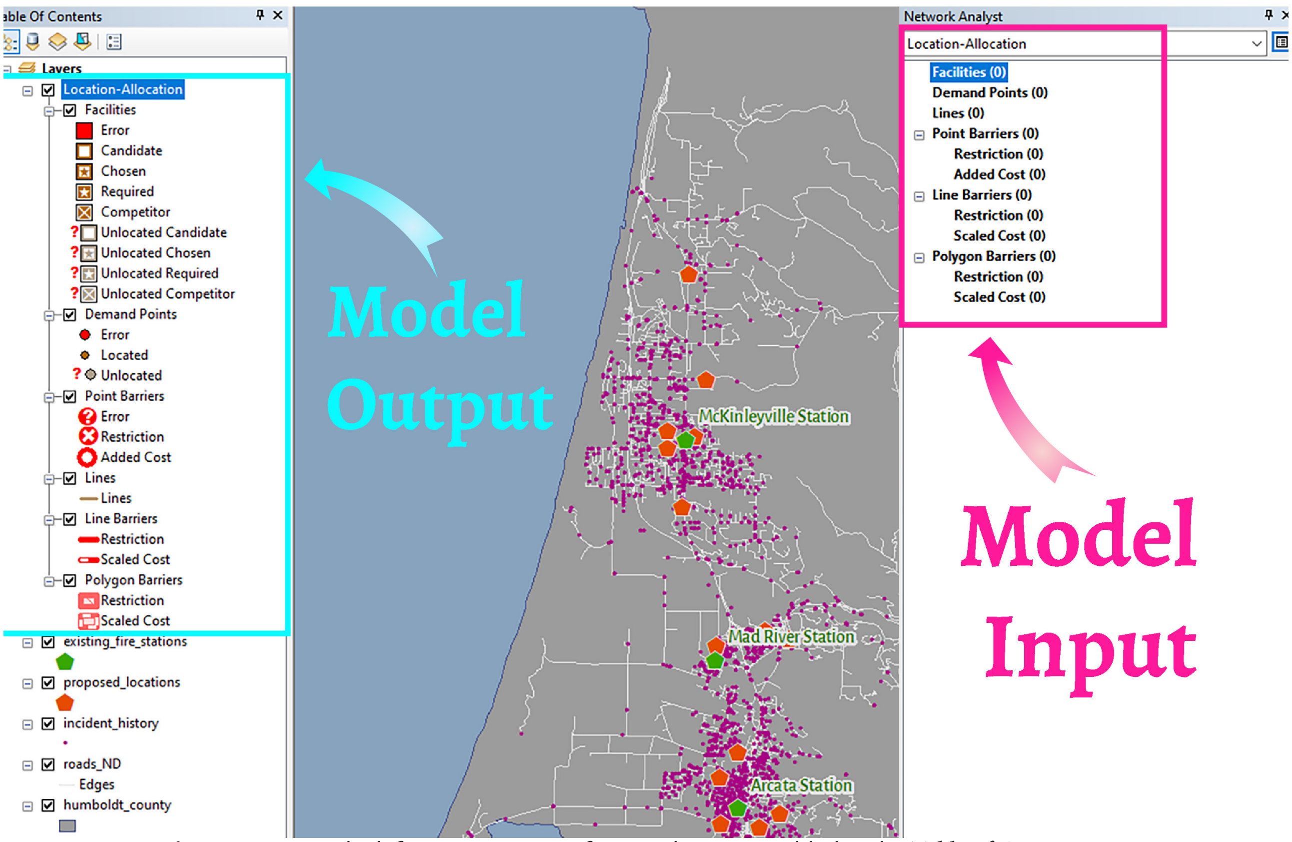


Figure 2.110: On the left, empty memory feature classes are added to the *Table of Contents*. On the right, a network allocation model is added to the *Network Analyst Window*.

Loading Facilities into the Model

Recall from earlier in the chapter, the **facilities** are a network class that represents places where services originate. They can be places where people travel to, such as a health clinic. Facilities can also be places from which where people travel, such as the delivery of retail goods from a warehouse to a set of retail stores. In a network allocation model, you can have multiple candidate facilities competing with one another. In this scenario, several proposed locations are under consideration as replacements for fire stations within the Arcata Fire Protection District. The proposed locations will represent the **facilities**.

Facilities also have attributes that influence cost or place limitations on the model. In this model, the **Facility Type** attribute determines how each facility will be considered when determining the solution. If all of the proposed sites are given a *Facility Type* of **Candidate**, then each site will be considered in the analysis. If one of the proposed sites is given a *Facility Type* of **Required**, then the model will force that location to be part of the final results. In this model, all of the proposed sites will be **candidates**. There are other facility attributes, such as *Weight* and *Capacity*. You **will not** be using these attributes in this model. Right-click on *Facilities* in the *Network Analyst Window* and select *Load Locations* (**Figure 2.111**).

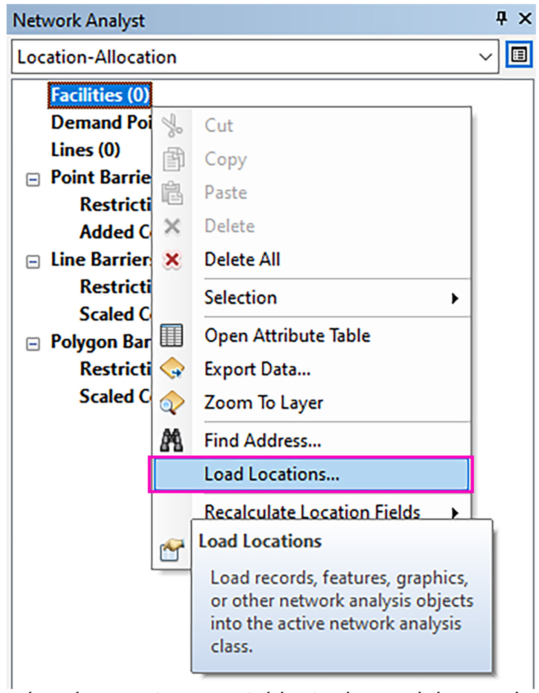


Figure 2.111: It is important to remember that any input variables in the model enter through the *Network Analyst Window*.

From the drop-down menu next to *Load From*, select the proposed locations layer. Facility parameters are defined under the *Location Analysis Properties* pane using two attributes, *Field*, and *Default Value*. Use the following values in the Load Locations Properties (**Figure 2.112**).

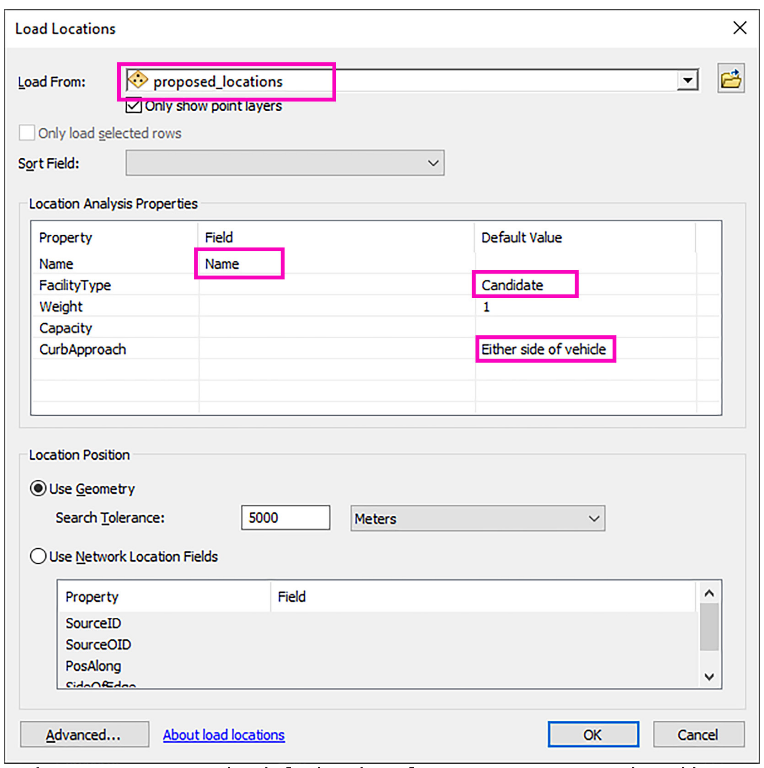


Figure 2.112: Keep the default values for any properties not listed here.

Once the values above are entered, *click OK*. The facilities should now be loaded in two places. The facilities in the *Network Analyst Window* represent the model parameters. Any modification of the model parameters, or **input** to the model, are entered here. In the *Table of Contents*, the memory feature class, also named Facilities, is now populated with the facility attributes, the **output** from the model.

Take a moment to open the attribute table for the Facilities memory feature class in the *Table of Contents* and inspect the information (**Figure 2.113**). The attribute table should look very similar to the proposed locations shapefile, but with some additional attribute fields. In many ways, memory feature classes look and behave like shapefiles. However, it is essential to remember that the information in the attribute table of a memory feature class is temporary. Any changes to the model in the *Network Analyst Window* will also change the attributes in the memory feature class. You must export the memory feature class to make any of the information permanent. Close the attribute table for the Facilities memory feature class. Next, you will load the Demand Points into the model.

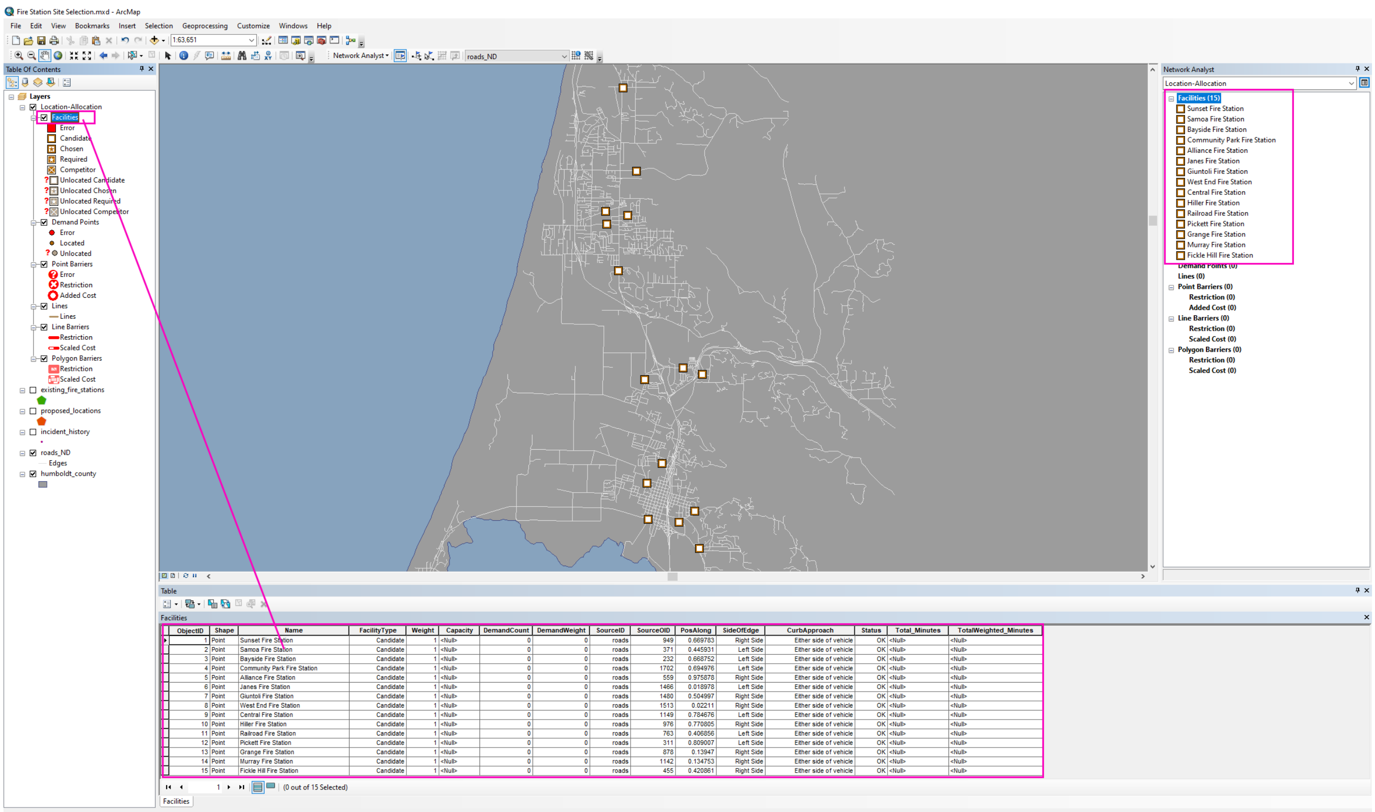


Figure 2.113: On the right, the model parameters (input) for the Facilities are added to the *Network Analyst Window*. On the left, empty memory feature classes in the *Table of Contents* are populated with attributes (output) from the model. Notice that all of the potential sites are using the Candidate symbology. Once the model is run, the three chosen locations will change to the *Chosen* symbology.

Loading the Demand Points into the Model

Recall from earlier in the chapter that the **demand points** are places served by the facility in some way. The exact nature of the demand points will depend on the service being provided. In this scenario, the **history of incident responses** serves as *demand points* for the proposed fire stations. The demand points have attributes that influence cost or place limitations on the model.

In this model, the demand points need to be **weighted** so that certain types of incidents will have a higher priority. By default, each demand point has a value of **1** for weight. When determining which fire station will be chosen, the weights for all of the demand points within its time window cutoff are summed. The sum of the weights is called **allocated demand weight**. It represents the **total demand allocated** to the chosen facility. However, you will change the weights of certain types of calls by adding a *weight* field to the attribute table then populating the results with the following weights:

» Assign a weight of **3** for incidents involving building fires, wildland fires, brush fire, and trash fires. In the incident history attribute table, the incident types are **111**, **141**, **142**, and **151**.

» Assign a weight of **2** for incidents involving medical assistance and emergency medical services. In the incident history attribute table, the incident types are **311** and **3211**.

» Assign all other incident types a value of **1**.

Start by adding a weight field to the incident history layer (**Figure 2.114**).

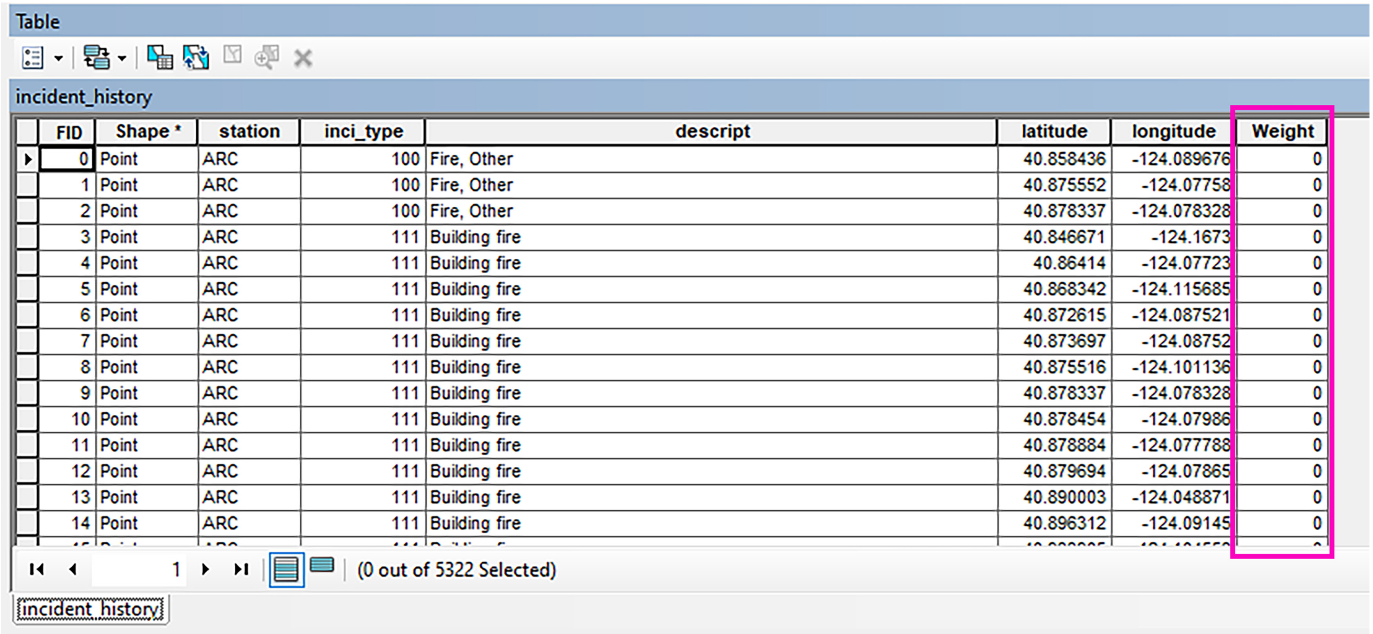


Figure 2.114: You may leave the field type as a short integer because this field will only have a value of 1, 2, or 3.

Use an *attribute query* to select the records for each incident type (**Figure 2.115**).

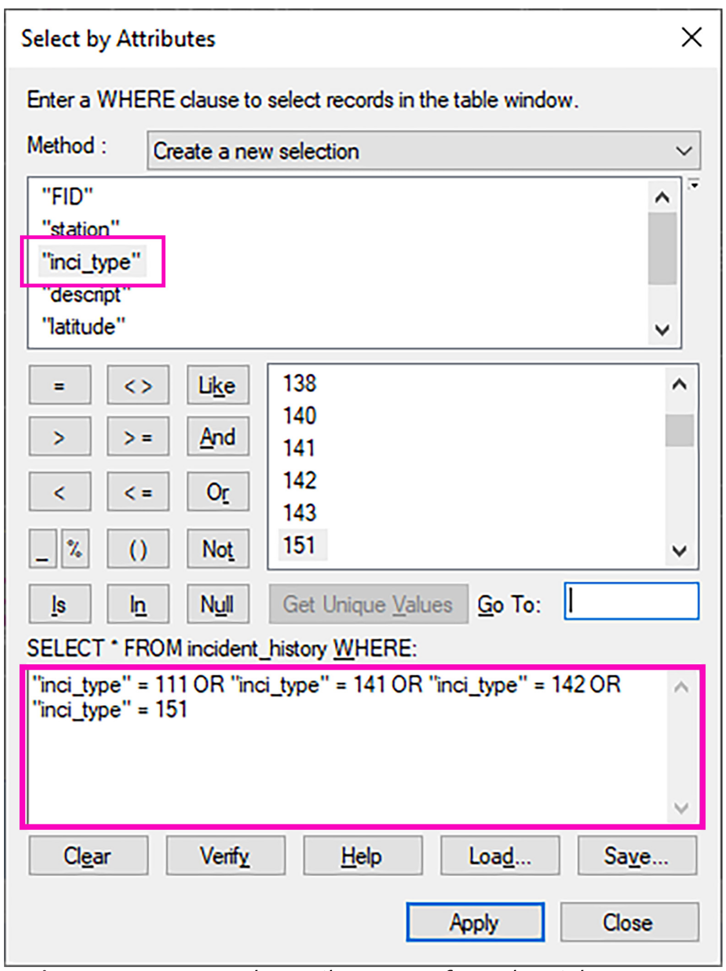


Figure 2.115: Repeat the attribute query for each weight category.

Then use the *Field Calculator* to populate weight value for the selected records (**Figure 2.114**). Remember, you can access *Field Calculator* by right-clicking on the field name, *Weight*.

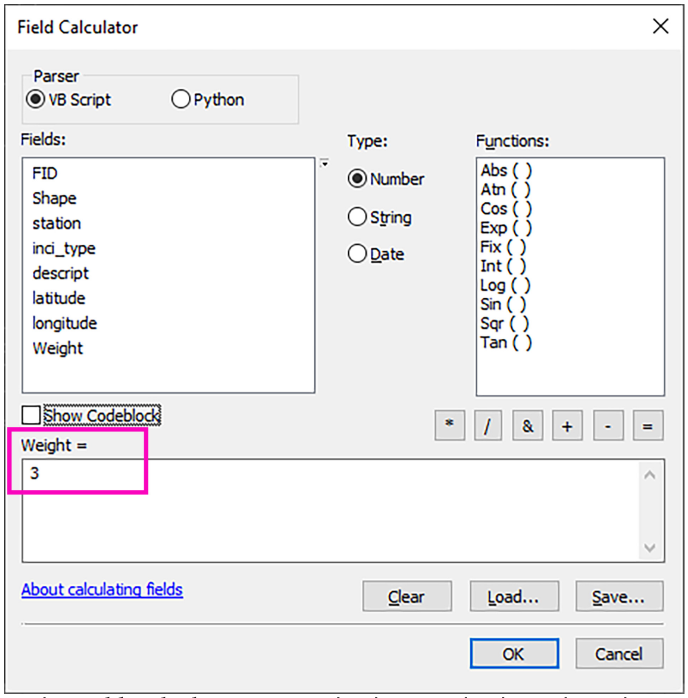


Figure 2.116: In the *Field Calculator*, you need only enter the desired number to set the value.

Once the demand points are assigned the appropriate weights, you are ready to load them into the model. Right-click on *Demand Points* in the *Network Analyst Window* and select *Load Locations*. From the drop-down menu next to *Load From*, select the incident history layer (**Figure 2.117**). Facility parameters are defined under the *Location Analysis Properties* pane using two attributes, *Field*, and *Default Value*. Use the following values in the Location Analysis Properties. When ready, *click OK*.

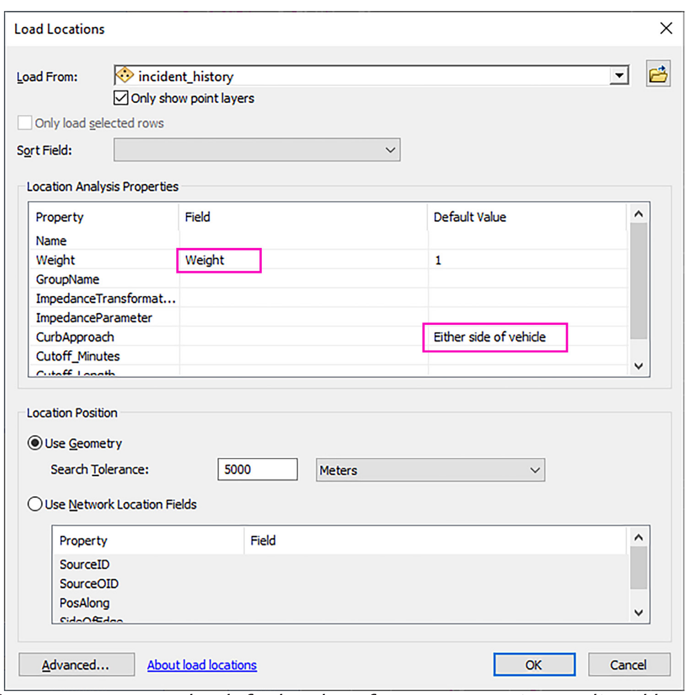


Figure 1.117: Keep the default values for any properties not listed here.

Note: Loading the locations may take a while. Be prepared to wait for approximately 20 minutes or more, depending on the speed of your computer.

Adjusting the Analysis Setting of the Model

The network allocation model also has global properties that must be set. So far, you have been working with attributes for individual model parameters such as facilities and demand points. Here you will adjust the analysis settings of the **entire** network allocation model. These settings are **global**, meaning they will influence all of the model parameters.

Click on the *Location-Allocation Properties* button in the *Network Analyst Window* (**Figure 2.118**).

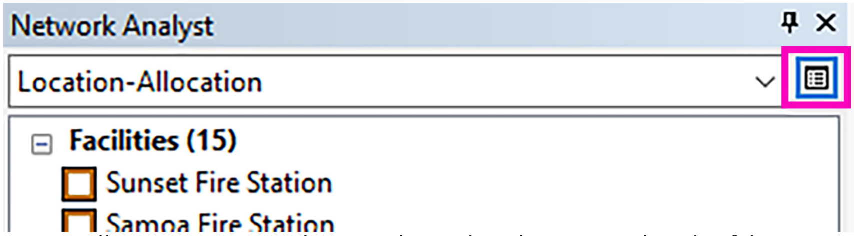


Figure 2.118: The Location-Allocation Properties button is located on the upper right side of the *Network Analyst Window*.

Navigate to the *Analysis Settings* tab. Verify that the *Impedance* is set to **Minutes (Minutes)** (**Figure 2.119**). There are two choices for the *Impedance* property, time and distance. The **Impedance** property specifies the cost of traveling along the edges of the network. By setting the property to minutes, the model will define the least cost path as the *quickest* path, not the shortest one. When calculating a least-cost path along the length of a network, the direction of travel can affect distance and travel time. In this scenario, travel will always originate *from* the fire stations *to* the location of the incident. Verify that the *Travel From* setting is **Facility to Demand**. Leave all other settings in the *Analysis Settings* tab as default.

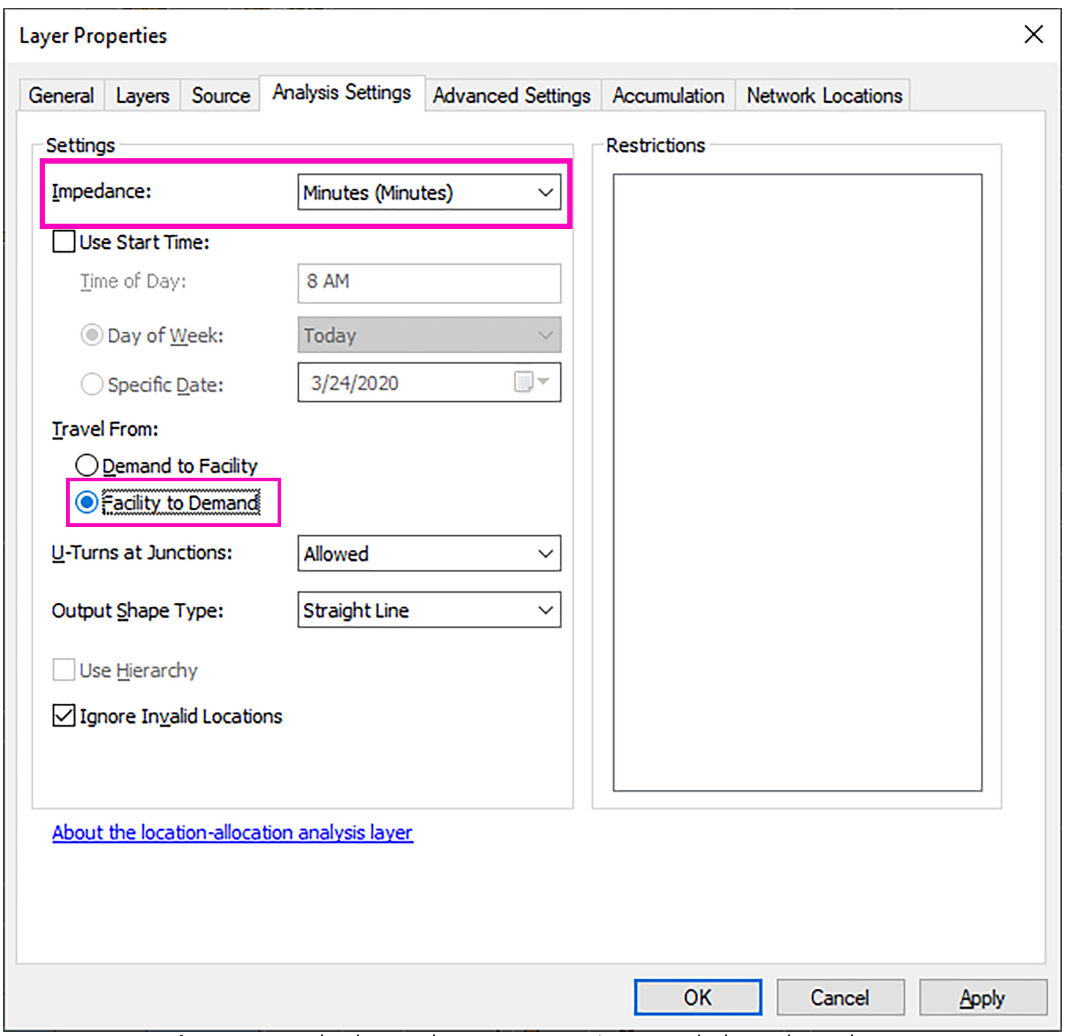


Figure 2.119: Check to make sure your settings match those shown here.

Navigate to the *Advanced Settings* tab. In the *Advanced Settings* tab (**Figure 2.120**), there are several *problem types* to choose from. The **Problem Type** settings determine how the model will calculate the allocated weight to the chosen facilities. The methods you use largely depend on the needs of the problem you are trying to solve. For example, a common *Problem Type* for emergency responders is **Maximize Coverage**. The premise is that emergency services are required to arrive at all demand points within a specified response time. This method is a straightforward calculation of allocated demand. Any demand points that fall outside of the impedance cutoff, the response time window, are not considered in the calculation. *All* of the demand points that fall within the time window cutoff are counted. Also, the *full weight* of the demand point is used in the calculation. If two or more facilities are competing with one another, the weight from a demand point only goes to the *nearest* facility. The demand weight is summed up for each candidate facility to determine the *allocated demand*. The facility with the **highest allocated demand** is chosen.

Verify that the *Problem Type* is set to **Maximize Coverage**. The *Facilities to Choose* setting specifies the total number of fire stations that can be chosen by the model. In this scenario, the Arcata Fire Protection District is interested in choosing only three sites from the list of proposed locations. Verify that the *Facilities to Choose* is set to **3**.

The **Impedance Cutoff** setting defines the boundaries of a service area originating from each candidate facility measured by the least-cost path along the length of the network. Since the *Impedance* setting in the *Analysis Settings* tab was set to *minutes*, the number entered here will define a service area based on a response time window. Set the *Impedance Cutoff* to **2**. Any path to a demand point with a higher cost than two minutes will not be considered. Leave all other settings as default and *click OK*.

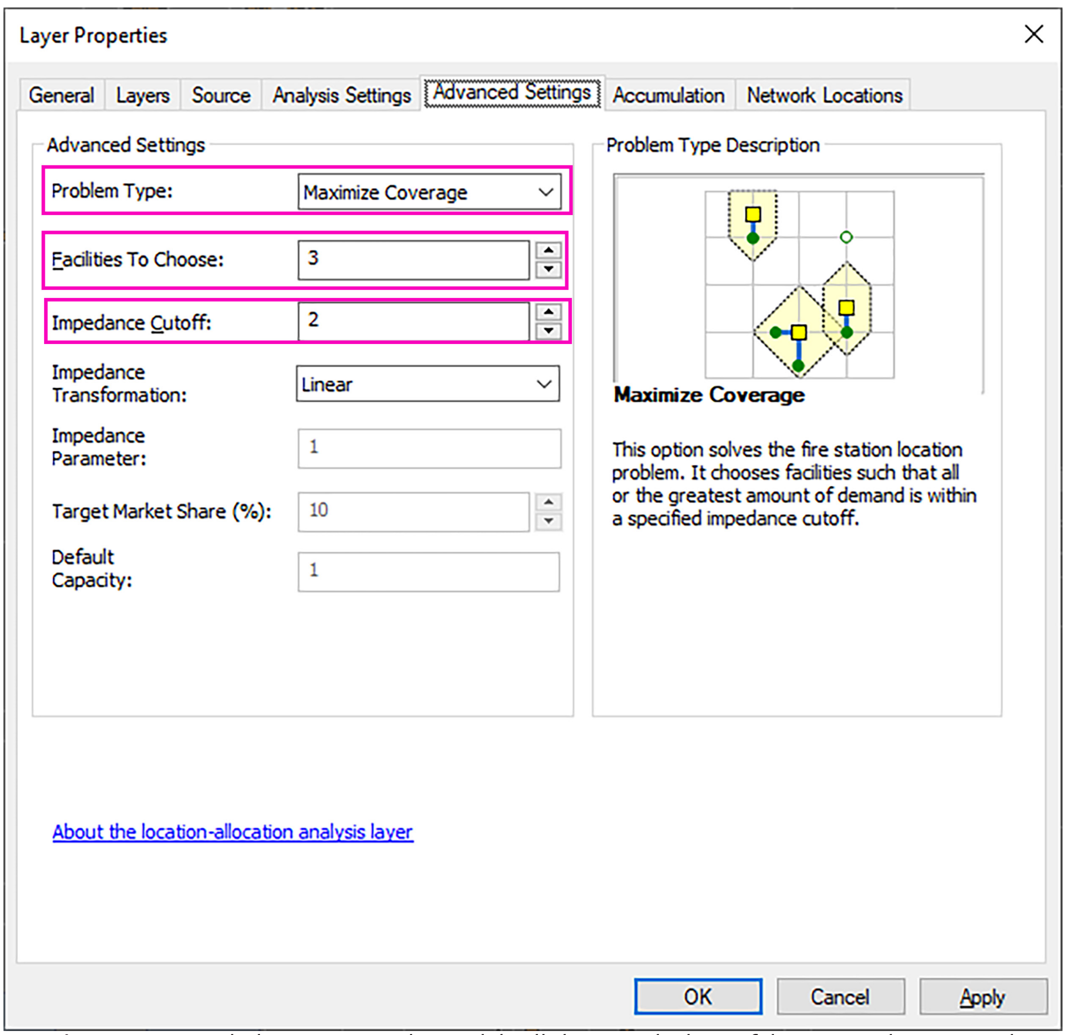


Figure 2.120: With these settings, the model will choose only three of the potential sites. Travel will be calculated from each of the potential sites to the demand points. However, any demand points farther than two minutes away from a proposed location will not be counted.

Running the Network Allocation Model

One of the advantages of creating a network allocation model via the *Network Analyst* extension is the ability to run the model multiple times while making small adjustments in between. In this activity, you run several models, adjusting the parameters each time. For the first run of the model, all of our parameters were set in the previous steps. To run the model currently loaded in the *Network Analyst Window*, click on the *Solve* button in the *Network Analyst Toolbar* (**Figure 1.21**).

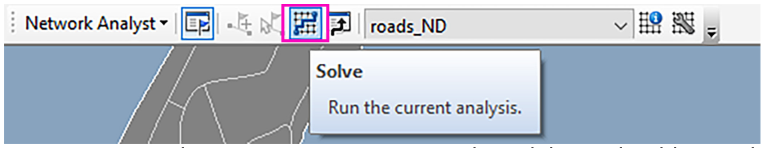


Figure 2.121: Every time one wishes to run or re-run a network model, one should press the *Solve* button.

When the *Network Analyst* completes the first run, it writes the results into the temporary memory layers in the *Table of Contents*. Take a moment to inspect the results. Of the proposed locations, you should see three fire stations chosen. Each of the three chosen locations has lines radiating outward to demand points within the two minute travel time window. It is a useful visualization for determining the extent of a two-minute time window.

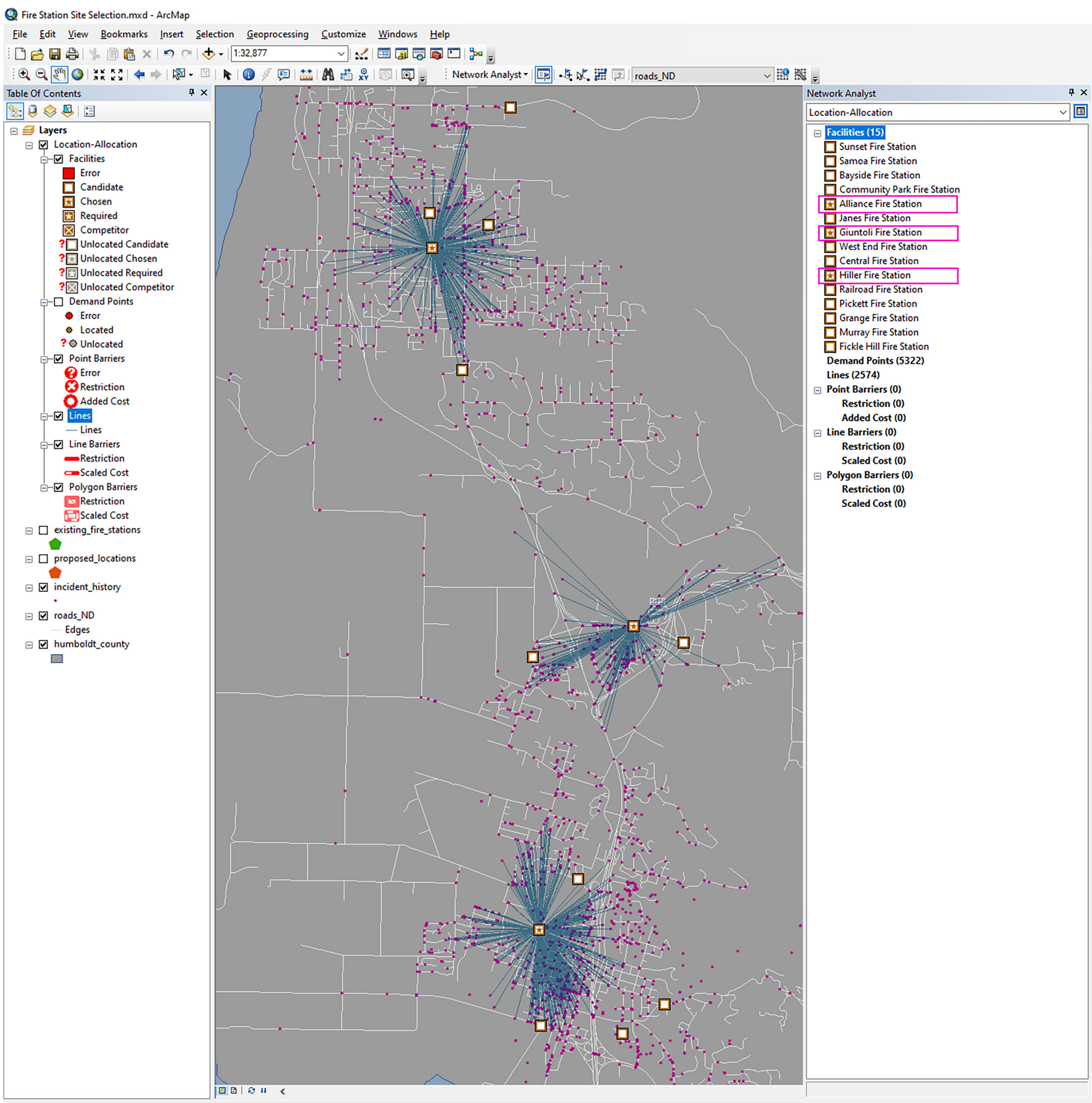
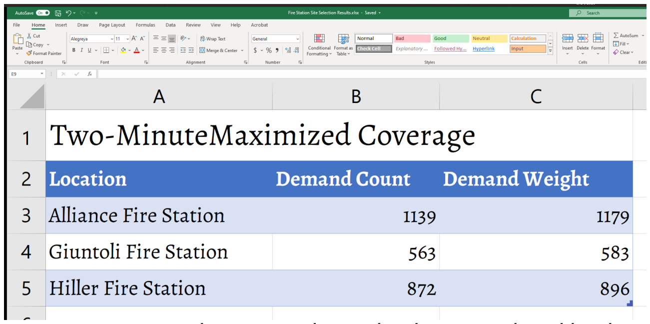


Figure 2.122: The lines radiate outward from the chosen facility to the demand points that can be reached within two minutes. The image may not match your results as the data is updated by the instructor over time.

Open the attribute table from the Facilities layer in the *Table of Contents*. Use an attributes query to select the records with a *Facility Type* of Chosen. Open a blank excel workbook and create a table recording the following results from your first model:

- The name of the three chosen locations
- The demand count for each location
- The demand weight for each location

Title the table, *Two-minute Maximized Coverage* (**Figure 2.123**). Save your excel file in your *final* folder for later use.



	A	B	C
1	Two-MinuteMaximized Coverage		
2	Location	Demand Count	Demand Weight
3	Alliance Fire Station	1139	1179
4	Giuntoli Fire Station	563	583
5	Hiller Fire Station	872	896

Figure 2.123: The image may not match your results as the data is updated by the instructor over time.

Using the Maximize Attendance Problem Type

When using the *Maximize Coverage* problem type, demand weight is allocated evenly among demand points, regardless of position within the response time window. For example, assuming each point carries the same weight, a demand point adjacent to the facility will allocate the same amount of weight as one that is farthest away, near the outer edge of the impedance cutoff. Maximize Coverage is an optimal problem type when demand points are based on population, such as Census block centroids. It guarantees service to the highest number of demand points served within a limited response time window. However, in this scenario, the demand points **are not based on the population**. Instead, demand points consist of a history of incident responses. The demand points are less uniform, with **clustering of incidents** over specific areas. This clustering might be an important indicator of where services might be needed more often. If so, another problem type, *Maximize Attendance*, could suit the situation better.

Maximize Attendance calculates allocated demand based on both weight and distance to the candidate facility. Demand points *closer* to a facility will allocate a *more significant* proportion of weight. Demand points *farther* from a facility will allocate *less* weight in proportion to the distance away. As a result, clusters of demand points near a facility will affect the allocated weight.

Click on the *Location-Allocation Properties* button in the *Network Analyst Window*. Navigate to the *Advanced Settings* tab (**Figure 2.124**). Verify that the *Problem Type* is set to **Maximize Attendance** and *click OK*.

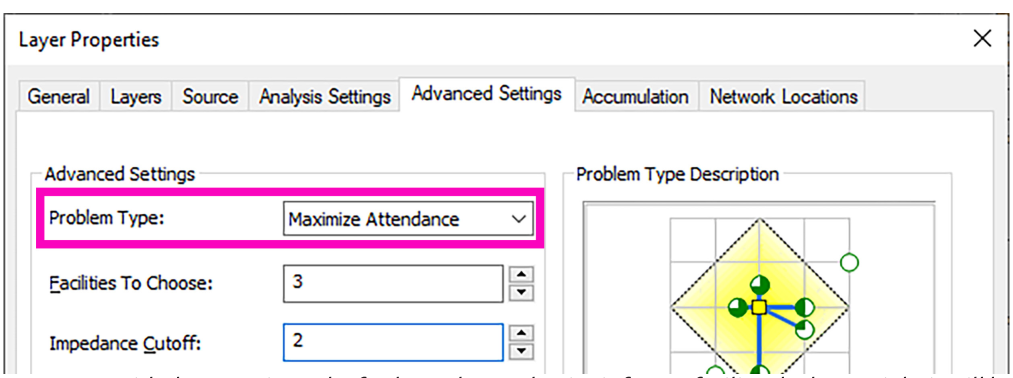


Figure 2.124: With these settings, the farther a demand point is from a facility, the less weight it will have.

Once the *Problem Type* and *Impedance Cutoff* are changed, run the model by clicking on the *Solve* button in the *Network Analyst Toolbar*. The new model will run, and the new results will overwrite the previous results in the memory feature classes in the *Table of Contents*. Record the results on a **second table** in Microsoft Excel titled *Two-Minute Maximized Attendance*.

Skill Drill: Adding Existing Fire Stations to the Model

After running the second model, three of the proposed locations were chosen based on a two-minute response time window using the Maximized Attendance problem type. In this model, you will also include the **existing fire stations** among the candidate facilities. The purpose is to determine whether or not any of the proposed locations is a better choice than the original fire stations.

Right-click on Facilities in the *Network Analyst* window and select *Load Locations*. Use the same methods and settings in previous steps to add the existing fire station layer to the list of Facilities. When this is done, the list of facilities should contain **both** the proposed locations as well as the existing locations. Once the existing fire stations are added to the list of facilities, run the model by clicking on the *Solve* button in the *Network Analyst Toolbar*. The new model will run, and the new results will overwrite the previous results in the memory feature classes in the *Table of Contents*. Record the results on a **third** table in Microsoft Excel titled *Two-Minute Maximized Attendance with Existing Fire Stations*.

Skill Drill: Increasing the Impedance Cutoff to Five Minutes

The Arcata Fire Protection District prefers a two-minute response time. However, a response time of five minutes is acceptable. For this last model, change the impedance cutoff from two minutes to **five minutes**. Click on the *Location-Allocation Properties* button in the *Network Analyst Window*. Navigate to the *Advanced Settings* tab and change the *Impedance Cutoff* setting to five minutes. Record the results on a **fourth** table in Microsoft Excel titled *Five-Minute Maximized Attendance with Existing Fire Stations*.

Skill Drill: Creating a Map of the Results

You should be familiar with the steps needed to create a map layout of your results. Design three maps for use as figures in a report or summary (**Figure 2.125**). Ideally, the map should be designed at a size of approximately 6 or 7 inches wide. Include a north arrow, a scale bar, and a legend. The map should include the results of your final model with both proposed and existing fire station locations using a five-minute response time. When your map layout is complete, export the map as a PNG file with a resolution of 300dpi. Save the file in your *final* folder.

Figure 2.125: Actual results may vary as the instructor changes the data over time.

Chapter 3: Modeling Overland Paths

Overland paths model movement or routes through places with no infrastructure or systematic constraints. This chapter covers the elements that make up an overland path model. Modeling overland paths is often based on modeling which path has the least cost. For example, determining the path for a new fire road in a national park might consider cost factors such as slope, vegetation, and proximity to sensitive areas such as streams or wildlife habitat. When modeling the large-scale movement of wildlife over diverse topography, cost factors might include elevation, slope, vegetation, proximity to water sources, and natural barriers to movement. Cost factors for modeling the spread of a wildfire from the point of origin might include slope, wind direction, and fuel sources.

Overland paths can be used in a variety of applications such as the following:

- Determining the path for a new road
- Modeling large-scale movement of wildlife over diverse topography
- Predicting the spread of a wildfire from the point of origin

Components of an Overland Path Model

Four primary factors are necessary components for modeling overland paths. These components are usually generated in the form of raster surface models. These surface models are then used to calculate a least-cost path:

- Distance
- Direction
- Allocation
- Cost

Modeling Distance

In a *distance surface model*, **distance** is calculated using a raster dataset in which cells are assigned a value based on the distance to the nearest *source feature*, sometimes called the *origin*. A **source feature** is a point of origin where the overland path begins. For example, the source feature for a wildfire might be a campsite or the location of a lightning strike. A source feature doesn't have to be a point and can also be an area or a line. For example, the source feature for a new fire road could be an established road segment along which the fire road would begin. Or the source feature for wildlife dispersal might include a specific region, such as the existing habitat of a population of mountain lions.

When creating a distance surface model, distance is calculated from the center of one cell in a raster to another using the Pythagorean theorem. This method is sometimes called **Euclidean distance** (**Figure 3.01**). It finds the shortest distance, a straight line, from each cell in the raster to the center of a source cell. The result is a distance surface model in which the value of each cell is based on the distance to the source feature. The farther away the cell is from the source feature, the higher the cell value.

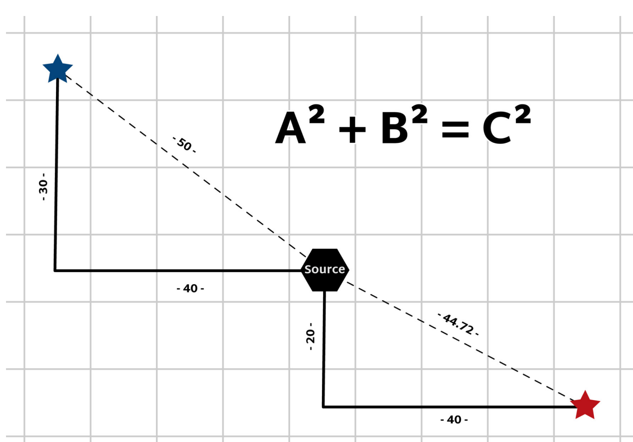


Figure 3.01: Euclidean distance is measured from each cell in the raster to the center of a source cell.

Modeling Direction

In a *direction surface model*, the **direction** is calculated using a raster dataset in which cells are assigned a value based on the direction of the *source feature* to that cell (**Figure 3.02**). The direction value is expressed as an **azimuth** in degrees ranging from 0° to 360°. North is given a value of 360°, while the source feature is given a value of 0°. All of the cell values in this raster point the way **back** to the source feature.

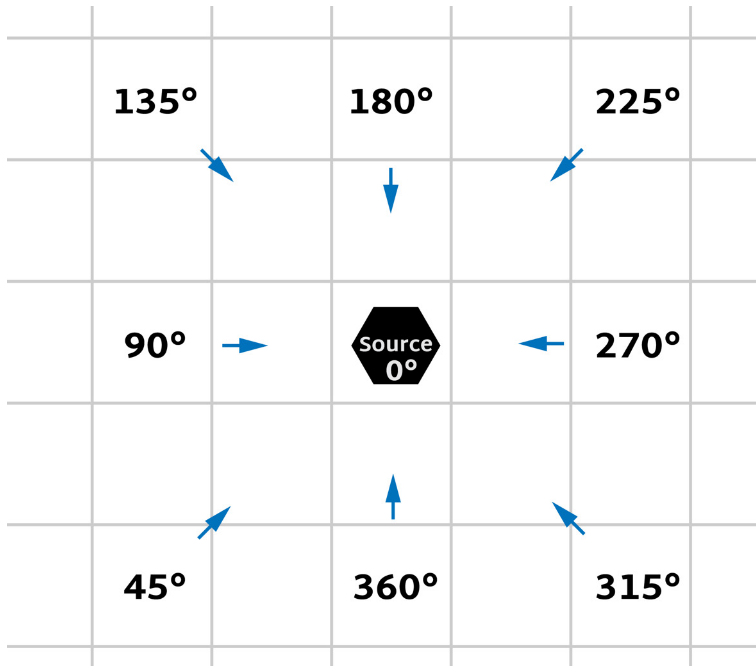


Figure 3.02: The cell values indicate the direction back to the source feature.

Modeling Allocation

Allocation is another component of modeling overland paths and is used when there is **more than one source feature**. It means that individual cells are assigned values based on specific sources. For example, if you are creating a *distance surface model* with multiple points of origin, the distance value assigned to each cell depends on the *nearest* source feature (**Figure 3.03**). This result is due to **allocation**.

2.82	2.24	2	2.24	2.82	2.82	2.24	2	2.24	2.82	2.82	2.24	2	2.24	2.82	3.6
2.24	1	1	1	2.24	2.24	1	1	1	2.24	2.24	1	1	1	2.24	3.16
2	1	1	1	2	2	1	2	1	2	2	1	3	1	2	3
2.24	1	1	1	2.24	2.24	1	1	1	2.24	2.24	1	1	1	2.24	3.16
2.82	2.24	2	2.24	2.82	2.82	2.24	2	2.24	2.82	2.82	2.24	2	2.24	2.82	3.6

Figure 3.03: In this distance surface model, each hexagon represents a source feature.
Distance values in the raster dataset are based on the nearest source feature.

If you created a *direction surface model* with multiple origins, the direction values would only point back to the nearest source feature (**Figure 3.04**). This result is also due to **allocation**.

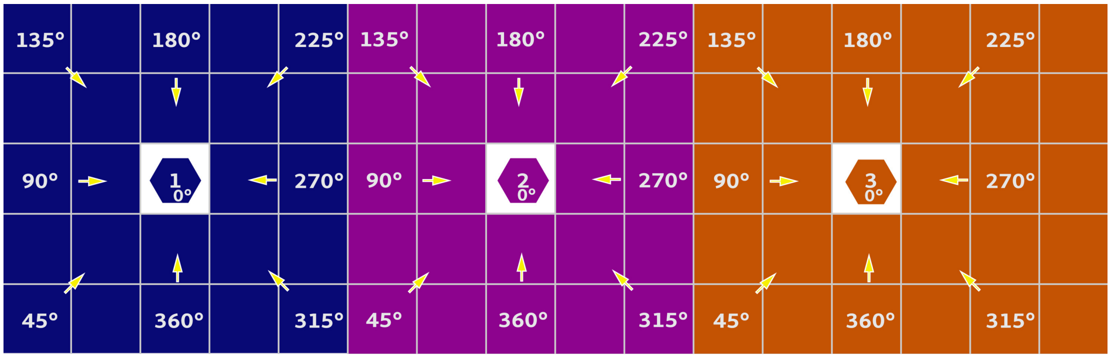


Figure 3.04: In this direction surface model, each hexagon represents a source feature.
Direction values in the raster dataset are based on the nearest source feature.

You can also create an **allocation surface model** where each cell is assigned to the nearest source feature based on the straight-line distance between them. Each source feature will be assigned an integer value, and all of the cell values in an allocation surface model will be based on that number (**Figure 3.05**).

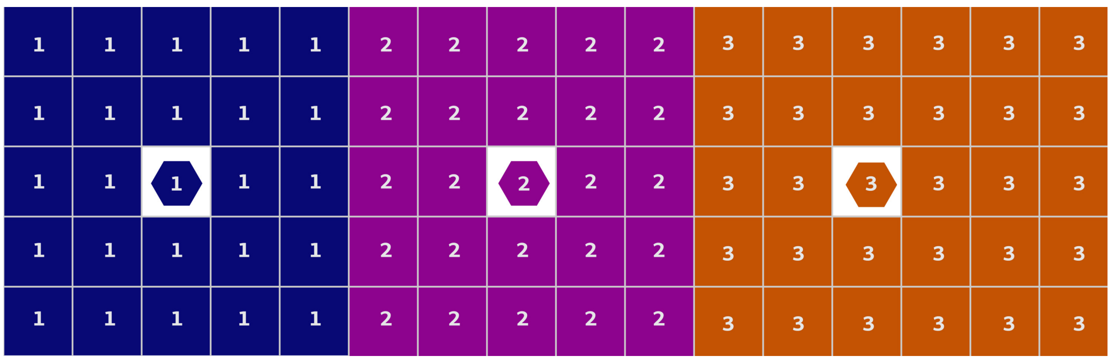


Figure 3.05: In this allocation surface model, each hexagon represents a source feature.
Values in the raster dataset are based on the nearest source feature.

Modeling Cost

A **cost surface model** represents some factor or a combination of factors that affect travel across an area. The goal is often to find the path with the least cost, which can be evaluated in several ways. For example, steep slopes may increase construction costs for a new road. In this case, your cost factor is the slope, and the cost unit might be U.S. dollars per meter. The cost can also represent the amount of energy it takes to move through an area. For example, the cost factor for elk migration routes might also be the slope. However, your cost units might be measured in calories or the oxygen consumption of the animal as it moves over the terrain. The cost can also represent the likelihood of movement through an area. For example, the cost factor for elk migration routes might also include proximity to water sources. In this case, the cost factor is the distance to the nearest water source, and cost units are measured in terms of distance units, such as meters.

Determining which factors affect costs in your study and ranking cost surface values can take time and research. The following are some examples of cost factors:

- Distance
- Direction
- Money
- Time
- Elevation
- Slope
- Land Cover
- Land Use Category

Total Cost Surface

A **total cost surface model** is used when you have **multiple** cost factors that you wish to combine into a single cost surface model. When you need to consider multiple cost factors, uniform cost units must be used. The values of the cost units must be converted into a **relative cost scale**, which is a standard scale that represents the costs between different cost factors. A relative cost scale could be as simple as a scale of 1 through 10. The number 1 might represent a low cost or low high likelihood of travel. The number 10 might represent a high cost or even a prohibitive factor. For the results to be meaningful, the relative scale must be the same among the different cost factors. Each cost category must be given the same relative scale, where the numbers roughly represent the same level of difficulty or likelihood of travel.

For example, if one wanted to create a total cost surface for an elk migration scenario which factored in slope and the proximity to water, a problem would arise because each of these cost surface models uses different **cost units**. The cost units for slope usually are in degrees (**Figure 3.06**). The cost units for the proximity to water would be distance. Because these costs use different units, it is challenging to combine them into a total cost surface.

Slope	Relative Cost Scale	Proximity to water	Relative Cost Scale
45	10	5 km	10
40	9	3 km	9
35	8	2 km	8
30	7	1 km	7
25	6	800 m	6
20	5	500 m	5
15	4	350 m	4
10	3	250 m	3
5	2	150 m	2
< 5	1	< 100 m	1

Figure 3.06: Slope and proximity to water typically have different cost units. Setting up a relative scale addresses this issue.

Through research, data collection, ground-truthing, a **relative cost scale** could be developed for this scenario. A relative cost scale means that a value of five on a cost surface model based on the *slope* would represent the **same** level of difficulty, energy, or likelihood of travel as a value of five in a cost surface model based on the *proximity to water*. Though each cost surface model would be based on different cost factors, they would share the same **cost units**. Once this happens, each cost surface model can be added together to create a total cost surface model (**Figure 3.07**).

Slope Cost Surface

1	1	5	9
2	1	3	8
2	2	4	7
3	4	5	3

+

Water Proximity Cost Surface

1	1	2	2
2	2	3	3
3	3	4	4
2	2	4	5

=

Total Cost Surface

2	2	7	11
4	3	6	17
5	5	8	15
5	6	6	7

Figure 3.07: Because slope and water proximity costs are converted into a shared relative scale, it is possible to combine them into a total cost surface model.

Least-Cost Path Model

The Least-Cost Path Model takes a cost surface model and evaluates all potential routes between a source feature and a destination. By keeping track of the accumulated cost units as it moves across the surface, it finds the path with the least cost.

There are several steps for modeling a least-cost path:

- Specify the origin
- Specify the destination
- Create the total cost surface model
- Create a cost-distance surface model
- Create a cost-direction surface model (back-link)
- Calculate the least-cost path

As mentioned previously, the **origin**, also referred to as the **source feature**, is a point of origin where the overland path **begins**. The origin can be a point, line, or polygon feature. A **destination feature** is a point where the overland path will **end**. A destination is similar to the origin in that it can be a point, area, or line feature. The feature type of the destination does not need to match the feature type of the origin. A key point to remember is that most of the surface models, such as distance, direction, and cost, are created relative to one or more source features, **not the destination**. This allows you to place a destination feature anywhere on one of these surfaces and have the GIS use the cell values to calculate distance, direction, and cost all the way back to the point of origin.

Also mentioned earlier, a **total cost surface model** combines all of the cost factors you wish to consider for your model into one surface using a uniform set of cost units. Developing a relative scale of cost units between different cost factors might require research, data collection, and ground-truthing and could be the most challenging aspect of your model to defend. A total cost surface model can be used to create a modified distance surface called a **cost-distance surface model**. However, in this model, the cell values are **not** linear geographic distance units such as meters or feet. Instead, each cell is assigned values based on cost units **weighted by distance**. It stores the *cumulative* cost moving outward from the origin.

The **cost-direction surface**, often called a **back-link**, stores information on the least-cost path from **each cell** to the **origin**. Each cell is assigned values between 0 and 8. The cell which contains the source feature is assigned a value of **zero**. The values **1** through **8** indicate a direction, start to the right and moving clockwise (**Figure 3.08**)

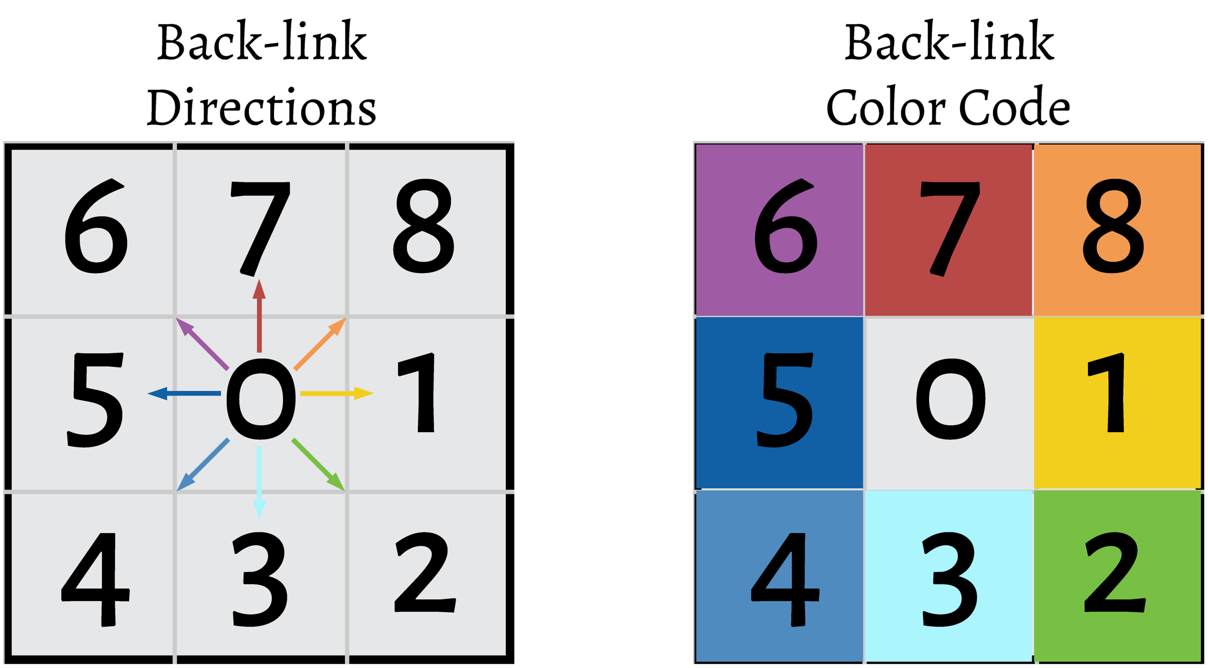


Figure 3.08: The numbers indicate the direction the least cost path will take.

The back-link is a raster dataset that is generated from a total cost surface model. The GIS creates a back-link by using the cumulative cost to *identify* which neighboring cell has the least cost to travel through to reach the source feature (**Figure 3.09**).

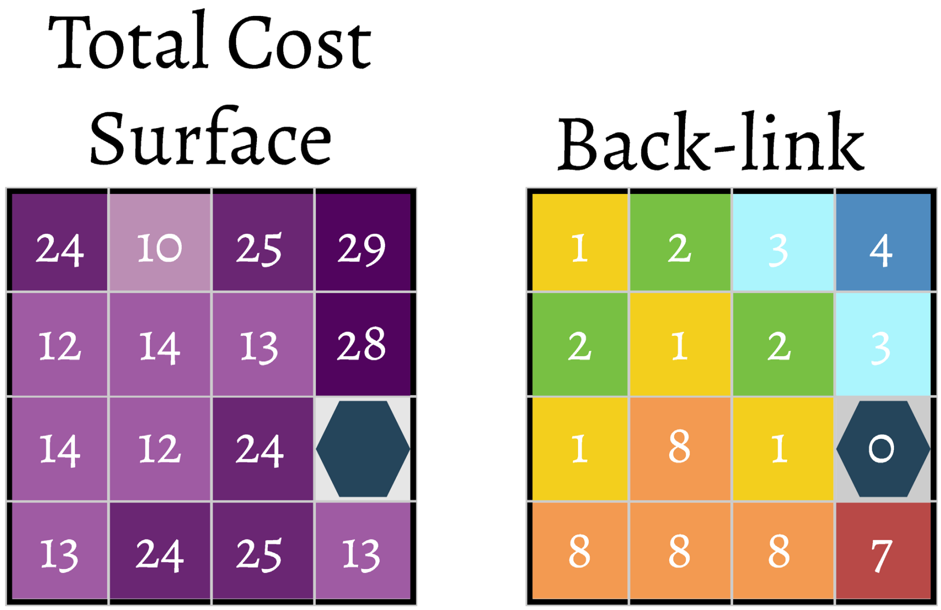


Figure 3.09: One can place a destination feature anywhere on the total cost surface raster dataset, and the GIS determines the least-cost-path back to the origin using the back-link.

One can place a destination feature **anywhere** on the total cost surface raster dataset. The GIS traces the route with the least cost **back to the source feature** using the back-link. The process stops when it reaches the first back-link cell with a value of zero, the origin. The result is a raster layer modeling the least-cost path (**Figure 3.10**).

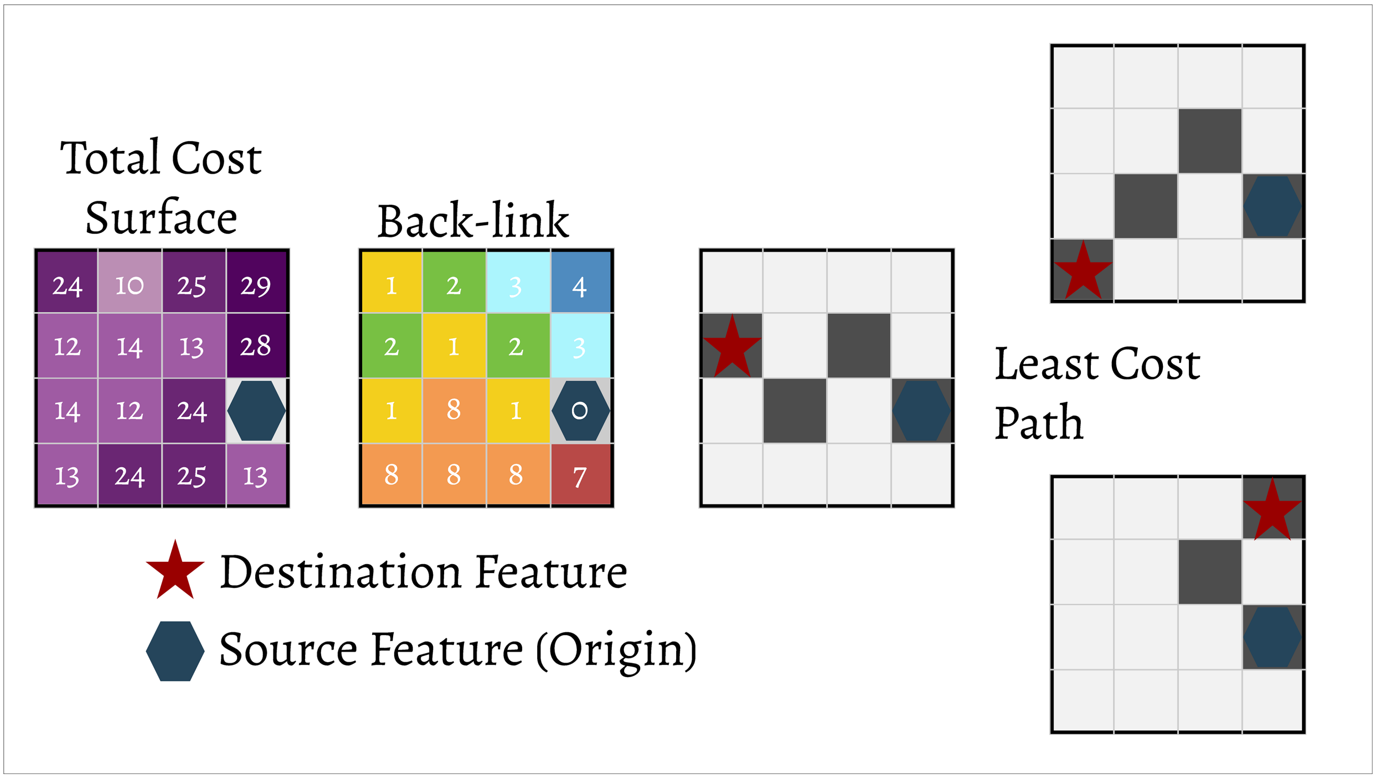


Figure 3.10: As the least-cost path is generated, each cell along the path is assigned a value based on the cost units and is written onto a new raster layer.

Tutorial: Tracking Creatures of Bavarian Folklore Using a Least-Cost Path Model

In this tutorial, you will explore modeling overland paths using a least-cost path model. You will create a series of distance, directional, allocation, and cost surface models, which will allow you to create the least-cost paths and wildlife corridors.

Estimated time to complete this tutorial: 6 hours

Learning Outcomes

Readers should be able to accomplish the following outcomes by the end of this tutorial:

- Review how to acquire data from a public source
- Geocode an address
- Review adding XY data
- Review Data Management Tools: project, define projection
- Review *Spatial Analyst* tools: Slope, *Raster Calculator*, & Reclassify
- Create a remap table
- Create a composite cost surface raster
- Create a cost distance raster
- Model least-cost path
- Model a migration corridor

Scenario

In this scenario, a dedicated group of graduate students from Humboldt State University spent last summer in the Redwoods State Park near the town of Orick. While maintaining a high level of beer consumption, they collected data on an elusive creature described in Bavarian folklore as the wolpertinger (**Figure 3.11**). Bavarian folklore portrays the wolpertinger as a fanged animal with the head of a rabbit, the body of a squirrel, the wings of a pheasant, and the antlers of a deer. The wolpertinger is thought only to be visible to persons who have consumed large amounts of beer.

Figure 3.11: Illustration by Albrecht Durer, 1502

During their stay in Redwoods State Park, the HSU students documented many sightings, located several wolpertinger dens, and later discovered an annual migratory behavior. It seems that in the fall, around late September and early October, sightings of the creature significantly increase near towns and villages. The wolpertinger prefers areas of high elevation and steep slopes to avoid predators. They also prefer areas covered by a dense forest canopy. Wolpertingers also avoid crossing deep rivers and lakes during periods of migration.

You will predict the overland movement of the wolpertinger using the following cost factors in your analysis:

» Elevation
» Slope
» Tree canopy density
» Hydrology

Conduct this analysis using the **Universal Transverse Mercator (UTM)** system along with the **North American Datum of 1983 (NAD83)**. Humboldt County lies in **Zone 10** of the UTM system. The data frame and all of your data must be in this spatial reference system at the start of your analysis. Create working copies of your data in this spatial reference system using the *Project* tool and the *Project Raster* tool in ArcMap as needed.

Setting up Your Workspace

In a typical workflow, you work on geospatial data using a local hard drive. When done, you compress your data and back up your work to your cloud storage so that you can retrieve the files from anywhere. When referring to a **local hard drive**, it means you are working on data physically located on the computer in front of you.

In contrast, some computers also include networked drives. **Networked drives** link to cloud storage and save the data elsewhere. Examples include services like OneDrive or Google Drive. For this tutorial, use the **desktop** as your local hard drive location. You may also use an external USB drive if you plan to work in multiple places.

You must avoid using networked drives while you work.
They increase the processing time and can cause technical glitches.

In this book, you use a particular folder structure. Start by creating your workspace folder on the local hard drive. A **workspace** is a folder or series of folders that contain all of your project files. The top-level folder in your workspace should indicate the activity or the project on which you are working. Organize all of your work within the workspace folder. On your **desktop**, create a new folder and give it a descriptive name, such as **LeastCostPathModel**. Be sure there are no spaces. You may use underscores instead of spaces. Inside this folder, create the following three subfolders: *original*, *working*, and *final*. Having a standardized folder structure helps to keep a project organized, primarily when you are working with multiple partners. The folder structure you see here (**Figure 3.12**) is the standard used in each of the tutorials presented in this book.

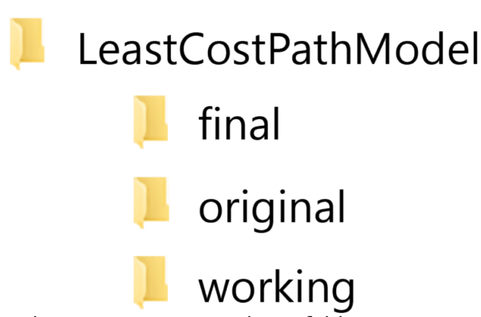


Figure 3.12: This diagram represents a basic folder structure used in this book.

As the name indicates, use the **original folder** for storing original, unaltered data. As you are working on a project, if, for some reason, your working version of the data gets lost or corrupted, you can go back to your *original* folder and find a fresh copy of the data. Use the **working folder** for data that you *create* or *alter* while working on your project. Use the **final folder** for storing any output you produce as a result of your work, such as images, maps, tables, or reports. Setting up a standard folder structure for a project is good practice and a habit you should develop.

Disable Background Geoprocessing

In the ArcGIS software, the *Background Geoprocessing* setting is often turned on by default. This setting allows users to continue to work while *a* tool is running in the background. However, sometimes this setting will stop tools from running or cause other unforeseen problems. To reduce that chances of the ArcGIS software crashing during this exercise, turn this setting off. In ArcMap, **open a new blank map document**. Open the *Geoprocessing options* from the *Geoprocessing* menu. Under *Background Geoprocessing*, uncheck the box next to the word *Enable* (**Figure 3.13**).

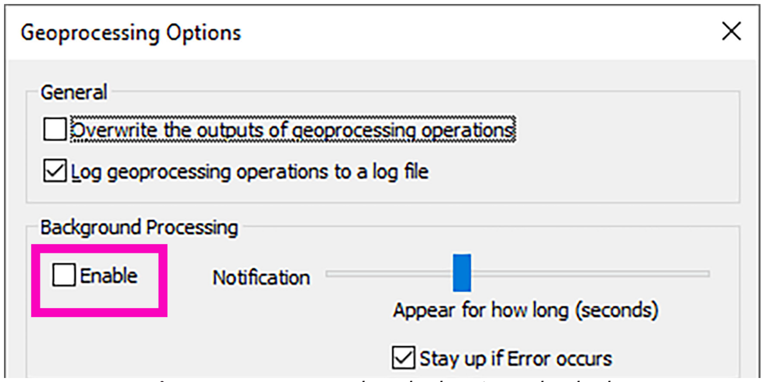


Figure 3.13: Be sure that the box is unchecked.

Spatial Analyst Extension

The steps in this activity involve using the *Spatial Analyst* extension. After launching ArcMap, make sure this extension is activated.

Skill Drill: Creating a Basemap for the Study Area

To start, you must give your study area a spatial context. To do so, you will download data from the ***National Historical Geographic Information System (NHGIS)***[1]. The (NHGIS) is a website that provides historical data and GIS boundary files from the United States Census. The data is free to use, which makes this site an excellent resource. To access the data, you need to create an account. Once you have created an account, log in to the website. Under *Apply Filters*, click the button that says *Geographic Levels* (**Figure 3.14**).

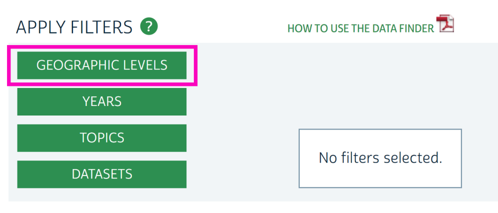


Figure 3.14: The NHGIS website allows you to apply filters to narrow down the search results.

When the *Geographic Levels* window appears, click the plus sign next to *County (by State)* (**Figure 3.15**). Then, click *Submit*.

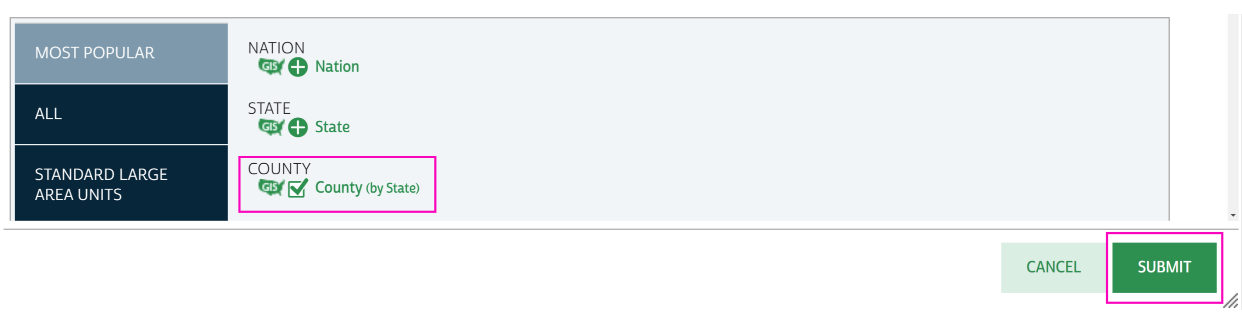


Figure 3.15: The geographic levels define which geographic extents to which the data is aggregated.

1 URL: https://www.nhgis.org/

Open the *GIS Files* tab and find the very last entry. Currently, the 2017 TIGER/Line shapefile is the most recent (**Figure 3.16**). Click the plus sign on the left.

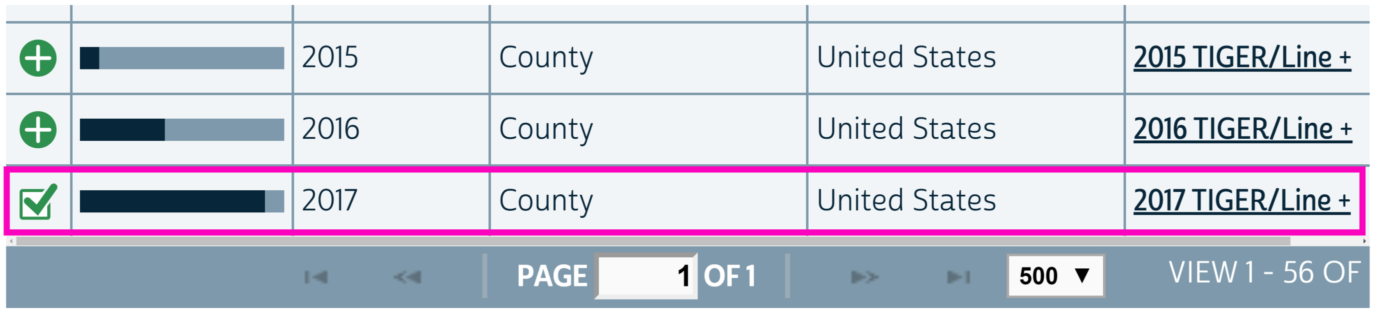


Figure 3.16: The *GIS Files* tab provides shapefiles at a resolution you define with the *Geographic Levels* filter. The NHGIS provides State, County, Census Tract, and Census block-level data.

On the upper right side of the NHGIS webpage, click the button that says *Continue* (**Figure 3.17**).

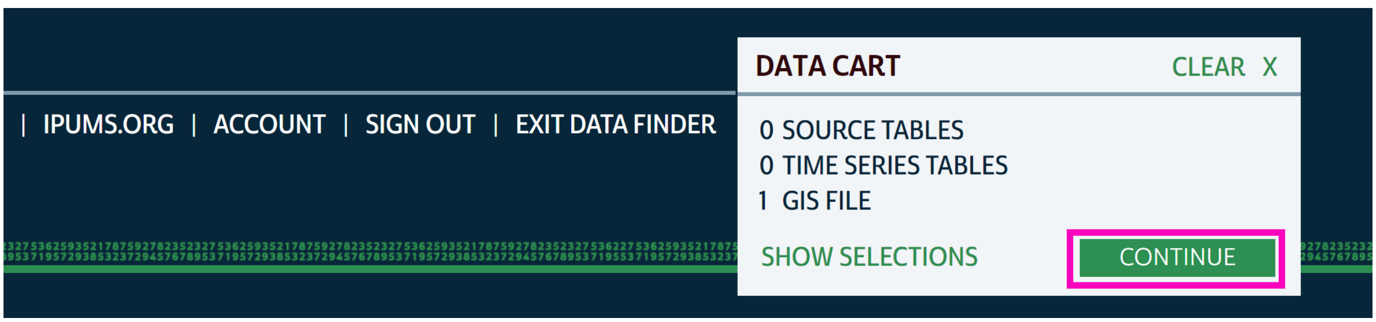


Figure 3.17: Press the Continue button repeatedly until it disappears.

Click continue again until you reach the *Review and Submit* page (**Figure 3.18**). Type in a short description of your order, then click *Submit*. You will get an email when your data is ready. Click the link in the email.

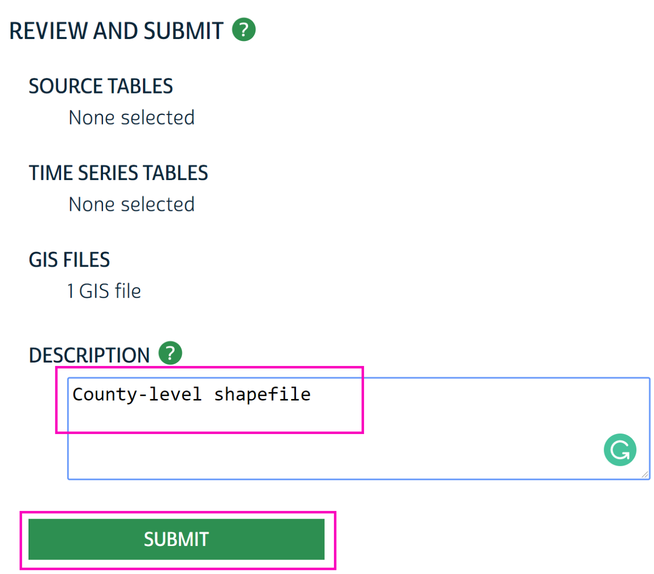


Figure 3.18: The Review and Submit page typically have more options from which to choose. There are no additional options with only GIS files in the cart.

If you do not get an email within about five minutes, refresh the NHGIS Extracts History webpage.
If, for some reason, the website is down or the data is no longer available,
you can download a backup copy of the data for this tutorial using the link provided in Appendix A.

Use the link under *Download GIS Data* and save the file to your *original* folder. Once ready, use the 7zip software to decompress the file. The NHGIS double zips their shapefiles, so you must open the extracted folder and locate the second zip file. Delete the original compressed zip files when done. In ArcMap, refresh your *original* folder in the *Catalog Window*. Use the *Project* tool to create a layer that uses the **NAD 83 UTM Zone 10** spatial reference (**Figure 3.19**). Save the output to your *working* folder. Once ready, add the new shapefile to your *Table of Contents*.

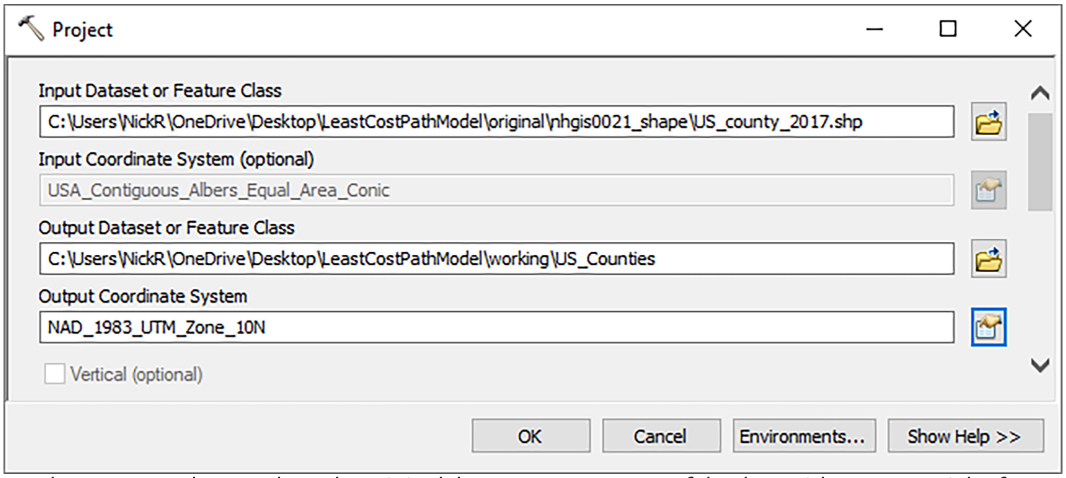


Figure 3.19: The *Project* tool never alters the original data. It creates a copy of the data with a new spatial reference system.

The data frame adopts the spatial reference of the **first** layer added to the *Table of Contents*. The first layer added to ArcMap must use the spatial reference **NAD 83 UTM Zone 10**. In future steps, specific tools and settings will rely on the data frame spatial reference.

If you already added a layer with a different spatial reference, close ArcMap.
Then reopen a new blank map document and add the projected U.S. county layer.

The U.S. county layer is extensive. For this tutorial, only the California counties Humboldt and Del Norte are required. Select Humboldt County and Del Norte County, then export the selected features as a new shapefile. Add the Humboldt Del Norte shapefile to the data frame. When done, remove the U.S. county layer (**Figure 3.20**).

Figure 3.20: Humboldt County and Del Nort County serves as the basemap layer for the study area.

Skill Drill: Adding the Den Locations as XY Data

In this scenario, HSU students recorded the location of several wolpertinger dens using hand-held GPS receivers. Information about these dens, including location and the *identification* number, is available in a ***comma-separated value (CSV) file***[1]. Download the file and save it to your *original* folder (**Figure 3.21**).

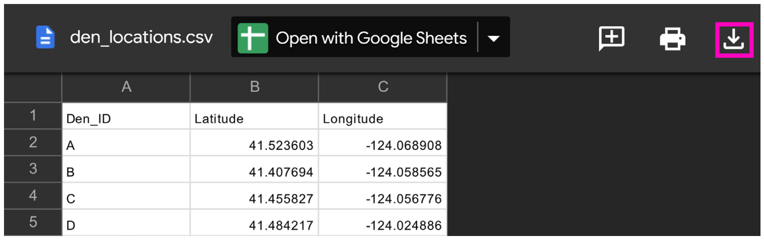


Figure 3.21: For the best results, use the Chrome browser to download the CSV file. Click the download button on the upper right.

It is important to remember that each geodetic datum as a unique set of latitude and longitude values. For example, the latitude and longitude values for your home using the North American Datum of 1983 will be different than the latitude and longitude values using the North American Datum of 1927. Likewise, the World Geodetic Datum of 1984 will use yet another set latitude and longitude values to define the location of your home. Like most latitude and longitude values of which you obtain from the internet or from a GPS receiver, the decimal degrees in this CSV table are currently in the geographic spatial reference system **WGS 1984**. When adding XY data, the ArcGIS software only reads the decimal degrees. It is unaware to which spatial reference system these decimal degree values belong. Always specify the datum *source* when adding XY data.

In the *Catalog Window*, expand the *original* folder and right-click on the business CSV file. Select *Create Feature Class*, then choose *From XY Table* (**Figure 3.22**).

In ArcMap, you may need to refresh your original folder to see the CSV file.

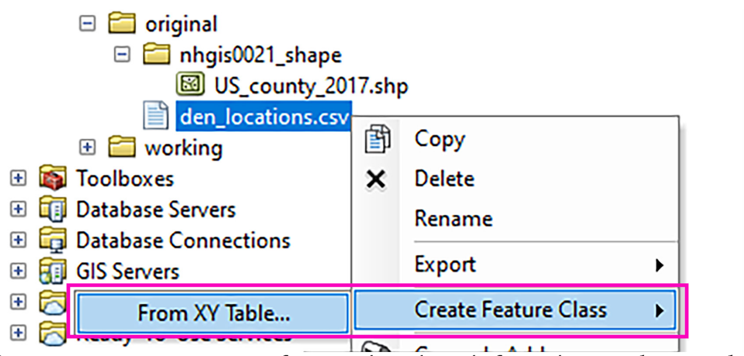


Figure 3.22: You can create a feature class directly from the *Catalog Window*.

When the *Create Feature Class from XY Table* window opens, click the button that says *Coordinate System of Input Coordinates*. Open the *Geographic* folder, then the *World* folder. Locate **GCS WGS 1984** and set that as the spatial reference (**Figure 3.23**).

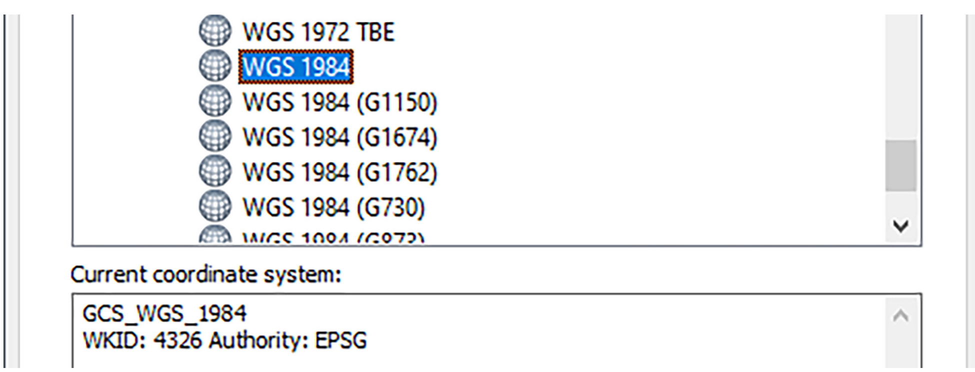


Figure 3.23: The source datum for the latitude and longitude coordinates in the CSV file is WGS 1984.

1 URL: https://bit.ly/wolpertinger-dens

On the *Create Feature Class from XY Table* window, click the yellow file folder icon and save the output to the *working* folder. Name the file *den_locations*. Next, click the *Advanced Geometry Options* button (**Figure 3.24**).

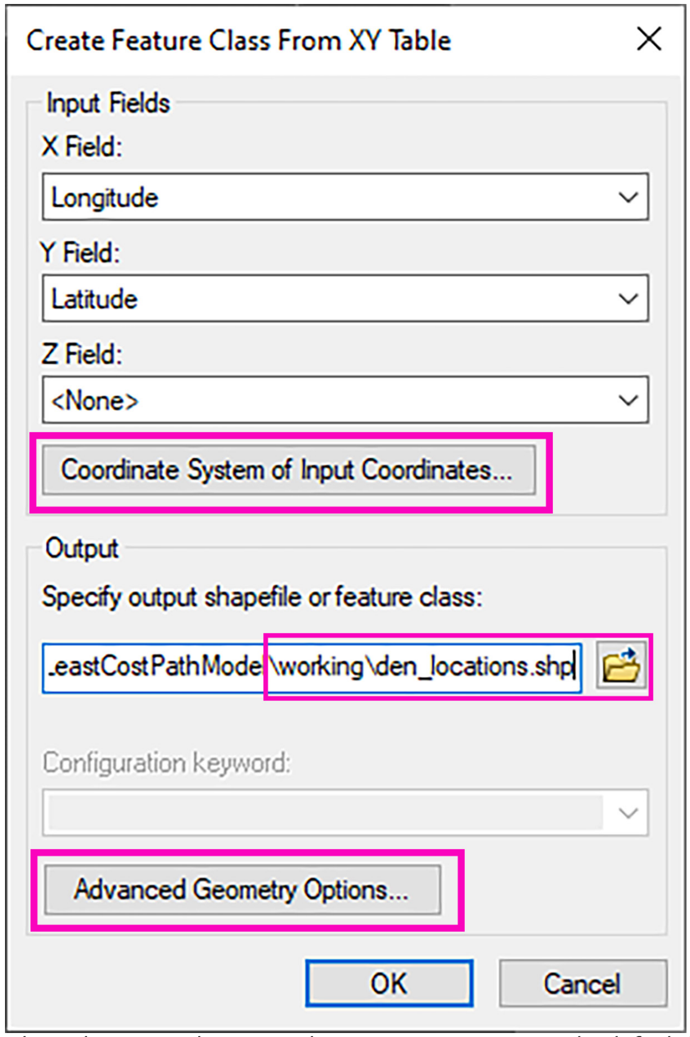


Figure 3.24: Always browse to the output destination. Never accept the default file location.

When the *Advanced Geometry Options* window appears, click the radio button next to *Use the map's spatial reference* (**Figure 2.31**). This option has the effect of transforming the output shapefile so that it uses **NAD 83 UTM Zone 10** as its spatial reference (**Figure 3.25**).

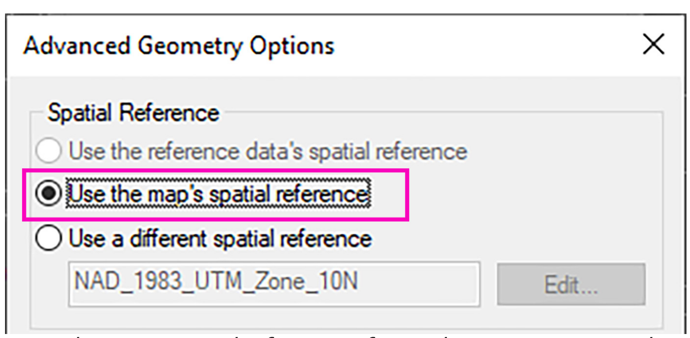


Figure 3.25: The map's spatial reference refers to the *Data Frame* window, which currently uses **NAD 83 UTM Zone 10** as the display projection.

Add the new den_locations layer to the *Table of Contents* (**Figure 3.26**).

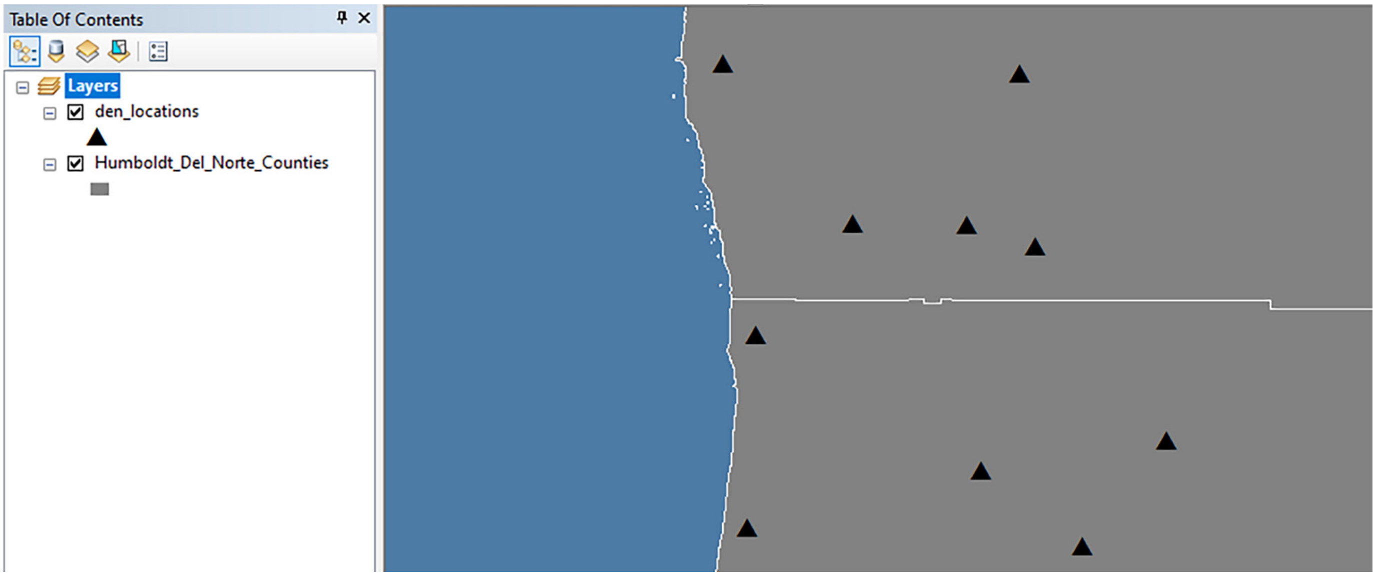


Figure 3.26: Only the business layer with the correct spatial reference should appear in the *Table of Contents*.

Skill Drill: Geocoding an Address and Creating a CSV Table to Import As XY Data

In this scenario, you will predict the overland movement of the wolpertinger during the fall migration from their den locations to the town of Orick. As a point of reference, the U.S. Post Office will serve as a point location representing the relative center of town. The US Post Office in Orick is located at the following address:

121147 US-101, Orick, CA 95555

Use **Google Maps** to geocode the address and obtain the latitude and longitude values for this address in decimal degrees. Google Maps uses **WGS 1984** as the datum for latitude and longitude values. In Microsoft Excel, create a CSV table with the following field headers: name, latitude, longitude. You may use *Orick* as the name attribute for the point feature (**Figure 3.27**). Enter the latitude and longitude values in the appropriate fields. *Save as* a CSV table in your *original* folder. Be sure to *close* Excel when you are done saving the file.

In ArcMap, you may need to refresh your original folder to see the CSV file.

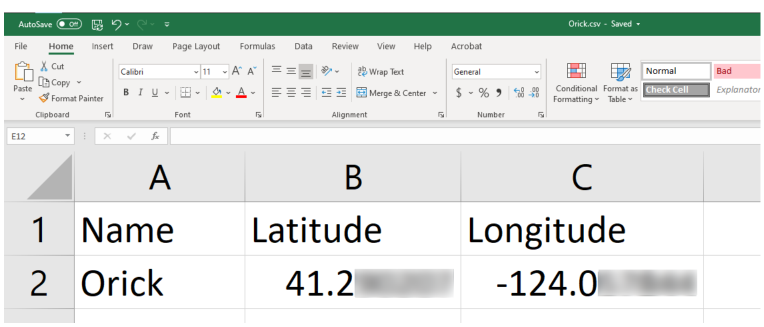


Figure 3.27: You will need to enter the latitude and longitude values acquired from Google Maps manually.

In the *Catalog Window*, expand the *original* folder and right-click on the Orick CSV file. Select *Create Feature Class*, then choose *From XY Table*. Follow the same workflow as for the businesses CSV to create a file that uses **NAD 83 UTM Zone 10**. Add the new file to your *Table of Contents* (**Figure 3.28**).

Figure 3.28: The town of Orick is indicated here as a yellow star.

Go to the ArcMap properties window and set the Pathnames to *Store relative pathnames to data sources* (**Figure 3.29**).

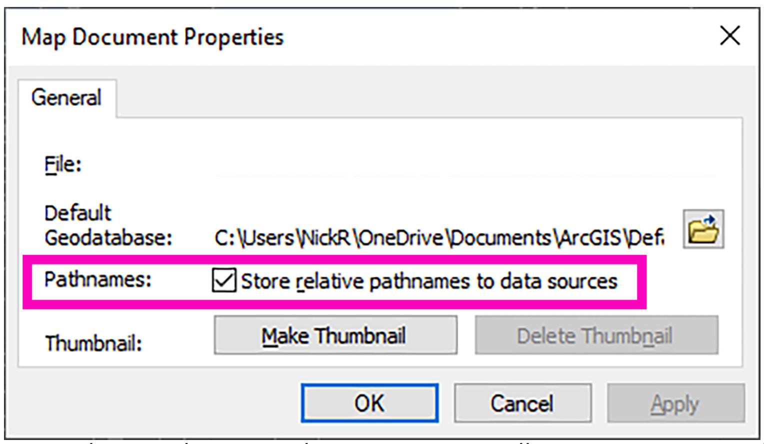


Figure 3.29: The store relative pathnames to data sources setting allows ArcMap to remember the location of the data sources relative to its current position. *This image has been modified for clarity.*

Take a moment to save your map document. Be sure to give the map document a meaningful name. Take care to save the map document in your workspace folder, *LeastCostPathModel* (**Figure 3.30**).

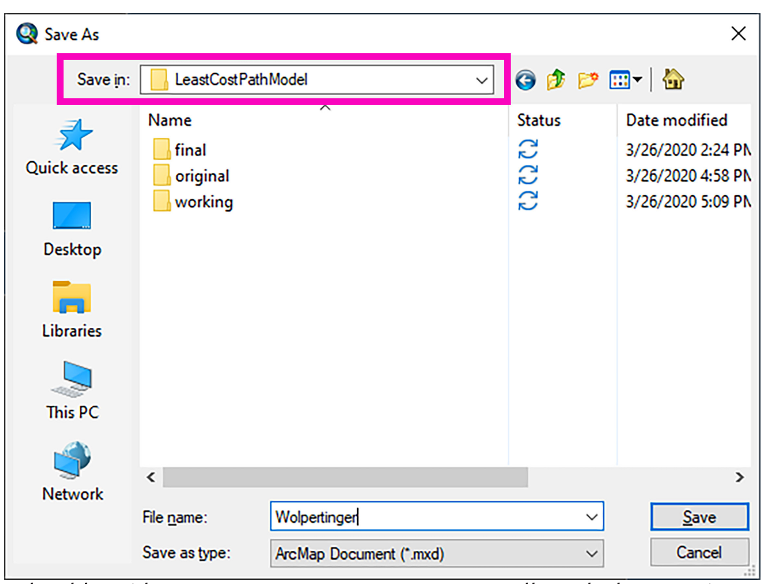


Figure 3.30: Generally, you should avoid spaces. However, empty spaces are allowed when naming map document files (.mxd)

Defining the Study Area

Raster processing can place significant demands on a computer's processing power and memory. To save time during the analysis, you will limit raster processing to the areas around the den locations and the town of Orick. In ArcMap, activate the *Draw* toolbar (**Figure 3.31**).

Figure 3.31: The Draw toolbar has many tools from which to choose. In this image, the *Rectangle* tool is highlighted.

Use the *Rectangle* tool to draw a rectangle around the den locations and the town of Orick (**Figure 3.32**). On the Draw toolbar, use the drop-down menu under the word *Drawing* and choose *Convert Graphics to Features*.

Note: Only the selected graphic gets converted. Be sure that the rectangle is selected.

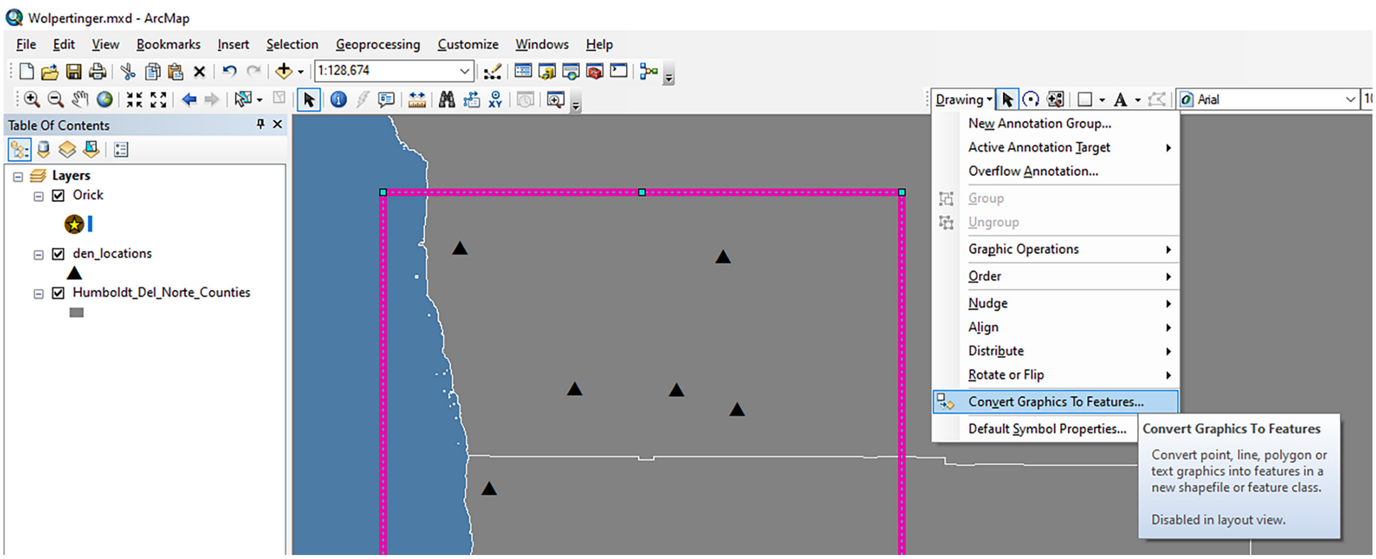


Figure 3.32: One may use the option to convert graphics to features, such as shapefiles and geodatabase feature classes

On the *Convert Graphics to Features* window, find the radio buttons under *Use the same coordinate system as*, and check the radio button next to *the data frame* (**Figure 3.33**). Be sure to save the output to your *working* folder. Call the file *study_area*. Check the box next to *Automatically delete graphics after conversion*. When ready, *click OK*. Add the new layer to the map.

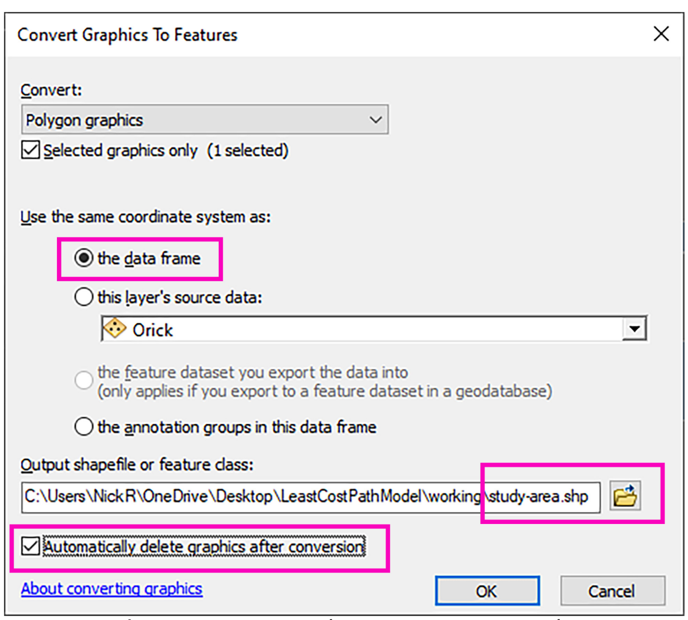


Figure 3.33: Be sure that your settings match.

Skill Drill: Acquire Elevation Data from the USGS National Map Viewer

To complete the least-cost path analysis, you will need an elevation layer. Navigate to the ***USGS National Map Viewer***[1]. On the right, zoom in and center the map to the approximate region of interest. On the left, check the box next to *Elevation Products*. Check the box next to *1/3 arc-second DEM*. Then click, *Find Products* (**Figure 3.34**).

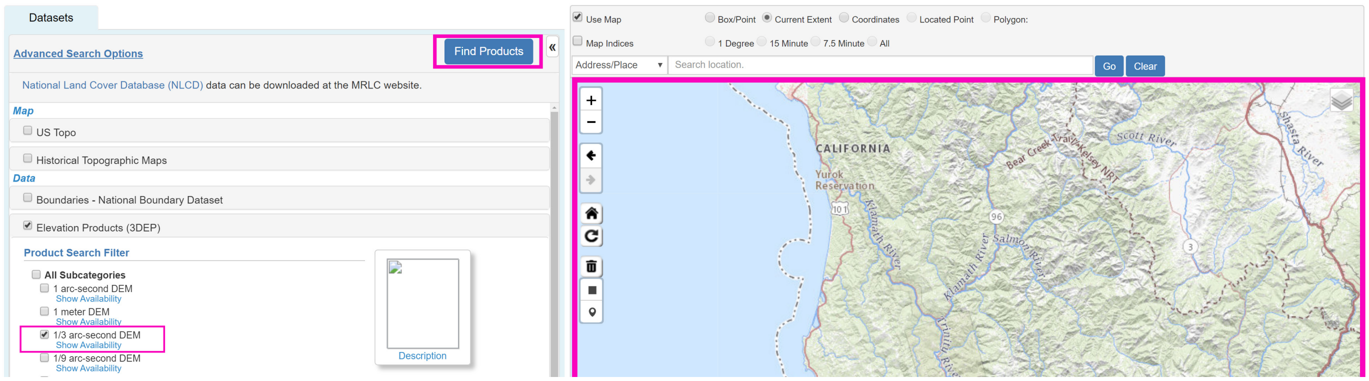
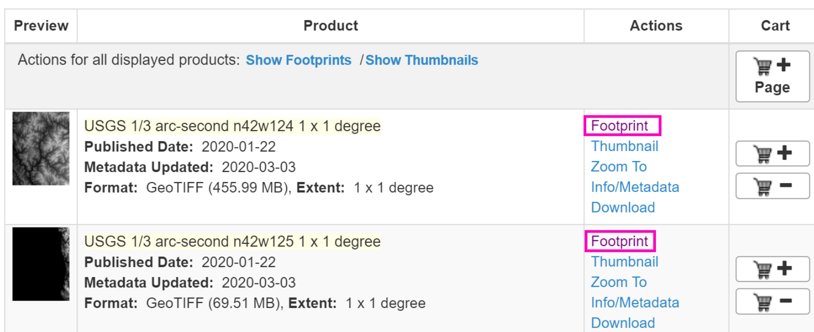


Figure 3.34: The website may change over time. Your screen may not match exactly.

Use the *Footprint* link to determine which digital elevation models cover the study area (**Figure 3.35**).



Preview	Product	Actions	Cart
	Actions for all displayed products: Show Footprints / Show Thumbnails		Page
	USGS 1/3 arc-second n42w124 1 x 1 degree **Published Date:** 2020-01-22 **Metadata Updated:** 2020-03-03 **Format:** GeoTIFF (455.99 MB), **Extent:** 1 x 1 degree	Footprint Thumbnail Zoom To Info/Metadata Download	+ / −
	USGS 1/3 arc-second n42w125 1 x 1 degree **Published Date:** 2020-01-22 **Metadata Updated:** 2020-03-03 **Format:** GeoTIFF (69.51 MB), **Extent:** 1 x 1 degree	Footprint Thumbnail Zoom To Info/Metadata Download	+ / −

Figure 3.35: Toggle the footprints on and off to check the extent of the digital elevation model.

When one clicks the *Footprint* link, blue bounding boxes appear on the map indicating the extent of the digital elevation model (**Figure 3.36**).

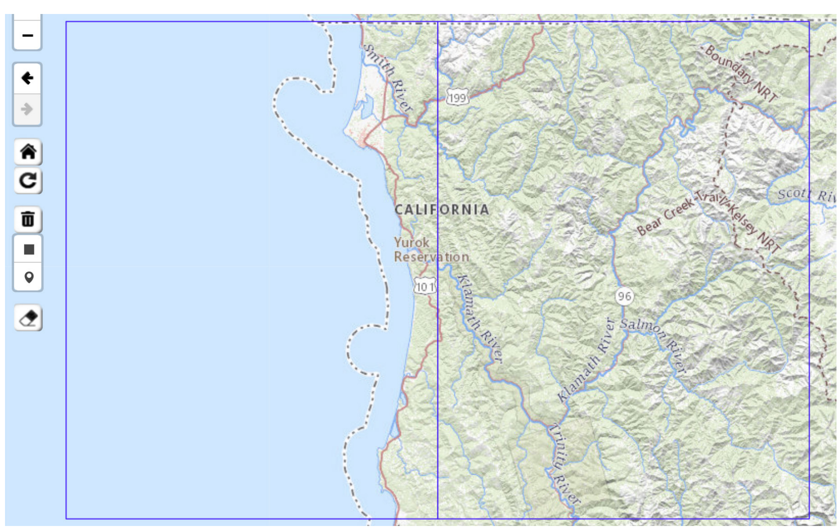


Figure 3.36: Currently, two digital elevation models cover the study area, each located side by side.

1 URL: https://viewer.nationalmap.gov/basic/

Once you locate the two digital elevation models for the study area, click the *Download* link for each. Save the files to your *original* folder. To continue, you will need to combine these two digital elevation models into one raster dataset. You can do this by using the *Mosaic to New* Raster tool, located in the *Raster* Dataset toolbox. The *Mosaic to New Raster* tool has settings that require particular attention. For the Input Rasters, use the yellow file folder icon to browse to your *original* folder. Add the two USGS TIF files. The *Output Location* is looking for a **folder**, not a file name (**Figure 3.37**). Readers often get confused with this setting. Enter the name *mosaicDEM.tif* under *Raster Dataset Name with Extension*.

Under *Spatial Reference for Raster*, choose **NAD 83 UTM Zone 10** N. The *Pixel Type* settings is also very important. By default, the output of *this* tool is an 8 BIT pixel depth. If the original datasets have a more substantial pixel depth, you will lose quite a bit of information. Always make sure the *Pixel Type* setting matches the original data. In this instance, you want to change the *Pixel Type* to **32 Bit Floating Point**. A digital elevation model has only **one** band, like most monochromatic raster datasets. Enter the number **1** under the *Number of Bands*. Next, click the *Environments* button.

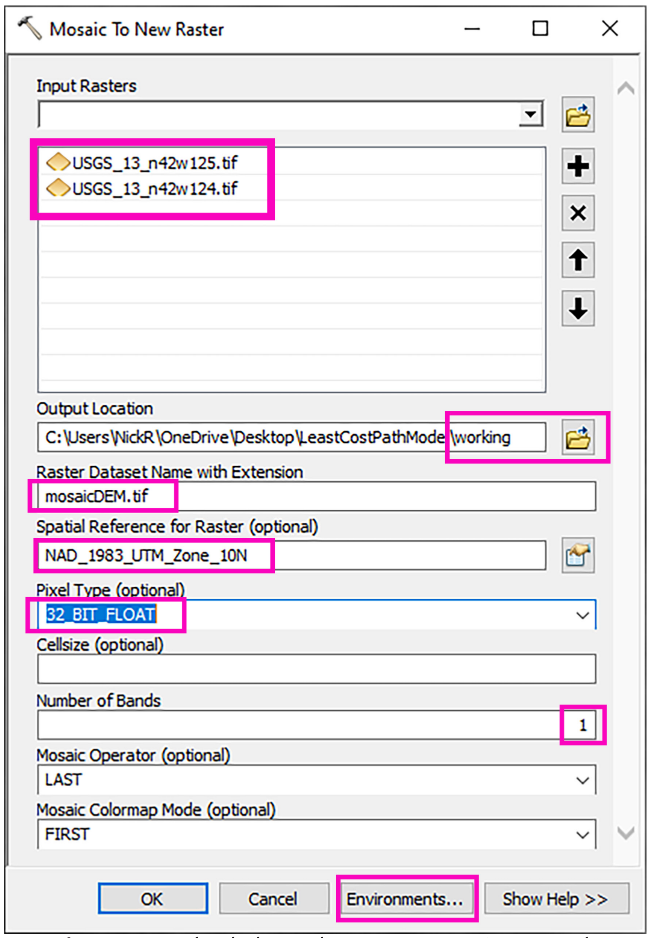


Figure 3.37: Check the make sure your settings match

On the Environment Settings window, expand the *Processing Extent* options (**Figure 3.38**). Use the drop-down menu and select *Same as layer study area*.

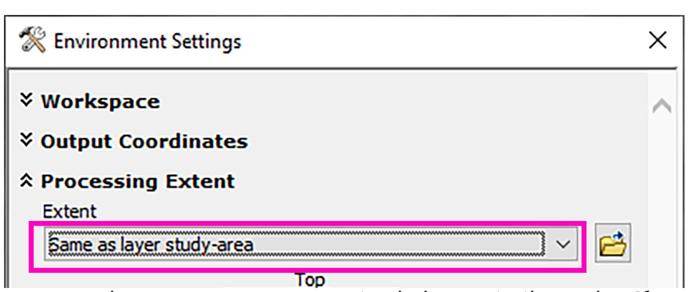


Figure 3.38: The *Processing Extent* option behaves similar to the *Clip* tool.

You may leave all other settings as default. Once you are ready, *click OK*. The result will be a single, seamless digital elevation model, clipped to the study area (**Figure 3.41**)

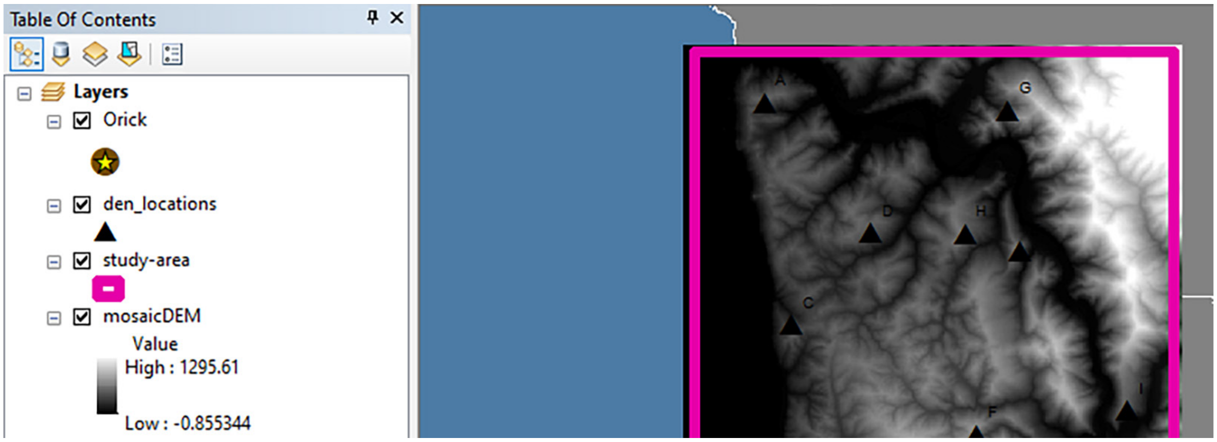


Figure 3.41: The mosaicDEM layer fits within the study area, reducing your computer's processing requirements.

The digital elevation model needs one more step. The Pacific Ocean is not part of the study area. It can be excluded from the digital elevation model using the *Extract by Mask* tool.

If you have not done so already, be sure to activate the Spatial Analyst Extension. Otherwise, the Extract by Mask tool will not be available to you.

For the *Input raster*, choose the *mosaicDEM* layer. For the *Input raster or feature mask data*, choose the layer representing the Humboldt and Del Nort Counties. Use the yellow file folder icon and browse to your *working* folder. Save the file as *DEM.tif*. When ready, *click OK*.

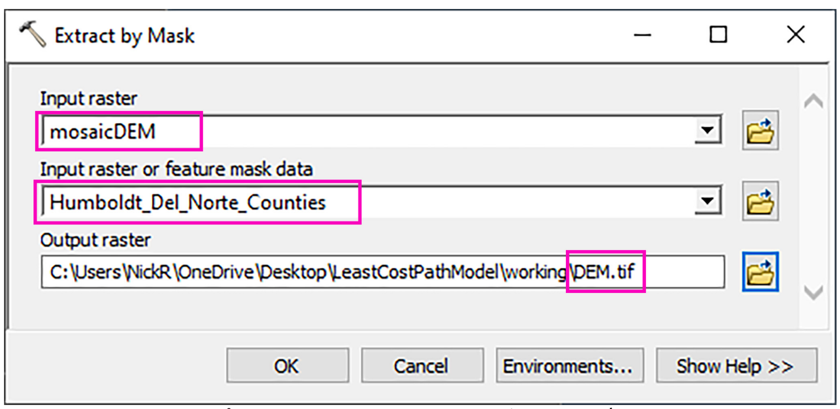


Figure 3.40: Be sure your settings match.

The result is a digital elevation model that excludes the Pacific Ocean.

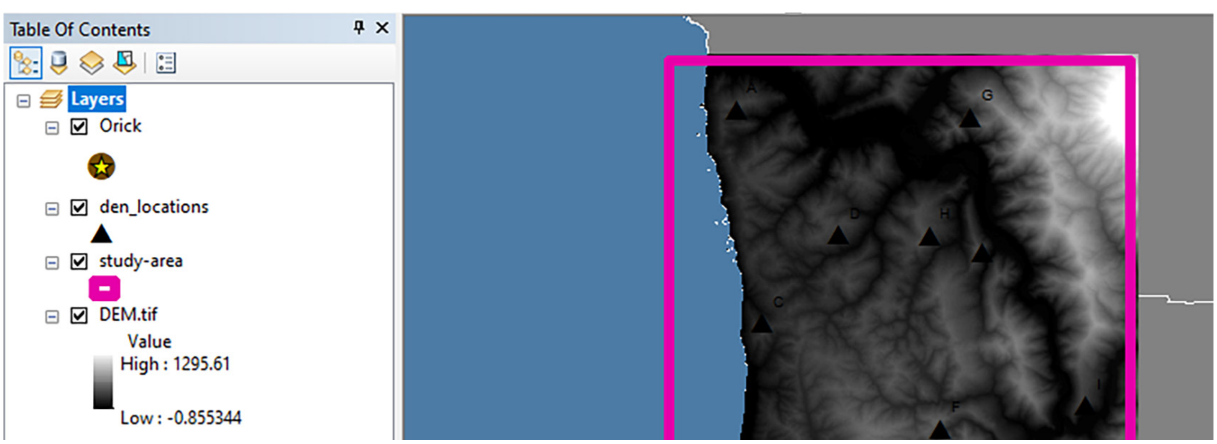


Figure 3.41: *Extract by mask* is useful for converting raster datasets into complex shapes.

Changing Global Environment Settings for Raster Processing

When using the *Mosaic to new raster* tool, you explored some of the options in the *Environment* settings. These settings applied only to that instance. After running the *Mosaic to new raster* tool, those *Environment* settings disappear. However, it is possible to create **global** *Environment* settings for the map document that applies to nearly all tools and geoprocessing. The benefit is that one can enter the settings once instead of for each geoprocessing function.

In ArcMap, select Geoprocessing from the main menu, then select Environments (**Figure 3.42**).

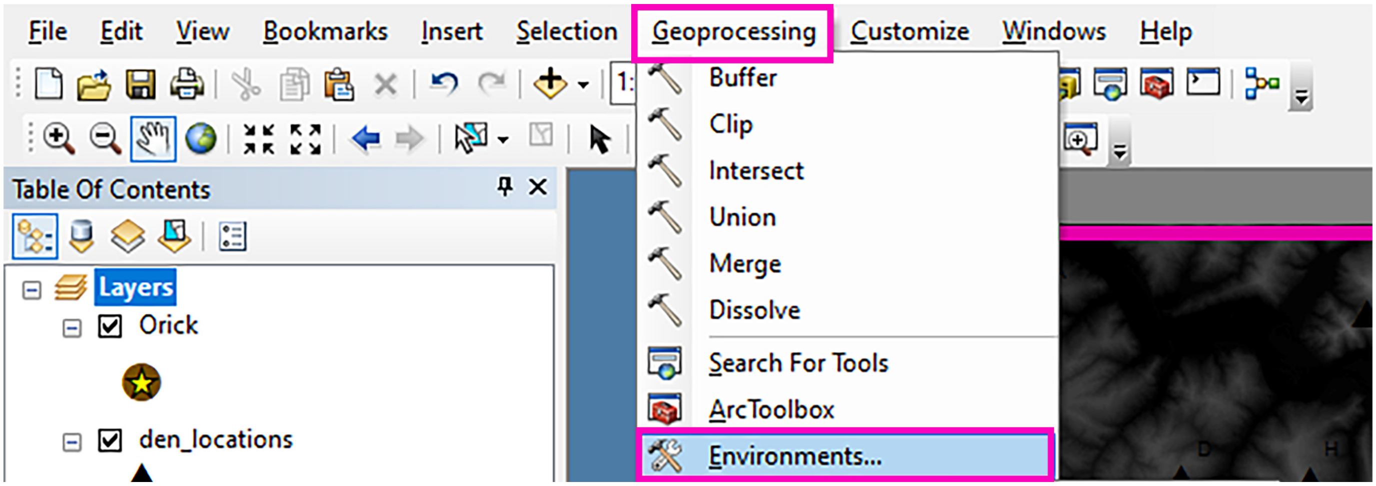


Figure 3.42: Locate the Environments options under the Geoprocessing menu.

This step will open up the *Environment Settings* window. Here you will define the *Processing Extent* **globally**. Any changes you make here will affect the processing extent for *nearly* all the tools you will use later. Under *Processing Extent*, use the drop-down menu to select *Same as layer study-area* (**Figure 3.43**). Under *Snap Raster*, select your DEM layer.

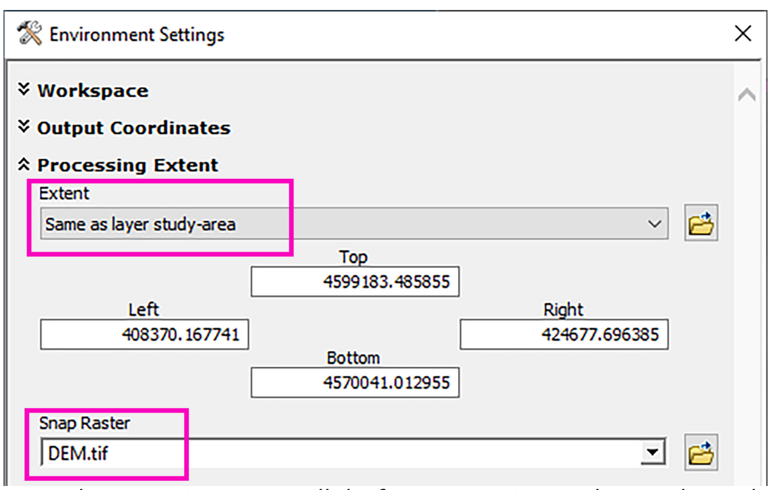


Figure 3.43: The processing extent will clip future operations to the DEM layer. The Snap Raster will align the cells in a raster dataset output to the DEM layer as well.

Under Raster Analysis, set the *Cell Size* setting to *Minimum of Inputs* (**Figure 3.44**). By default, the *coarsest* resolution of any input layers is used as the output for raster processing. By changing this setting, the *smallest* cell size of any input rasters will be used to define the output.

For example, suppose the DEM layer has a cell size of 10 meters, and the tree canopy layer has a cell size of 30 meters. If you multiply these two raster layers together using the *Raster Calculator* tool, the default output would have a 30-meter resolution. After changing the *Cell Size* setting to *Minimum of Inputs*, the output would use the 10-meter resolution of the DEM layer instead of the 30-meter resolution of the tree canopy. Also, since you changed the processing extent and the Snap Raster to match the DEM, the output will match the size and cell alignment of the DEM layer.

Under the *Mask* option, use the drop-down menu to choose the DEM layer (**Figure 3.44**). This setting should help to maintain the California shoreline for any future raster datasets created. When ready, *click OK*.

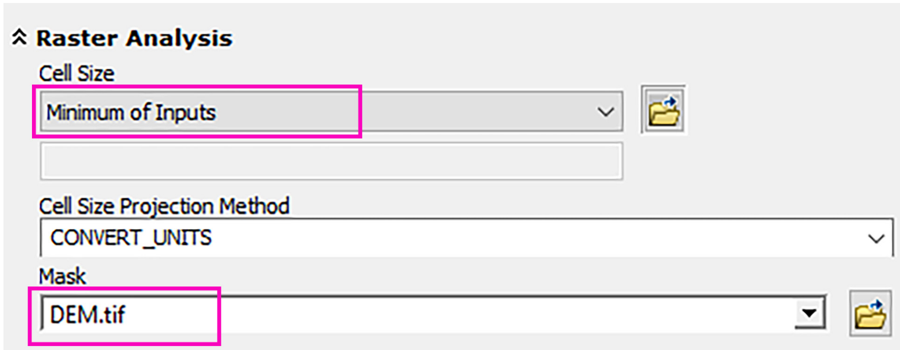


Figure 3.44: Be sure your settings match.

With these global environment settings, the raster processing you perform in later steps will be constrained to the study area. Any output raster datasets will maintain the finest resolution and have cells that align with the DEM layer. Additionally, any output raster datasets will have a value of **NoData** for regions that fall in the Pacific Ocean. Take a moment to save your map document before continuing on to the next step.

Skill Drill: Acquire Hydrography Data from the USGS National Map Viewer

To complete the least-cost path analysis, you will also need a layer representing waterbodies. Return to the ***USGS National Map Viewer***[1]. Locate the *NHD Plus High-Resolution* dataset under the *Datasets* tab. For the Data Extent, choose HU-4. Then, click on the map to select Subregion 108 (**Figure 3.45**). When ready, click on the button that says *Find Products*.

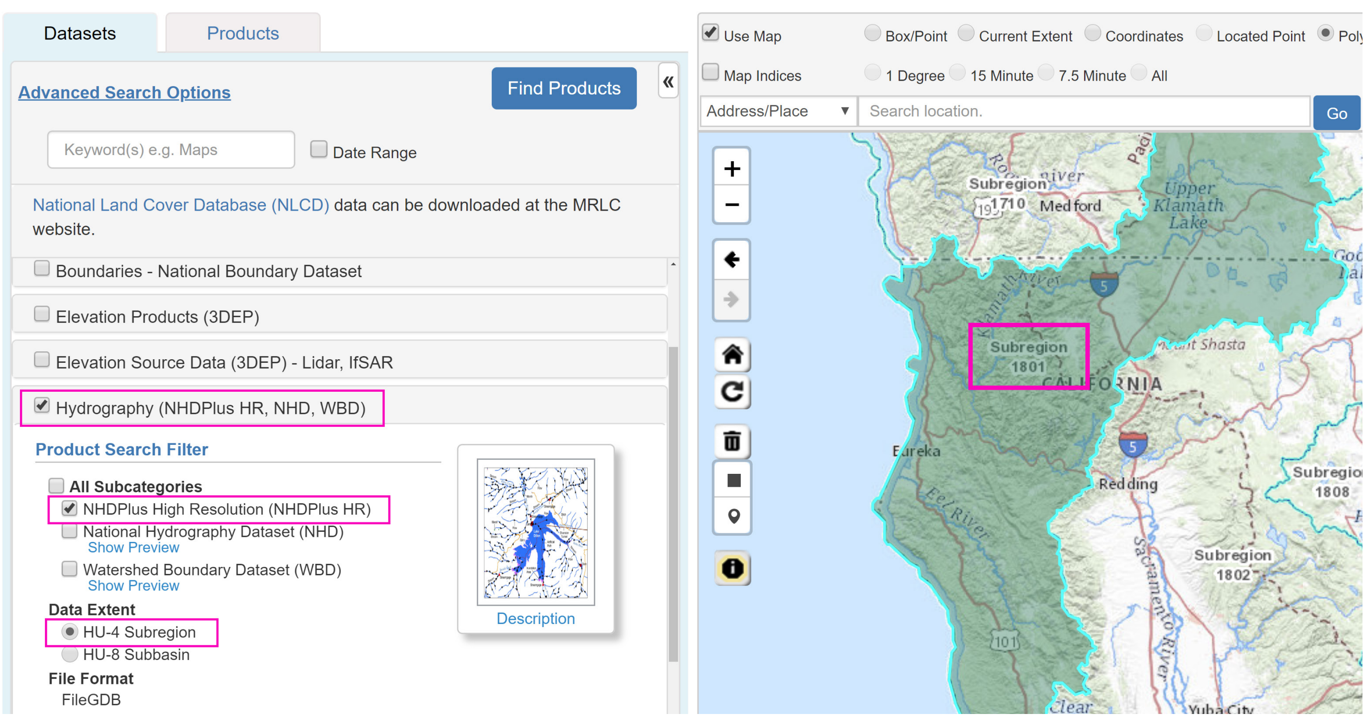


Figure 3.45: Subregion 1801 includes the Klamath Basin Watershed and the Eel River Watershed.

1 URL: https://viewer.nationalmap.gov/basic/

On the Find Products page, use the link that says, *Download 2 Raster*. Save the file to your *original* folder. Use ***7zip***[1] to decompress the 7z file. Delete the 7z file when done.

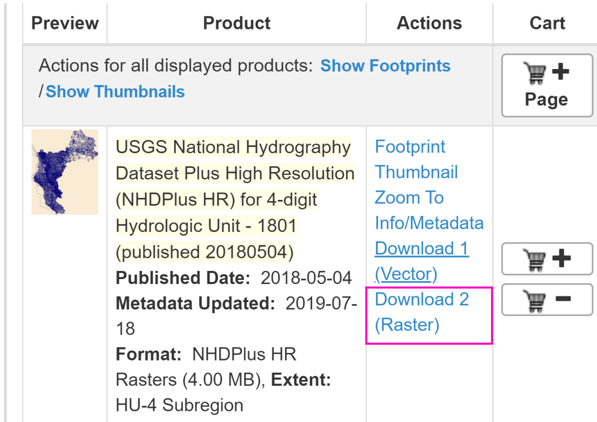


Figure 3.46: The raster dataset is extensive and may take several minutes to download.

In ArcMap, you may need to refresh your original folder to see the TIF files.

Before adding the data to ArcMap, use the *Project Raster* tool to create a new raster dataset that uses the **NAD 83 UTM Zone 10**N spatial reference system (**Figure 3.47**). For the Input Raster, browse to your *original* folder and choose the file named *hydrodem.tif*.

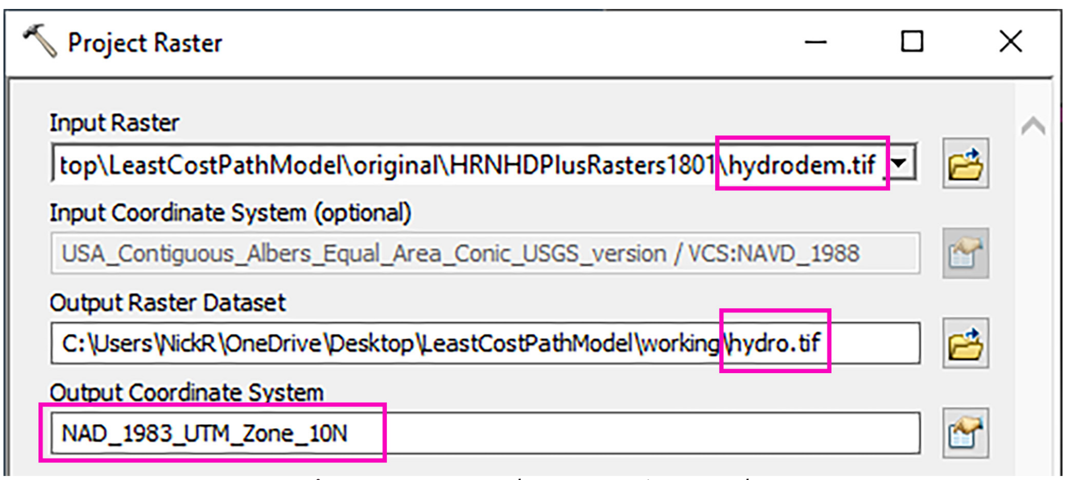


Figure 3.47: Be sure that your settings match.

For the *Output Raster Dataset*, browse to your *working* folder. Name the file, *hydro.tif*. When done, add the new raster dataset to the *Table of Contents* (**Figure 3.48**).

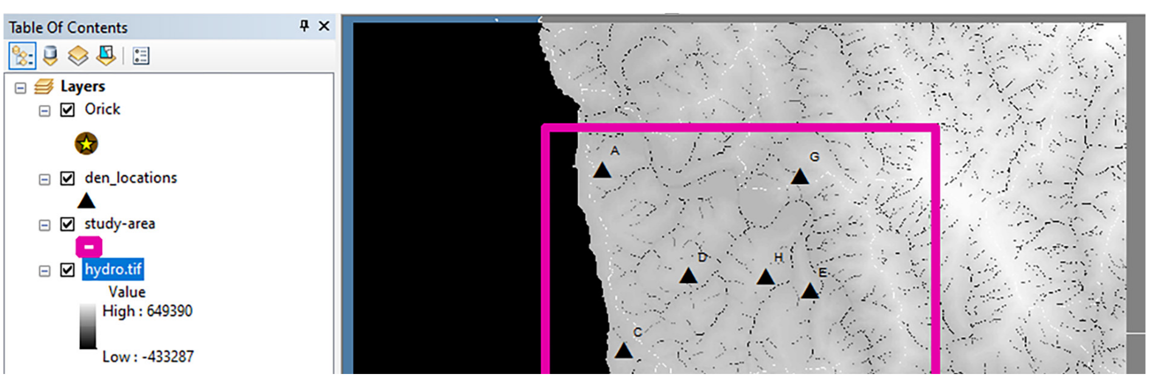


Figure 3.48: Though the results were not perfect, the global *Environment Settings* limited the spatial extent of the output raster dataset to just beyond the study area.

1 URL: https://www.7-zip.org/download.html

Skill Drill: Acquire Land Cover Data From the MRLC Consortium

In this step, you download tree canopy data from the ***Multi-Resolution Land Characteristics (MRLC)***[1] consortium. THE MRLC consists of several federal agencies that provide land cover information at the national scale. Readers may use this public domain data for free and without restrictions, which makes this website an excellent resource. Navigate to the MRLC website and click the button that says *Tree Canopy* (**Figure 3.50**).

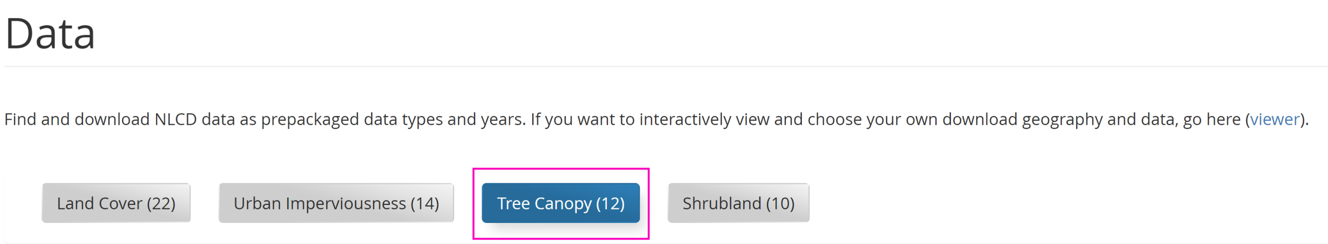


Figure 3.49: The MRLC offers many datasets for a wide variety of environmental, land management, and modeling applications.

On the Tree Canopy page, download the tree canopy cover for the continental United States (CONUS) (**Figure 3.50**). Save the file to your *original* folder and decompress it. Delete the original zip file when done.

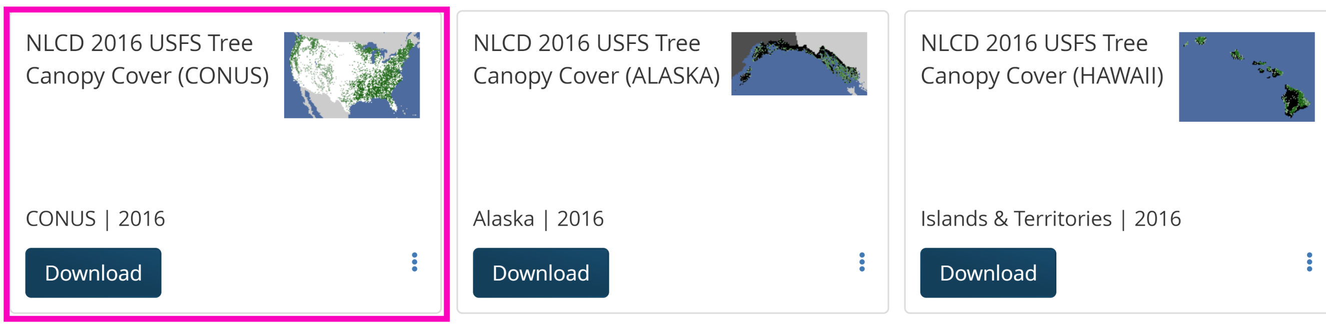


Figure 3.50: NLCD tree canopy cover is a 30 m raster dataset.

If, for some reason, the website is down or the data is no longer available,
you can download a backup copy of the data for this tutorial using the link provided in Appendix A.

Before adding the data to ArcMap, use the *Project Raster* tool to create a new raster dataset that uses the **NAD 83 UTM Zone 10**N spatial reference system (**Figure 3.51**).

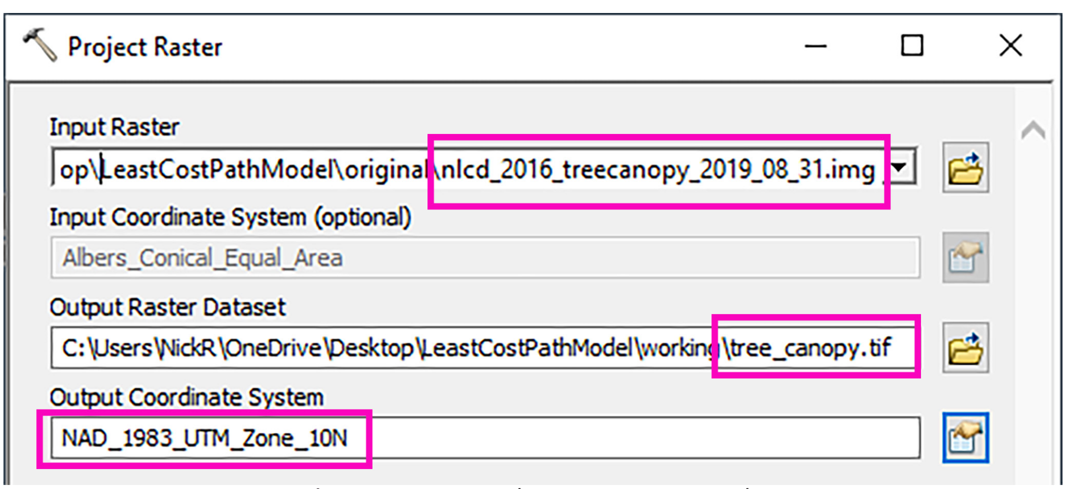


Figure 3.51: Be sure that your settings match.

1 URL: https://www.mrlc.gov/data

Save the output to your *working* folder. Call the file, *tree_canopy.tif*. When done, add the new raster dataset to the *Table of Contents* (**Figure 3.52**).

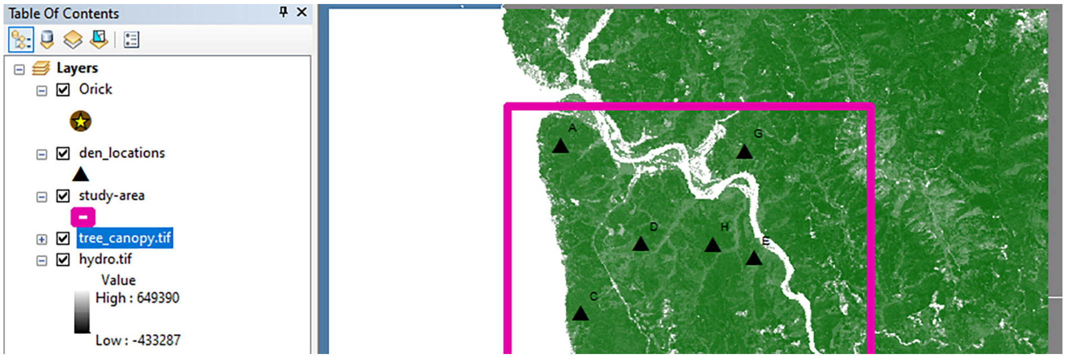


Figure 3.52:

Creating Cost Surface Models Using a Relative Cost Scale

A **cost surface model** represents some factor or a combination of factors that affect travel across an area. The goal is often to find the path with the *least* cost. In this scenario, the following cost factors have been *identified*:

» Elevation
» Slope
» Tree canopy density
» Hydrology

Since there are multiple cost factors in this analysis, your goal is to create a *total cost surface model*. A **total cost surface model** is used when you have **multiple cost factors** that you wish to combine into a **single cost surface model**. When you need to consider multiple cost factors, uniform **cost units** must be used. The values of the cost units must be converted into a **relative cost scale**, which is a universal scale that represents the costs between different cost factors.

A relative cost scale could be as simple as a scale of 1 through 10. The number 1 might represent a low cost or low high likelihood of travel. The number 10 might represent a high cost or even a prohibitive factor. For the results to be meaningful, the relative scale must be the same among the different cost factors. Each cost category must be given the same relative scale, where the numbers roughly represent the same level of difficulty or likelihood of travel.

In this example, you want to create a total cost surface to use for modeling the wolpertinger migration from den locations to the town of Orick. The problem is that each of the cost factors currently uses different units. Elevation uses meters, slope uses degrees, and tree canopy uses density, etc. To solve this, you will reclassify each of these layers to create a relative cost scale. Though the relative scale **has no units**, the values between the cost factors are meaningful when compared to each other. Typically, developing the relative scale between cost factors would involve extensive research. In this scenario, we will assume thorough research was conducted by the HSU students and their professors (**Figure 3.53**).

Relative Cost	Elevation	Slope	Tree Canopy	Water Features
1	>1000	>35	>90	No water
2	>800 to 1000	>30 to 35	>80 to 90	—
3	>600 to 800	—	>60 to 80	—
4	>400 to 600	>25 to 30	—	—
5	>200 to 400	—	>40 to 60	—
6	>100 to 200	>20 to 25	—	—
7	>1 to 100	—	>20 to 40	—
8	—	>10 to 20	—	—
9	—	>5 to 10	—	—
10	>-1 to 1	0 to 5	0 to 20	Water

Figure 3.53: This table defines the relative cost for each cost factor.

Creating a Remap Table to Reclassify Elevation

To assign a relative scale to a cost factor, you must use the Reclassify tool. The first step will be to reclassify the DEM layer. After careful study, the researchers evaluated which elevations the wolpertinger likes best. They believe the wolpertinger is most likely to travel through areas of high elevation. When using the *Reclassify* tool in the past, you may have entered the break values manually every time. For most purposes, manually entering break values works fine. However, there may be situations when you want to run a model multiple times and would like the ability to quickly and accurately change parameters. In this instance, a remap table is useful. A **remap table** is a specially formatted table used by the *Reclassify* tool to define how the data will be reclassified. The remap table must contain the fields **FROM**, **TO**, **OUT**, and **MAPPING**.

The FROM field defines the lower limit of a data range. In this instance, the lowest value in the digital elevation model is **-0.855344** (**Figure 3.54**).

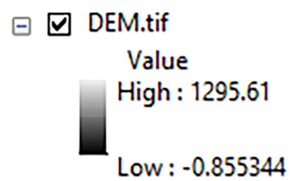


Figure 3.54: The DEM layer has a low value of -0.855344 and a high value of 1295.61

For the first data range in the remap table, a **-1** has been entered in the FROM field to ensure that each of the lowest pixel values is captured by the remap table (**Figure 3.55**). The TO field defines the upper limit of a data range. In this instance, the highest value in the digital elevation model is **1295.61**. By entering **1296** in the last row of the TO field, you ensure that each of the highest pixel values is captured by the remap table.

	A	B	C	D
1	FROM	TO	OUT	MAPPING
2	-1	1	10	ValueToValue
3	1	100	7	ValueToValue
4	100	200	6	ValueToValue
5	200	400	5	ValueToValue
6	400	600	4	ValueToValue
7	600	800	3	ValueToValue
8	800	1000	2	ValueToValue
9	1000	1296	1	ValueToValue

Figure 3.55: Be sure that your CSV table matches *precisely*, including upper and lowercase letters.

The OUT field defines the new values assigned to the data by the Reclassify tool. For example, the first row in **Figure 3.55** ranges FROM -1 TO 1. Since **10** is entered under the OUT field, any pixel within this range will have an output value of 10 assigned by the Reclassify tool.

Unfortunately, there is not much documentation about The MAPPING field or its specific purpose. However, it is **required** by the Reclassify tool. You must enter **ValueToValue** under this field for each data range. In Microsoft Excel, create the remap table shown in **Figure 3.55**. Save the table in your *working* folder as a comma-separated value (CSV) file. When done, be sure to close Microsoft Excel.

If you have not done so already, be sure to activate the Spatial Analyst Extention. Otherwise, the Reclassify tool will not be available to you.

In ArcMap, open the Reclassify tool. Enter the DEM layer as your Input Raster. The *Reclass* field should automatically populate. Click the *Load* button to load the remap table you created. Once loaded, you should see the Old values and New values in the *Reclassify* tool match up with the values in your remap table (**Figure 3.56**). Save the *Output Raster* to your *working* folder. As an extra precaution, check the box next to *Change missing values to NoData*. This step will ensure that any values not covered by the remap table will be removed. When ready, *click OK*.

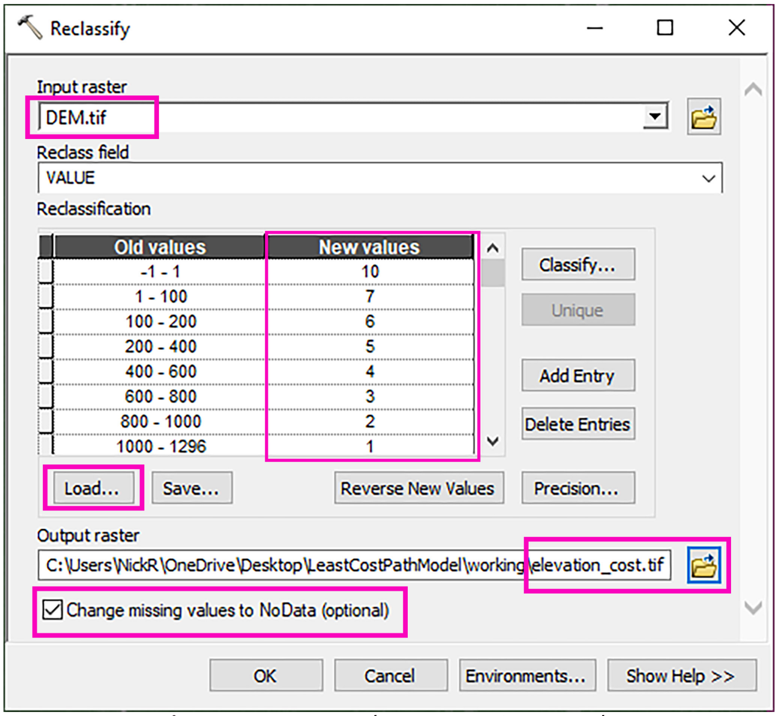


Figure 3.56: Be sure that your settings match.

The digital elevation model has now been reclassified using the remap table. Hopefully, you will have found using a remap table much easier than manually entering in the reclassification values. This layer now represents the relative cost or likelihood, on a scale of 1 through 10, that this species will traverse each pixel based on elevation (**Figure 3.57**).

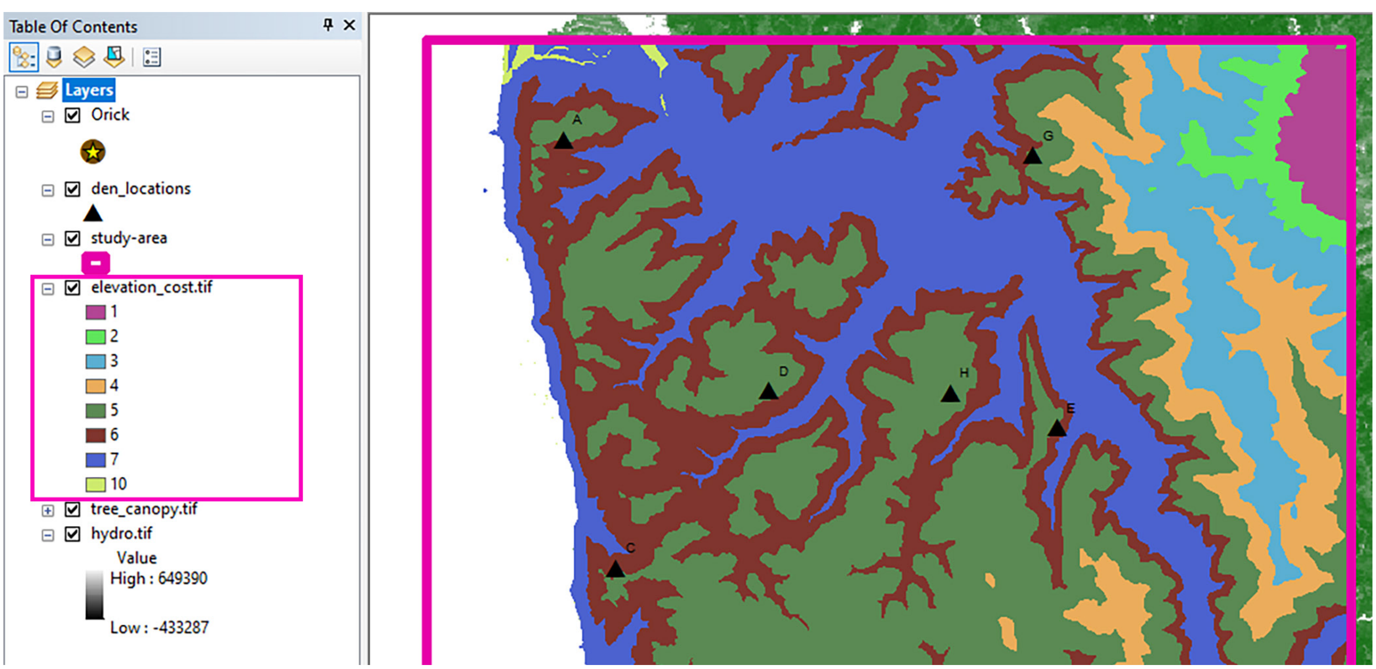


Figure 3.57: In this image, the ArcGIS software has assigned random colors to the pixel values between 1 and 10. Your results should look similar. Be sure there are **no** values beyond 1 through 10.

Skill Drill: Creating a Remap Table to Reclassify Slope

Your next cost raster will be based on the slope. The wolpertinger prefers to travel over steep slopes to avoid predators and to make use of a tendency to fly short distances. Start by creating a *slope raster* using the *Slope* tool located in the *Spatial Analyst* toolbox under Surface tools. For the input, use the DEM layer. Save the output to your *working* folder. Call the file *slope.tif*. Use *degrees* as the output measurement (**Figure 3.58**). When ready, *click OK*.

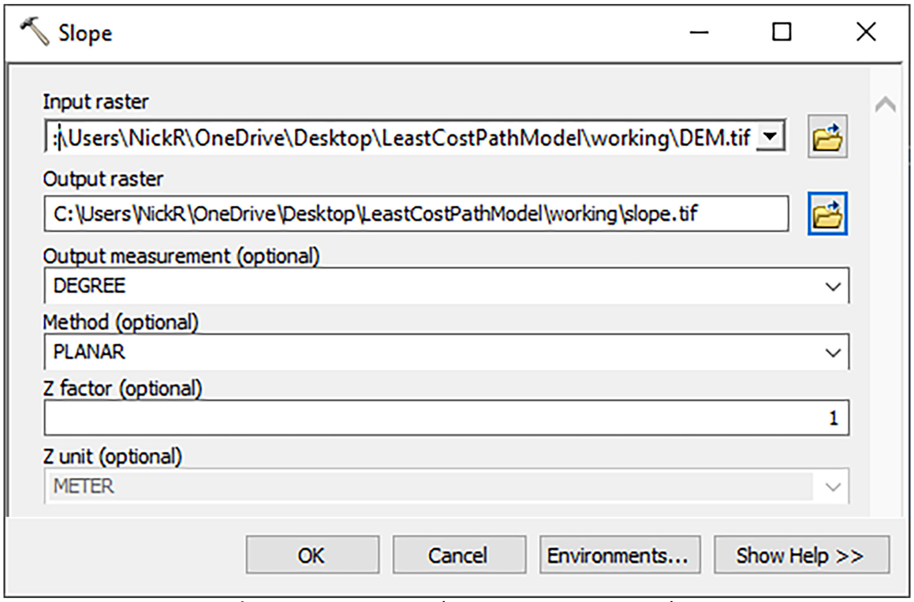


Figure 3.58: Be sure that your settings match.

The result is a raster dataset with pixel values that represent the degrees slope (**Figure 3.59**).

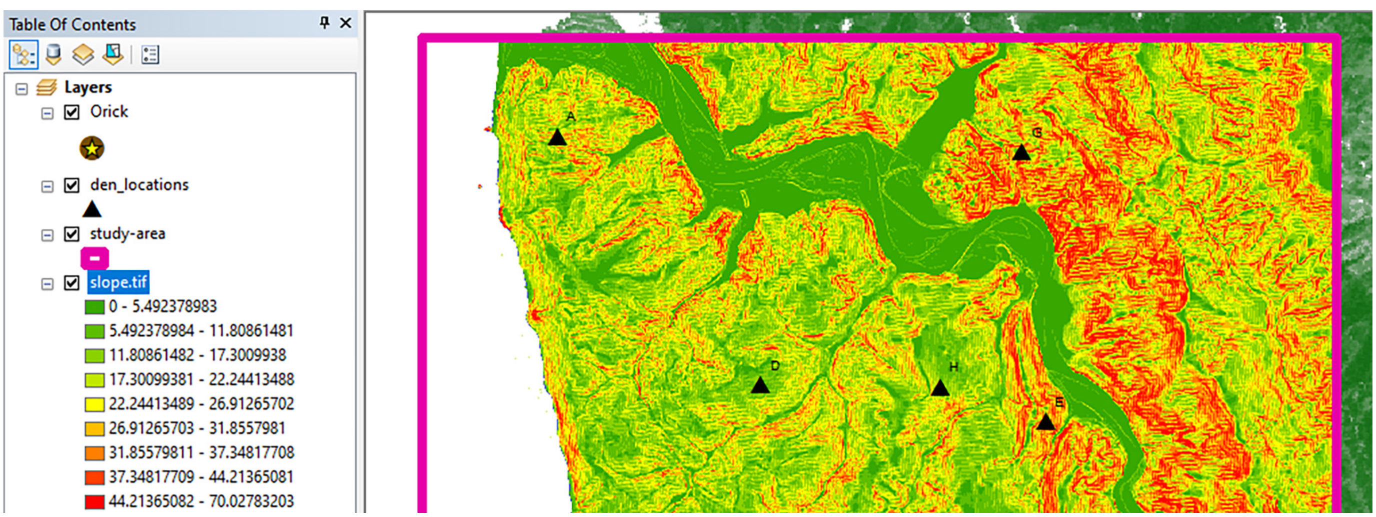


Figure 3.59: ArcMap may add the slope layer to the *Table of Contents* automatically.

Create a remap table with the following values to reclassify the slope layer (**Figure 3.60**). In Microsoft Excel, save the file to your *working* folder as a CSV table.

	A	B	C	D
1	FROM	TO	OUT	MAPPING
2	0	5	10	ValueToValue
3	5	10	9	ValueToValue
4	10	20	8	ValueToValue
5	20	25	6	ValueToValue
6	25	30	4	ValueToValue
7	30	35	2	ValueToValue
8	35	90	1	ValueToValue

Figure 3.60: Be sure that your CSV table matches *precisely*, including upper and lowercase letters.

Once the remap table has been created and saved as a CSV file, close Microsoft Excel. Use the skills learned previously to run the Reclassify tool. Load the slope remap table to obtain the new values (**Figure 3.61**). Save the output to your *working* folder. Name the file *slope_cost.tif*.

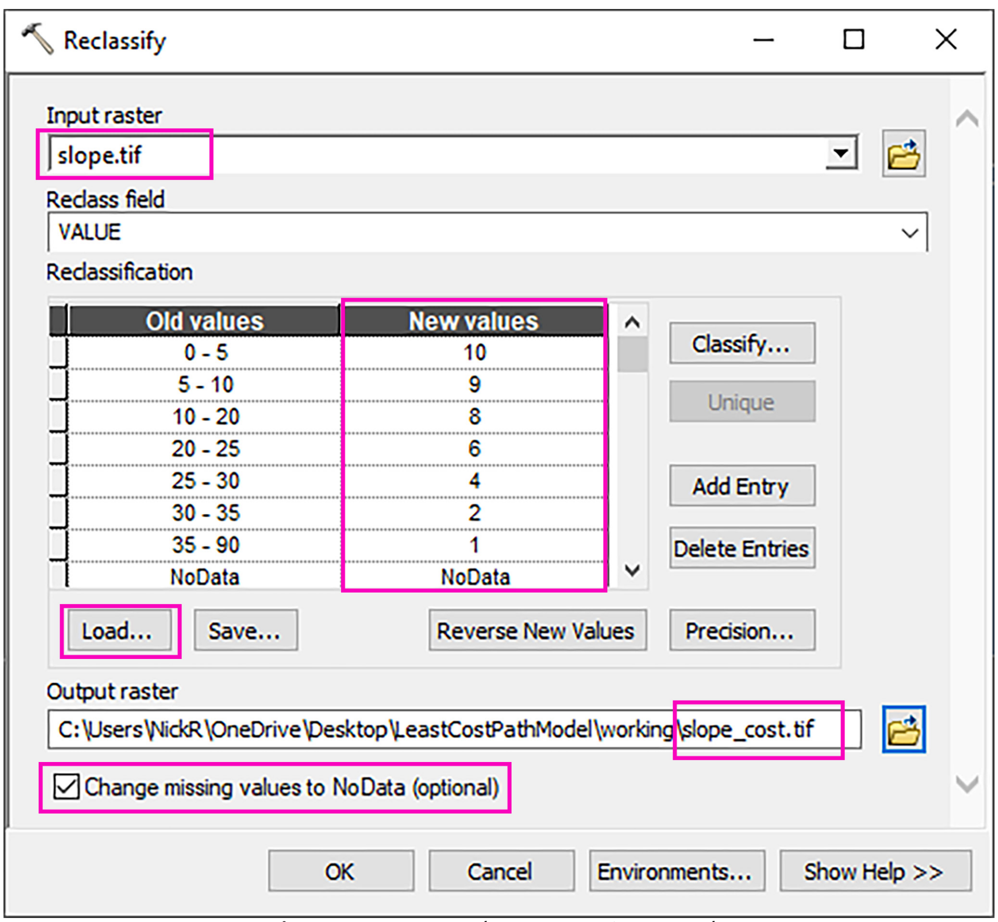


Figure 3.61: Be sure that your settings match.

The slope layer has now been reclassified using the remap table (**Figure 3.62**). This layer now represents the relative cost or likelihood, on a scale of 1 through 10, that this species will traverse each pixel based on the steepness of the slope.

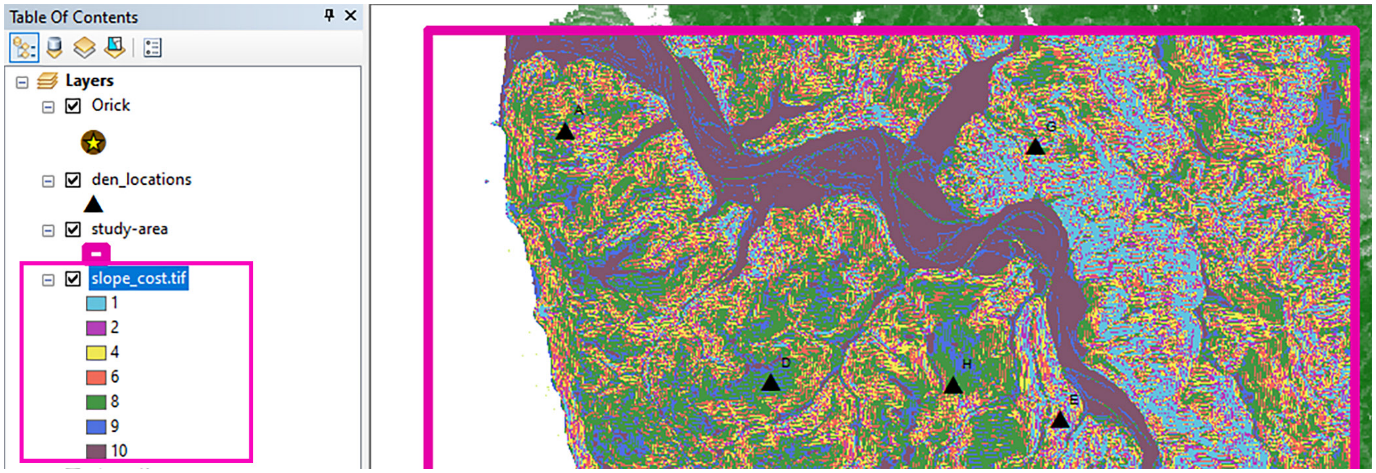


Figure 3.62: In this image, the ArcGIS software has assigned random colors to the pixel values between 1 and 10. Your results should look similar. Be sure there are **no** values beyond 1 through 10.

Take a moment to save your map document before continuing to the next step.

Skill Drill: Creating a Remap Table to Reclassify Tree Canopy Density

Your next cost raster will be based on tree canopy density. The wolpertinger prefers to travel under dense canopy. Use the following values to create a remap table for the tree canopy raster (**Figure 3.63**). Save the table to your *working* folder as a CSV file.

	A	B	C	D
1	FROM	TO	OUT	MAPPING
2	0	20	10	ValueToValue
3	20	40	7	ValueToValue
4	40	60	5	ValueToValue
5	60	80	3	ValueToValue
6	80	90	2	ValueToValue
7	90	100	1	ValueToValue

Figure 3.63: Be sure that your CSV table matches *precisely*, including upper and lowercase letters.

Using the skills you learned previously, use the *Reclassify* tool to create a layer called *canopy_cost.tif* (**Figure 3.64**).

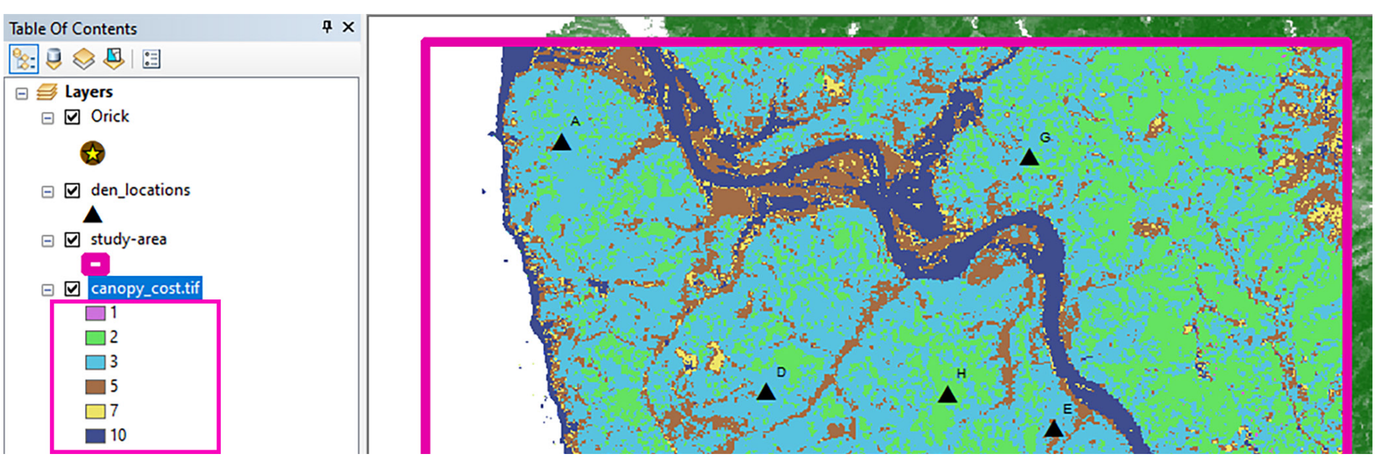


Figure 3.64: In this image, the ArcGIS software has assigned random colors to the pixel values between 1 and 10. Your results should look similar. Be sure there are **no** values beyond 1 through 10.

Converting the Hydrography Features to a Cost Surface Model

Your next cost raster will be based on hydrography. The wolpertinger prefers to avoid crossing water bodies, such as lakes and rivers. In this instance, a remap table is not necessary because the hydrography cost layer has only *two* values, **1** and **10**. You can enter these manually in the Reclassify tool.

In ArcMap, open the Reclassify tool. For the *Input Raster*, choose the *hydro.tif* layer. Then, click the button that says *Classify* (**Figure 3.65**).

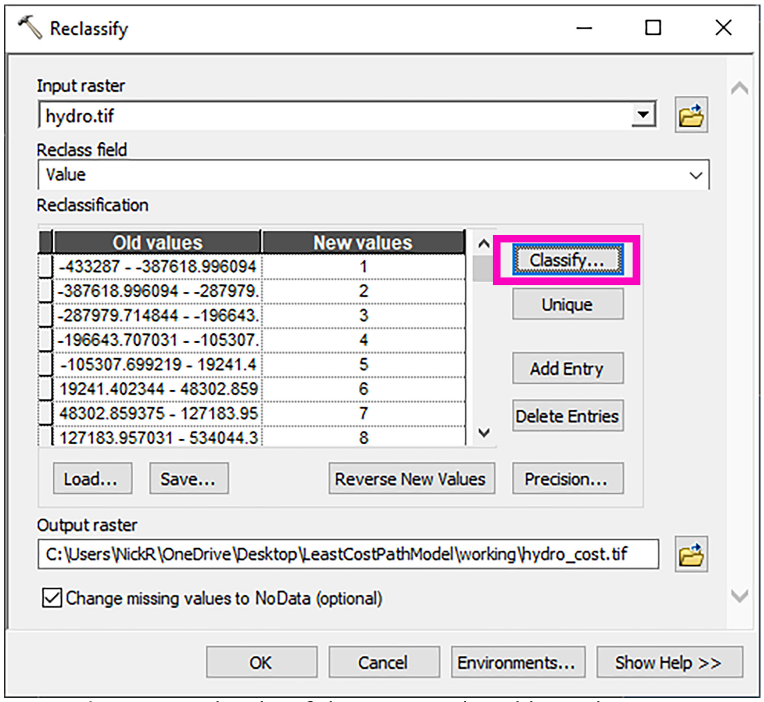


Figure 3.65: The Classify button provides additional options.

When the classification window opens, change the number of classes to **2** (**Figure 3.66**). For the first break value, enter **0**. Leave all other settings as they are and *click OK*.

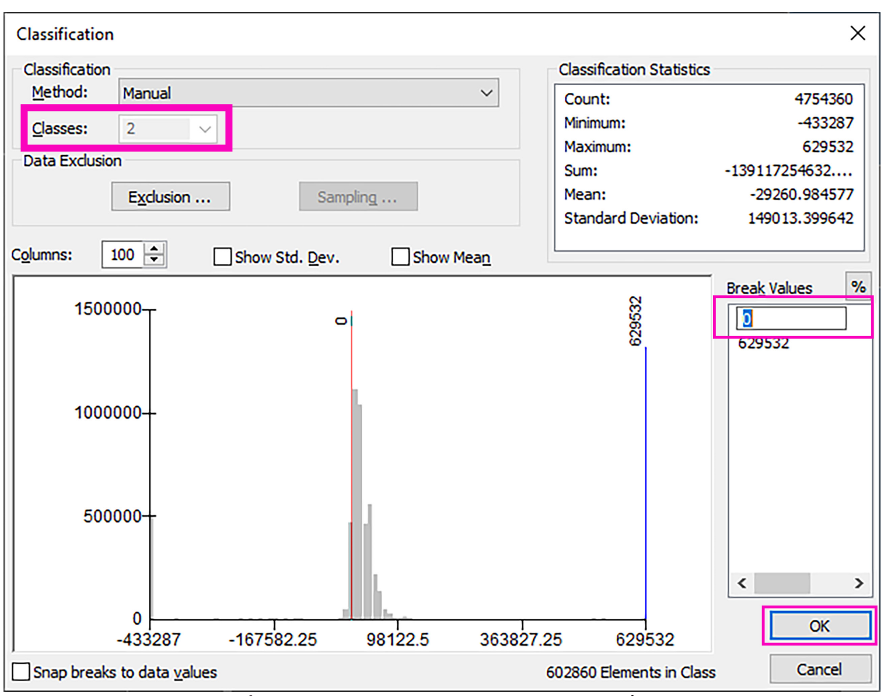


Figure 3.66: Be sure your settings match.

For the Output Raster, save the file to your *working* folder. Name the file, *hydro_cost.tif* (**Figure 3.67**). As an extra precaution, check the box next to *Change missing values to NoData*. This step will ensure that any values not covered by the remap table will be removed. When ready, *click OK*.

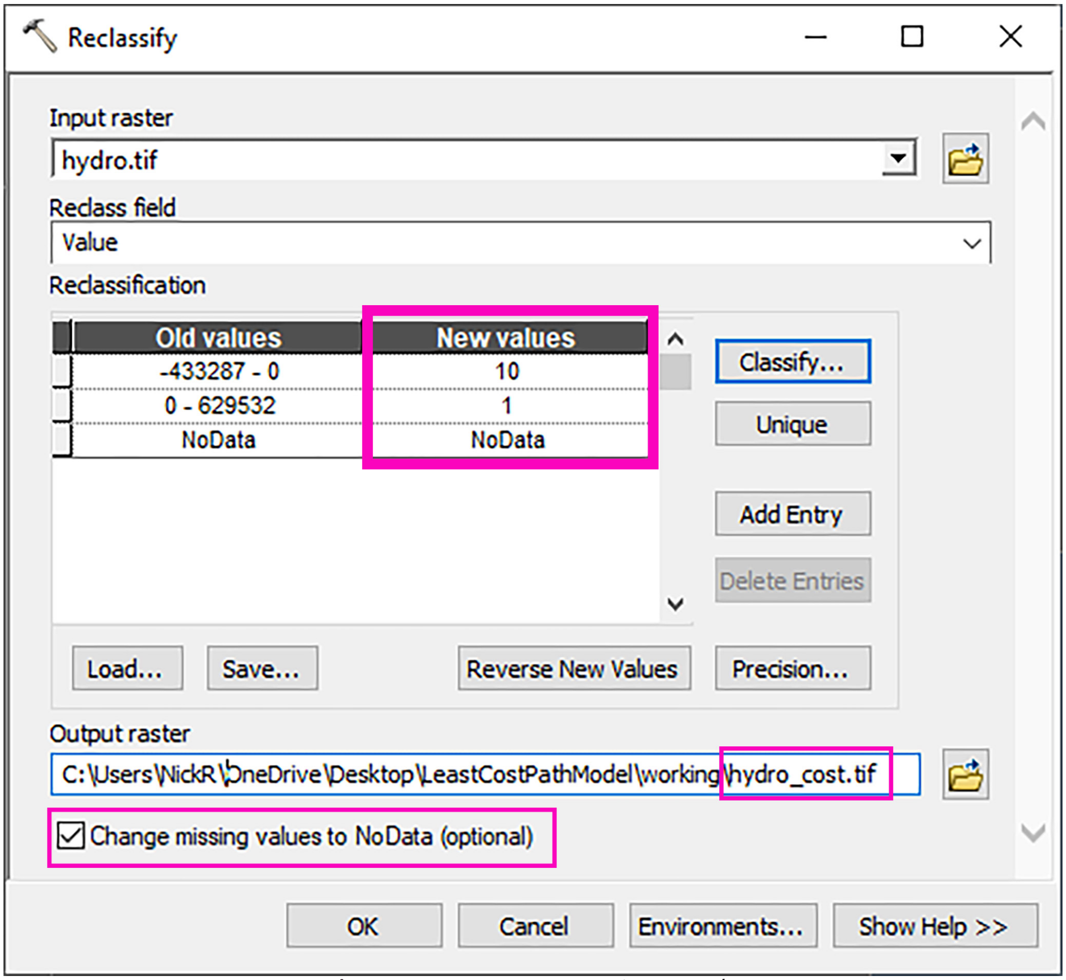


Figure 3.67: Be sure your settings match.

The result is a layer where any cells that represent water features have a value of **10**. In contrast, everything else has a value of **1**.

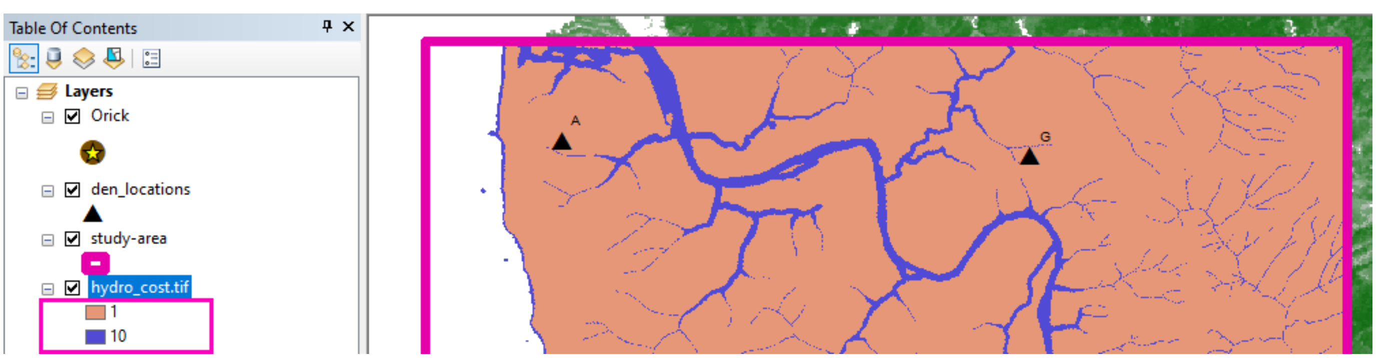


Figure 3.68: In this image, the ArcGIS software has assigned random colors to the pixel values between 1 and 10. Your results should look similar. Be sure there are **no** values other than 1 **and** 10.

Creating a Total Cost Surface Model

As mentioned earlier, a **total cost surface model** combines all of the cost factors you wish to consider for your model into **one** raster dataset using a relative set of cost values. The question remains about what mathematical function to use to combine the cost surface models. Like developing the relative scale, the methods for combining cost surface models into one total cost surface might require research, data collection, and ground-truthing.

In this scenario, you will assume thorough research was conducted by the HSU students and their professors on how to combine the cost surfaces into one total cost surface model. Use the *Raster Calculator* tool to multiply each of the cost surface rasters together (**Figure 3.69**). Save the results to your *working* folder. Name the file, *total_cost.tif*.

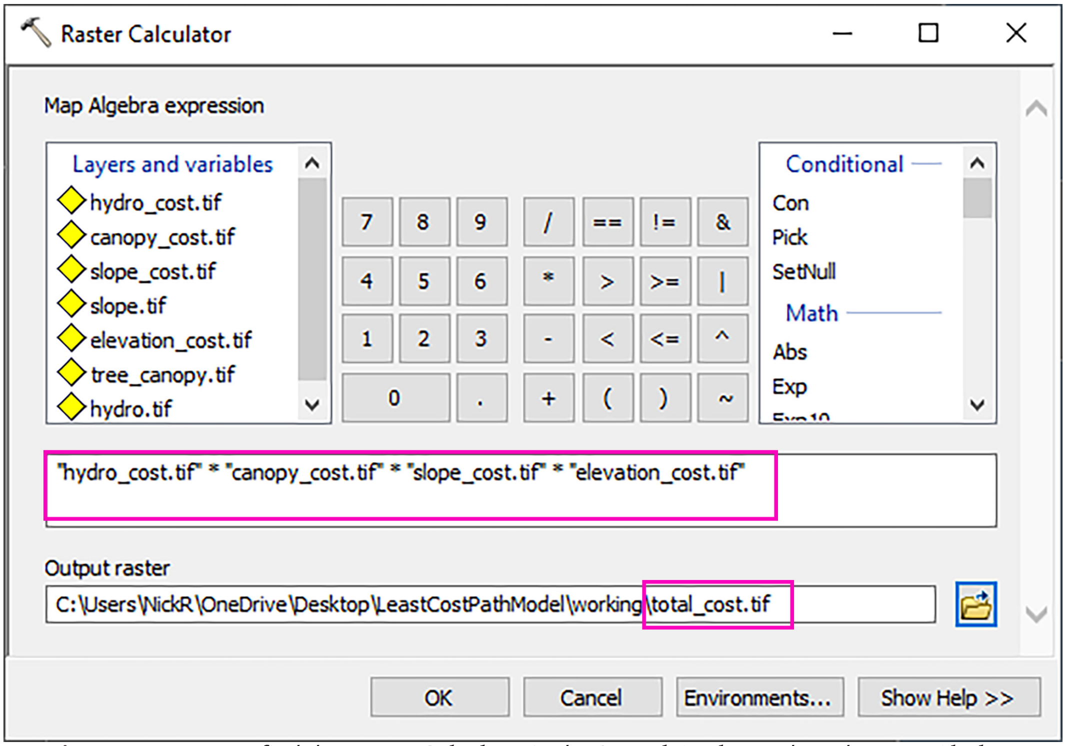


Figure 3.69: You can find the *Raster Calculator* in the *Spatial Analyst* tools under *Map Algebra*.

Remove the individual cost surface layers from the *Table of Contents* when you are done, leaving only the total cost raster for the next step (**Figure 3.70**). Take a moment to save your map document before continuing to the next step.

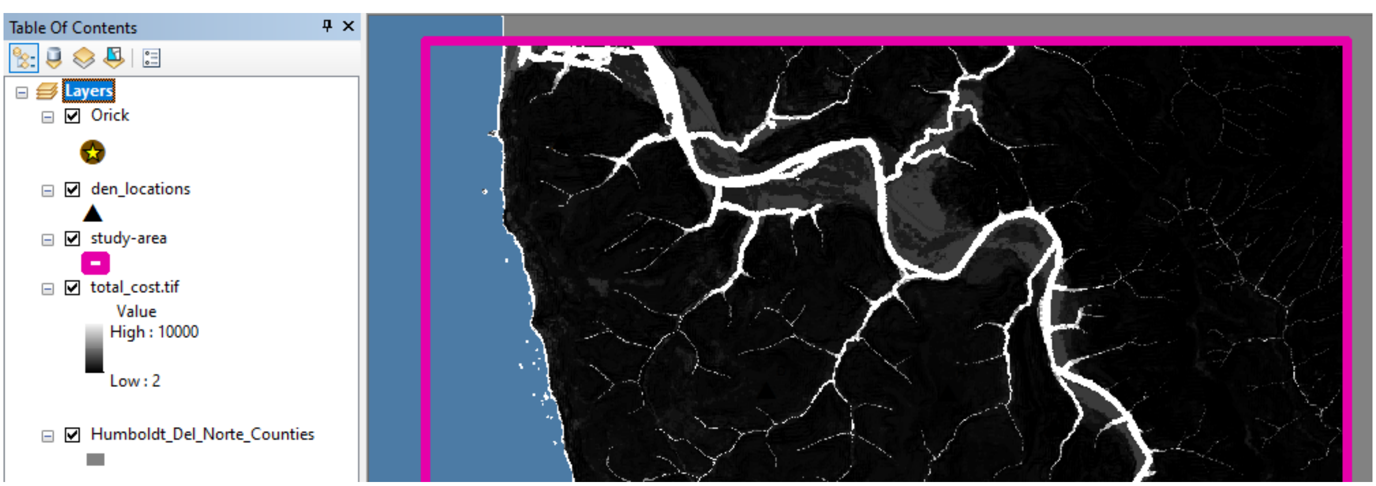


Figure 3.70: The upper limit of the total cost values should not exceed **10,000**.

Clearing away the other raster datasets from the Table of Contents may improve the performance of your computer. It also helps to prevent confusion while working.

Creating a Cost-Distance Surface Model

In this scenario, you want to model movement between the wolpertinger dens and the town of Orick as they migrate into town during the Oktoberfest season. To do this, you need to know how the cost increases as the wolpertingers move in both directions, away from the town of Orick and away from the dens. You will use the *Path Distance* tool, which calculates the accumulative cost over a cost surface. The *Path Distance* tool calculates the **Euclidean distance** between cells and origin features. However, it also incorporates changes in elevation when determining distance. As distance increases from origin feature, cells are assigned a **cumulative** travel cost based on both distance and cost surface values. It stores the cumulative cost moving outward from the *origin*. In this step, you will use the *Path Distance* tool twice. You will run it first using the town of Orick as the *source feature*, then again using the den locations layer as the *source feature*. By doing so, you will calculate the cost distance in both directions.

In ArcMap, open the *Path Distance* tool, located in the *Spatial Analyst* toolbox under *Distance*. Under *Input Raster or feature source data*, use the layer representing the town of Orick (**Figure 3.71**). For the *Output distance raster*, navigate to your *working* folder. Name the output *Orick_Cost_Distance.tif*. For the *Input Cost raster*, choose your total cost surface model. For the *Input surface raster*, browse to your *working* folder and choose the *DEM.tif*. This step allows the *Path Distance* tool to account for changes in elevation when calculating distance. For the *Output back-link raster*, navigate to your *working* folder and name the output *Orick_back-link.tif*. When ready, *click OK*.

There are many additional options in the Path Distance tool.
This tutorial keeps things simple, but readers should feel free to explore additional options later.

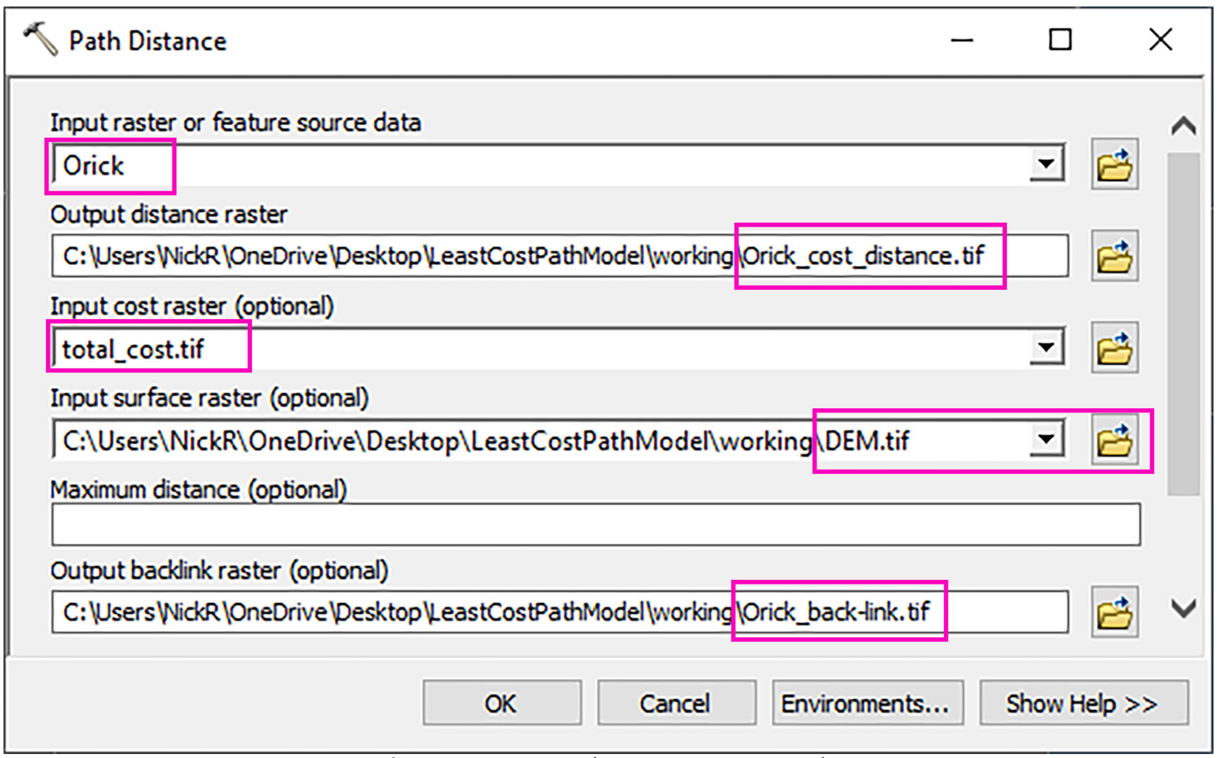


Figure 3.71: Be sure that your settings match.

Depending on the speed of your computer, it may take up to 10 minutes for the Path Distance tool to run.
To avoid crashes, try not to click on ArcMap while the Path Distance tool is running.

The result will be two new raster layers added to ArcMap, one cost-distance model, and one back-link raster dataset (**Figure 3.72**). The **back-link** is a **cost-direction surface** that contains cells with values ranging from 0 through 8. These cells track the direction along the least-cost path from anywhere on the surface back to the *origin* (*source feature*). In a later step, you will use both the cost-distance model and the back-link to determine the *least-cost path*.

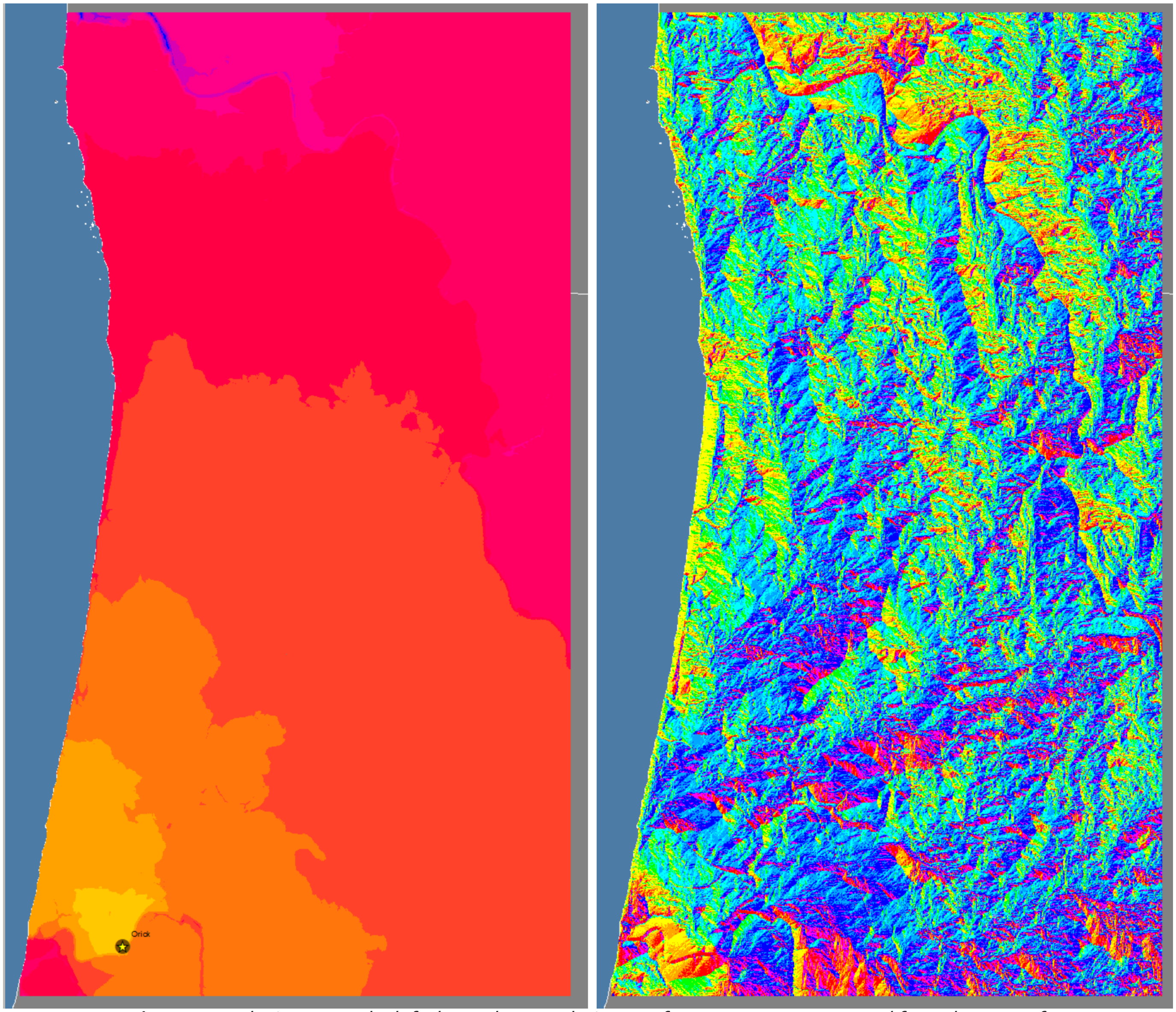


Figure 3.72: The image on the left shows the cumulative cost from as you move outward from the town of Orick. The image on the right shows the back-link raster generated by the *Path Distance* tool. These two layers **will not appear side by side** in ArcMap and are arranged this way for demonstration purposes only.

Skill Drill: Creating a Second Cost-Distance Surface Model

Using the same methods, open the *Path Distance* tool again to create a cost-distance surface utilizing the **den locations** for the *Input raster or feature source data*. Because there are multiple den locations, the *Path Distance* tool uses **allocation** to determine which specific den to use as an origin. For the *Output Distance Raster*, name the file, *Den_cost_distance.tif* (**Figure 3.73**). You **will not** need to create a back-link raster for the den locations.

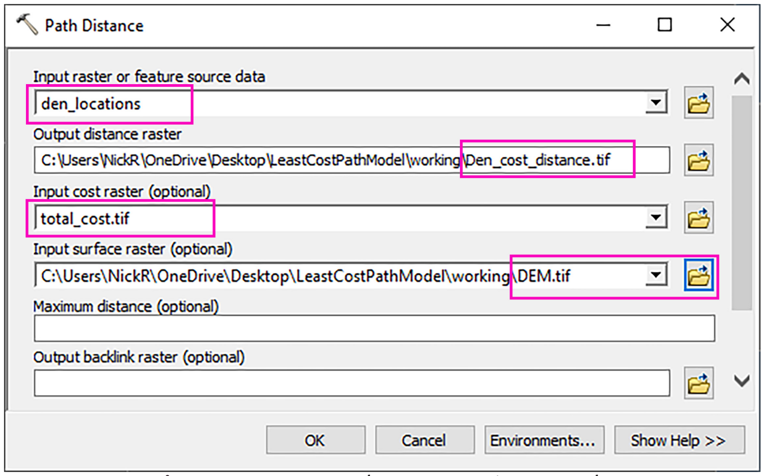


Figure 3.73: Be sure that your settings match.

When you are done with this step, you should have two cost-distance models, one for the den locations (**Figure 3.74**) and one for the town of Orick. Next, you will use both cost-distance models to create a migration corridor.

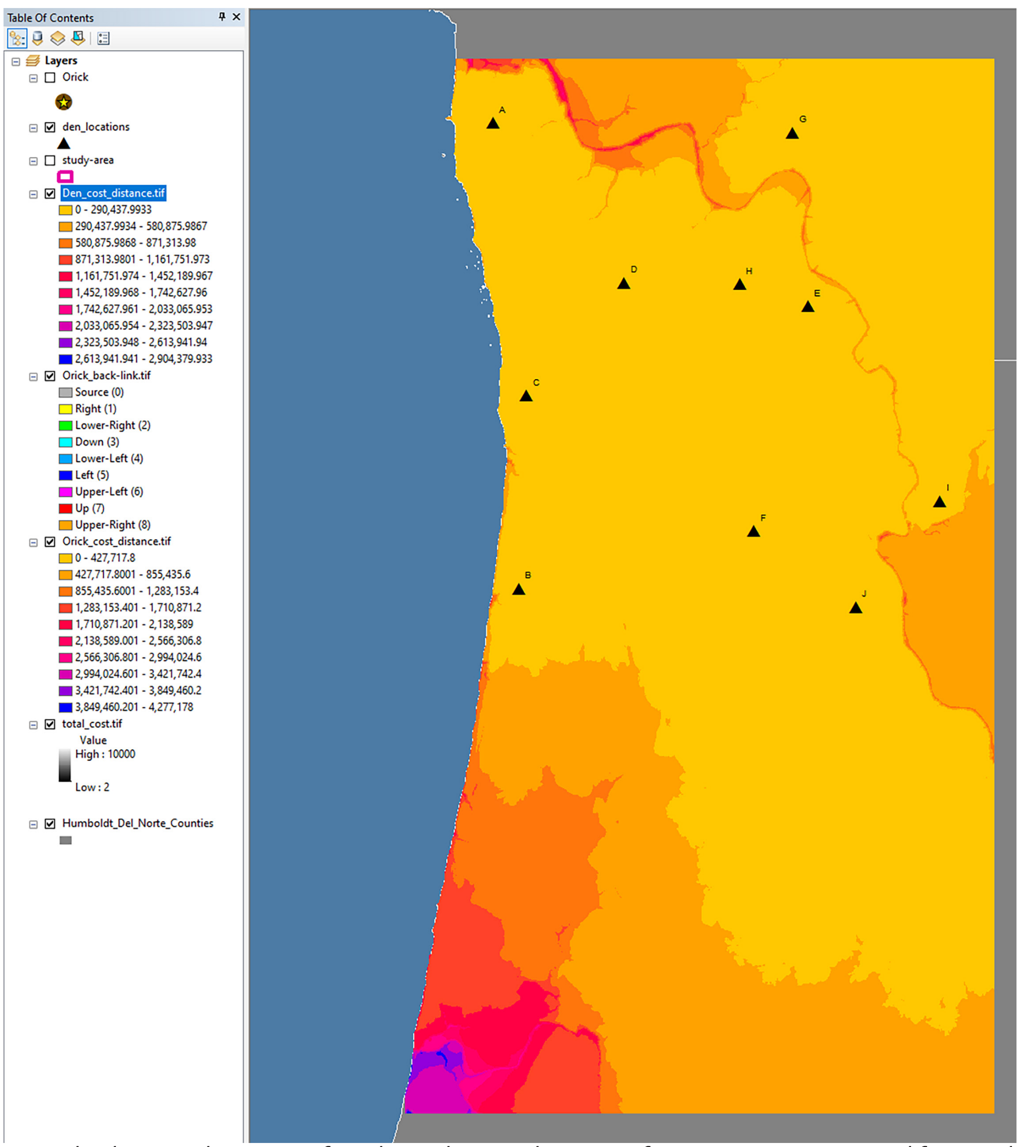


Figure 3.74: The den cost-distance surface shows the cumulative cost from as you move outward from each of the den locations. The *Path Distance* tool uses **allocation** to determine which particular den to use as an origin.

Creating a Migration Corridor

Corridors are useful in the early stages for many projects when you want to define a general location rather than a specific path or when you need to delineate a path wider than a single cell. To create a corridor for the wolpertinger, you will combine the two cost-distance surface models by **adding** them together. You will then *symbolize* the cells to observe patterns and delineate a corridor.

There are a couple of ways to create a migration corridor in ArcMap. You could use the *Raster Calculator* to add the two layers together. However, the *Distance* toolbox has a *Corridor* tool that does the same thing (**Figure 3.75**). For the *Input cost distance raster 1*, choose the Orick cost-distance surface. Choose the den cost-distance surface for *Input cost distance raster 2*. Save the output to your *working* folder. Name file *corridor.tif*. When ready, *click OK*.

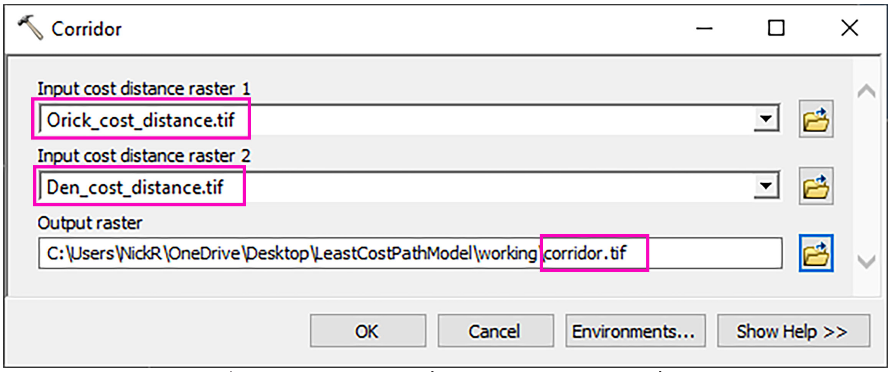


Figure 3.75: Be sure that your settings match.

The initial result may seem unremarkable. To observe patterns and delineate a corridor, you will have to change the *symbology* in the layer properties. There are many options to chose from in the *Symbology* tab. I have found using a **histogram equalization** stretch type with the *Temperature* color ramp works well for observing patterns created by the *Corridor* tool.

You may choose a different method if you find one that works better for you.

Open the layer properties for the corridor layer and navigate to the *Symbology* tab (**Figure 3.76**). On the left, choose the Stretched symbology method. For the *Stretch Type*, choose *Histogram Equalize*. For the *Color Ramp*, choose the *Temperature* color scheme. When you are ready, *click OK*.

You can view the color ramp labels by right-clicking on the color ramp and un-checking Graphic View.

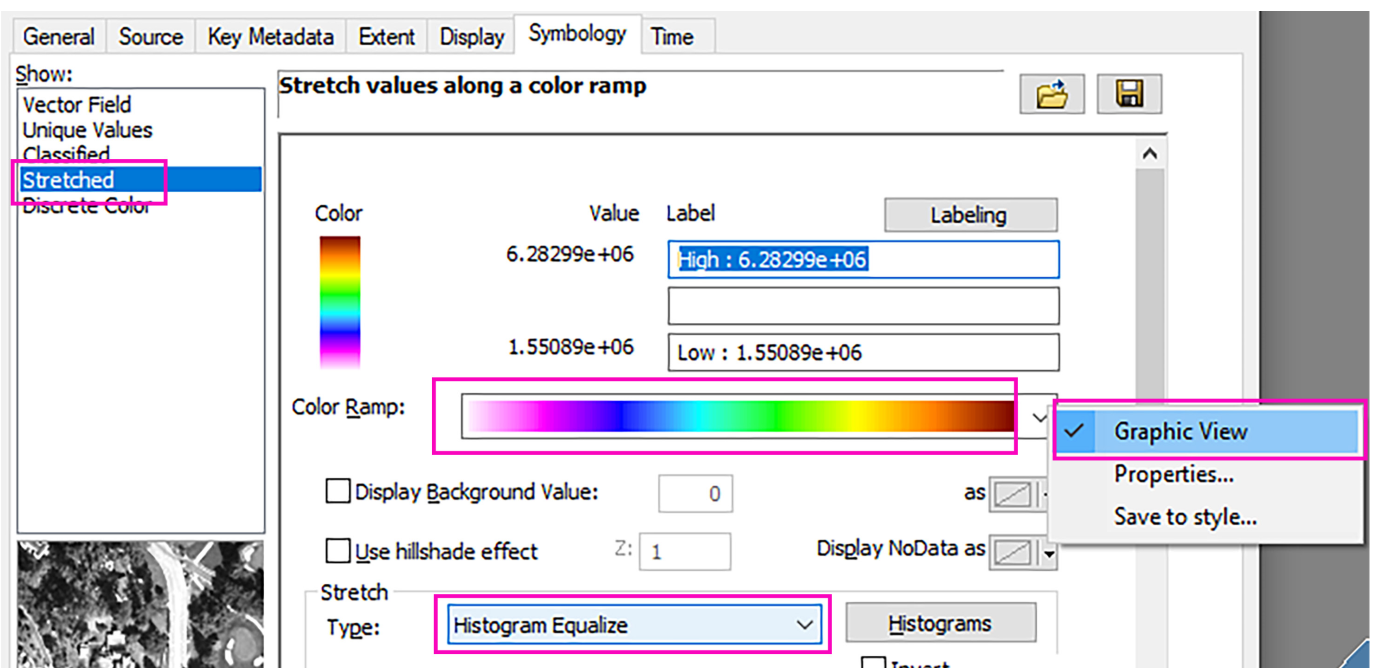


Figure 3.76: Be sure that your settings match.

Areas in **white** and very light **pink** represent areas of **low** relative costs. These are the areas this species will likely move through based on the given cost factors. In the next step, you will narrow the path even further by determining the least-cost path from each den location to the town of Orick.

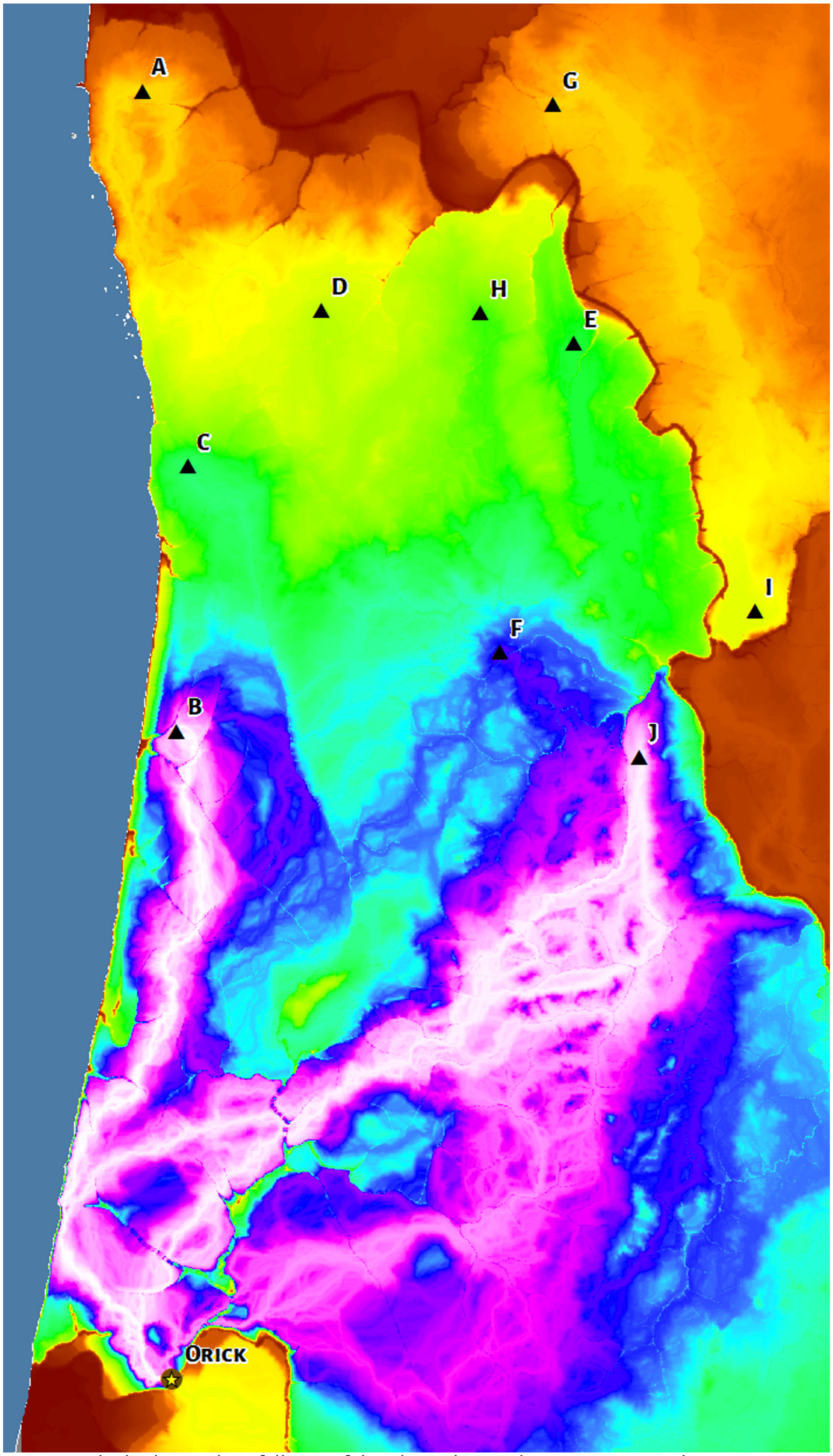


Figure 3.77: As you can see, the highest values fall east of the Klamath River because crossing the river generated a very high cost.

Determining the Least-Cost Path

Using the cost-distance surface model and the back-link raster, the GIS can find the least-cost path by starting at the destination cell and tracing the **least-cost path** back to the origin. As the least-cost path is generated, each cell along the path is assigned a value based on the cost units and is written onto a new raster layer. The process stops when it reaches the first back-link cell with a value of zero, the *origin*.

The result is a raster layer modeling the least-cost path. In this step, you will determine the least-cost path from each of the individual den locations to the town of Orick. To prepare, you will have to separate each of the den locations so that they may be entered into the Cost Path tool, **one at a time**. There are several ways to do this, including selecting each den and creating new shapefiles. However, a straightforward way is to use a *Definition Query* in the layer properties.

Open the layer properties for the den locations layer and navigate to the *Definition Query* tab (**Figure 3.78**). Click on the *Query Builder* button. The *Query Builder* should look very familiar as it is nearly identical to the *Select by Attribute* tool. Create a query where the **Den_ID** is equal to the letter **A**. Then *click OK* and OK again to set the layer properties. On the map, you should see that the den locations layer only contains den A. With this *Definition Query* in place, ArcMap will treat this shapefile as having only one point feature, **den A**.

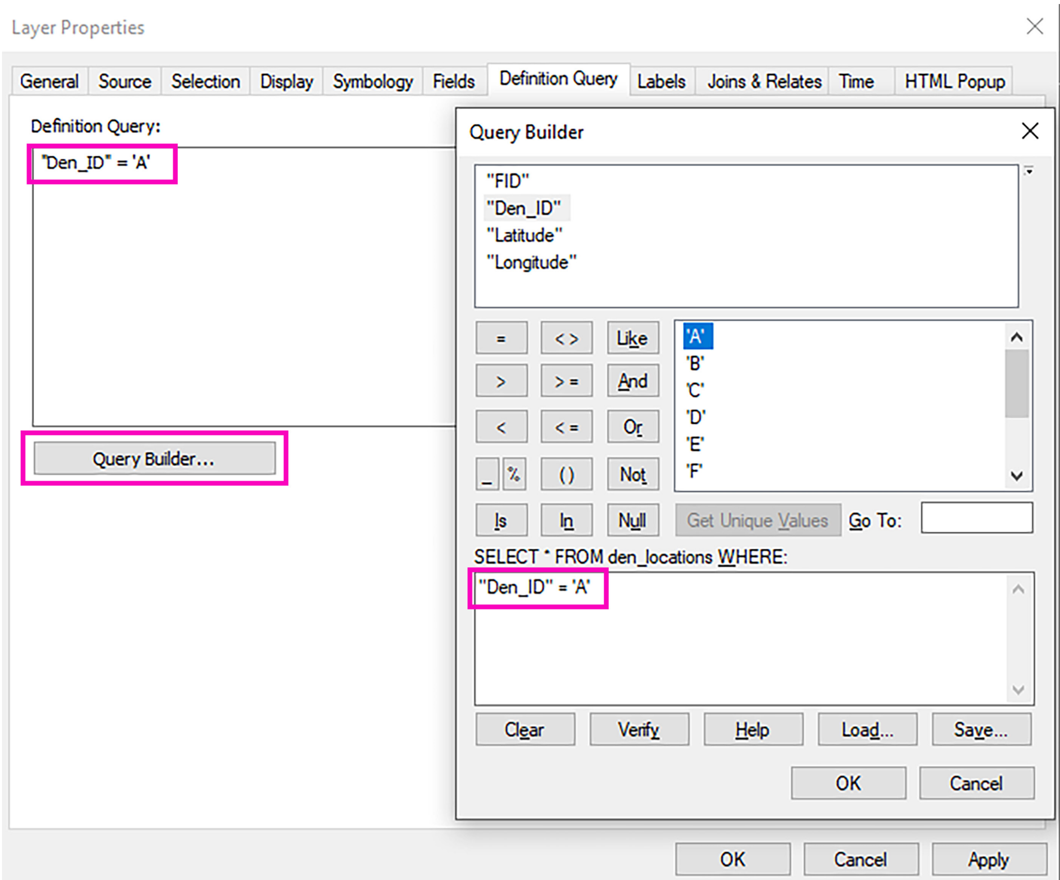


Figure 3.78: A Definition Query is a quick way to make a temporary change to a shapefile without altering the original data or creating new data.

In ArcMap, open the *Cost* Path tool, located in the *Spatial Analyst* toolbox under *Distance*. Under *Input raster or feature destination data*, choose the den locations layer (**Figure 3.79**). Under the *Input Cost distance raster*, load the Orick cost distance surface. Under the *Input cost, back-link raster,* load the Orick back-link raster. For the *Output raster*, browse to your *working* folder and name the file *LCP_DenA.tif*. Leave all other settings as default and *click OK*.

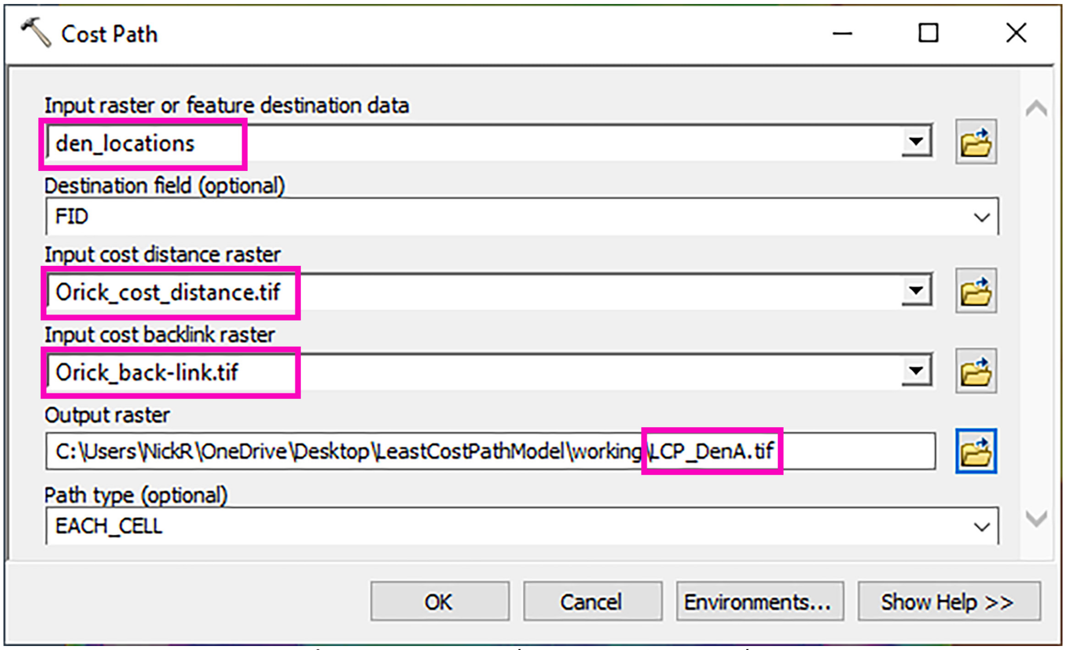


Figure 3.79: Be sure that your settings match.

Converting Raster to Polyline

The result will be a raster layer with the least-cost path from **den A** to the town of Orick. Because the least-cost path is precisely 1 pixel wide, it will be very hard to see. To compensate, you will need to convert the *LCP_DenA.tif* into a shapefile of the same name. Open the attribute table for the *LCP_DenA.tif* and add a new field (**Figure 3.80**). Name the field *Cost*. Change the *Type* to *Long Integer*. When ready, *click OK*.

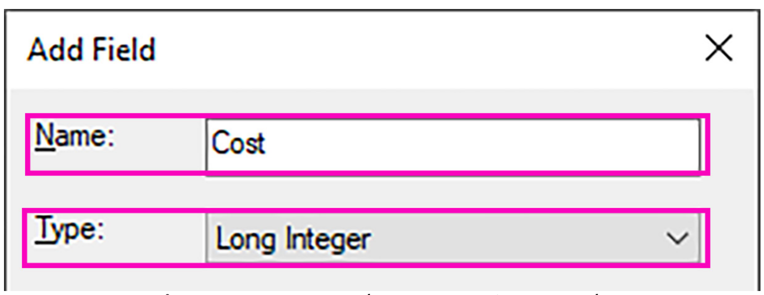
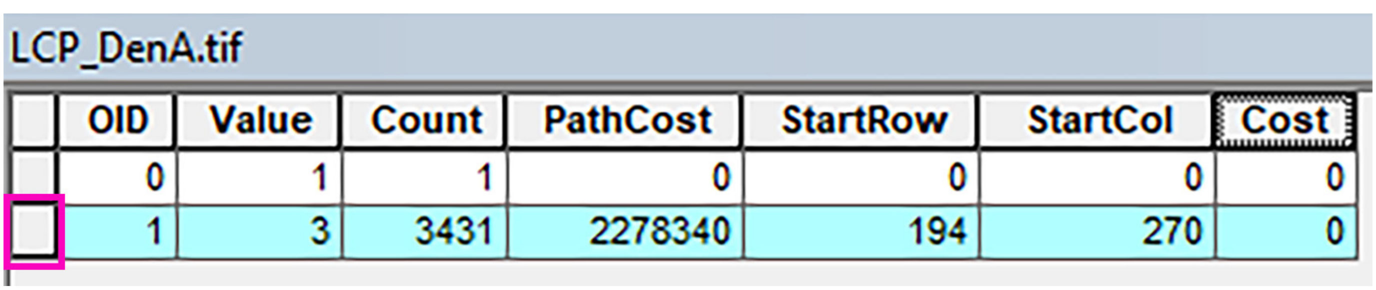


Figure 3.80: Be sure that your settings match.

Click the gray box on the left of the attribute table to select the row that has a value of 3 (**Figure 3.81**).



LCP_DenA.tif

	OID	Value	Count	PathCost	StartRow	StartCol	Cost
	0	1	1	0	0	0	0
	1	3	3431	2278340	194	270	0

Figure 3.81: You must highlight the correct row before using the *Field Calculator*.

Right-click on the *Cost* field and open the *Field Calculator*. Enter the value from the *PathCost* field into the white box. When ready, *click OK*.

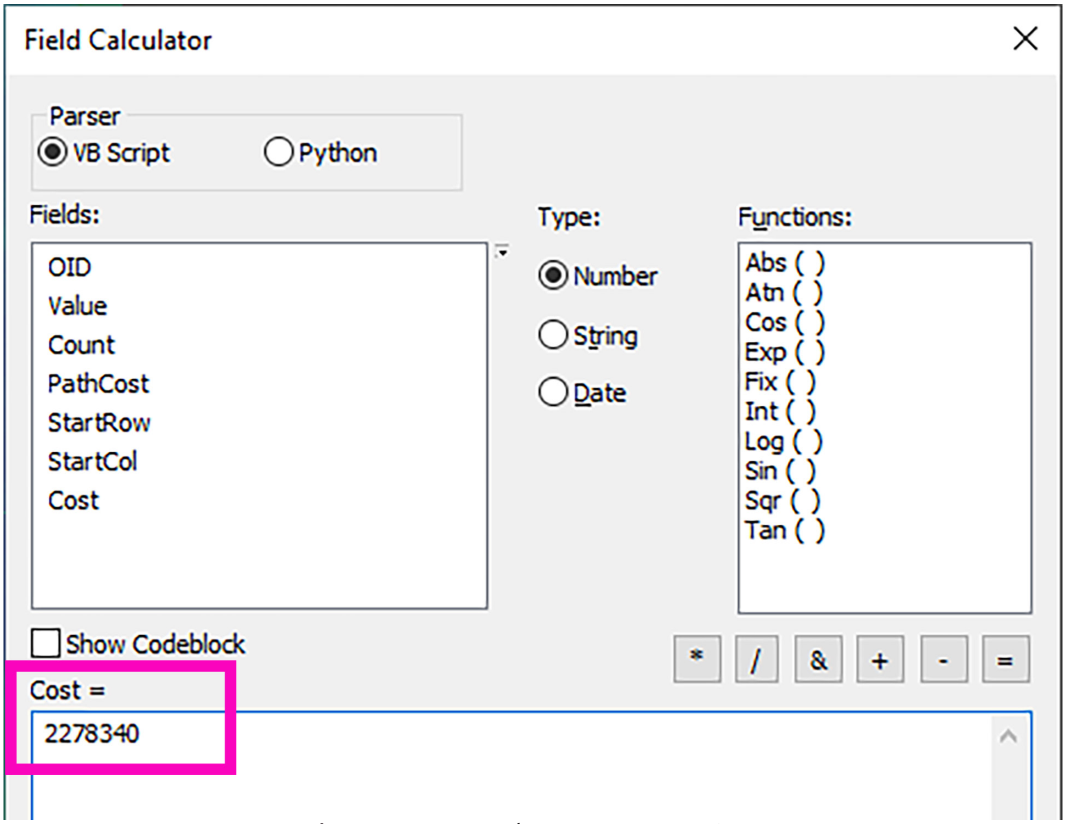


Figure 3.82: Exact values may vary over time.

Locate the *Raster to Polyline* tool in the *Conversion* toolbox under *From Raster*. For the *Input raster*, choose the *LCP_DenA.tif*. Use the drop-down menu under *Field* and choose the *Cost* field. For the *Output polyline features*, browse to your *working* folder. Call the file *LCP_DenA*. For *Background value*, choose *NODATA*. When ready, *click OK*.

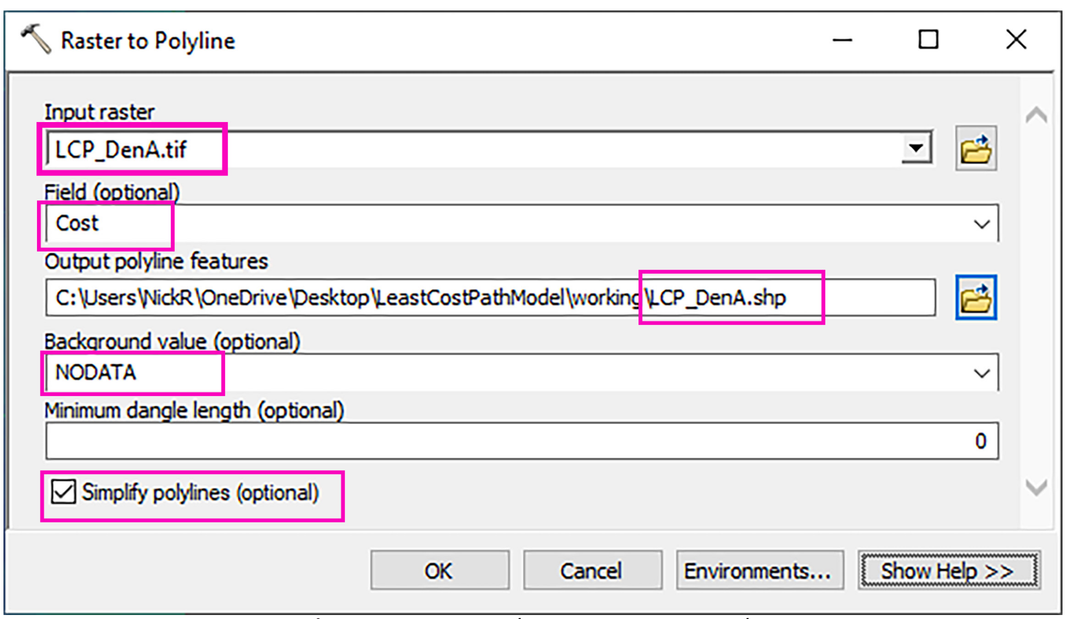


Figure 3.83: Be sure that your settings match.

You may remove the original *LCP_DenA.tif* raster from the *Table of Contents*. As you can see, the least-cost path is not always the shortest (**Figure 3.84**).

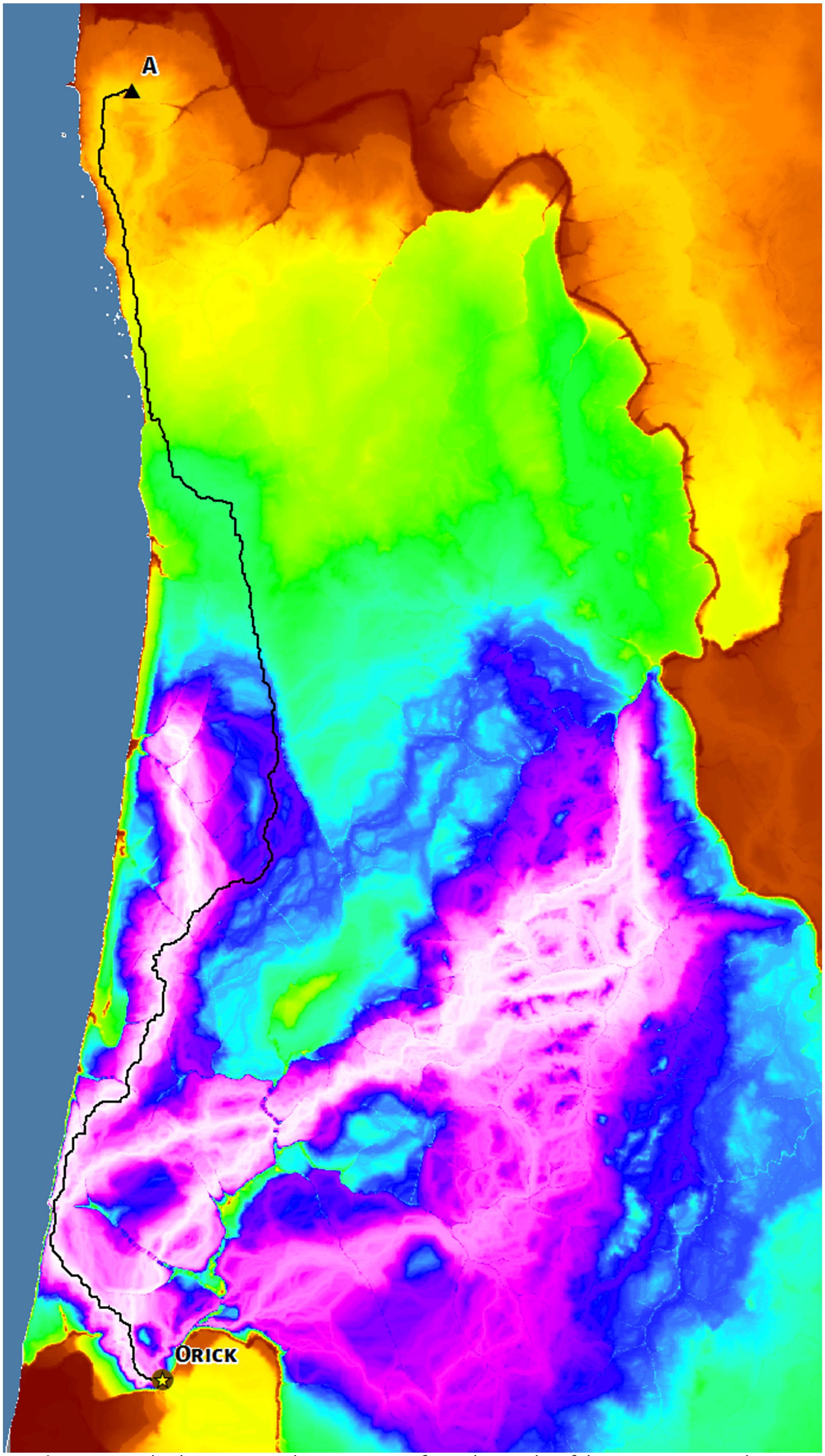
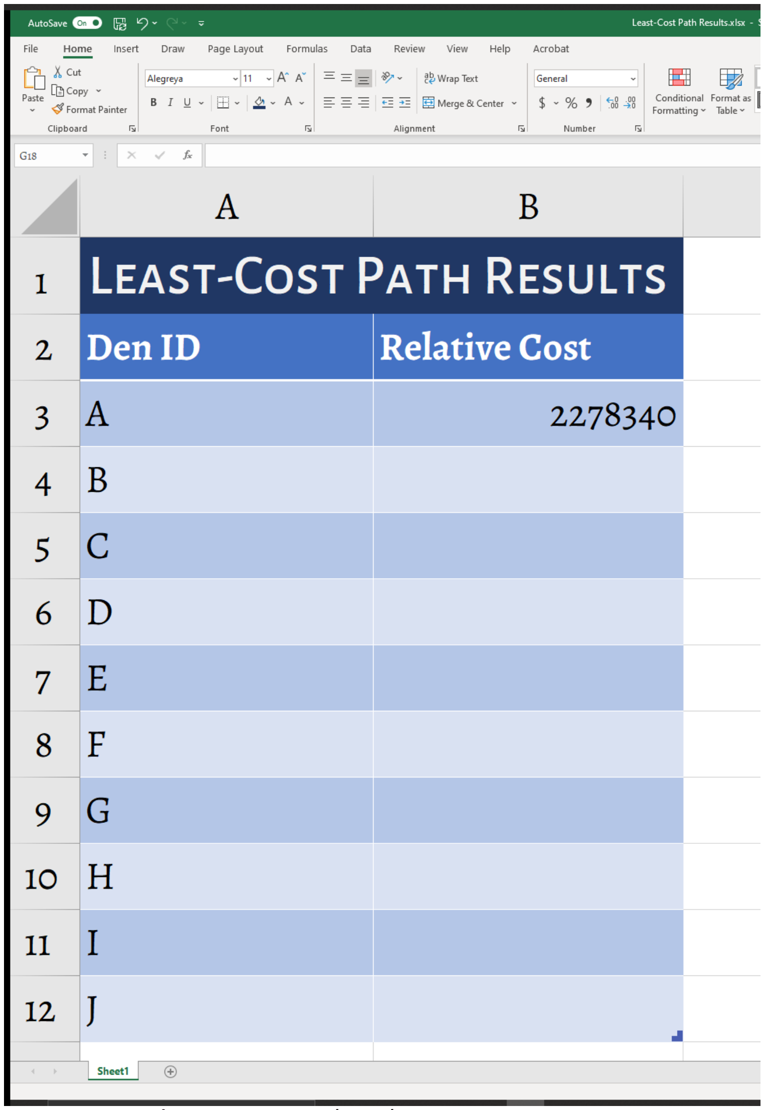


Figure 3.84: This least-cost path seems to reinforce the results of the migration corridor as the path seems to mostly follow the areas colored white and light pink.

Record the Cost value for Den A in a table in Microsoft Excel (**Figure 3.85**). Save the file to your *final* folder for later reference.



Least-Cost Path Results	
Den ID	Relative Cost
A	2278340
B	
C	
D	
E	
F	
G	
H	
I	
J	

Figure 3.85: Actual results may vary over time.

Skill Drill: Determine the Least-Cost Path for each of the Den Locations

Using the skills you learned previously, determine the least-cost path for each of the remaining den locations. Record the cost value for each den in your Microsoft Excel table. Convert each of the output TIF files from the *Cost Distance* tool to a polyline shapefile using *Raster to Polyline*. Remove each of the TIF files from the *Table of Contents* so that only the migration corridor remains.

Skill Drill: Creating a Map of the Results

You should be familiar with the steps needed to create a map layout of your results. Design a map for use as a figure in a report or summary. Ideally, the map should be designed at a size of approximately 6 or 7 inches wide. Include a north arrow, a scale bar, and a legend. The map should include the results of each least-cost path model along with your migration corridor. Include the den locations and the town of Orick. For context, you may also want to add a hillshade under your migration corridor. When your map layout is complete, export the map as a PNG file with a resolution of 300dpi. Save the file in your *final* folder.

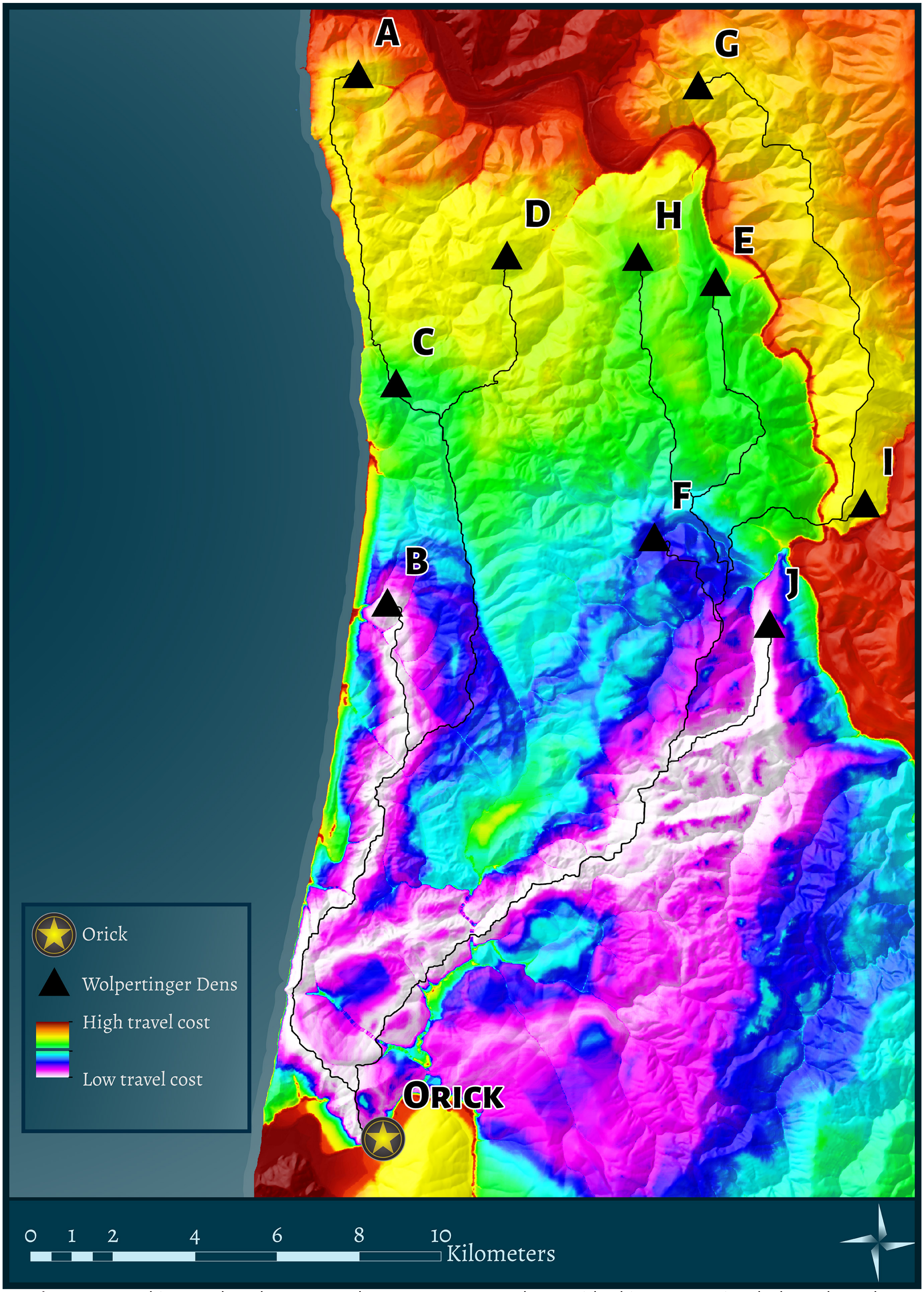


Figure 3.86: In this map, the color ramp used was a temperature scheme with white representing the low values. The least-cost path layers were given a contrasting color to remain visible. A hillshade was added to provide context.

Chapter 4: Modeling Surfaces

This book defines a **surface model** as a depiction of a geographic event, trend, or feature, based on a sample of data, and represented by a continuous vector or raster dataset. A surface model may depict both physical and thematic phenomena, such as crime density hot spots or elevation. However, it is essential to remember that a surface model is merely a *representation* or best guess of whatever phenomenon or feature being depicted. In terms of accounting for error and uncertainty, one should never mistake a surface model for reality. Many novice GIS users forget or are unaware of this fact. For example, a **digital elevation model (DEM)** is most often derived from a set of points, measured at various locations. Some digital elevation models can be incredibly precise, such as those derived from Lidar data. However, they are still derived from sample data and are often generalized to a cell-size resolution of 10 or 30-meters. Further, GIS analysts use DEMs as input datasets for the creation of many other surface models, such as slope or aspect. The result is essentially an educated guess based on another educated guess.

Density

There are two primary surface modeling techniques, density, and interpolation. **Density** is a function that distributes the quantity or magnitude of a phenomenon over a unit of area to create a continuous surface. Examples of themes modeled by a density surface include crime hot spots or disease hot spots represented by magnitude over an area. Density allows you to create a surface from known quantities of a thematic phenomenon that is spread across a region. It is based on both the magnitude and the spatial relationship of the sample data. Density surfaces are used to interpret patterns and trends within the data, such as clustering and hot spots. Density does not try to predict specific values at precise locations. Instead, it measures the *frequency* or *likelihood* of a phenomenon over an area. For example, a density surface modeling crime **does not** represent *actual* crimes occurring at each location, but only the frequency of crimes (**Figure 4.01**)

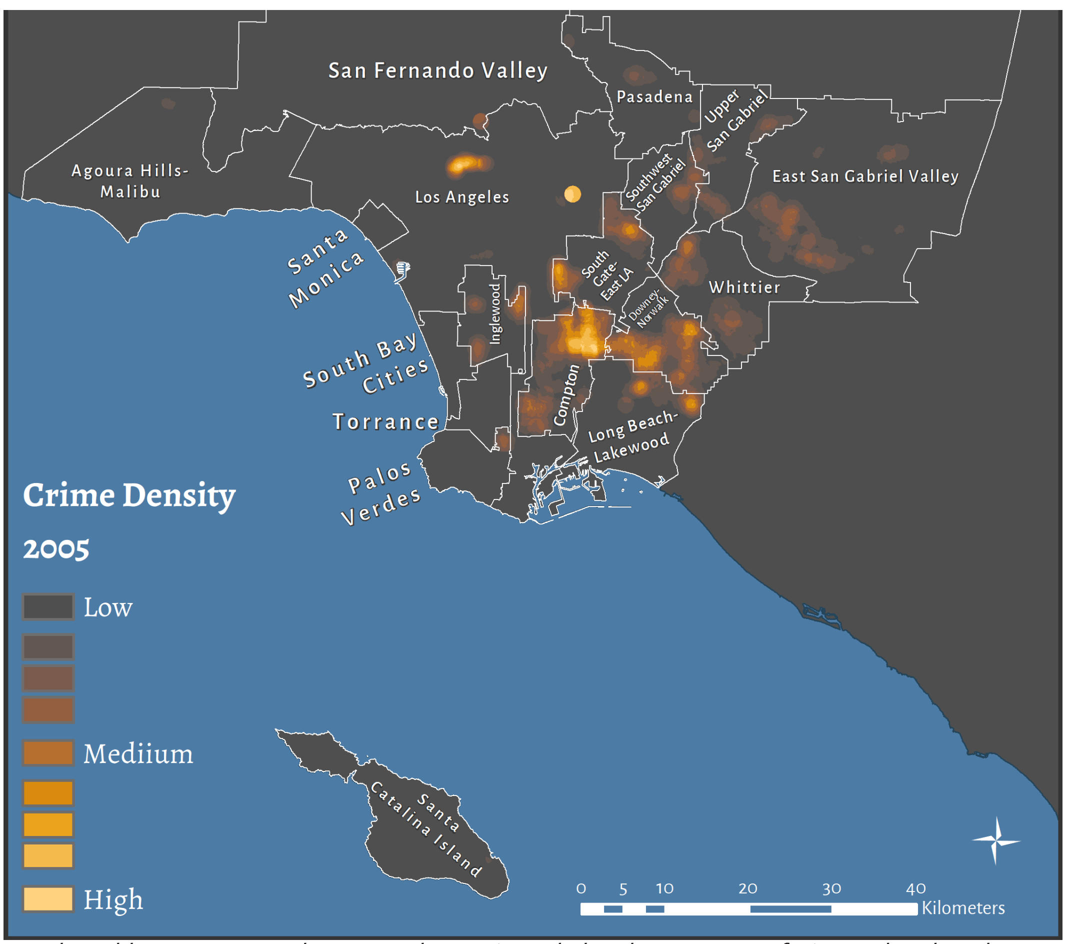


Figure 4.01: The gold areas represent hots spots that are intended to show *patterns* of crime rather than the actual crime.

The methods used to create a density surface vary. This book touches on two different methods, simple and kernel. A **density surface model** calculates the density of a phenomenon by using known values within a user-defined search neighborhood, sometimes called a search radius. When using the **simple method**, the cell values in a density surface model are a function of the **number of data points** in the search neighborhood **divided by the area** of the search neighborhood (**Figure 4.02**).

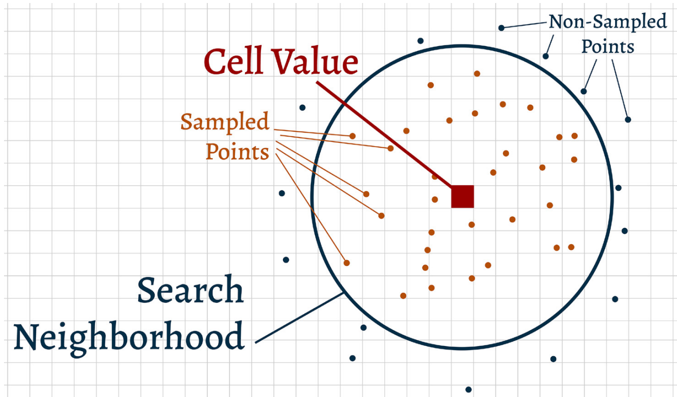


Figure 4.02: The orange dots represent the sampled points that are summed up and divided by the search neighborhood to determine the value of the cell at the center.

The search neighborhood you define will have the most significant impact on the resulting surface. A more extensive search neighborhood will result in a more generalized raster surface. If the search neighborhood is too small, the results will obscure the patterns in the data. Because the size of the search neighborhood depends on both the dataset and the phenomenon in question, it is the **most arbitrary factor** when creating a density surface model. Therefore, the results of a density surface model are **highly subjective**. Determining the most appropriate search neighborhood for your dataset can take time, research, a good understanding of the data, and some **trial and error**.

When using the **Kernel Method**, the cell values in a density surface model are a function of fitting a bell-shaped surface, called a kernel, over each point (**Figure 4.03**). Over the center of the data point, the kernel has a value of **1**. As the surface of the kernel progresses towards the outer edge of the search neighborhood, the value gradually decreases to a value of **0**.

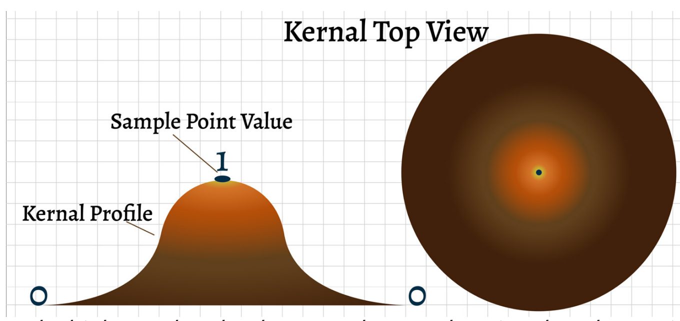


Figure 4.03: The highest values land nearest the sample point, then decay with distance.

The Kernel Method calculates the density values by adding the kernel surfaces together where they overlap the center of each cell (**Figure 4.04**). The sum of the kernel surfaces is then divided by the area of the search neighborhood.

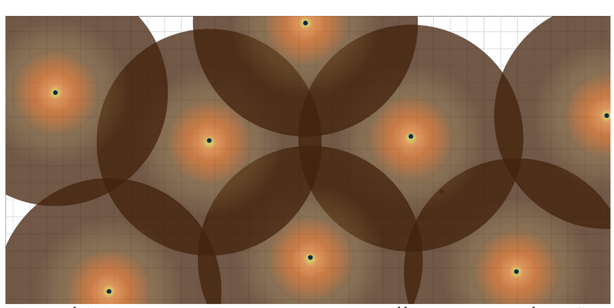


Figure 4.04: This image uses transparency to illustrate overlapping values.

Interpolation

Interpolation is the process of estimating unknown values that fall between known values. Like Density, interpolation also allows you to create a surface from a series of known locations and values. Unlike density, interpolation tries to predict **real values** at specific locations. For example, an interpolated surface modeling elevation tries to predict *specific* elevation values in between the measured locations (**Figure 4.05**).

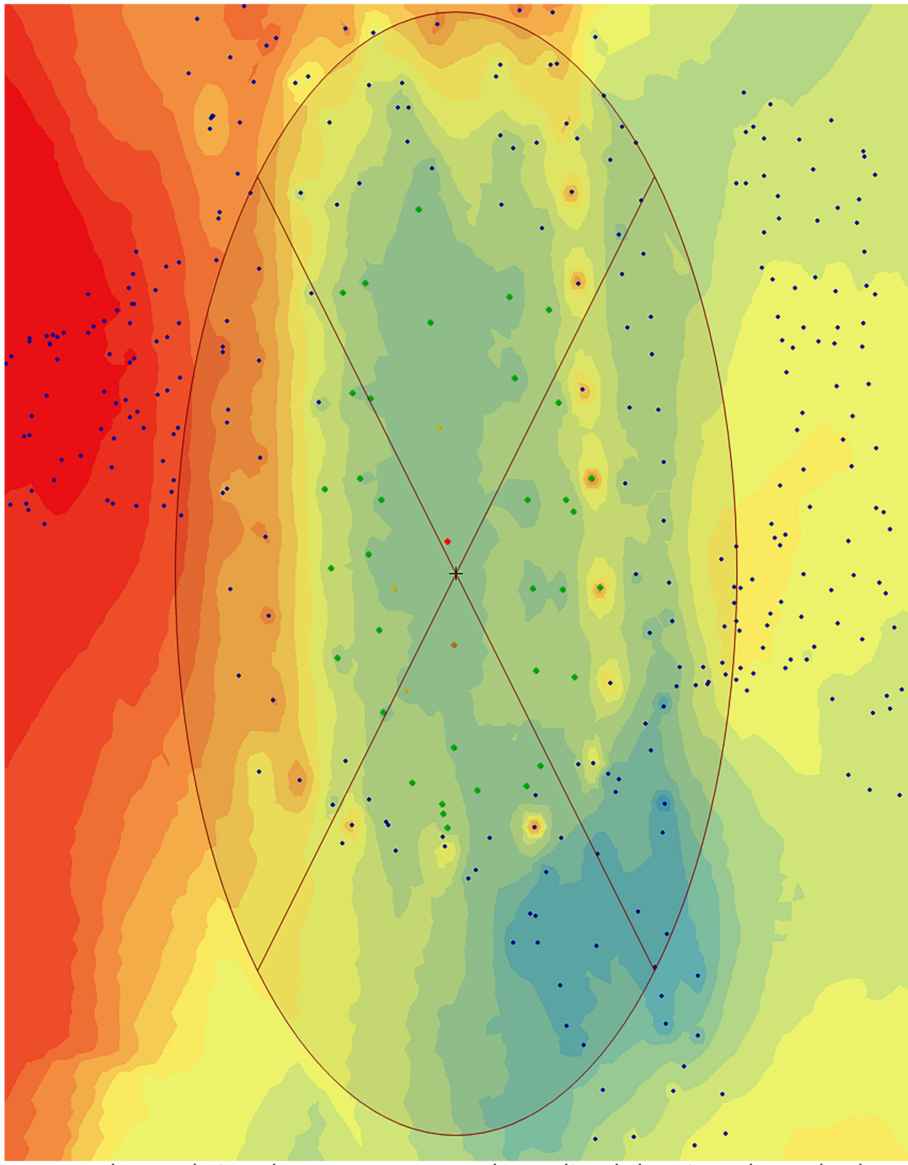


Figure 4.05: The crosshair at the center represents the predicted elevation value at that location.

Tobler's first law of geography states, "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Modeling surfaces using interpolation relies heavily on this premise. Values in between measured locations are spatially dependent. We expect them to be similar to measured values nearby. The mathematics underpinning the various interpolation methods is a complicated subject and beyond the scope of this book. This book will only be introducing basic concepts. Generally speaking, there are two common types of interpolation methods, deterministic and geostatistical. **Deterministic interpolation** uses the measured values at each point to determine values between. When using deterministic interpolation, the values between the measured locations are always entirely and precisely known because they are derived by the measured values. **Geostatistical interpolation** makes predictions using statistics that include **spatial autocorrelation**, a measure of how values that are clustered together are similar. Geostatistical interpolation can also provide standard error values to indicate the accuracy of predictions.

Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) is a *deterministic* interpolation method that uses known values within a user-defined search neighborhood to predict the value at a specific location (**Figure 4.06**). The known values are weighted by distance, so the local influence diminishes as it reaches the edges of the search neighborhood.

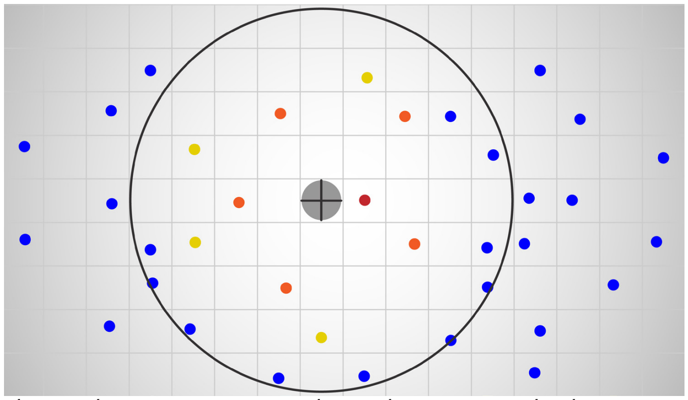


Figure 4.06: The crosshair at the center represents the prediction point. The dots represent measured values.

The **search neighborhood**, an ellipsoidal region centered on the prediction point, can be modified by a GIS analyst in several ways. The major and minor semi-axis can be lengthened, shortened, and rotated to alter the shape, area, and orientation of the ellipsoid (**Figure 4.07**). This will restrict the number of measured values the GIS will use when calculating the value at the prediction point.

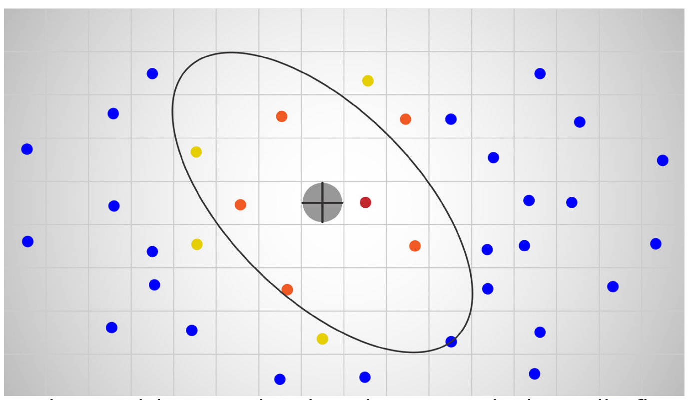


Figure 4.07: The change in shape and direction alters how the measured values will influence the prediction point.

You can also place a limit on the minimum and the maximum number of locations to use within a search neighborhood to modify the results. Also, you can divide the search neighborhood into sectors (**Figure 4.08**). When you do this, any minimum and maximum restrictions are applied to each sector *individually*.

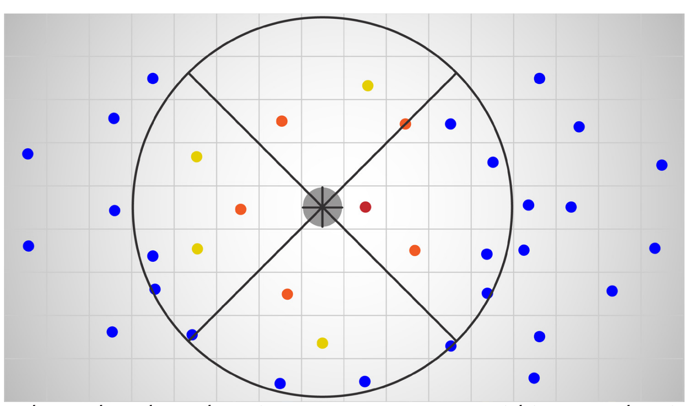


Figure 4.08: The wedge-shaped sections represent sectors within a search neighborhood.

When using IDW interpolation, the influence of nearby data points in the search neighborhood decays exponentially with distance (**Figure 4.09**).

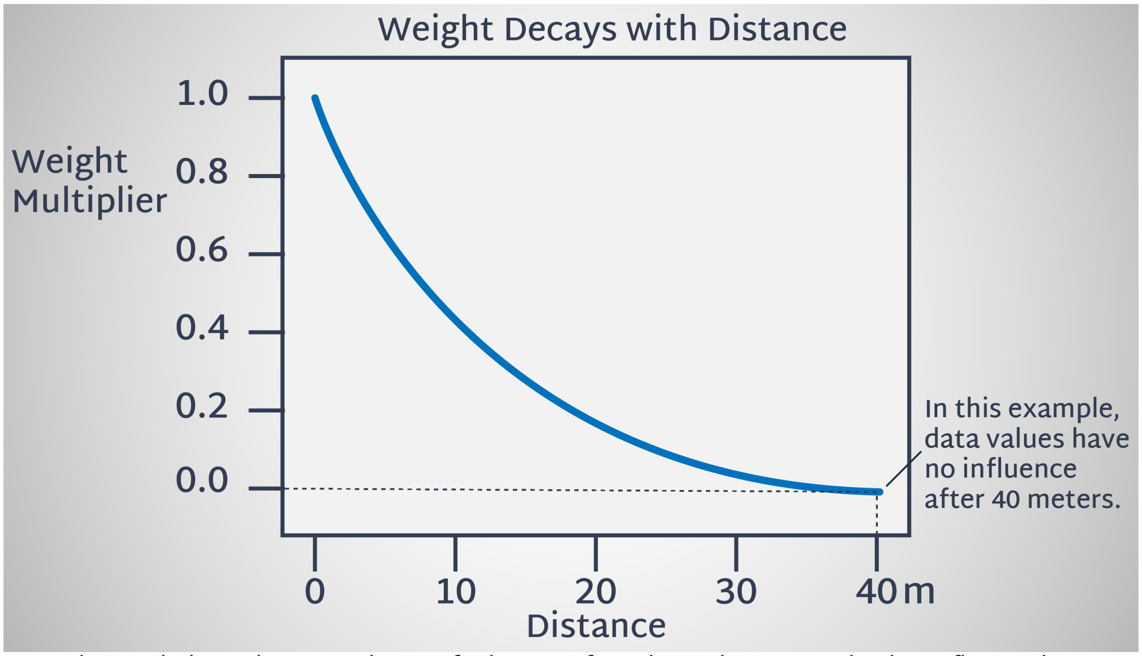


Figure 4.09: This graph shows that a sample point farther away from the prediction point has less influence than one nearby.

You can modify the rate of decay by adjusting the **power (p) variable**. Using a higher power value will increase the rate of decay. The result will mean that local data points will have a more substantial influence. This will increase variation locally and result in a rougher surface overall. Using a lower power will have a smoothing effect as the prediction point is influenced more evenly among the data points at varying distances (**Figure 4.10**).

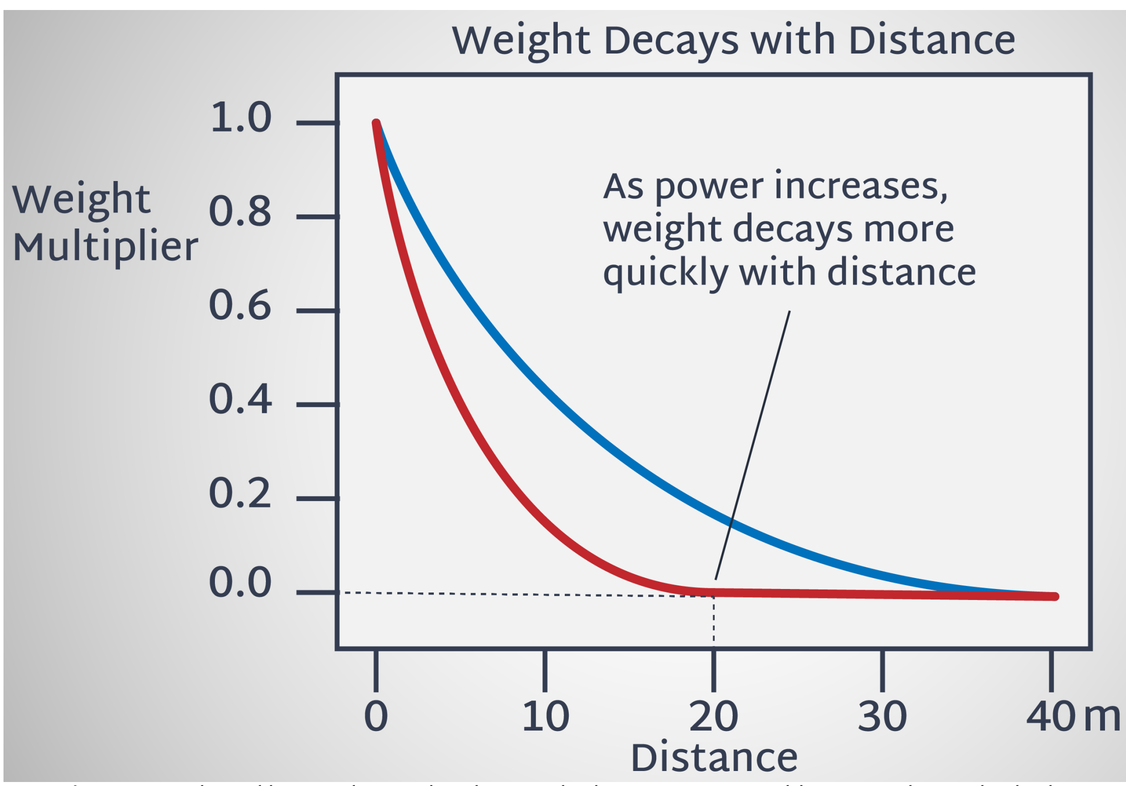


Figure 4.10: The red line on the graph indicates a higher power (p) variable. Notice that with a higher power, the sample points no longer influence the prediction point after 20 meters.

IDW is a *deterministic* interpolator that makes predictions based on averages of sampled points. The highest and lowest values on the prediction surface cannot exceed the highest or lowest values of known data points (**Figure 4.11**). You should use IDW when the phenomenon you are modeling is **not** expected to rise above or drop below known values, and when the distance between points is the primary factor influencing the data.

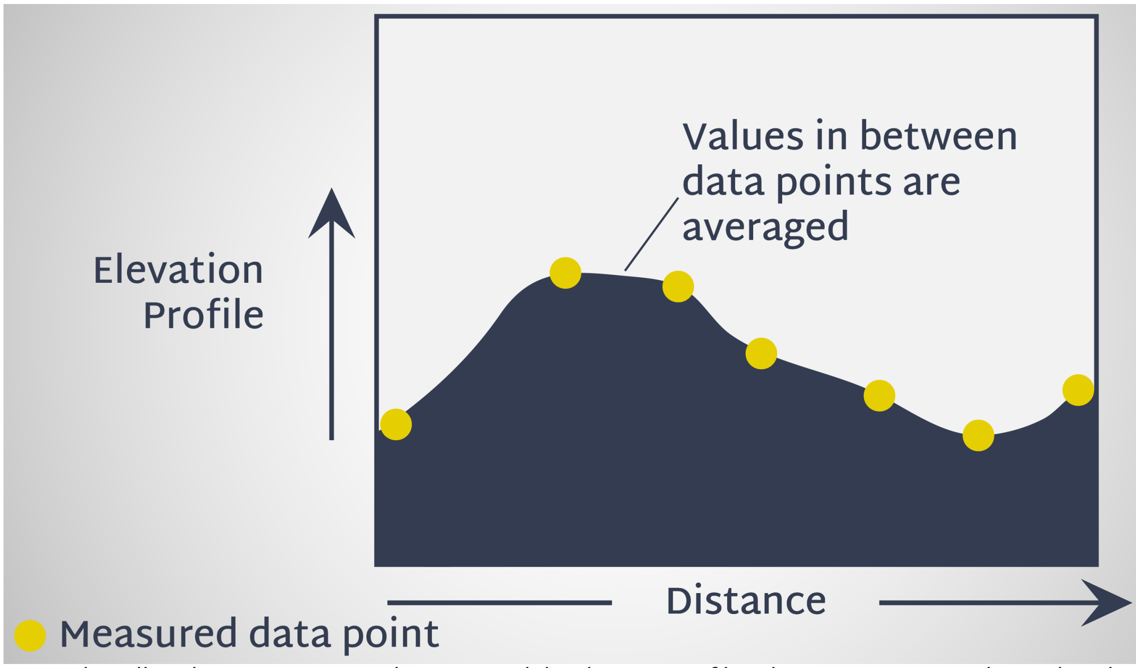


Figure 4.11: The yellow dots represent sample points, and the elevation profile in between represents the predicted surface.

Spline

Spline is a *deterministic* interpolation method where the predicted values are estimated using a function that minimizes the total curvature of the surface. The result is a smooth surface that passes precisely through all of the measured data points (**Figure 4.12**).

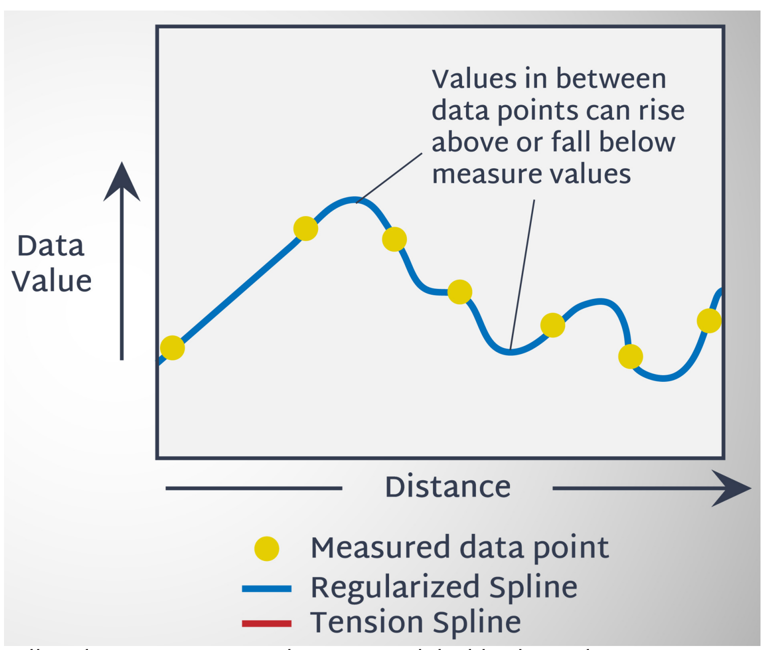


Figure 4.12: The yellow dots represent sample points, and the blue line in between represents the predicted surface. Spline allows the predicted surface to rise above and drop below the sample points

There are two types of spline interpolation methods, **regularized splines** and **tension splines** (**Figure 4.13**). While the mathematical details are beyond the scope of this book, you should be aware that *tension splines* produce a smoother flatter surface by forcing the predicted values to stay closer to the measured data points. Also, unlike IDW, spline interpolation can predict values **above and below sample points**. Spline interpolation methods are best used for gently varying surfaces and for predicting values above and below measured points.

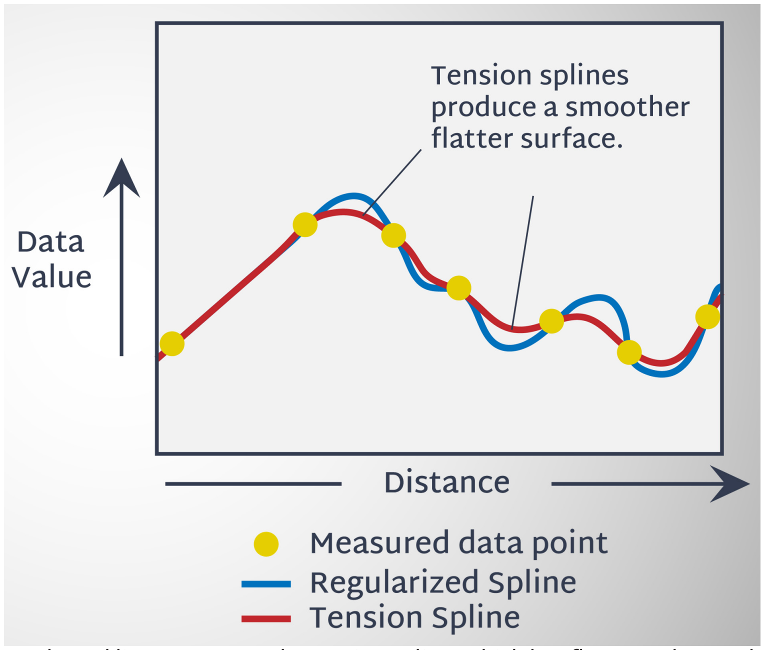
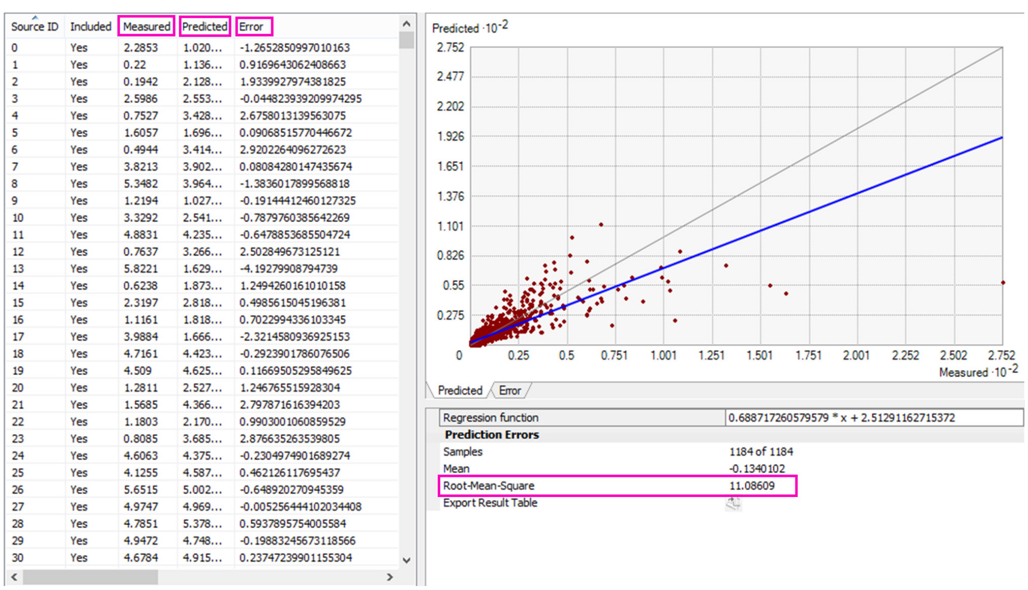


Figure 4.13: The red line represents the tension spline, which has flatter and smoother curves.

Cross-Validation

Deterministic methods like IDW and Spline do not provide a standardized measure of uncertainty. This can pose a problem when you need to defend the methods used in your model. Fortunately, the GIS can provide some measure of uncertainty, even among deterministic methods, through the use of cross-validation. When performing a **cross-validation** operation, the GIS uses all of the known data points to validate predictions. Each known data point is removed from the model one at a time. After one of the data points is removed, the GIS tries to predict the value at that location. The actual values and the predicted values are compared for each data point, and the differences are recorded (**Figure 3.14**).



Source ID	Included	Measured	Predicted	Error
0	Yes	2.2853	1.020...	-1.2652850997010163
1	Yes	0.22	1.136...	0.9169643062408663
2	Yes	0.1942	2.128...	1.9339927974381825
3	Yes	2.5986	2.553...	-0.044823939209974295
4	Yes	0.7527	3.428...	2.6758013139563075
5	Yes	1.6057	1.696...	0.09068515770446672
6	Yes	0.4944	3.414...	2.9202264096272623
7	Yes	3.8213	3.902...	0.08084280147435674
8	Yes	5.3482	3.964...	-1.383601789956818
9	Yes	1.2194	1.027...	-0.191444124601 27325
10	Yes	3.3292	2.541...	-0.787976038564 2269
11	Yes	4.8831	4.235...	-0.6478853685504724
12	Yes	0.7637	3.266...	2.502849673125121
13	Yes	5.8221	1.629...	-4.19279908794739
14	Yes	0.6238	1.873...	1.249426016101 0158
15	Yes	2.3197	2.818...	0.4985615045196381
16	Yes	1.1161	1.818...	0.7022994336103345
17	Yes	3.9884	1.666...	-2.3214580936925153
18	Yes	4.7161	4.423...	-0.292390178607 6506
19	Yes	4.509	4.625...	0.11669505295849625
20	Yes	1.2811	2.527...	1.246765515928304
21	Yes	1.5685	4.366...	2.797871616394203
22	Yes	1.1803	2.170...	0.9903001060859529
23	Yes	0.8085	3.685...	2.876635263539805
24	Yes	4.6063	4.375...	-0.2304974901689274
25	Yes	4.1255	4.587...	0.462126117695437
26	Yes	5.6515	5.002...	-0.648920270945359
27	Yes	4.9747	4.969...	-0.005256444102034408
28	Yes	4.7851	5.378...	0.5937895754005584
29	Yes	4.9472	4.748...	-0.19883245673118566
30	Yes	4.6784	4.915...	0.2374723990 1155304

Figure 3.14: The GIS can provide some measure of uncertainty using cross-validation.

These values are then used to calculate the **root mean square error (RMSE)**, a statistical measure of how closely your model predicts the actual values (**Figure 3.15**). To calculate the root mean square error, the difference between each predicted and actual value is squared and then summed together. The sum of squared values is then divided by the number of data points, resulting in the mean squared error. Finally, the square root of the mean squared error is calculated to provide the *root mean square error*. You won't have to worry about the math. The GIS will make these calculations for you. A model with the best fit will have an **RMSE closest to zero**.

$$\text{RMSE} = \sqrt{\sum\left(\frac{\left(Y_{predicted}-Y_{referenced}\right)^2}{N}\right)}$$

Figure 3.15: The root mean square error formula.

Kriging

Kriging refers to a group of *geostatistical* interpolation techniques that can provide a standardized measure of uncertainty in their predictions. Like some of the deterministic interpolation methods, it is based on the assumption that things that are closer together are more alike than those farther away. There are numerous options to choose from when using the Kriging method, most of which are beyond the scope of this book. Instead, you will primarily rely on the default settings in the ArcGIS software and use Cross-Validation to compare the results. Unlike deterministic interpolation methods, Kriging predicts values by modeling the *statistical correlation* between pairs of known points. This relationship between the values of data points and the distance between then is known as **spatial autocorrelation**.

To create a surface model using Kriging, the GIS will use the measured data points to create a semivariogram (**Figure 4.16**). In a **semivariogram**, each measured data point is paired with every other measured data point. The difference in value between each paired data point is recorded along with the distance between them. The expectation is that there will be small differences in measured values between pairs that are close together. Pairs that have greater distances between them are expected to have more significant differences in measured values. Each of these pairs is then plotted on the graph, which makes up the *semivariogram*. The GIS will then define the surface model by finding the best fit through the points in the semivariogram.

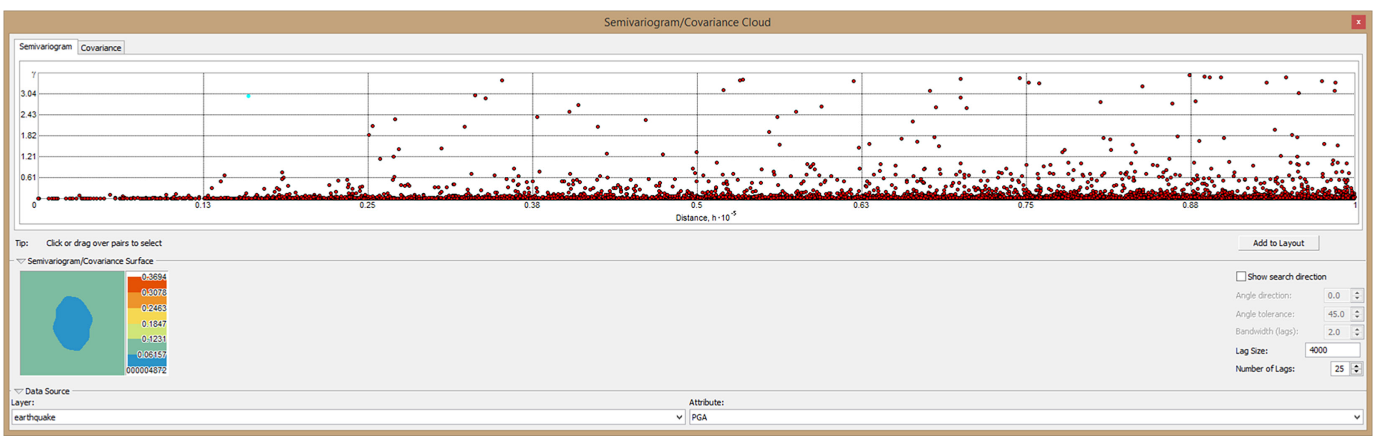


Figure 4.16: A semivariogram cloud.

Tutorial: Exploring Patterns of Crime In Los Angeles County Using A Density Surface Model

A **surface model** is an approximation of a surface based on a set of sample points. It represents a phenomenon that can be measured continuously across a surface. Elevation, slope, temperature, precipitation, population, crime, and disease are all examples of surface models that you may be familiar with or have seen before. There are two primary surface modeling techniques used to create surfaces, **density**, and **interpolation**.

Density allows one to create a surface from point locations. However, unlike interpolation, density does not try to predict specific values at a location but instead communicates geographic trends. For example, a map of cancer density in a region does not predict that someone developed cancer at a specific location. Instead, it communicates the patterns and perhaps the likelihood of cancer over a region. **Density** is a function that distributes the quantity or magnitude of a phenomenon over a unit of area. It is derived from the spatial relationship of the locations and magnitude of the measured quantities.

Estimated time to complete this tutorial: 6 hours

Learning Outcomes

Readers should be able to accomplish the following outcomes by the end of this tutorial:

- Demonstrate how to add feature classes and feature datasets to a file geodatabase
- Discuss the importance of the subjective decisions made when running a density analysis
- Develop a search neighborhood based on a manual inspection of the data
- Develop a search neighborhood based on spatial autocorrelation
- Analyze clustering patterns in point data
- Create a density surface using the *Simple Method*
- Create a density surface using the *Kernel Method*
- Create a density surface based on weighted feature attributes

Scenario

In this scenario, the Los Angeles County Sheriff's Department is interested in analyzing patterns of crime to make administrative decisions on how to allocate resources. You will attempt to assist the Sherriff's department by making a series of density surface models from a history of incident responses.

Conduct this analysis using the **Universal Transverse Mercator (UTM)** system along with the **North American Datum of 1983 (NAD83)**. Los Angeles County lies in **Zone 11** of the UTM system. All of your data must be in this spatial reference system at the start of your analysis. Create working copies of your data in this spatial reference system using the *Project* tool in ArcMap as needed.

Setting up Your Workspace

In a typical workflow, you work on geospatial data using a local hard drive. When done, you compress your data and back up your work to your cloud storage so that you can retrieve the files from anywhere. When referring to a **local hard drive**, it means you are working on data physically located on the computer in front of you.

In contrast, some computers also include networked drives. **Networked drives** link to cloud storage and save the data elsewhere. Examples include services like OneDrive or Google Drive. For this tutorial, use the **desktop** as your local hard drive location. You may also use an external USB drive if you plan to work in multiple places.

You must avoid using networked drives while you work.
They increase the processing time and can cause technical glitches.

In this book, you use a particular folder structure. Start by creating your workspace folder on the local hard drive. A **workspace** is a folder or series of folders that contain all of your project files. The top-level folder in your workspace should indicate the

activity or the project on which you are working. Organize all of your work within the workspace folder. On your **desktop**, create a new folder and give it a descriptive name, such as **LeastCostPathModel**. Be sure there are no spaces. You may use underscores instead of spaces. Inside this folder, create the following three subfolders: *original*, *working*, and *final*. Having a standardized folder structure helps to keep a project organized, primarily when you are working with multiple partners. The folder structure you see here (**Figure 4.17**) is the standard used in each of the tutorials presented in this book.

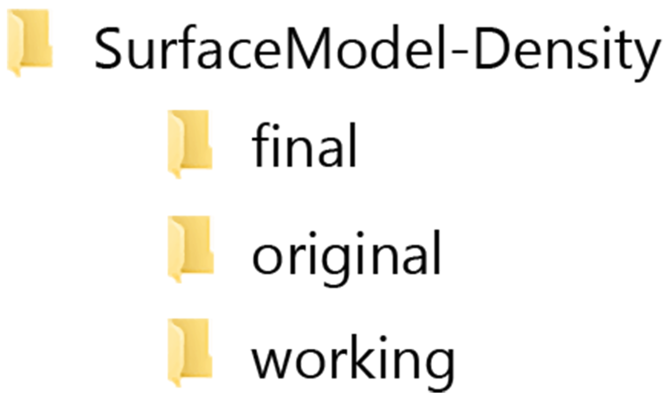


Figure 4.17: This diagram represents a basic folder structure used in this book.

As the name indicates, use the **original folder** for storing original, unaltered data. As you are working on a project, if, for some reason, your working version of the data gets lost or corrupted, you can go back to your *original* folder and find a fresh copy of the data. Use the **working folder** for data that you *create* or *alter* while working on your project. Use the **final folder** for storing any output you produce as a result of your work, such as images, maps, tables, or reports. Setting up a standard folder structure for a project is good practice and a habit you should develop.

Disable Background Geoprocessing

In the ArcGIS software, the *Background Geoprocessing* setting is often turned on by default. This setting allows users to continue to work while *a* tool is running in the background. However, sometimes this setting will stop tools from running or cause other unforeseen problems. To reduce that chances of the ArcGIS software crashing during this exercise, turn this setting off. In ArcMap, **open a new blank map document**. Open the *Geoprocessing options* from the *Geoprocessing* menu. Under *Background Geoprocessing*, uncheck the box next to the word *Enable* (**Figure 4.18**).

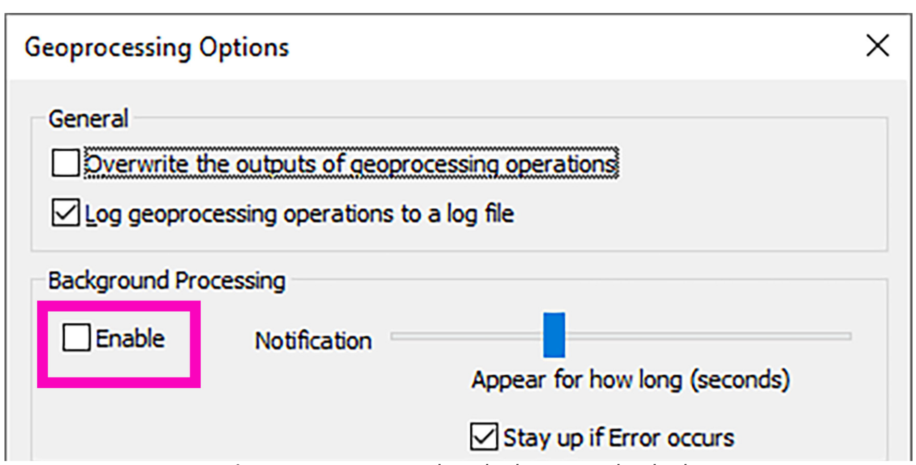


Figure 4.18: Be sure that the box is unchecked.

Spatial Analyst Extension

The steps in this activity involve using the *Spatial Analyst* extension. After launching ArcMap, make sure this extension is activated.

Skill Drill: Downloading the Tutorial Data

The Los Angeles County Sheriff's Department publishes crime data for download through the Los Angeles County GIS Data Portal. However, the crime data they provide does not contain XY values, such as Latitude and Longitude coordinates. For this reason, the author has prepared a series of feature datasets for download, which has been prepared and curated for use in a GIS. Use this link to download the ***Crime Data Geodatabase***[1]. Download the file and save it to your *original* folder (**Figure 4.19**). Decompress the file using ***7Zip***[2].

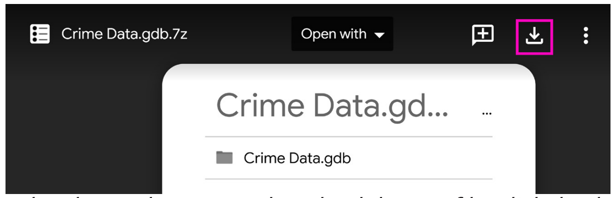


Figure 4.19: For the best results, use the Chrome browser to download the CSV file. Click the download button on the upper right.

Skill Drill: Creating a Working Version of the Geodatabase

A geodatabase is like a container that can hold many types of geospatial datasets. It is a way to organize and manage related datasets. There are several different types of geodatabases. The one covered here is called a **file geodatabase**, which can store individual datasets up to one terabyte (TB) in size.

Recall that the *original* folder is used for unaltered data, and the *working* folder is used for data that you alter or create. In this tutorial, you will be modifying file geodatabase. Thus, you must create a copy of the *Crime Geodatabase* in your *working* folder. In the *Catalog Window*, right-click the *Crime Geodatabase* and choose *Copy*. Then, right-click on your *working* folder and choose *Paste*. Rename version in the *working* folder *Crime Data Analysis.gdb* (**Figure 4.20**).

In ArcMap, you may need to refresh your original folder to see the geodatabase.

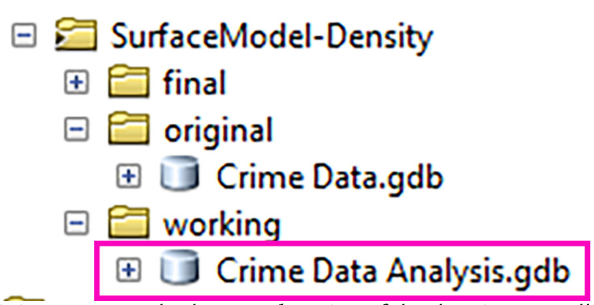


Figure 4.20: Maintaining an untouched *original* version of the data is an excellent habit to develop.

When saving files, ArcMap uses a default geodatabase as an output location. As a result, many readers encounter a situation where they forget to specify the output location, and data gets saved to the **default geodatabase**. This mistake may lead to lost data or other unforeseen problems. In this step, you set the default output location to the *Crime Data Analysis Geodatabase*. From the *File* menu, open the *Map Document Properties* window. Next to *Default Geodatabase*, click the yellow file folder icon to browse to your *working* folder (**Figure 4.21**).

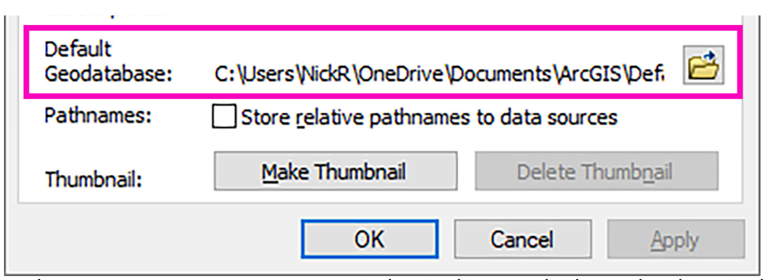


Figure 4.21: The Map Document Properties window indicates which geodatabase is the default.

1 URL: bit.ly/Crime_GDB
2 URL: https://www.7-zip.org/download.html

In the *Default Geodatabase* window, select the *Crime Data Analysis Geodatabase* and *click Add* (**Figure 4.22**).

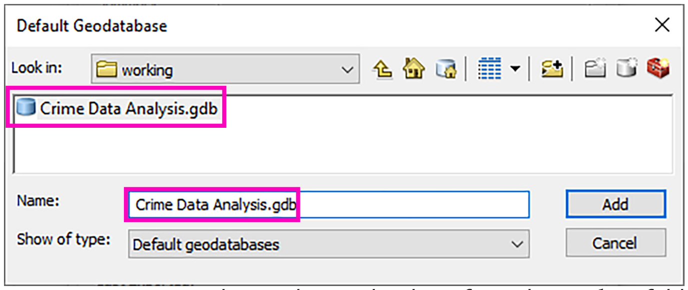


Figure 4.22: Be sure to choose the geodatabase from the *working* folder.

When you return to the Map Document Properties window, be sure to check the box next to *Store relative pathnames to data sources* (**Figure 4.23**). When ready, *click OK*. Then, take a moment to save your map document to your workspace folder.

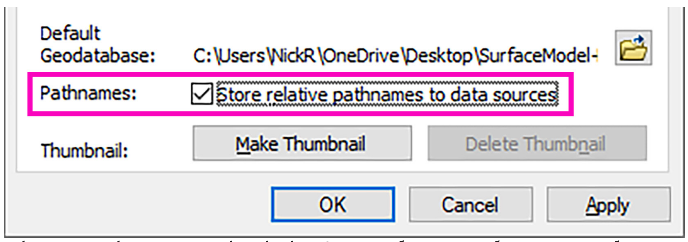


Figure 4.23: Always make sure to check the *Store relative pathnames to data sources* option.

Skill Drill: Adding a Geodatabase Feature Class to the Map

In the *Catalog Window*, expand the *Crime Data Analysis Geodatabase*. You should see a couple of feature datasets. A **feature dataset** is a collection of feature classes that use the same spatial reference system. Expand the *Geographic Regions* feature dataset, locate the feature classes, *California* and *Los Angeles County Subdivisions* (**Figure 4.24**).

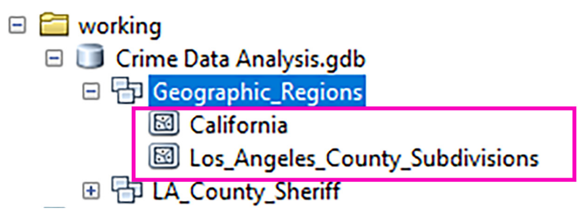


Figure 4.24: You can import several different data types into a file geodatabase, a feature class, a table, and a raster dataset.

In many ways, a **feature class** works similar to a shapefile. Like a shapefile, it is a collection of geographic features stored in vector format that have the same geometry type, such as point, line, or polygon. For most readers, the most noticeable differences relate to the file structure and file size. When stored in a file geodatabase, a feature class can save a maximum of **one terabyte** of data. A stand-alone shapefile has a size limit of only **two gigabytes (GB)**. Add the *California* and *Los Angeles County Subdivisions* feature classes to the *Table of Contents* (**Figure 4.25**).

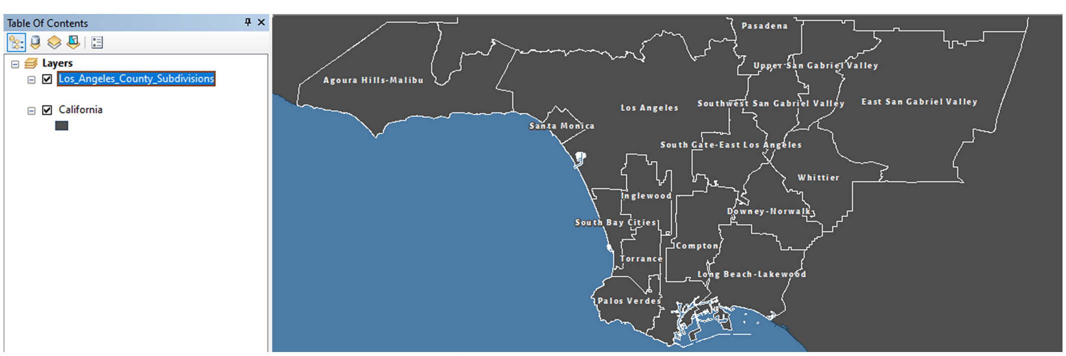


Figure 4.25: The California layer in the background provides a spatial context for Los Angeles County.

Changing Global Environment Settings for Raster Processing

Change the global environment settings so that the raster processing you will perform in later steps will be constrained to a specific geographic extent. In this instance, you want all raster processing to match the extent of the Los Angeles County boundary. In ArcMap, select *Geoprocessing* from the main menu, then select *Environments* (**Figure 4.26**).

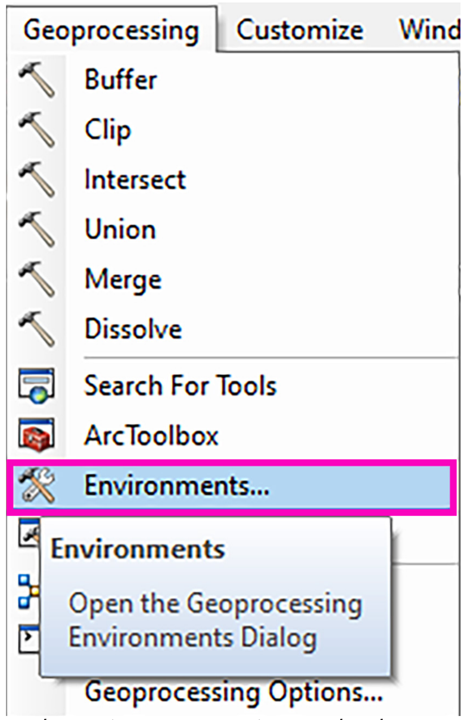


Figure 4.26: Locate the Environments options under the Geoprocessing menu.

This step will open up the *Environment Settings* window. Here you will define the *Processing Extent* **globally**. Any changes you make here will affect the processing extent for *nearly* all the tools you will use later. Under *Processing Extent*, use the drop-down menu to select *Same as layer Los_Angeles_County_Subdivisions* (**Figure 4.27**).

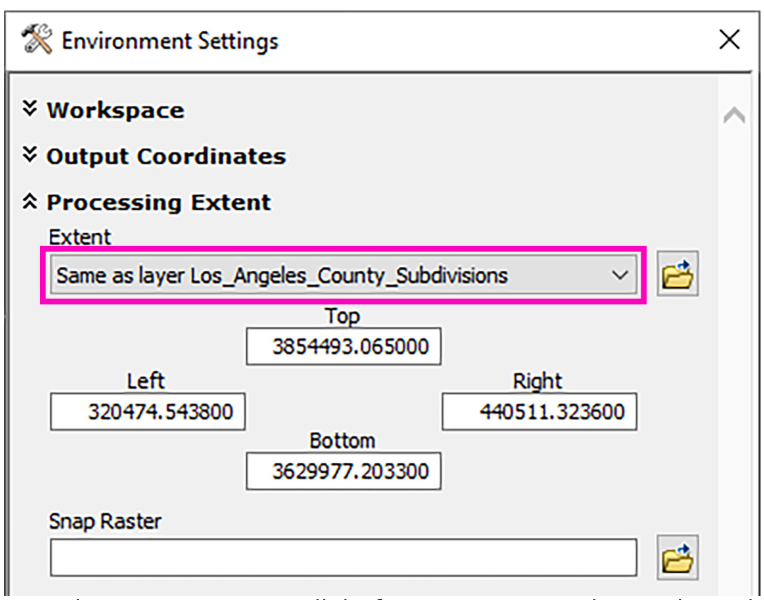


Figure 4.27: The processing extent will clip future operations to the DEM layer. The Snap Raster will align the cells in a raster dataset output to the DEM layer as well.

Under *Raster Analysis,* set the *Cell Size* setting to *Minimum of Inputs.* Set the *Mask* settings to the *Los Angeles County Subdivisions* (**Figure 4.28**). By default, raster outputs are rectangular. With this setting, the output of the analysis will fill only the shape of Los Angeles County. Leave all other settings as default and *click OK*.

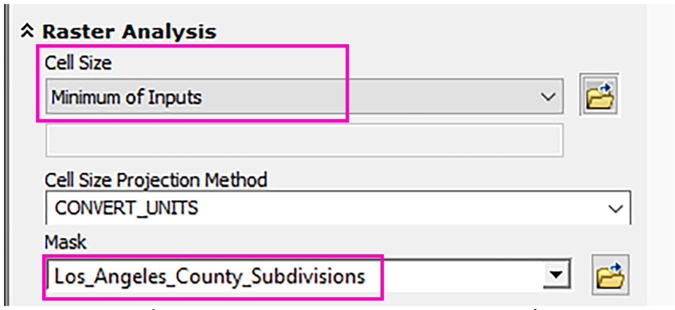


Figure 4.28: Be sure your settings match.

Developing a Search Neighborhood

A **density surface model** calculates the density of a phenomenon by using known values within a user-defined **search neighborhood**. When using the **Simple Method**, the cell values in a density surface model are a function of the **number of data points** in the search neighborhood **divided by the area** of the search neighborhood. The search neighborhood you define will have the most significant impact on the resulting surface. A broader search neighborhood will result in a more generalized raster surface. If the search neighborhood is too small, the results will obscure the patterns in the data. Because the size of the search neighborhood depends on both the dataset and the phenomenon in question, it is the **most arbitrary factor** when creating a density surface model. Therefore, the results of a density surface model are **highly subjective**. Determining the most appropriate search neighborhood for your dataset can take time, research, a good understanding of the data, and some trial and error.

In this instance, you have a history of crime events in Los Angeles County by year, ranging from 2005 to 2015. Start by adding the 2005 dataset to the map and make a visual inspection of the data to determine if there are any visible patterns. In the *Catalog Window*, expand the LA County Sheriff feature dataset. Add the 2005 crime data to the map (**Figure 4.29**).

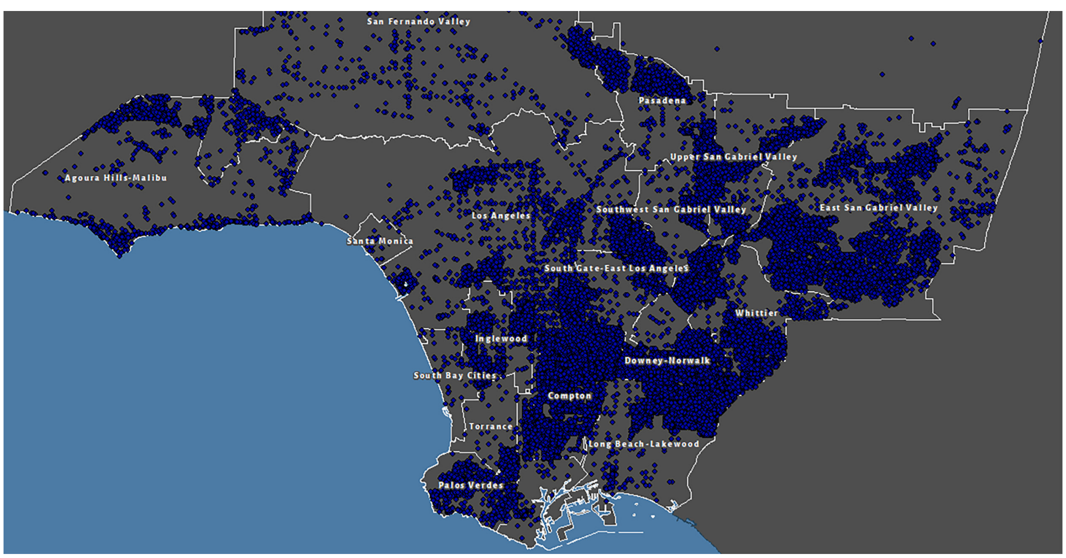


Figure 4.29: The LA Crime 2005 feature class contains over 190,000 data points representing different crimes.

At first glance, the number of points for the 2005 crime dataset may be overwhelming. With so many data points, it is hard to find the nuances that relay useful information. However, the human brain is remarkable when it comes to finding patterns. One exercise that will help to define a search neighborhood is to manually circle what appears to the eye to be clusters of crime events. In ArcMap, activate the *Draw* toolbar. Use the drop-down menu next to the *Rectangle* tool and activate the *Ellipse* tool (**Figure 4.30**).

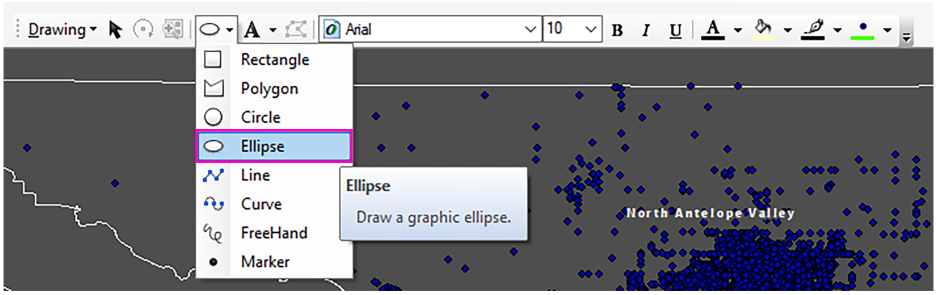


Figure 3.30: The draw toolbar has many tools from which to choose. In this image, the *Ellipse* tool is highlighted.

Zoom in and out of the data frame as needed, and several draw ellipses around what you see as the most distinct clusters of data (**Figure 4.31**). Change the fill and outline colors so that you can see the crime events. Resize and rotate[1] the ellipses as needed. Don't spend more than five or ten minutes delineating clusters. The goal is to get a general sense of the size of the most distinct clusters to help determine the size of the search neighborhood.

Don't over-think this exercise! Do what feels natural to you.

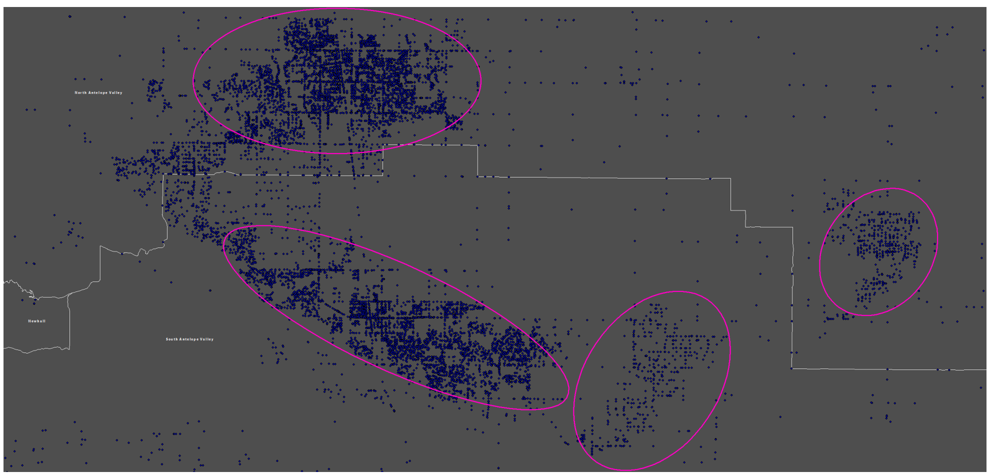


Figure 4.31: Several ellipses drawn over distinct clusters of crime events.

Once you are satisfied that you have a good sample of cluster sizes. Use the *measure* tool from the *Tools* toolbar to find the largest diameter of each cluster. Click one end of the ellipse, then double-click the other end to take a measurement (**Figure 4.32**). In this example, many of the most significant clusters fell between **ten** to **twenty thousand** meters across the widest dimension.

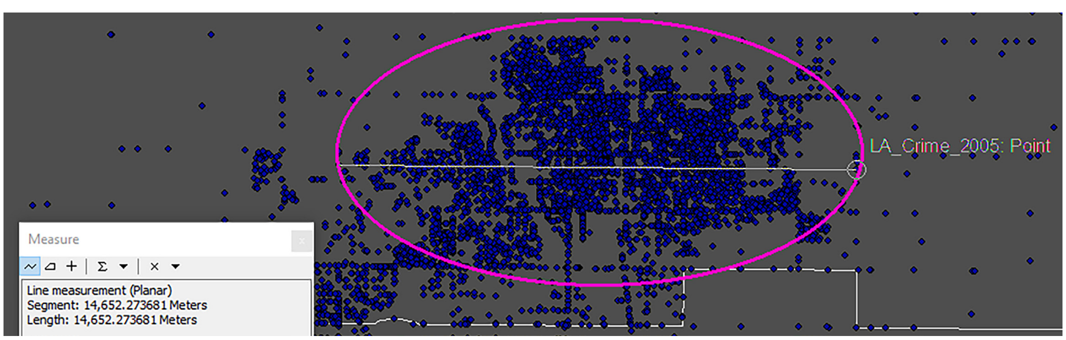


Figure 4.32: Locate the *Measure* tool on the *Tools* toolbar.

A search neighborhood between ten to twenty thousand meters would be too large to find any nuances within the cluster. As a rule of thumb, **divide** the average width of the clusters **by ten** as a starting point for defining the search neighborhood. Based on the measurements in this example, the first search neighborhood will be **1,000 meters**. You may delete your ellipses before moving on to the next step.

1 You can find the *Rotate* tool to the left of the *Ellipse* tool on the *Draw* toolbar..

Creating a Density Surface Model Using the Simple Method

The *Point Density* tool uses the *Simple Method* for creating a density surface. From the *Spatial Analyst* tools under *Density*, open the *Point Density* tool (**Figure 4.33**). For the *Input Point Features*, choose the *LA Crime 2005* feature class. Some tools, like *Point Density,* will not save to a geodatabase. Instead, save the *Output raster* to your *working* folder. Call the file *Crime_2005_Simple_1000m.tif*.

You will import one of the raster datasets you create into your geodatabase at a later time.

For the Output cell size, enter **30**. The smaller the cell size, the smoother the surface will look. However, the file size will also be more substantial. In this example, **30 meters** is a good balance given the size of Los Angeles County. Under *Neighborhood*, choose *Circle* as the shape. Next, to the *Radius*, enter **1000**. Next to *Area units*, change the unit of measurement to **SQUARE METERS**. Leave all other settings as default and *click OK*.

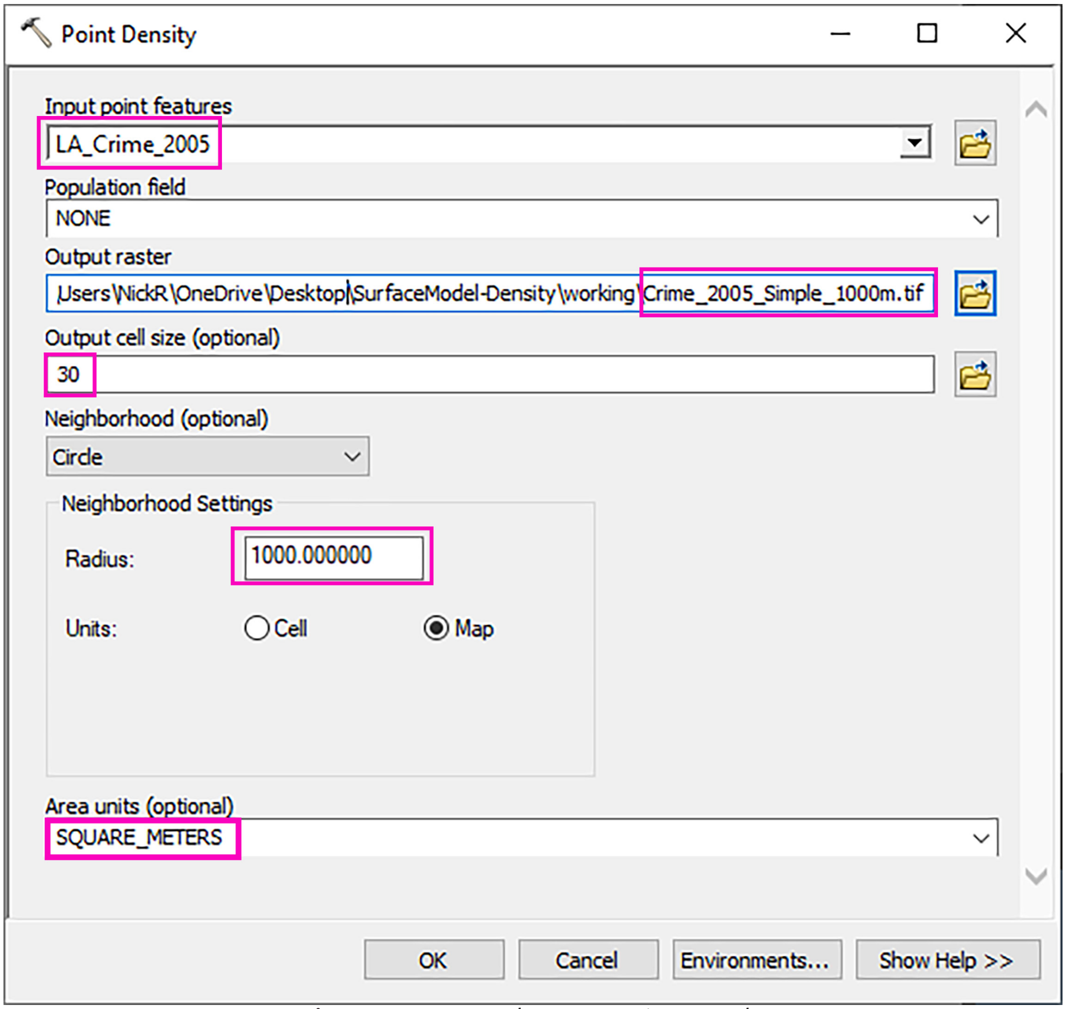


Figure 4.33: Be sure that your settings match.

The density surface model may appear at the bottom of the *Table of Contents*. Drag the density layer above the California layer, but below Los Angeles County. If you have not done so already, remove the fill color from the Lost Angeles County layer, so the density surface model is visible (**Figure 4.34**). Take a moment to compare the darker regions of blue with the crime events by checking and unchecking the *Crime 2005* layer in the *Table of Contents*. For your personal understanding, consider whether or not the dark blue regions match your mental clustering of the dataset. Were the results what you expected?

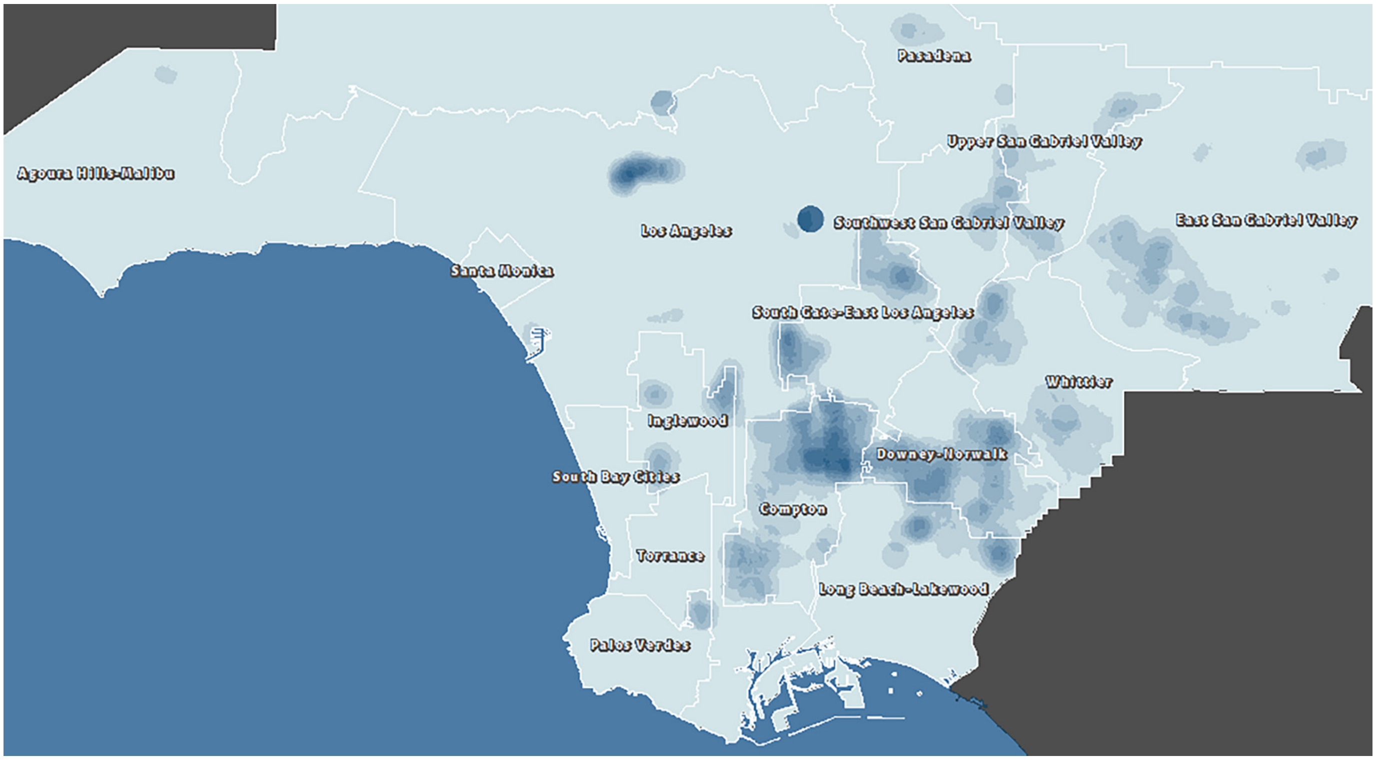


Figure 4.34: By default, a monochromatic blue color scheme gets used to symbolize the output.

The default color scheme for density surfaces is monochromatic. However, many cartographers often use diverging color schemes to represent hot spots, such as a blue-yellow-red color scheme. Diverging color schemes can sometimes reveal additional nuances within the density patterns. However, one must consider the map audience when choosing colors. Relying too much on color schemes that use both reds or greens may pose a problem to map readers with color-vision impairments. I recommend using color schemes that take advantage of brown, blue, yellow, teal, orange, white, and purple.

Open the layer properties for the *Crime_2005_Simple_1000m* layer and navigate to the *Symbology* tab. Choose the color scheme that you feel helps you to understand the clusters and patterns of the crime events from 2005 while keeping your color-impaired map readers in mind. In this example, I have chosen the dark brown, orange, and gold *Universe Mars 1* color scheme (**Figure 4.35**).

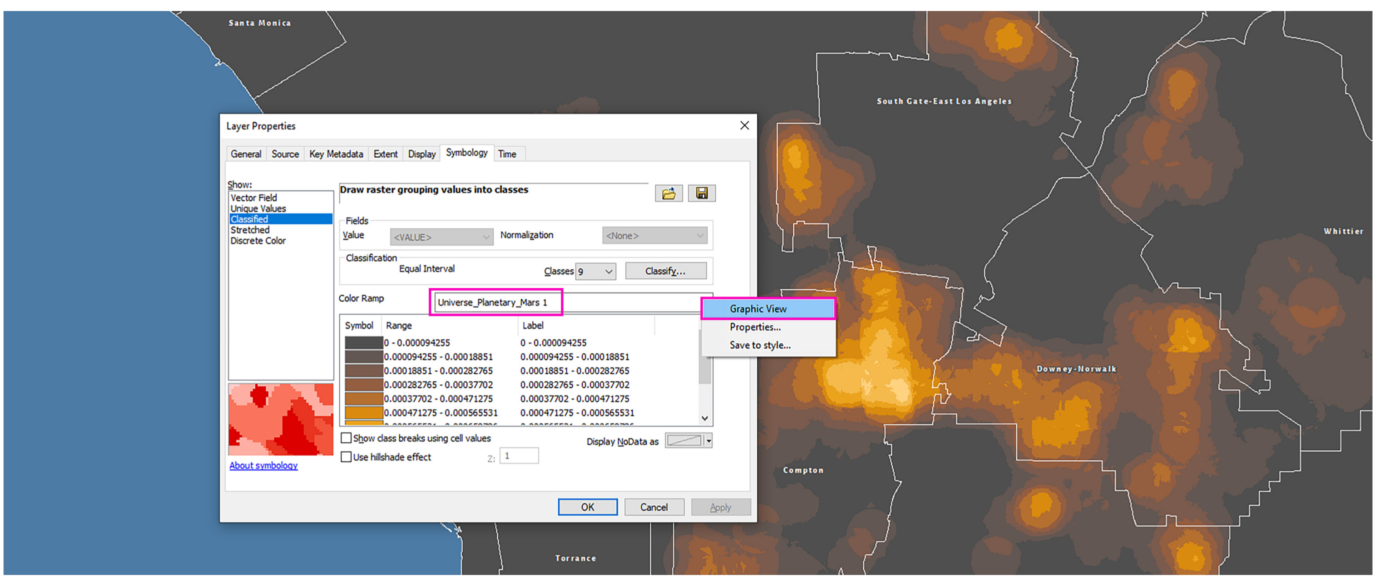


Figure 4.35: One can view the names of the color schemes by right-clicking on the color ramp and unchecking *Graphic View*.

Skill Drill: Using Different Search Neighborhoods

Repeat the steps for creating a density surface model using the *Simple Method*. First, use a search neighborhood of **1500** meters. Next, use a search neighborhood of **2000** meters. Save the output to your *working* folder. Change the color scheme for both so that they match your first density surface model. Compare each model and make a personal judgment on **which search neighborhood works best** to understand crime patterns in Los Angeles County (**Figure 4.36**).

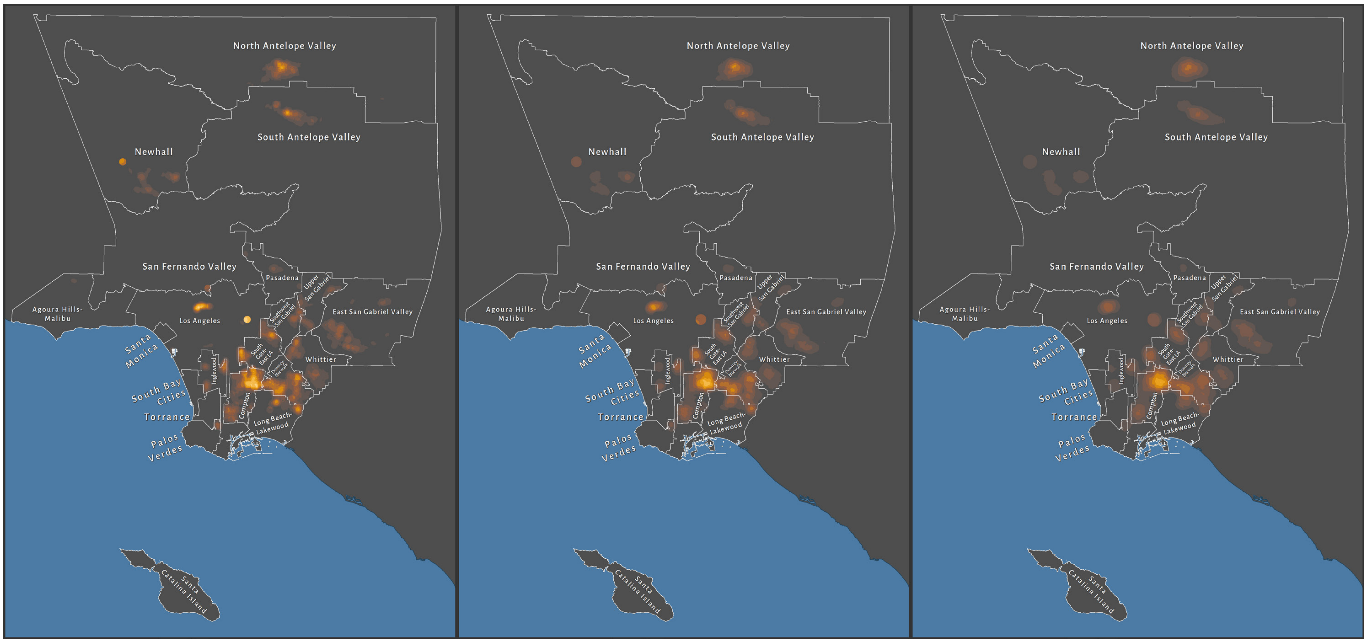


Figure 4.36: Each of the three simple density surface models side-by-side. From left to right, the search neighborhoods are 1000, 1500, and 2000 meters.

When done, right-click on the *Crime Data Analysis Geodatabase* and import the density surface model you like **best**.

Be prepared to explain your decision.

Creating a Density Surface Model Using the Kernel Method

When using the **Kernel Method**, the cell values in a density surface model are a function of fitting a bell-shaped surface, called a **kernel**, over each point. Over the center of the data point, the kernel has a value of **1** (**Figure 4.37**). As the surface of the kernel progresses towards the outer edge of the search neighborhood, the value gradually decreases to a value of **0**.

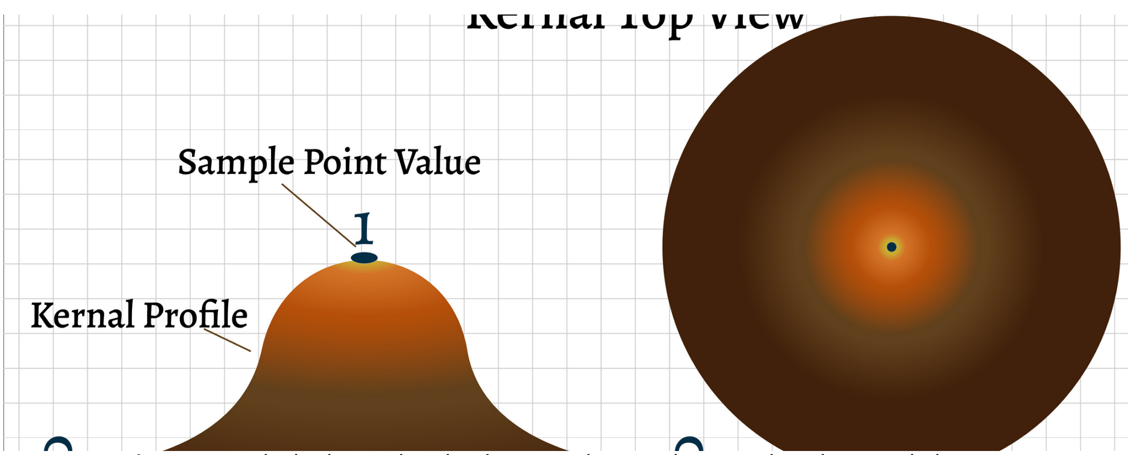


Figure 4.37: The highest values land nearest the sample point, then decay with distance.

The Kernel Method calculates the density values by adding the kernel surfaces together where they overlap the center of each cell (**Figure 4.38**). The sum of the kernel surfaces is then divided by the area of the search neighborhood.

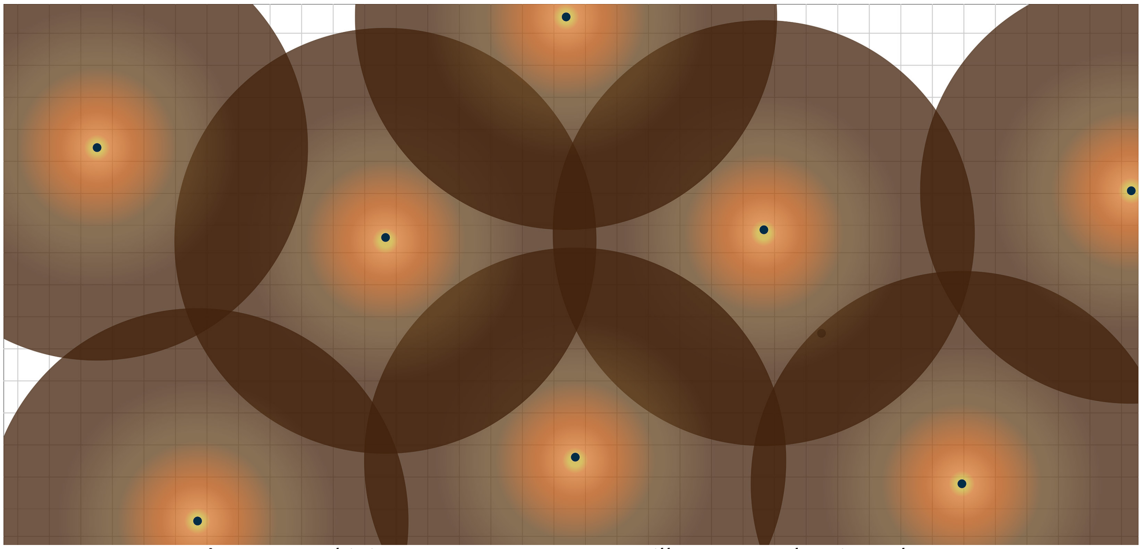


Figure 4.38: This image uses transparency to illustrate overlapping values.

From the *Spatial Analyst* Tools under *Density*, open the *Kernel Density* tool (**Figure 4.39**). For the *Input point or polyline features*, choose the *LA Crime 2005* feature class. Save the *Output raster* to your *working* folder. Name the file *Crime_2005_Kernel_1000m.tif*.

You will import one of the raster datasets you create into your geodatabase at a later time.

For the *Output cell size*, enter **30**. Next, to the *Search radius*, enter **1000**. As with the Point Density tool, you will eventually compare a search radius of **1000**, **1500**, and **2000 meters**. Next to *Area units*, change the unit of measurement to **SQUARE METERS**. Leave all other settings as default and *click OK*.

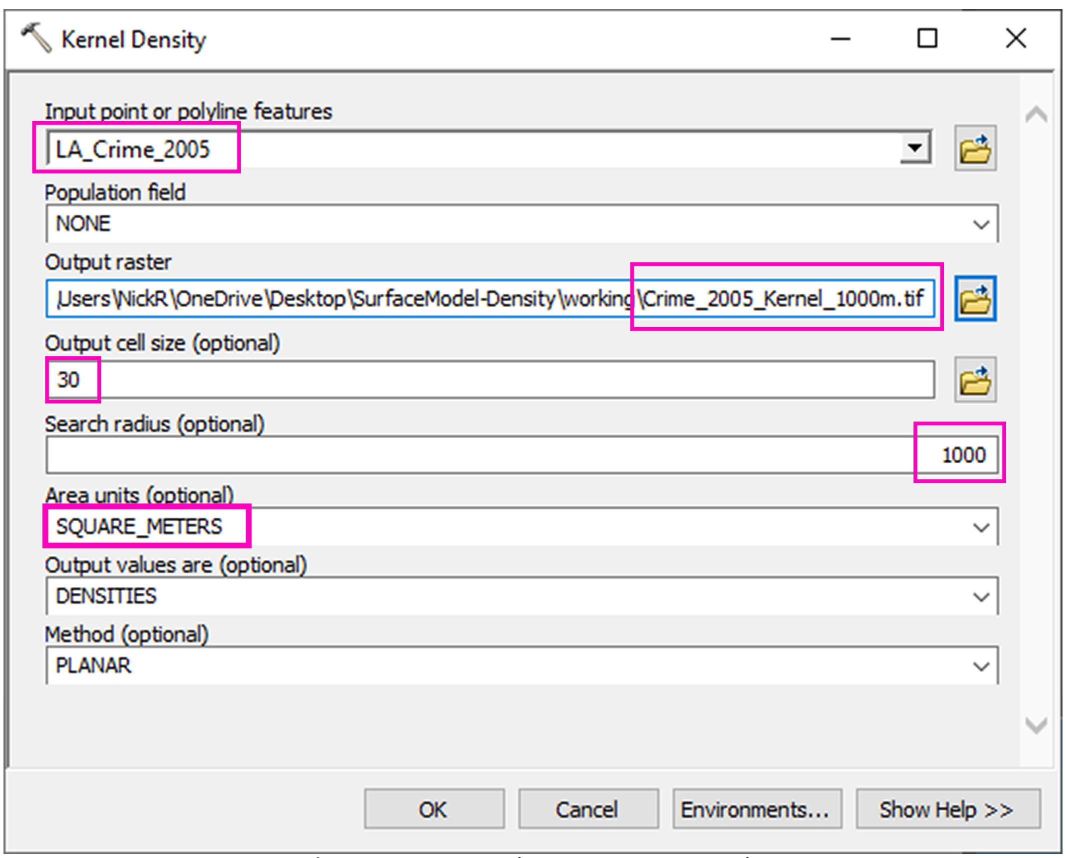


Figure 4.39: Be sure that your settings match.

The density surface model should appear in the *Table of Contents*. Drag the new density layer above the others if necessary layer, but below Los Angeles County (**Figure 4.40**). Match the symbology and color scheme with your previous density models. Compare the results of the Kernel Method for calculating density compared to the *Simple Method* by checking and unchecking layers in the *Table of Contents*.

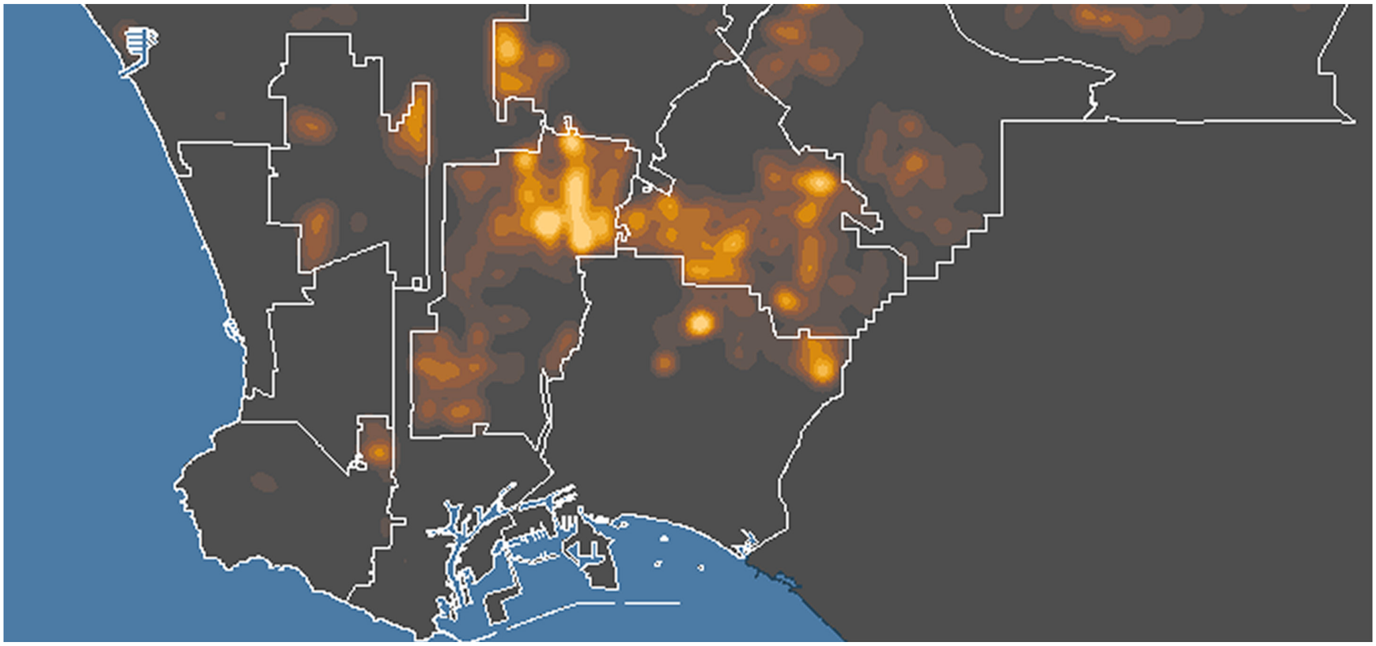


Figure 4.40: The *Kernel Method* for calculating density often results in a smoother surface than the *Simple Method.*

Skill Drill: Using Different Search Neighborhoods

Repeat the steps for creating a density surface model using the *Kernel Method*. First, use a search neighborhood of **1500** meters. Next, use a search neighborhood of **2000** meters. Save the output to your *working* folder. Change the color scheme for both so that they match your first density surface model. Compare each model and make a personal judgment on **which search neighborhood works best** to understand crime patterns in Los Angeles County (**Figure 4.41**).

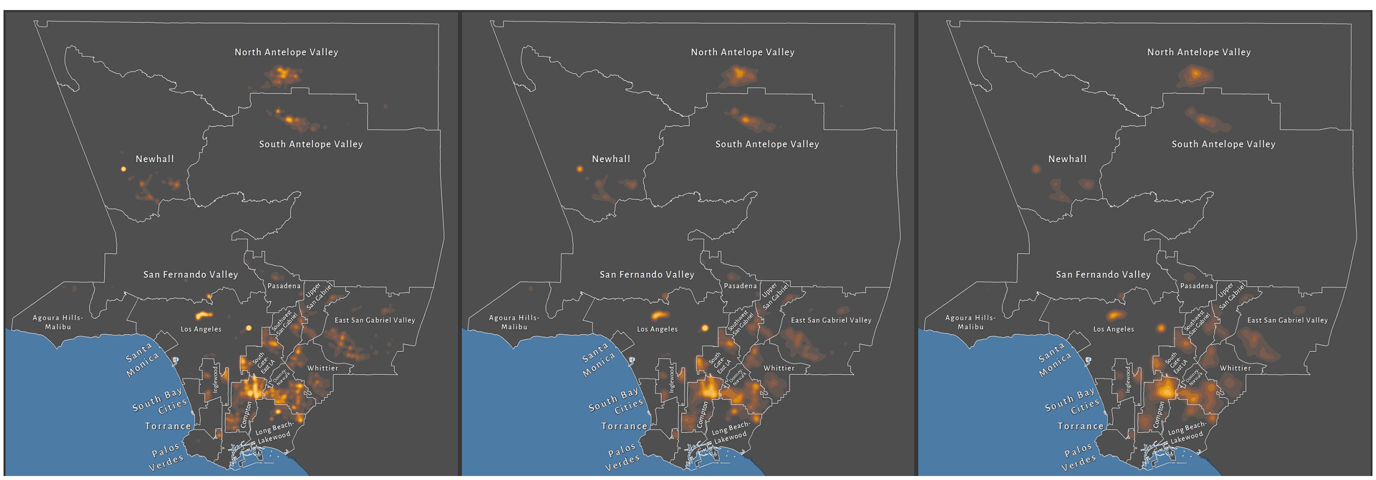


Figure 4.41: Each of the three simple density surface models side-by-side. From left to right, the search neighborhoods are 1000, 1500, and 2000 meters.

When done, right-click on the *Crime Data Analysis Geodatabase* and import the density surface model you like **best.**

Be prepared to explain your decision.

Creating a Crime-Category Subset of the Data

While it may be useful to map overall crime across Los Angeles County, the dataset can also provide answers to questions about specific types of crimes. First, you must find out how many individual crime categories there are in the *LA Crime 2005* dataset. Open the attribute table for the *LA Crime 2005* layer. Right-click the *CATEGORY* field and select *Summarize* (**Figure 4.42**).

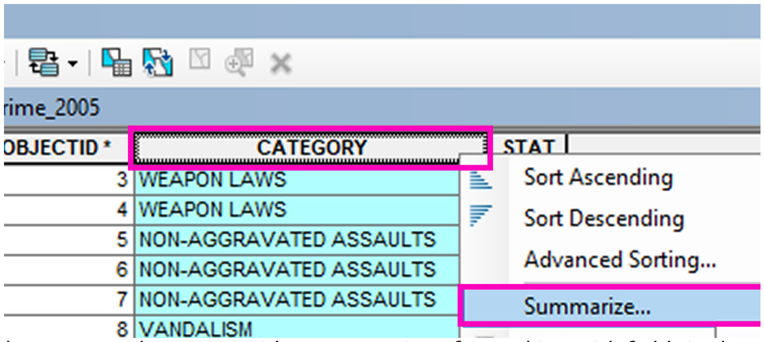


Figure 4.42: The contextual menu provides many options for working with fields in the attribute table

Under *Specify output table* in the *Summarize Window*, browse to your *Crime Data Analysis Geodatabase* (**Figure 4.43**). Name the file *Crime 2005 Categories*. Under *Save as type*, use the drop-down menu to select *File Geodatabase Tables*. When ready, *click Save*. Then, *click OK*.

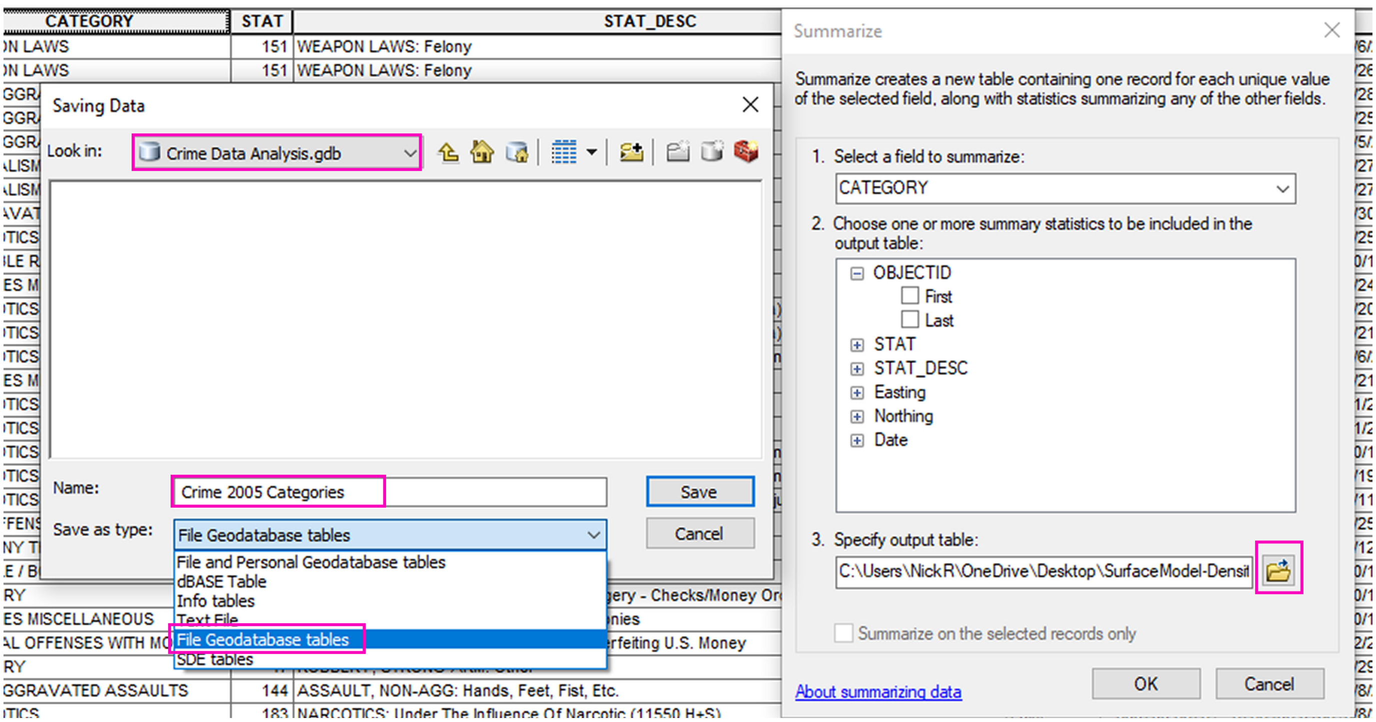


Figure 4.43: The summary table will be written to the *Crime Data Analysis Geodatabase*.

After the summarize is completed, *click No* when asked if you want to add the table to the map (**Figure 4.44**). You will only be looking at the table and will not use it for analysis.

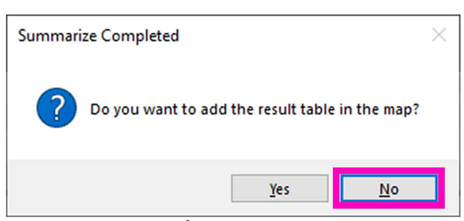


Figure 4.44:

In the *Catalog Window*, right-click on the *Crime 2005 Categories* table and choose *Item Description* (**Figure 4.45**)

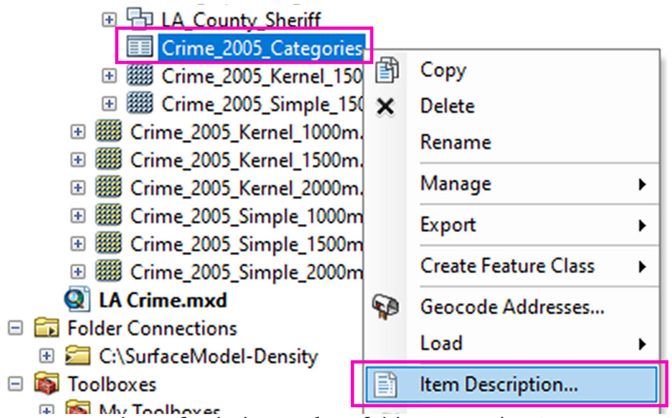


Figure 4.45: You may need to refresh the *working* folder to see the *Crime 2005 Categories* table

When the *Item Description* window opens, select the *Preview* tab. You should see a table with a list of each unique crime category (**Figure 4.46**). Review the different crime categories. Choose **one** category to use for the next step. When you have chosen your category, close the *Item Description* window.

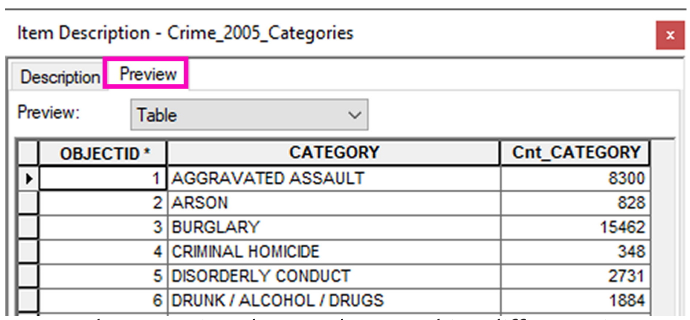


Figure 4.46: In the 2005 crime dataset, there are thirty different crime categories.

In ArcMap, run an attribute query to select all of the records for your chosen crime category. Export the selected records and save it as a feature class in the *LA County Sheriff* feature dataset. Name the file using the category you chose, followed by the number 2005. For example, if you chose grand theft auto as your crime category, then your feature class should be called *grand_theft_auto_2005*. If you chose arson as your crime category, then your feature class should be called *arson_2005*. When ready, add your crime category feature class to the *Table of Contents*. Use either the *Kernel Method* or the *Simple Method* to create a new density surface model for your chosen category. When you are satisfied with your density surface model, take a moment to save your map document.

You may need to tweak the search neighborhood to accommodate the changes in crime event distribution and clustering.

Skill Drill: Creating a Map of the Results

You should be familiar with the steps needed to create a map layout of the results of your density surface model for your chosen crime category. Design a map for use as a figure in a report or summary. Ideally, the map should be designed at a size of approximately **5 inches wide** and **7 inches tall**. Include a north arrow, a scale bar, and a legend. Change the legend labels for the crime density so that you **remove all of the numbers**. Instead, use the words low, medium, high to label the lowest, middle, and highest categories.

Hint: You can do this in the Symbology tab of the layer properties.

Remove the point layers for the crime events so that they do not obscure the results. Also, be sure to label each of the LA County Subdivisions (**Figure 4.47**).

It may be best to manually place the labels using the Text tool on the Draw toolbar.

When your map layout is complete, export the map as a PNG file with a resolution of **300 dpi**. Save the file in your *final* folder.

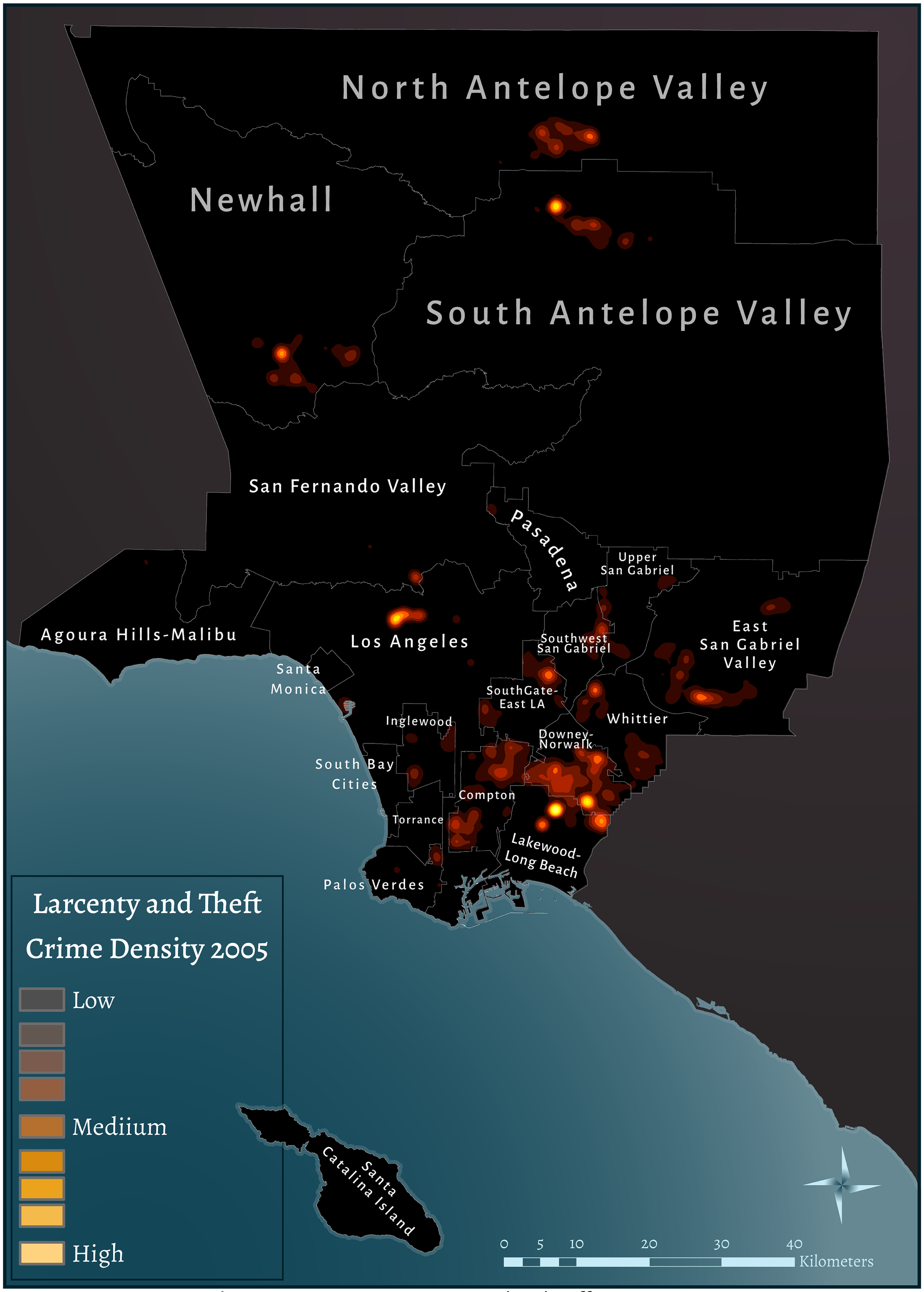


Figure 4.47: Data Source: Los Angeles Sheriff's Department

Skill Drill: Compare 2005 Crime Category to a Different Year.

Repeat the steps by choosing a different year. Recall that the *LA County Sheriff* feature dataset contains crime data from 2005 to 2015. Create a feature class for your chosen crime category and use the same methods to create a density surface model. Create a map of the results that uses the same symbology as your 2005 map. When your map layout is complete, export the map as a PNG file with a resolution of **300 dpi**. Save the file in your *final* folder.

Tutorial: Creating a Digital Elevation Model from GPS Data Using Interpolation Methods

Interpolation allows you to create a surface from point locations by predicting values between known values. For example, a temperature surface may be interpolated from a series of weather stations across the state of California. A surface model using interpolation attempts to **predict** the specific temperature values at **all locations** across the entire surface.

Estimated time to complete this tutorial: 6 hours

Learning Outcomes

Readers should be able to accomplish the following outcomes by the end of this tutorial:

- Review data collection methods using a GPS receiver
- Review downloading waypoints
- Review converting GPX files to Shapefiles
- Create a surface model using the Inverse Distance Weighted (IDW) method
- Create a surface model using the Spline method
- Create a surface model using the Kriging method
- Optimize inputs for interpolation methods
- Compare interpolation results

Scenario

In this scenario, you have been charged with the task of collecting elevation data to create a digital elevation model. You will generate points for the surface model by collecting elevation data using a GPS receiver. You will then compare the accuracy of the surface model using various interpolation methods.

Conduct this analysis using the **Universal Transverse Mercator (UTM)** system along with the **North American Datum of 1983 (NAD83)**. Determine the correct UTM zone for your location (**Figure 4.48**).

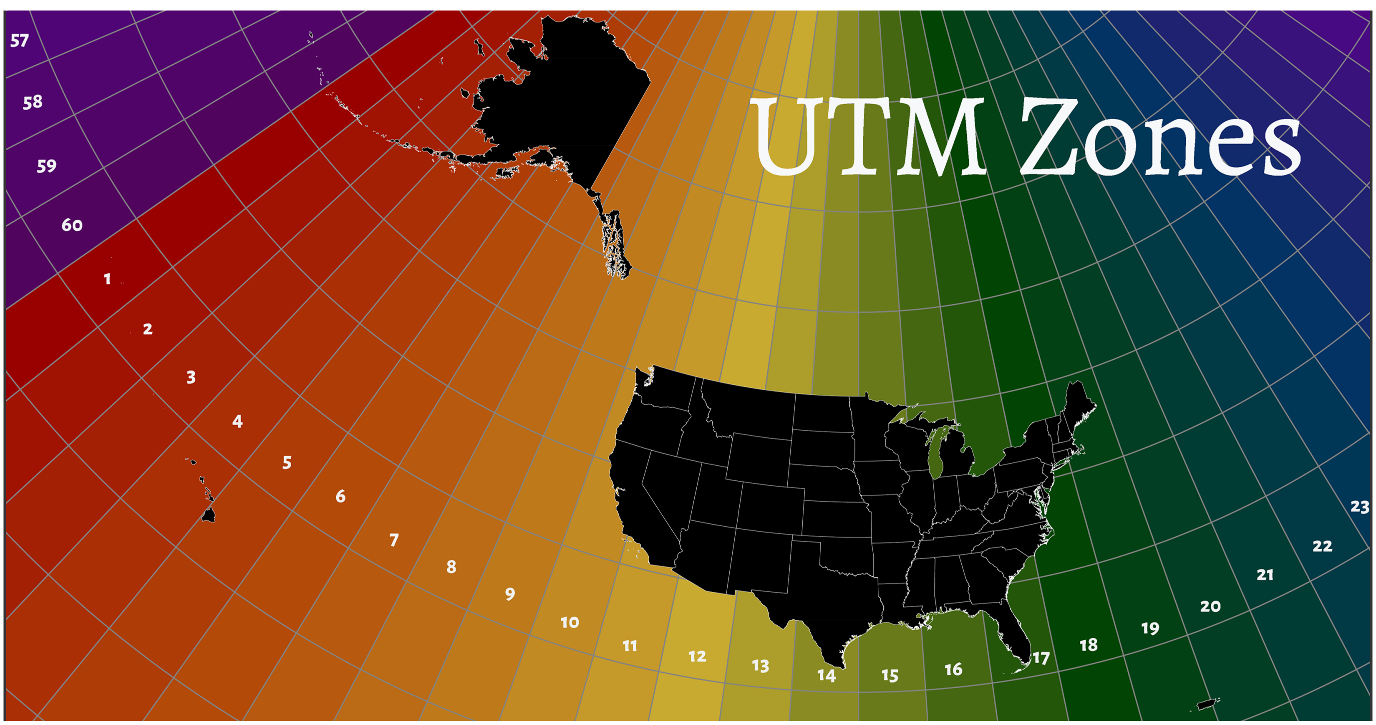


Figure 4.48: UTM Zones for the United States

Setting up Your Workspace

In a typical workflow, you work on geospatial data using a local hard drive. When done, you compress your data and back up your work to your cloud storage so that you can retrieve the files from anywhere. When referring to a **local hard drive**, it means you are working on data physically located on the computer in front of you.

In contrast, some computers also include networked drives. **Networked drives** link to cloud storage and save the data elsewhere. Examples include services like OneDrive or Google Drive. For this tutorial, use the **desktop** as your local hard drive location. You may also use an external USB drive if you plan to work in multiple places.

You must avoid using networked drives while you work.
They increase the processing time and can cause technical glitches.

In this book, you use a particular folder structure. Start by creating your workspace folder on the local hard drive. A **workspace** is a folder or series of folders that contain all of your project files. The top-level folder in your workspace should indicate the activity or the project on which you are working. Organize all of your work within the workspace folder. On your **desktop**, create a new folder and give it a descriptive name, such as **DEMInterpolation**. Be sure there are no spaces. You may use underscores instead of spaces. Inside this folder, create the following three subfolders: *original*, *working*, and *final*. Having a standardized folder structure helps to keep a project organized, primarily when you are working with multiple partners. The folder structure you see here (**Figure 4.49**) is the standard used in each of the tutorials presented in this book.

Figure 4.49: This diagram represents a basic folder structure used in this book.

As the name indicates, use the **original folder** for storing original, unaltered data. As you are working on a project, if, for some reason, your working version of the data gets lost or corrupted, you can go back to your *original* folder and find a fresh copy of the data. Use the **working folder** for data that you *create* or *alter* while working on your project. Use the **final folder** for storing any output you produce as a result of your work, such as images, maps, tables, or reports. Setting up a standard folder structure for a project is good practice and a habit you should develop.

Spatial Analyst and Geostatistical Extensions

In ArcMap, **open a new blank map document**. The steps in this activity involve using both the *Spatial Analyst* extension and the *Geostatistical Analyst* extension. After launching ArcMap, make sure both extensions are activated. You will also need to open the *Geostatistical Analyst Toolbar* and dock it near the top of your window for easy access (**Figure 4.50**).

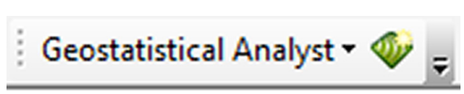


Figure 4.50: Activate the toolbar by right-clicking on the blank area along the top of the ArcMap window.

Disable Background Geoprocessing

In the ArcGIS software, the *Background Geoprocessing* setting is often turned on by default. This setting allows users to continue to work while *a* tool is running in the background. However, sometimes this setting will stop tools from running or cause other unforeseen problems. To reduce that chances of the ArcGIS software crashing during this exercise, turn this setting off. Open the *Geoprocessing options* from the *Geoprocessing* menu. Under *Background Geoprocessing*, uncheck the box next to the word *Enable* (**Figure 4.51**).

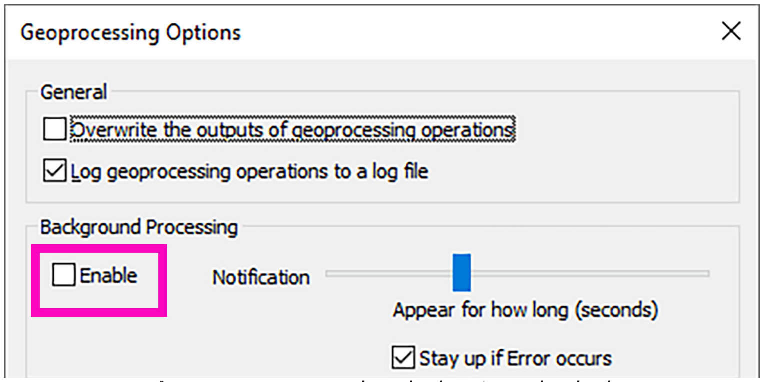


Figure 4.51: Be sure that the box is unchecked.

Skill Drill: Collecting Elevation Data Using a GPS Receiver

Choose a location available to you, which contains significant changes in elevation. A football or soccer stadium may be a good site to choose if there are no hills nearby. It contains varying elevation with **few GPS signal obstructions** (**Figure 4.52**).

Figure 4.52: In this image, GPS waypoints were collected in the Humboldt State University Redwood Bowl.

Readers that **do not have access to a GPS receiver**, or who cannot find a suitable location may use the waypoints depicted in **Figure 4.52**. Use the following link to download the data: ***GPS Waypoints file***[1]. Alternatively, some cell phone applications also allow users to mark and download waypoints. However, delving into specific smartphone applications for this purpose is beyond the scope of this tutorial.

1 URL: https://bit.ly/GPS_Waypoints

Decide where you will begin your data collection and where you will end. Plan how and where you will travel during data collection. For the best results, your waypoints should be distributed as evenly as possible across your area of interest. Avoid clustering your data points if possible. A useful method for doing this is to mark a waypoint about every **5 or 10 strides**. Using a GPS unit, try to collect between **300** and **500** waypoints with elevation data. Optimize the accuracy and efficiency of the GPS receiver if necessary. Specifically, you should make sure your GPS Unit is **WAAS** enabled (**Figure 4.53**). Also, be aware of any obstructions.

300 to 500 points may seem like a lot. However, the number of waypoints add up quickly as you walk around your location.

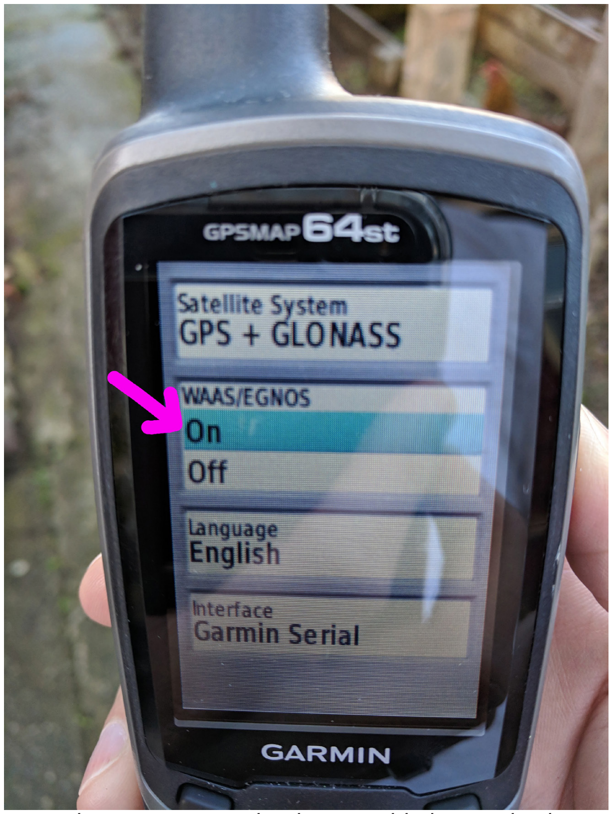


Figure 4.53: In this image, WAAS has been enabled using the device settings.

Skill Drill: Downloading GPS Data and Creating a Shapefile

Depending on the model of your GPS receiver or if you used a smartphone, there are different methods for downloading the waypoints. Most GPS receivers can plug into a computer and act like a standard USB storage device (**Figure 4.54**).

Figure 4.54: In this image, a Garmin GPSMap 64 is plugged into a computer tower.

Often, there will be a folder within the device directory called *waypoints*, *GPX*, or something similar which stores GPX files. A **GPX** file is a table that stores the waypoint X Y coordinates and other values, such as elevation. In ArcMap, you can use the *GPX to Features* tool to convert GPX files to shapefiles (**Figure 4.55**).

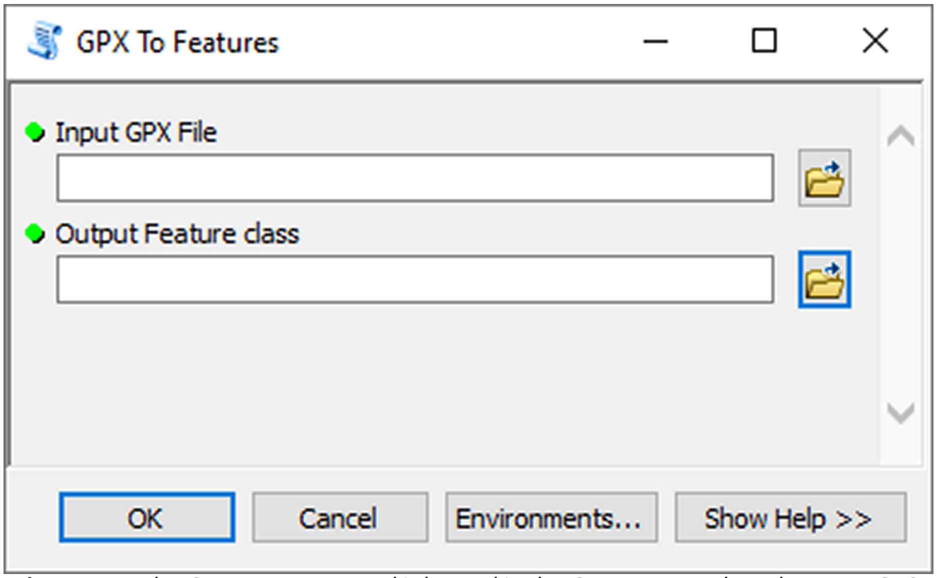


Figure 4.55: The *GPX to Features* tool is located in the *Conversion* tools under *From GPS*.

Once you have downloaded your waypoints and converted the GPX to a shapefile, use the *Project* tool to save a copy of the data to your *working* folder. Use the **NAD 83 UTM Zone 10** spatial reference. Add the projected waypoints to the *Table of Contents*.

You may choose a different UTM zone if you are not located in Zone 10.

Skill Drill: Downloading NAIP Imagery from the USGS Earth Explorer

Navigate to the ***USGS Earth Explorer website***[1] and log in. If you don't already have an account, you can register for one for free using the links provided near the top of the page. Use the map and your mouse to zoom to your region of interest. Click on the map to place a pin at your desired location. Then, click the *Data Sets* button (**Figure 4.56**).

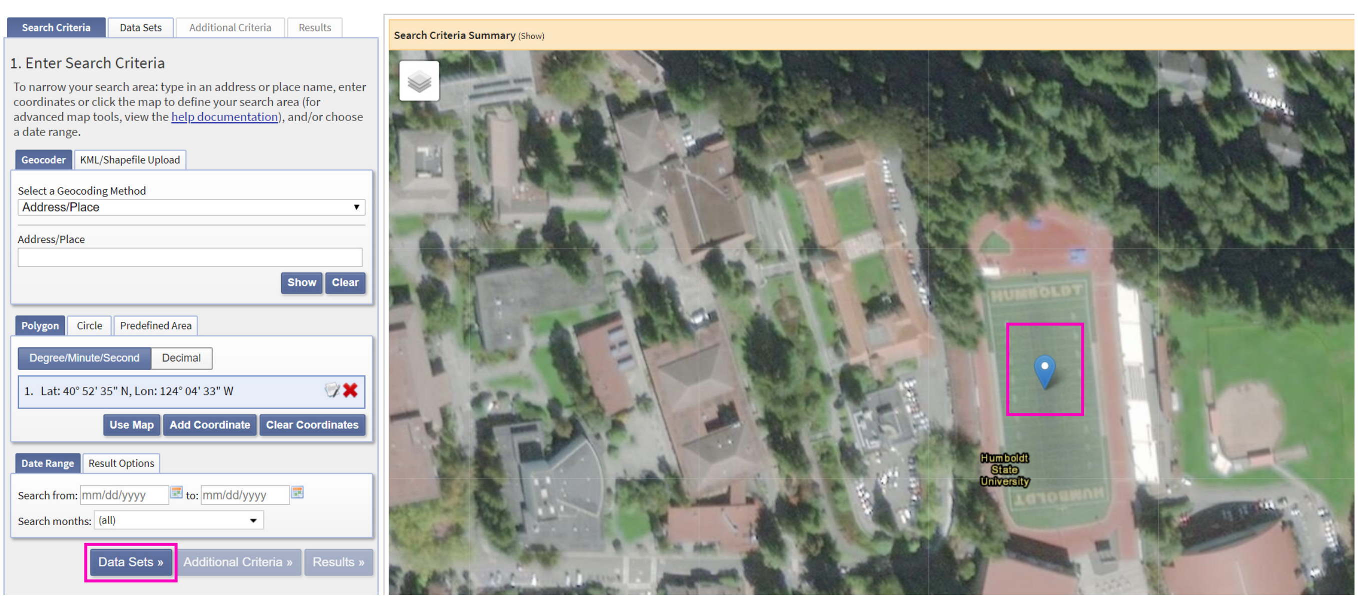


Figure 4.56: You can place a marker in an area of interest.

1 URL: http://earthexplorer.usgs.gov/

On the *Data Sets* tab, expand the *Aerial Imagery* and check the box next to **NAIP**. Then, click the *Results* button (**Figure 4.57**).

Note: The USGS changes the website from time to time. If your webpage does not match, do your best to locate the NAIP imagery.

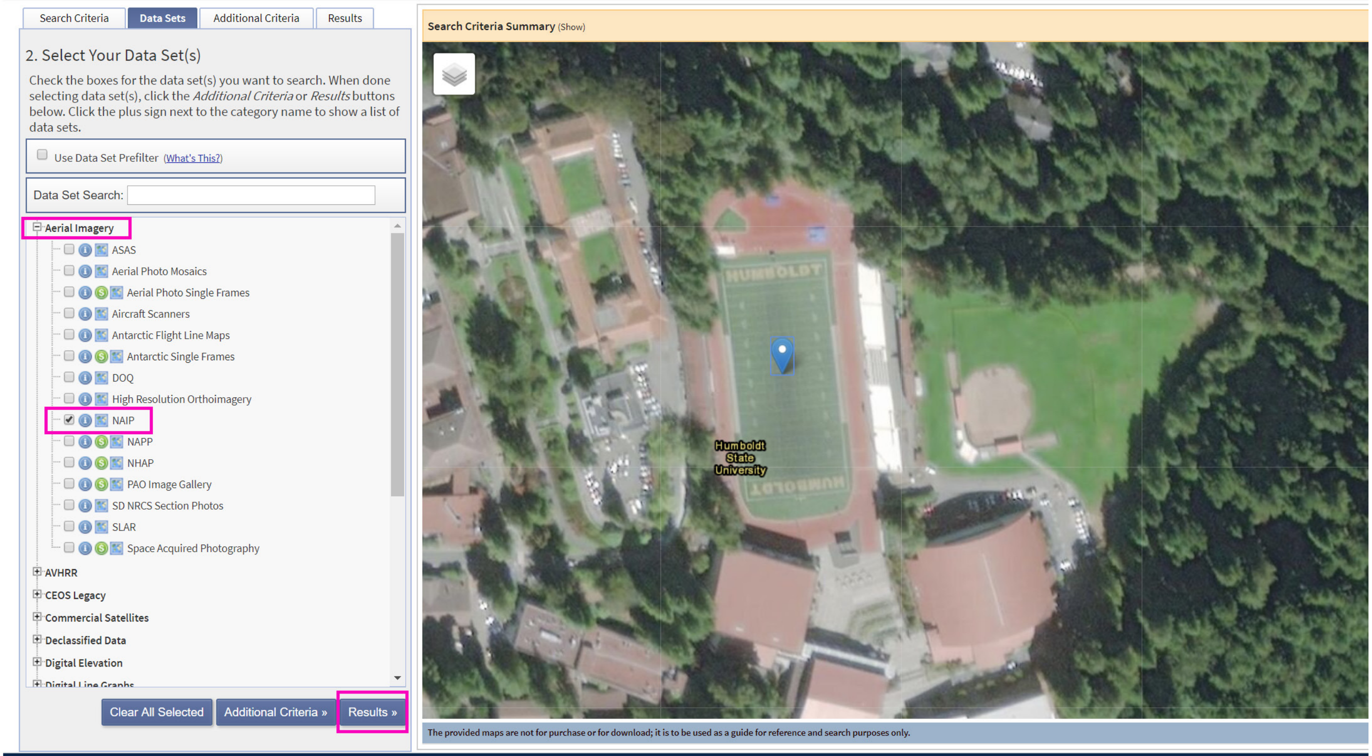


Figure 4.57: 6.40: The USGS Earth Explorer provides aerial imagery from many sources.

On the *Results* tab, you should see a list of NAIP imagery that covers your area of interest. The USGS lists the images in chronological order. Click the icon that looks like a foot. The map zooms to the image location, and the image footprint appears on the map (**Figure 4.58**). Once you have located a suitable image, click the download icon and save the file to your *original* folder. Decompress the zip file.

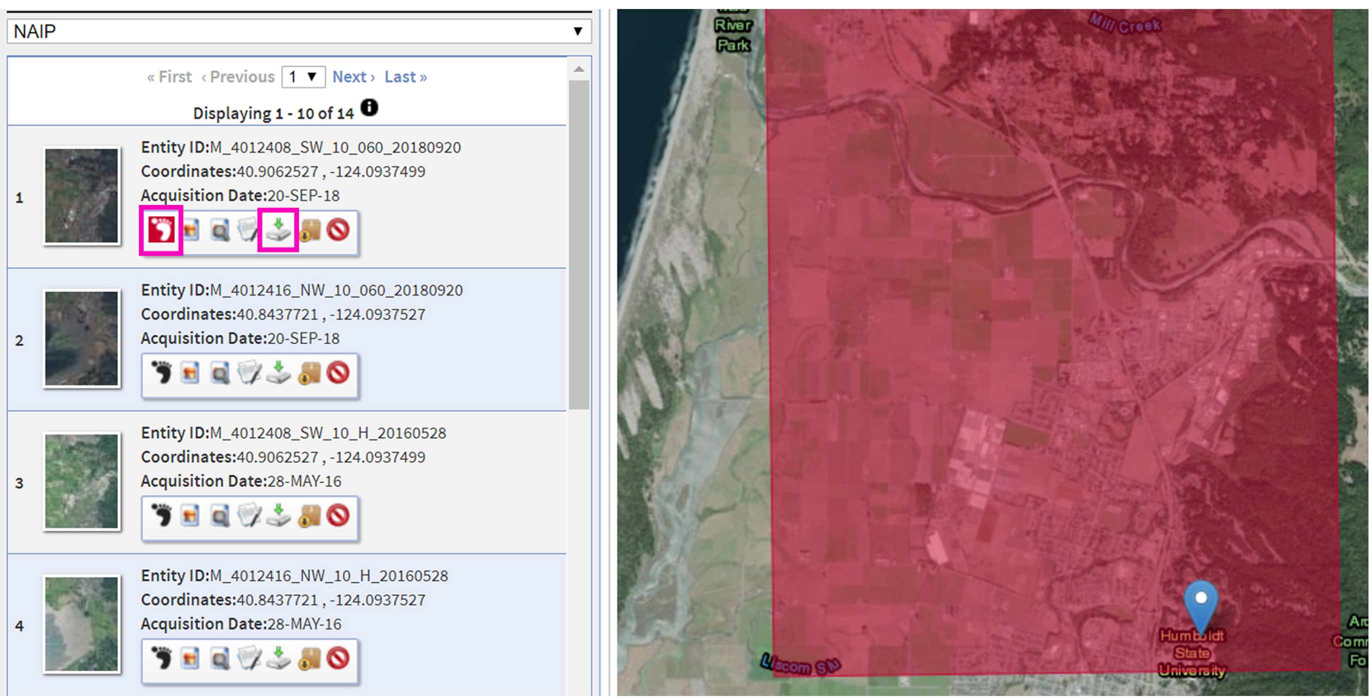


Figure 4.58: You should check the image footprints to make sure they cover your area of interest.

If necessary, use the *Project Raster* tool to create a new raster dataset that uses the UTM spatial reference. Once you are ready, add the NAIP image to your *Table of Contents*. Take a moment to set your map document properties to Store relative pathnames to data sources. Then save your map document to your workspace folder.

The NAIP imagery serves as an excellent reference to determine if your waypoints are in the correct location.

Creating an Elevation Model Using IDW Interpolation

You will start this interpolation exercise with the Inverse Distance Weighted (IDW) method. **Inverse Distance Weighting (IDW)** is a **deterministic interpolation** method that uses known values within a user-defined **search neighborhood** to predict the value at a specific location. The known values are weighted by distance, so the local influence diminishes as it reaches the edges of the search neighborhood. One of the assumptions of IDW interpolation is that local variation is one of the characteristics that define the dataset. In the best-case scenario for this method, sample points are evenly distributed throughout the area and are not clustered. The essential parameters for IDW are the **power** and the **search neighborhood** settings.

IDW is available as a stand-alone tool. However, deterministic methods like IDW and Spline do not provide a standardized measure of uncertainty. This lack can pose a problem when you need to defend the methods used in your model. The *Geostatistical Analyst* extension within the ArcGIS software can provide some measure of uncertainty, even among deterministic methods, through the use of **cross-validation**.

From the *Geostatistical Analyst* toolbar, select *Geostatistical Wizard* from the drop-down menu (**Figure 4.59**).

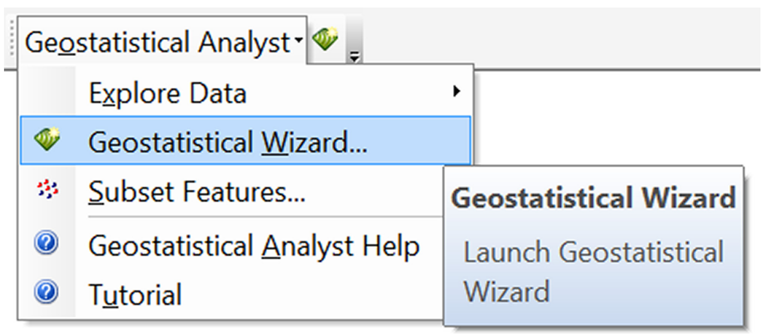


Figure 4.59: Activate the *Geostatistical Wizard*.

Under the *Methods* section, choose *Inverse Distance Weighting*, which is located under *Deterministic Methods* (**Figure 4.60**). For the *Source Dataset* option, select the shapefile you have created. In this example, the shapefile is called *GPS_Waypoints*. For the *Data Field* option, select the field that represents elevation values. Leave the *Weight Field* blank and *click Next*.

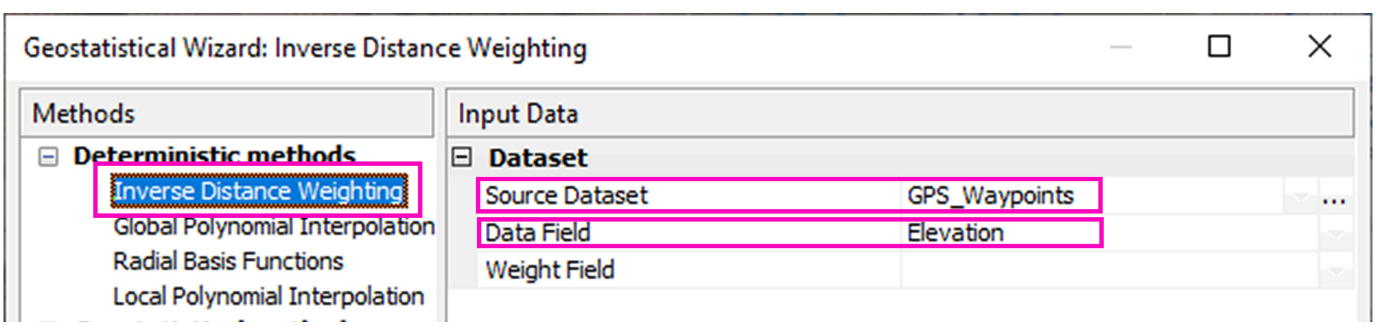


Figure 4.60: Be sure that your settings match.

By default, the Garmin GPSMap64st receiver uses this name altitude for the elevation field.
If you are using a different GPS receiver, the elevation field name might be different.

Maximize the *Geostatistical Wizard* window to view the results. You should now see a colored preview of your surface model along with a few model settings on the right (**Figure 4.61**). Take a moment to explore the data. Expand the list of weights to the right near the bottom of the *General Properties* pane. Then, click around on different places on the map. Notice how the values change and how the points nearby are symbolized on the map. The colored points indicate which known data points are used to calculate the value of the pixel under the crosshair.

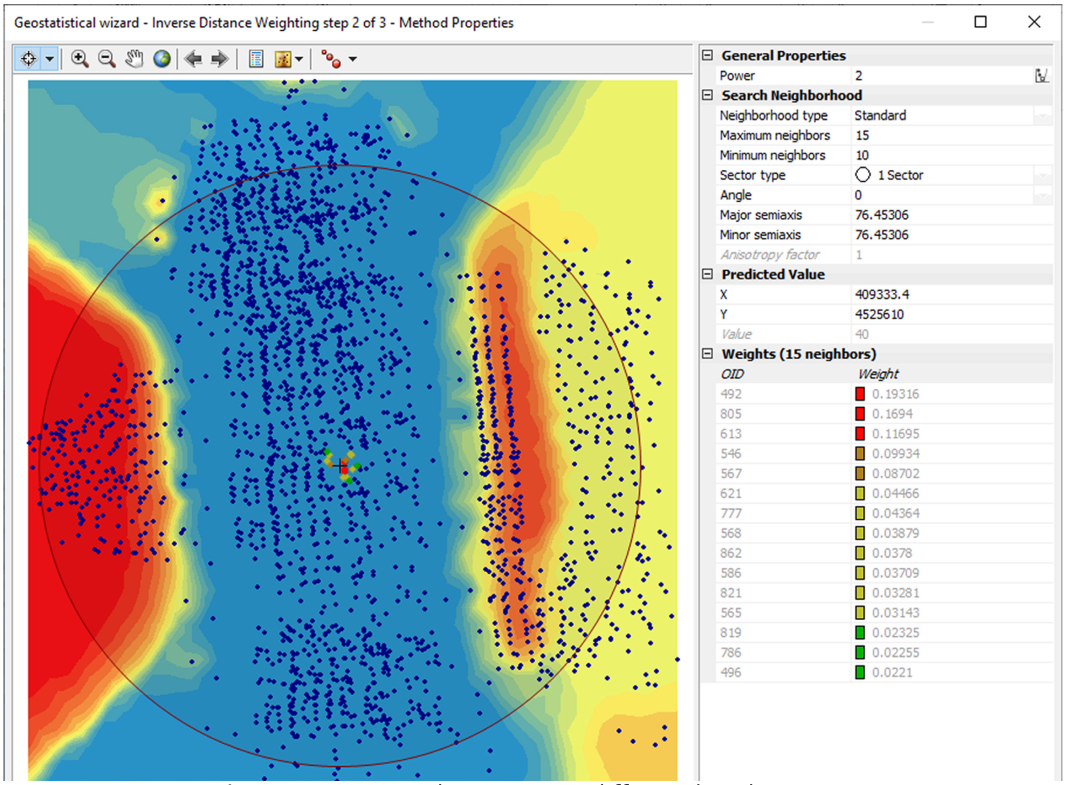


Figure 4.61: Your results may appear different than this image.

Open a blank Excel workbook and create a table recording the results from the default IDW surface model (**Figure 4.62**). Record the default IDW settings on your table, including *neighborhood type*, *maximum neighbors*, *minimum neighbors*, *sector type*, *angle*, *major semiaxis*, and *minor semiaxis*. Once you are done recording the data, *click Next*.

	A	B	C
1	**Setting**	**Default IDW**	**Optimized IDW**
2	Power	2	
3	Neighborhood Type	Standard	
4	Maximum Neighbors	15	
5	Minimum Neighbors	10	
6	Sector Type	1 Sector	
7	Angle	0	
8	Major semiaxis	76.45306	
9	Minor Semi Axis	76.45306	
10	Root Mean Square Error (RMSE)		

Figure 4.62: Your results may be different than the ones shown here.

On step 3 of the *Geostatistical Wizard* for IDW, **cross-validation** is performed. A list of error values appears on the left and a scatter plot on the right. Below the scatter plot, you will find statistics for prediction errors (**Figure 4.63**). Take note of the **Root Mean Square Error** and add it to the Excel table.

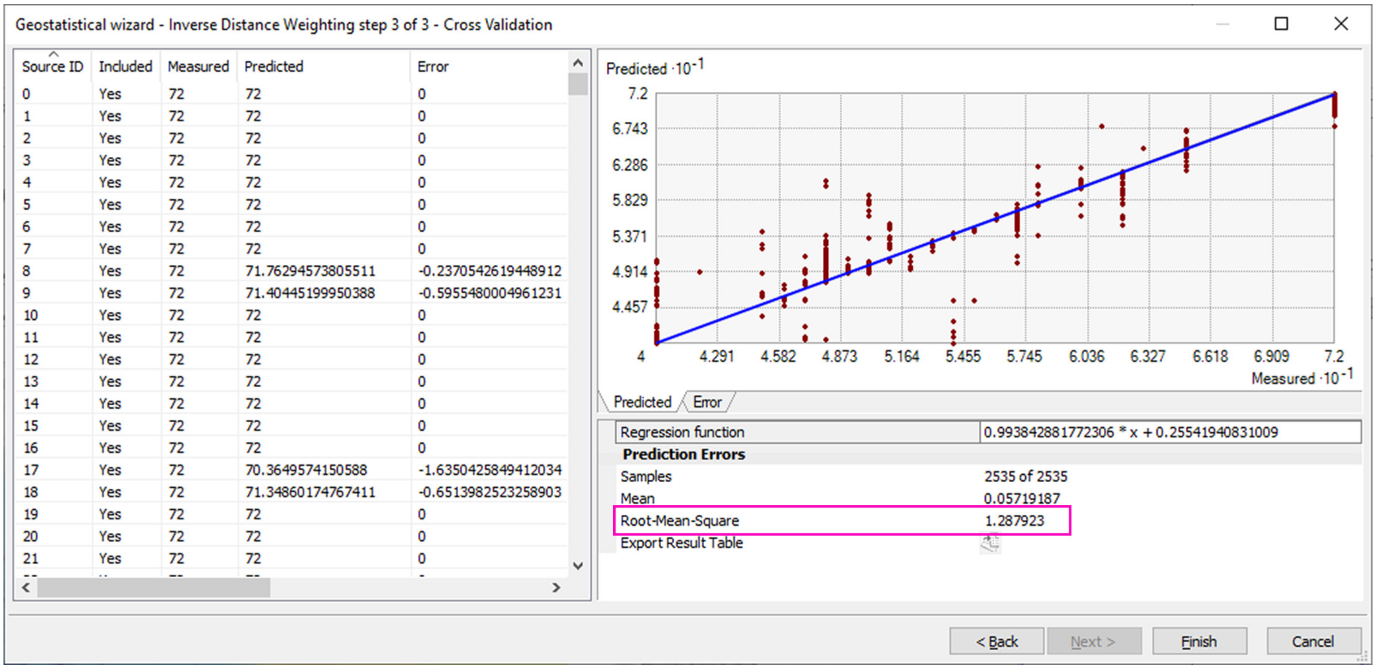


Figure 4.63: Your results may appear different than those found here.

Now you will attempt to optimize the IDW interpolation. On step 2 in the *Geostatistical Wizard*, spend about five minutes experimenting with the IDW settings mentioned above to try to improve the RMSE value. An excellent place to start is to click the *Optimize Power Value* button to the right of the *Power* setting (**Figure 4.64**). While you are experimenting, occasionally return to step 3 in the *Geostatistical Wizard* to check on the RMSE value.

Note: Do not spend more than a few minutes experimenting with the different settings. This part of the tutorial is for your personal exploration and understanding.

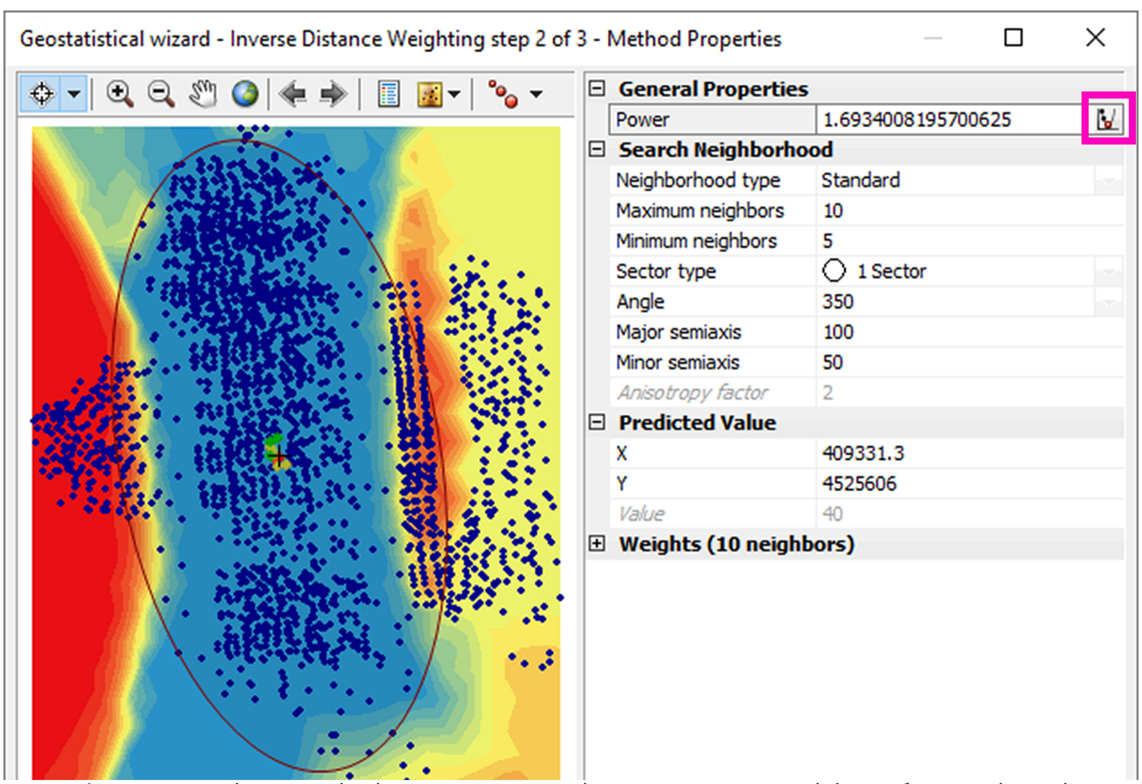


Figure 4.64: In this example, the *major semiaxis*, the *minor semiaxis*, and the *angle* were adjusted because the elevation of a football stadium is very dependent on **direction**.

Record your optimized values in your Microsoft Excel table. Save your Excel file in your *final* folder for later use. In the *Geostatistical Wizard*, click *Finish* and then *click OK*. A temporary geostatistical layer should now be added to your *Table of Contents* (**Figure 4.65**).

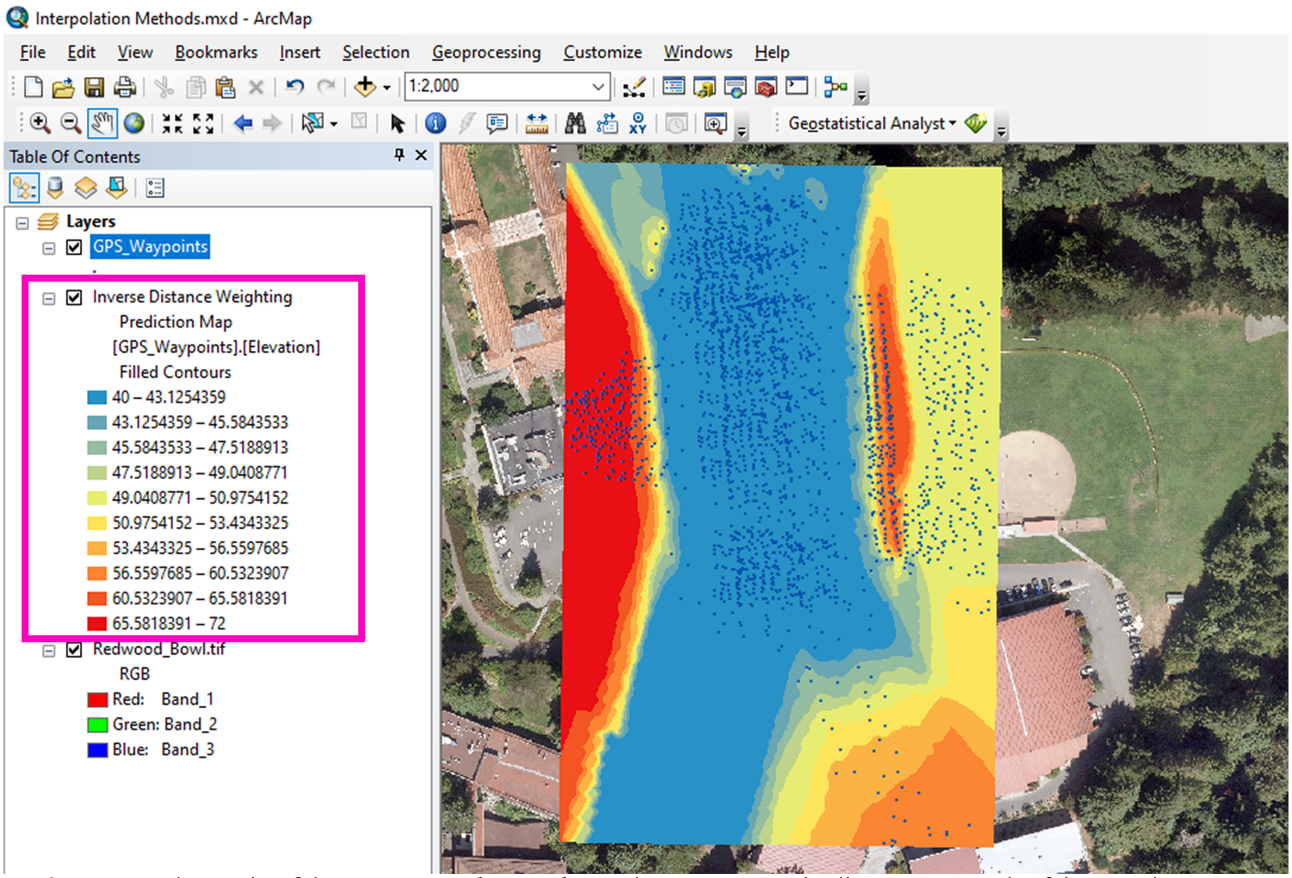


Figure 4.65: The results of the *Geostatistical Wizard* are only temporary and will not exist outside of the map document.

Creating an Elevation Model Using Spline Interpolation

Spline is a **deterministic interpolation** method where the predicted values are estimated using a function that minimizes the total curvature of the surface. The result is a smooth surface that passes precisely through all of the measured data points. Like IDW, the Spline interpolation is available as a stand-alone tool. However, deterministic methods like IDW and Spline do not provide a standardized measure of uncertainty. Just as with IDW, the *Geostatistical Analyst* extension within the ArcGIS software can provide some measure of uncertainty, even among deterministic methods, through the use of **cross-validation**.

From the *Geostatistical Analyst* toolbar, select *Geostatistical Wizard* from the drop-down menu. Under the *Methods* section, choose *Radial Basis Functions*, which is located under *Deterministic Methods* (**Figure 4.66**). For the *Data Field* option, select the field that represents elevation values. When ready, *click Next*.

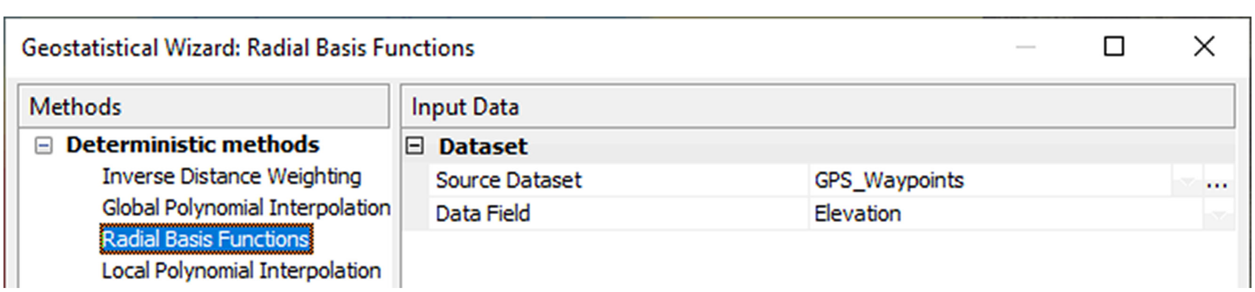


Figure 4.66: Be sure that your settings match.

Maximize the *Geostatistical Wizard* window to view the results. You should now see a colored preview of your surface model along with a few model settings on the right (**Figure 4.67**).

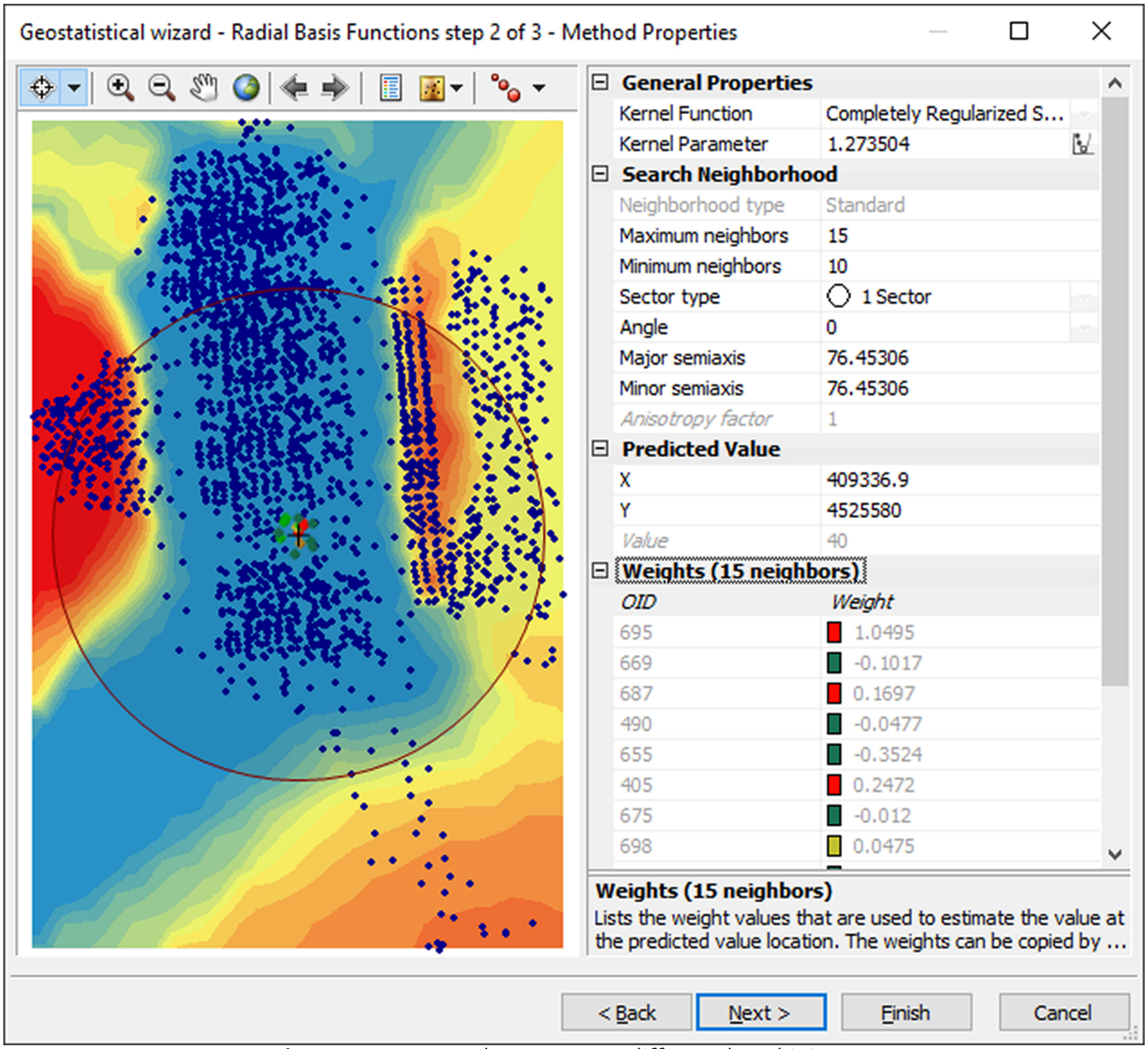


Figure 4.67: Your results may appear different than this image.

As before, you will record the default settings onto an Excel table (**Figure 4.68**). However, the settings here are slightly different. Instead of Power and Neighborhood Type, you have Kernel Function and Kernel Parameter. Once you are done recording the data, *click Next*.

	A	B	C
12	**Setting**	**Default Spline**	**Optimized Spline**
13	Kernel Function	Completely Regularized Spline	
14	Kernel Parameter	1.273504	
15	Maximum Neighbors	15	
16	Minimum Neighbors	10	
17	Sector Type	1 Sector	
18	Angle	0	
19	Major semiaxis	76.45306	
20	Minor Semi Axis	76.45306	
21	Root Mean Square Error (RMSE)		

Figure 4.68: Your results may be different than the ones shown here.

On step 3 of the *Geostatistical Wizard* for Spline, **cross-validation** is performed. A list of error values appears on the left and a scatter plot on the right. Below the scatter plot, you will find statistics for prediction errors (**Figure 4.69**). Take note of the **Root Mean Square Error** and add it to the Excel table.

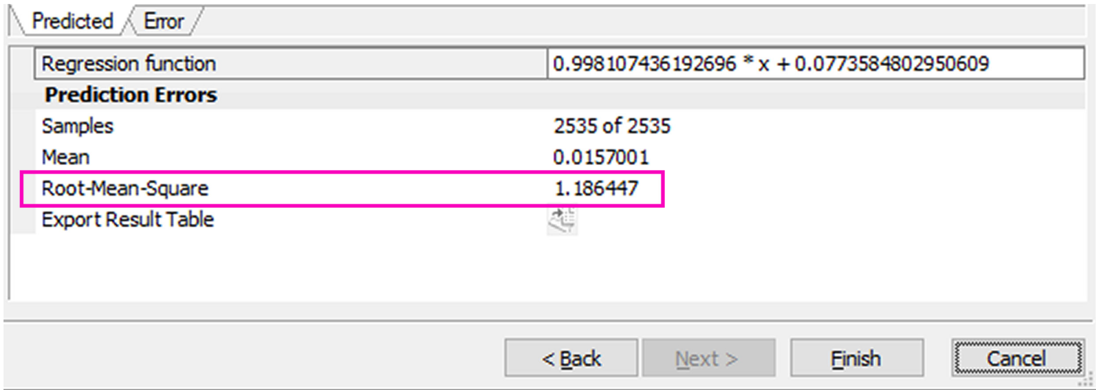


Figure 4.69: Your results will appear different than those found here.

Now you will attempt to optimize the Spline interpolation. On step 2 in the *Geostatistical Wizard*, spend about 5 minutes experimenting with the Spline settings mentioned above to try to improve the RMSE value. An excellent place to start is to click the *Optimize Kernel Parameter* button to the right of the *Kernel Parameter* setting (**Figure 4.70**). You may also want to compare regularized spline with other types of spline functions, such as the **tension spline** located under the *Kernel Function Settings*. While you are experimenting, occasionally return to step 3 in the *Geostatistical Wizard* to check on the RMSE value.

Note: Do not spend more than a few minutes experimenting with the different settings.
This part of the tutorial is for your personal exploration and understanding.

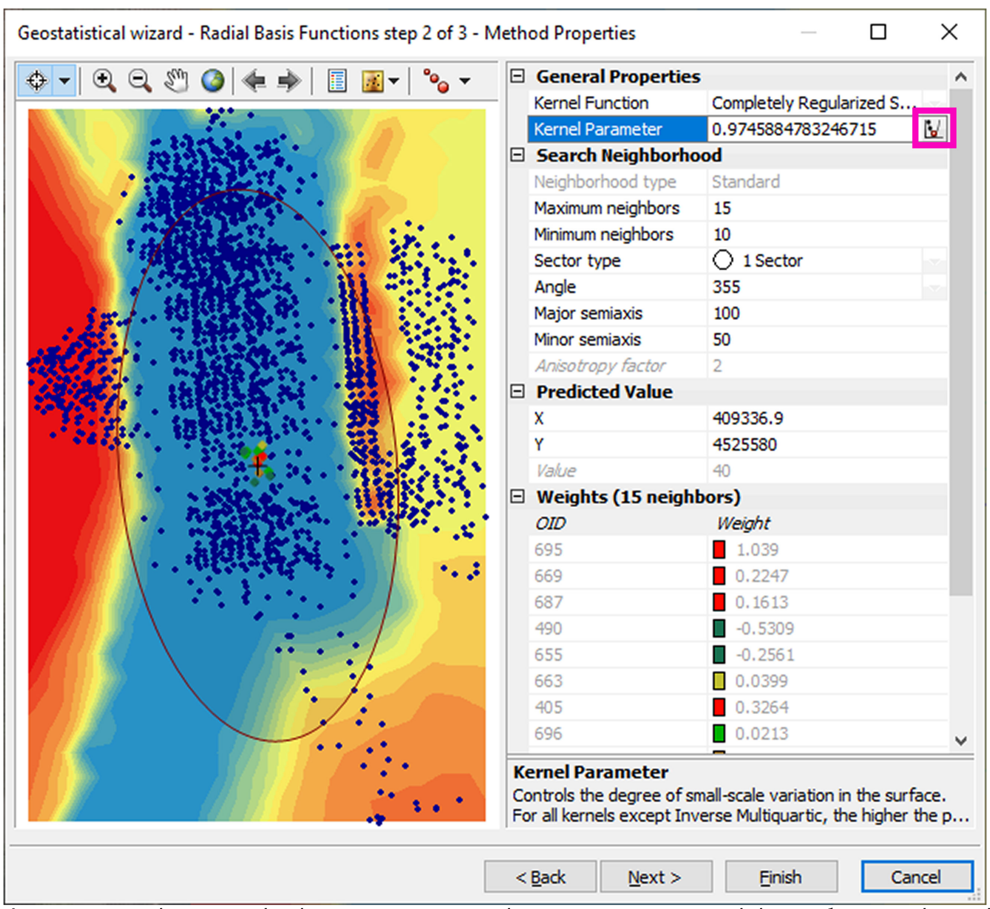


Figure 4.70: In this example, the *major semiaxis*, the *minor semiaxis*, and the *angle* were adjusted because the elevation of a football stadium is very dependent on **direction**.

Once you are satisfied, record the results in your Excel table. Save your excel file in your *final* folder for later use. Once you have recorded the optimized RMSE value, return to ArcMap. In the *Geostatistical Wizard*, click Finish and then *click OK*. A temporary geostatistical layer should now be added to your *Table of Contents* (**Figure 4.71**. Next, you will create a surface modeling using the *Kriging* method.

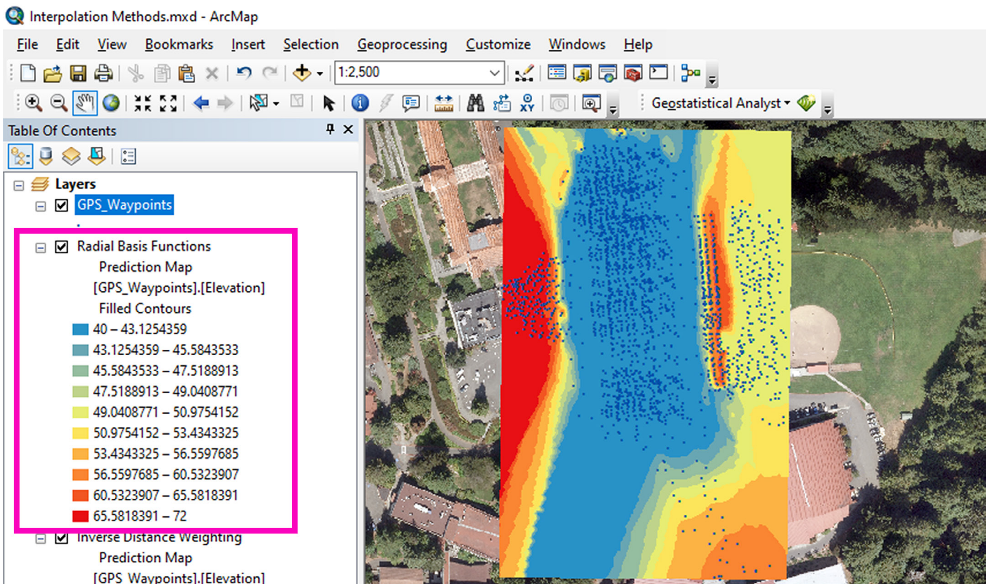


Figure 4.71: The results of the *Geostatistical Wizard* are only temporary and will not exist outside of the map document.

Creating an Elevation Model Using Kriging Interpolation

Kriging refers to a group of **geostatistical interpolation** techniques that can provide a standardized measure of uncertainty in their predictions. Like some of the deterministic interpolation methods, it is based on the assumption that things that are closer together are **more alike** than those farther away. There are numerous options to choose from when using the Kriging method, most of which are beyond the scope of this book. Instead, this chapter will primarily rely on the default settings and use cross-validation to compare the results. Unlike deterministic interpolation methods, prediction points are estimated by modeling the statistical correlation between pairs of known points. This relationship between the values of data points and the distance between them is known as **spatial autocorrelation**.

In this step, you will explore the use of a geostatistical interpolation method to compare with your two previous deterministic interpolation methods. The number of options available through the different Kriging methods can be overwhelming. The method we will use in this lesson will be **Ordinary Kriging**, which has only a limited number of options from which to choose.

From the *Geostatistical Analyst* toolbar, select *Geostatistical Wizard* from the drop-down menu. Under the Methods section, choose *Kriging/CoKriging*, which is located under *Geostatistical Methods* (**Figure 4.72**). For the Data Field option, select the field that represents the elevation in your dataset. Leave Dataset 2, Dataset 3, and Dataset 4 **alone**. These are used for cokriging, and we **will not** be using them for this lesson. *Click Next*.

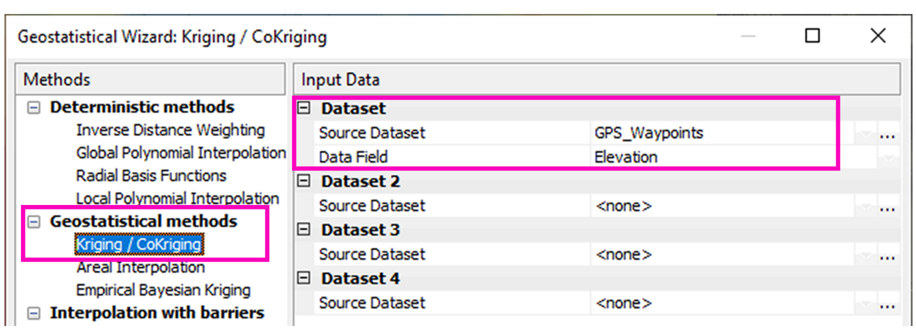


Figure 4.72: Be sure that your settings match.

On step 2 of the *Geostatistical Wizard*, you have the option to choose from a variety of Kriging types. From the list on the left, under *Kriging Type*, select *Ordinary* (**Figure 4.73**). Beneath that list, you will see some other options under *Output Surface Type*. On this list, choose *Prediction*. *Click Next*.

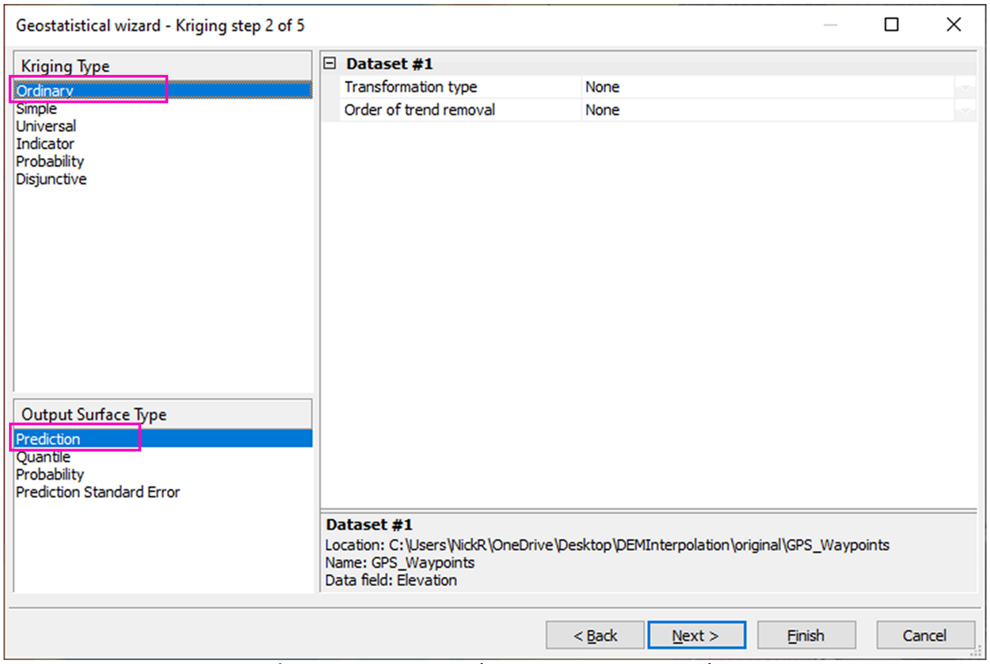


Figure 4.73: Be sure that your settings match.

Step 3 of the *Geostatistical Wizard* displays a graph called a **semivariogram** (**Figure 4.74**). In this semivariogram, each waypoint in the dataset gets paired with every other waypoint. The *Geostatistical Analyst* records the difference in elevation value between each paired data point along with the distance between them.

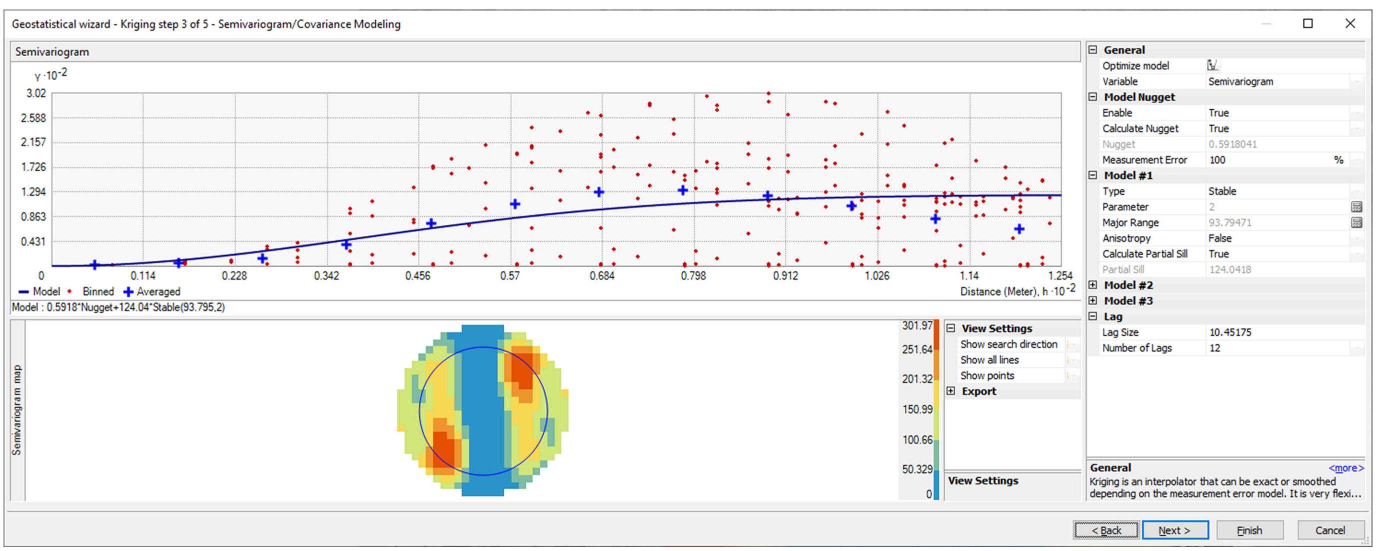


Figure 4.74: Your results may appear different.

The red dots in this graph represent a pair of waypoints. The pairs are grouped based on distance. These groups are called **bins**. The **blue crosses** represent the average value in each bin. The **blue line** represents the surface model. Ideally, the surface model line should pass through the center of the cloud of binned values and should also pass as closely as possible to the averaged values (blue crosses). The expectation is that there will be **small** differences in elevation values between pairs that are **close together**. Pairs that have **greater** distances between them are expected to have **more substantial** differences in elevation values.

At this time, you will accept the default settings. In step three of the *Geostatistical Wizard* window *click Next*. **Step 4** and **step 5** of the *Geostatistical Wizard* should look familiar based on the previous work with IDW and Spline interpolation (**Figure 4.75**). Take a moment to explore the data.

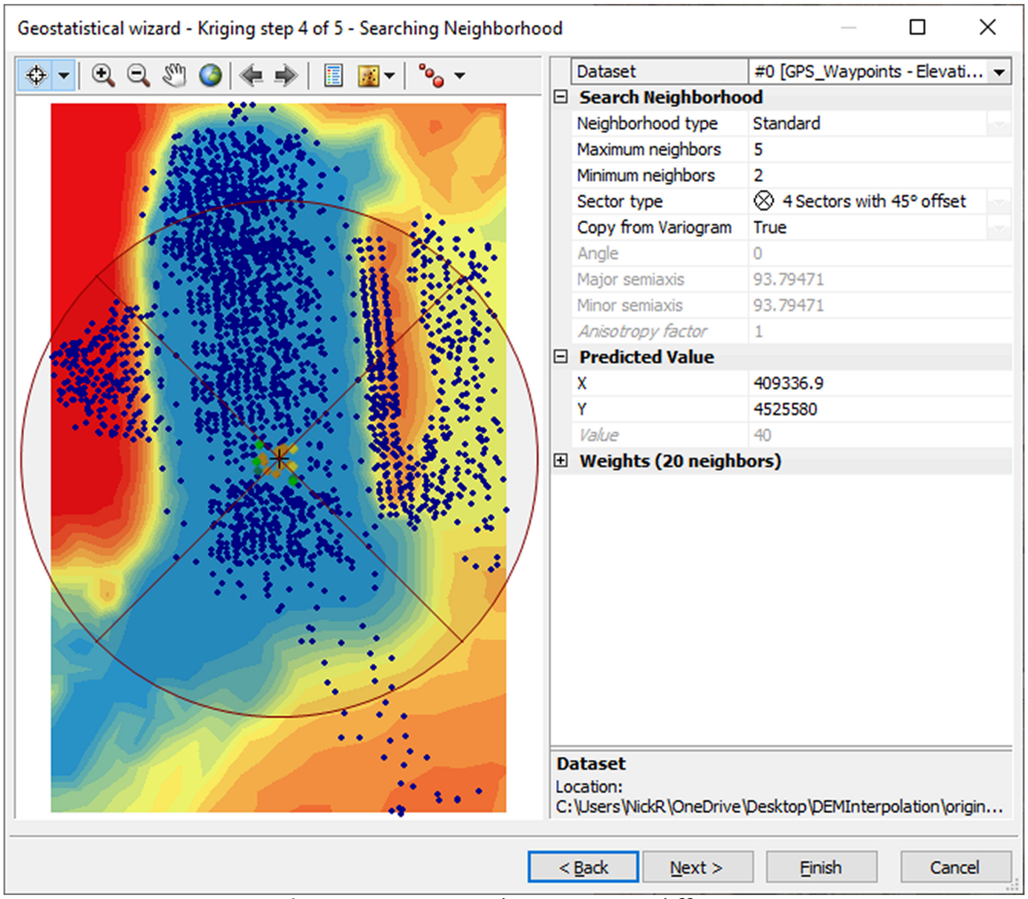


Figure 4.75: Your results may appear different.

As before, you will record the default settings and RMSE onto an Excel table (**Figure 4.76**). However, the settings here are slightly different. You may notice that the setting, *Copy from Variogram*, is set to *True*. This option means that many of the settings here will be determined by the settings in the semivariogram on **step 3** of the *Geostatistical Wizard*. While you are trying to optimize the model, you will go back and forth between step 3 and step 4. You can check the RMSE values on step 5 of the *Geostatistical Wizard*. For now, record the default values.

	A	B	C
23	Settings	Default Ordinary Kriging	Optimized Ordinary Kriging
24	Model Type	Stable	
25	Maximum Neighbors	5	
26	Minimum Neighbors	2	
27	Sector Type	4 Sectors with 45° offset	
28	Root Mean Square Error (RMSE)		

Figure 4.76: Your results may be different than the ones shown here.

You can find the Model Type on step 3 of the Geostatistical Wizard.

As mentioned before, Kriging interpolation can have an overwhelming number of options. The mathematics behind these options is beyond the scope of this book. However, experimenting with a few options may improve your surface model. Spend about five minutes to try to improve the RMSE value. An excellent place to start your optimization is to try the different *model types*. You can locate this parameter back on Step 3 of the *Geostatistical Wizard*. By default, it should be set to *Stable*. Once you are satisfied, record the results in your Excel table. Save your excel file in your *final* folder for later use.

Once you have recorded the optimized RMSE value, return to ArcMap. In the *Geostatistical Wizard*, click *Finish* and then *click OK*. A temporary geostatistical layer should now be added to your *Table of Contents* (**Figure 4.77**). In the next step, you will choose the best interpolation method from which to base your digital elevation model.

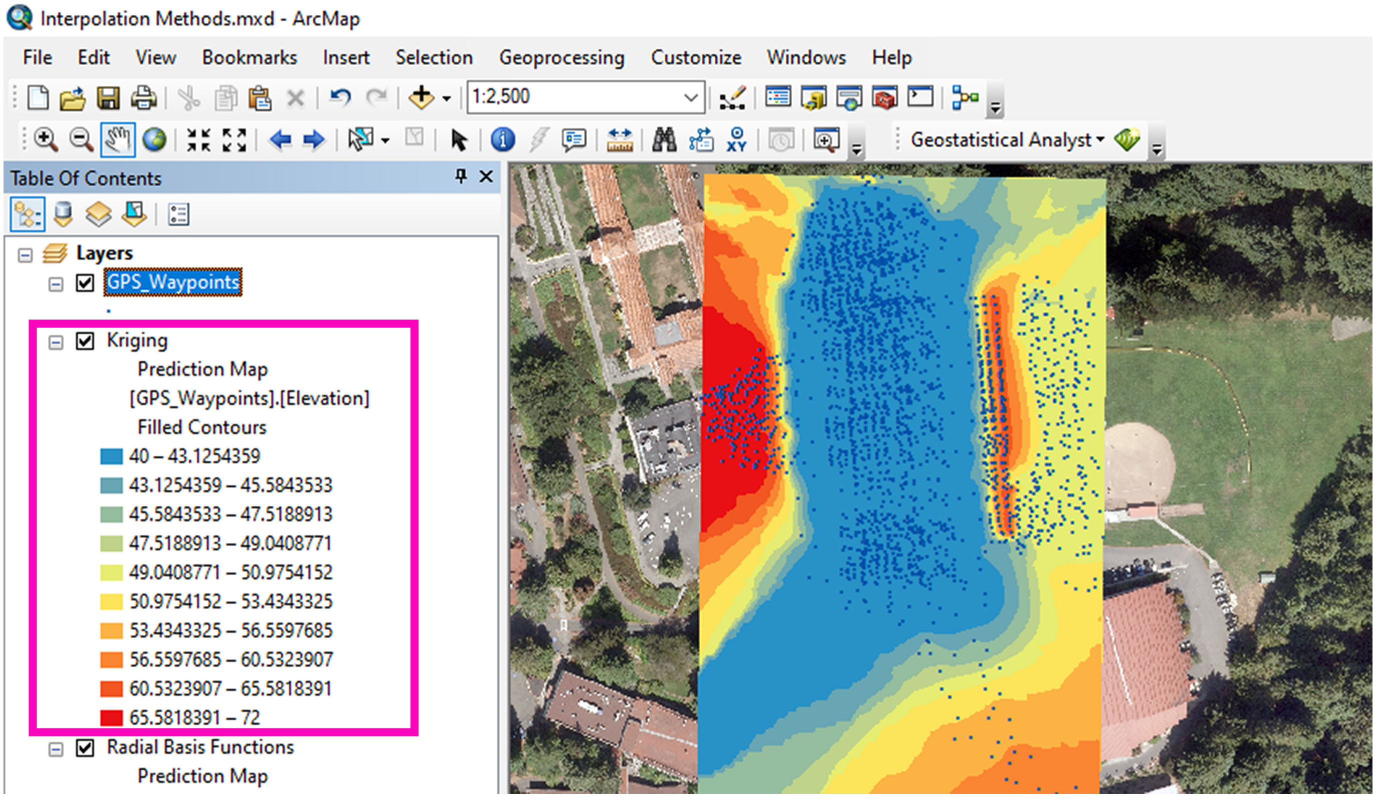


Figure 4.77: This layer is a temporary representation of your surface model.

Creating a Digital Elevation Model from a Geostatistical Layer

In this step, you will choose one of the surface models with the best RMSE value. Then you will export the temporary geostatistical layer and create a **permanent raster file**. Before exporting, there are a few steps you will need to take to prepare. The first step is to determine which of the surface models to chose. Start by right-clicking on one of your geostatistical layers. From the menu select, *Compare* (**Figure 4.78**).

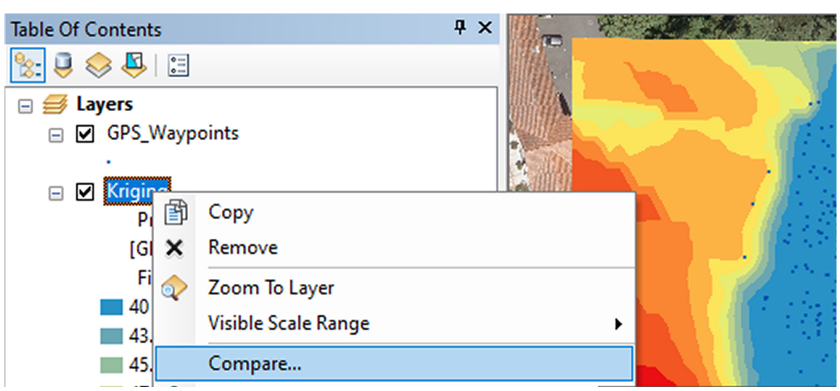


Figure 4.78: The Compare function is useful when choosing from multiple interpolation models.

This step opens the *Cross-Validation Comparison* window (**Figure 4.79**). Here you can compare the results of any geostatistical layers currently loaded in the *Table of Contents* by selecting them from the drop-down menu on the upper right. Take a moment to compare the RMSE values. Find the surface model with the **best RMSE**. This surface model will be the one on which you will base your digital elevation model.

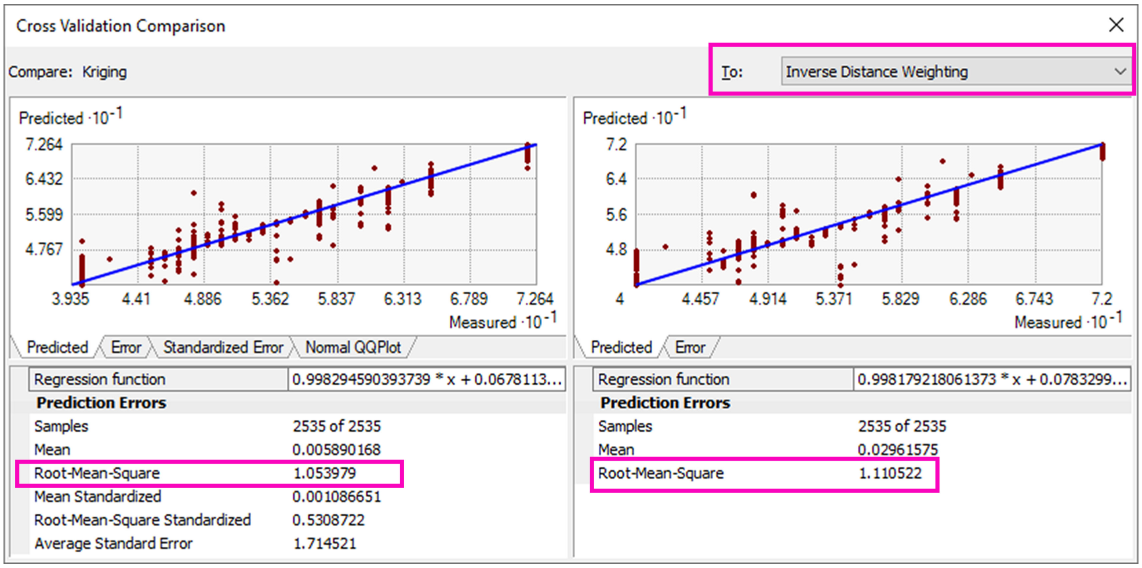


Figure 4.79: The *Cross-Validation Comparison* window provides a means for choosing between two interpolation models.

Right-click on your chosen geostatistical layer in the *Table of Contents*. Select *Data*, then *Export to Raster* (**Figure 4.80**).

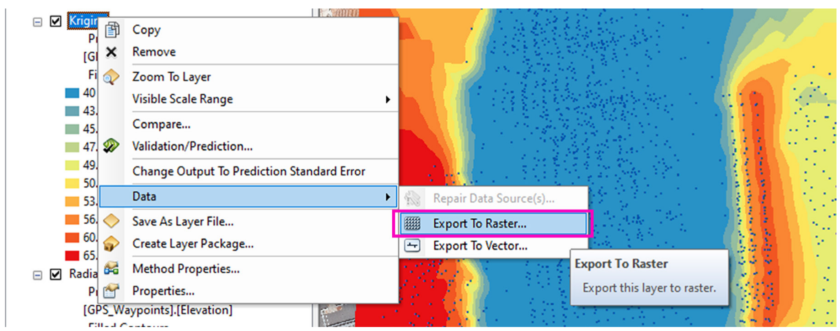


Figure 4.80: Right-click on the best layer for creating a digital elevation model (DEM)

The *GA Layer to Grid* tool opens (**Figure 4.81**). Save your *Output surface raster* to your *working* folder. Name the file *DEM.tif*. The *GA Layer to Grid* tool will automatically try to determine the cell size based on the dimensions height and width of the dataset. You may accept the default settings. Leave all other settings as default and *click OK*.

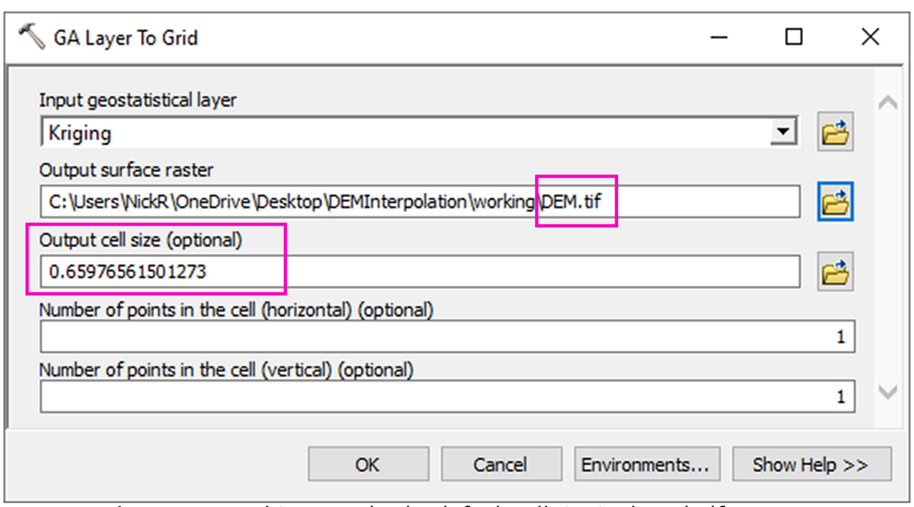


Figure 4.81: In this example, the default cell size is about half a meter.

This cell size may not always be appropriate. Determining the proper cell size is up to you and mainly a judgment call. Feel free to experiment with different cell sizes if you wish.

A digital elevation model will now appear in your *Table of Contents*. The result may appear very similar to your original geostatistical layer. This visualization is misleading since the values across the surface have a continuous range rather than discrete categories. Open the properties for the DEM and select the *Symbology* tab (**Figure 4.82**). Choose *Stretched* from the options on the left and *click OK*.

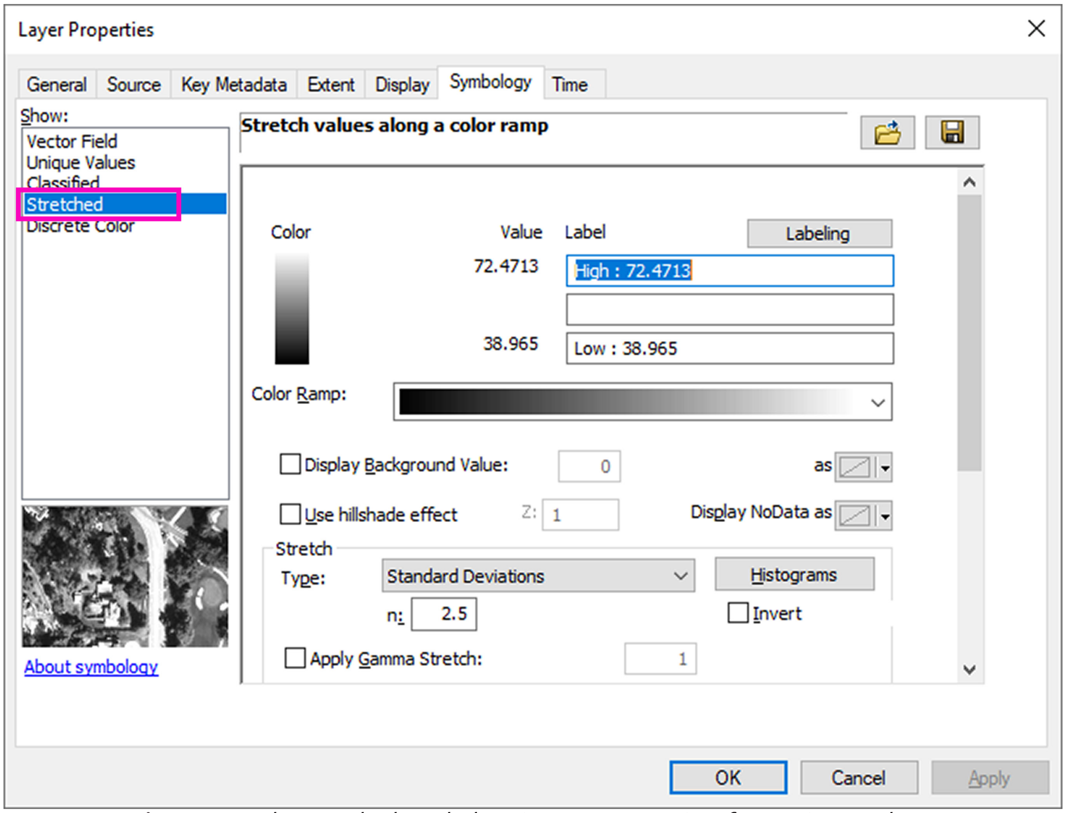


Figure 4.82: The stretched symbology is more appropriate for continuous data.

The digital elevation model should not look more familiar (**Figure 4.83**).

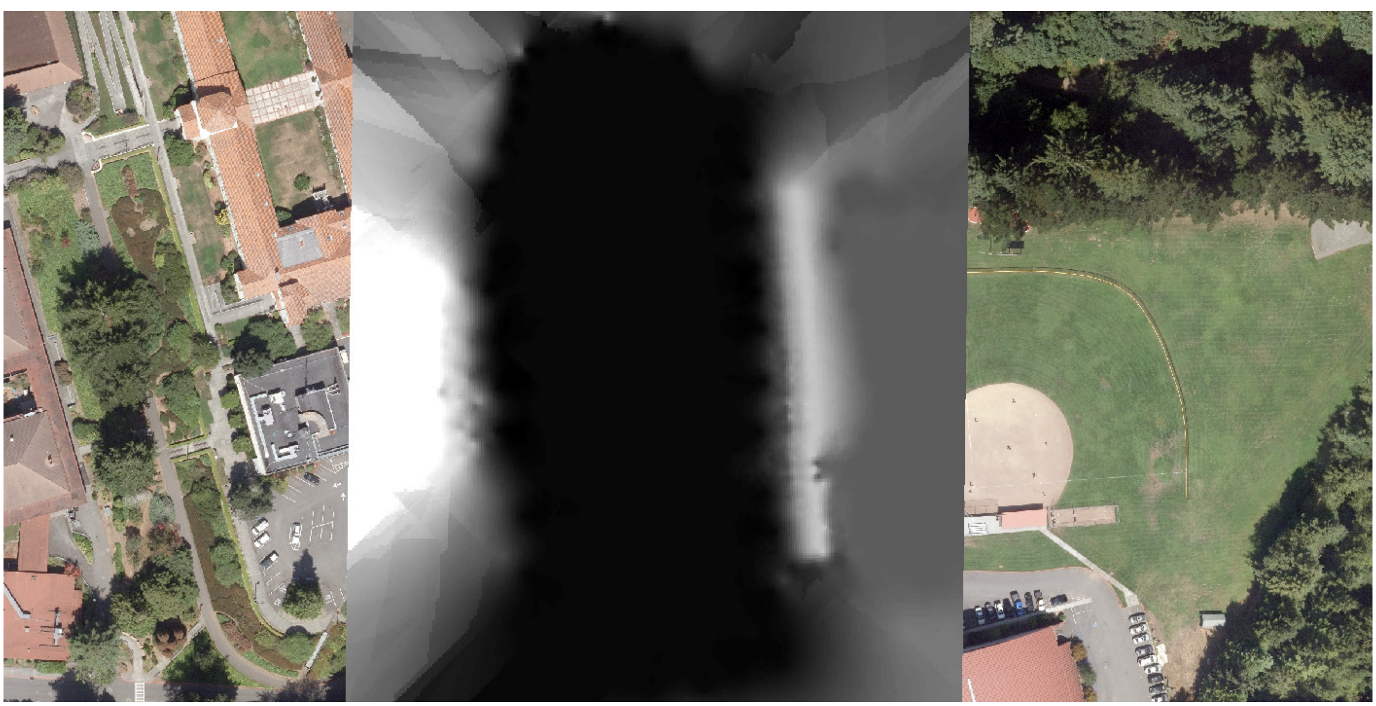


Figure 4.83: Most digital elevation models appear black and white.

To test your digital elevation model, create a hillshade. Within the *Spatial Analyst* tools under *Surface*, open the *Hillshade* tool. Use the default settings and save the results to your *working* folder. Call the file *Hillshade.tif* (**Figure 4.84**).

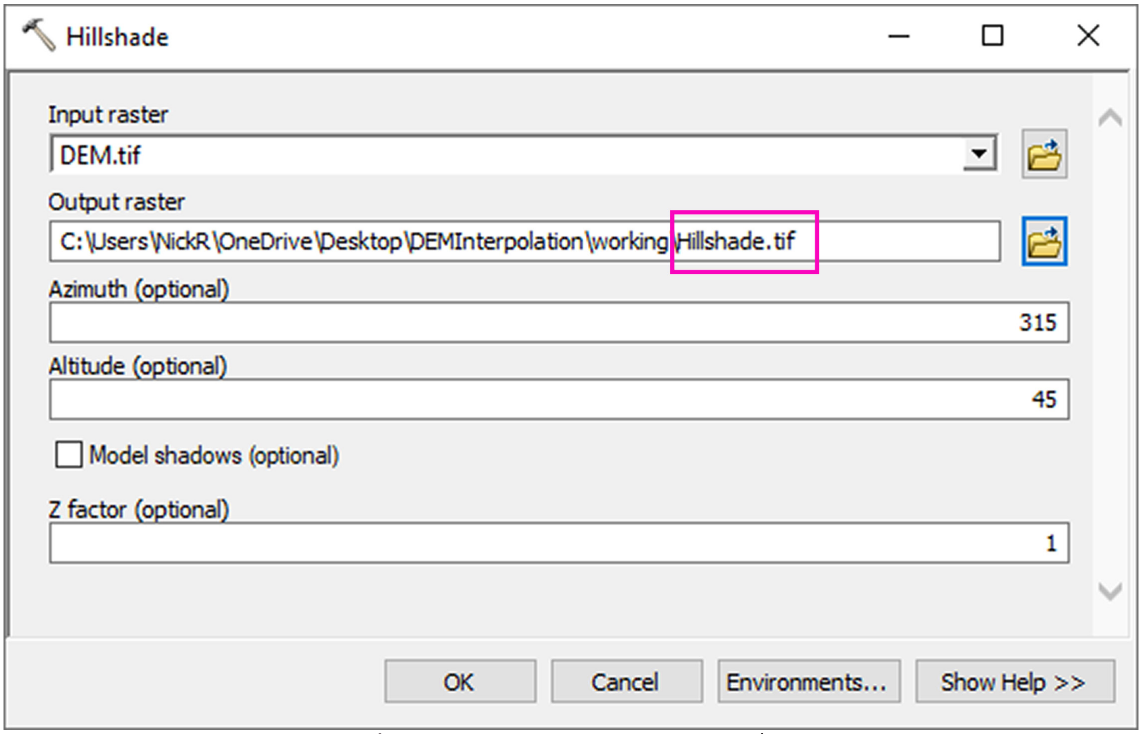


Figure 4.84: Be sure your settings match.

Add the hillshade layer to the *Table of Contents* (**Figure 4.85**). Take a moment to look at the results. Does the terrain appear to be similar to what you expected based on your personal experience while collecting the elevation data?

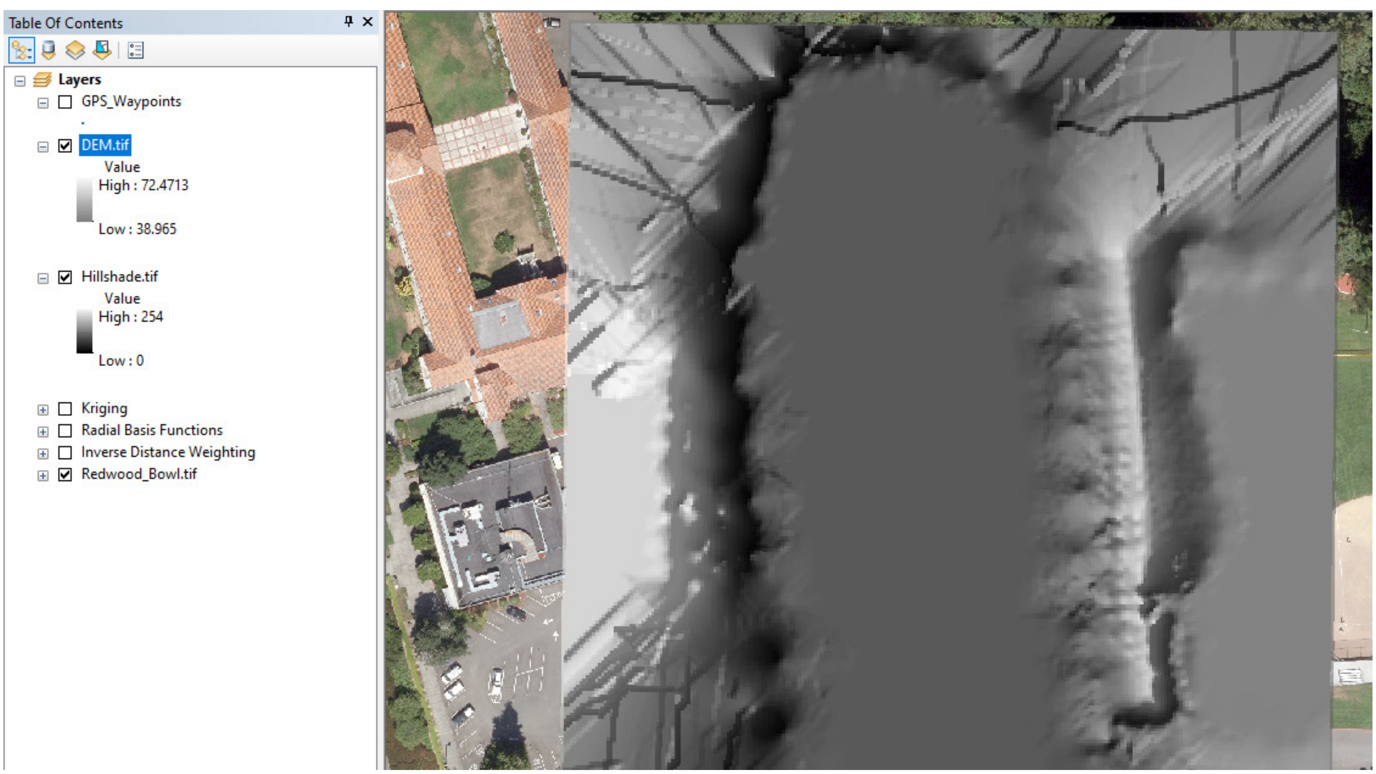


Figure 4.85: In this example, the DEM was given a transparency value of 50% to improve the appearance of the hillshade.

Skill Drill: Creating a Map of the Results

You should be familiar with the steps needed to create a map layout of your results. Design a map for use as a figure in a report or summary (**Figure 4.86**). Ideally, the map should be designed at a size of approximately 5 inches wide by 7 inches tall. Include a north arrow, a scale bar. You don't need a legend for this map. Include the original waypoints over your hillshade. You may change the color ramp of the digital elevation model as well as the transparency to overlay on top of the hillshade. When your map layout is complete, export the map as a PNG file with a resolution of 300 **dpi**. Save the file in your *final* folder.

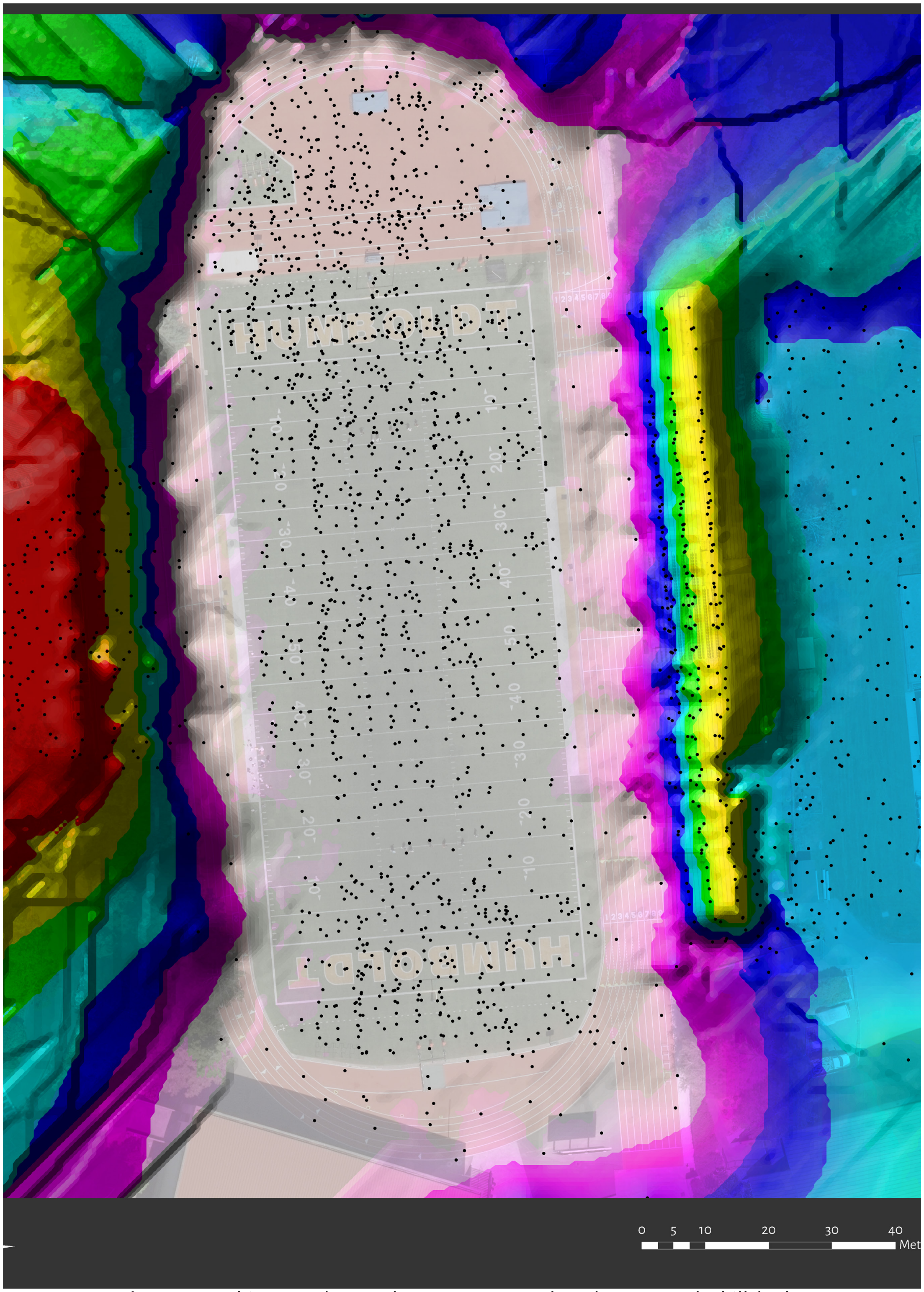


Figure 4.86: This example uses the temperature color scheme over the hillshade.

Appendix A

Chapter 1 Data

Advanced Geospatial Analysis Review Data	*https://bit.ly/Advanced_Review*

Chapter 2 Data

Organic Waste Business Data	*http://bit.ly/OrgWaste_business_data*
Organic Waste Original Folder Backup	*https://bit.ly/Org_Waste_backup*
Fire Station Proposed Locations Data	*https://bit.ly/Proposed_Fire_Stations*
Incident History	*https://bit.ly/ArcataFireData*
Fire Station Site Selection Original Folder Backup	*https://bit.ly/Fire_Station_Original_backup*

Chapter 3 Data

Wolpertinger Den Locations	*https://bit.ly/wolpertinger-dens*
Wolpertinger Original Folder Backup	*https://bit.ly/wolpertinger_original_backup*

Chapter 4 Data

LA Crime 2005-2015	*https://bit.ly/Crime_GDB*
DEM Interpolation GPS Waypoints	*https://bit.ly/GPS_Waypoints*
DEM Interpolation Original Folder Backup	*https://bit.ly/DEM_interpolation_original_backup*

Index