2013

# Introduction to Geographic Information Science using ArcGIS V10

Charles M Schweik, *University of Massachusetts - Amherst*
Bethany Bradley
Richard Smith

# INTRODUCTION TO GEOGRAPHIC INFORMATION SCIENCE

# USING ARCGIS V10

## Coursepack

September, 2013 version

Compiled by Charlie Schweik
University of Massachusetts, Amherst

Material contained within was written by either
Rick Smith (Texas A&M University Corpus Christie),
Bethany Bradley (UMass Amherst) or
Charlie Schweik (UMass Amherst)

NOTE:

The datasets needed for the lab exercises in this document can be found at
https://udrive.oit.umass.edu/cschweik/NRC585

## Acknowledgements and Copyright Licenses

Material in this coursepack falls under one of two licenses.

First, much of the material used in this lab manual was developed by my colleague Bethany Bradley or myself. I'd like to acknowledge the hard work by Bethany and thank her for the use of these excellent learning resources. Any material developed by either Bethany or myself falls under the following license:



Second, I'd like to acknowledge the work and support of Professor Phil Davis at the GeoTech Center and Rick Smith at Texas A&M University Corpus Christie who worked with us and gave us permission to make a "hybrid" course using some of their lesson and lab material. Their material falls under this Copyright:

# Table of Contents

# Reading: Geographic Information Systems

## INTRODUCTION

### LESSON OBJECTIVES

By the end of this lesson, you will be able to:

1. Discuss the geospatial technology industry and its sectors.

2. Describe the major technological systems used within the geospatial industry.

### LEARNING SEQUENCE

| | |
|---|---|
| Required Reading | Read the following:<br>Geographic Information Systems<br>• What are Geographic Information Systems?<br>• Components of GIS<br>• Thinking about Space |
| Resources | View the following:<br>1. Geospatial Revolution |

## INSTRUCTION

# What are Geographic Information Systems?

### What is "GIS"?

A GIS is a computer-based system to aid in the collection, maintenance, storage, analysis, output, and distribution of spatial and non-spatial data and information. These key words are the action words typically associated with a GIS. Let's go through each word one by one, to get a broad understanding of what a GIS is.

#### Collection

Collection refers to the ability of a GIS to ingest many different types of information. In fact, using many different formats, and many different types of data, are one of the main functions and benefits of a GIS.

### Maintenance

Maintenance refers to the task of keeping the collected data up-to-date. The world is not a static place, therefore often times, data collected about the world needs to be updated. A GIS can support dynamic data that changes often. The flexibility of a GIS to collect and maintain static and dynamic data is another strength of GIS.

### Storage

Storage refers to the structure coding of information that is stored in a digital format on the computer. The data can be stored in simple or complex data structures. The GIS is well suited to collect, maintain, and store data sets ranging in size from very small, such as a few kilobytes, all the way up to multiple terabytes and beyond.

### Analysis

The fourth word, analysis, speaks to the analytical abilities of a GIS. When a GIS analyzes the spatial information that has been collected maintained and stored within, it turns that data into information. Analysis techniques are from a wide range of fields, and assist the user of a GIS –by uncovering spatial patterns in the data.

### Output

Output refers to maps, reports, and statistics. Typically, the output of a GIS is a map. A map is a visual representation of the spatial phenomenon known as reality. Through the collection maintenance storage and analysis of reality, a GIS can produce a visualization and model of the world.

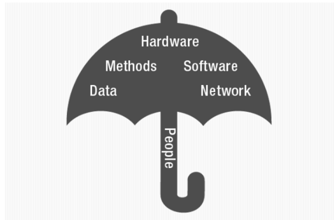
## Distribution of Information

A GIS allows for easy sharing of spatial and non-spatial data and information around the world headed both in a visual or tabular format. “Spatial and non-spatial data and information” refer to the fact that a GIS can hold all information whether it has a spatial component or not. The ability of a GIS to hold all types of information often makes the GIS the center of all information.

Ron Abler stated “GISs are simultaneously the telescope, the microscope, the computer, and the Xerox machine of regional analysis and synthesis of spatial data.” This quote really hits on many of the aspects of a GIS such as its ability to focus on a small or large area, to analyze that area, to store that information in the computer, and to copy and distribute that information like a Xerox machine.

## GIS as an Umbrella

GIS can be thought of as an umbrella composed of six parts that, when working together, produce a system that has many capabilities.




### What Falls Under the GIS Umbrella?

The six parts of a GIS are hardware, software, network, methods, data, and people. The people part of a GIS, is the part that operates and maintains the GIS. By successfully completing this course, you will begin your path as the people part of the GIS, and will be able to operate a GIS to collect, maintain, store, analyze, output, and distribute spatial and non-spatial information.

With the six parts of a GIS working together, the live capabilities fall underneath the umbrella of GIS. Capability such as: spatial data services, spatial data analysis, mapmaking, data collection, programming, and much, much more. As will be discussed throughout this first module, GIS can truly be applied to any situation at any location.


# Components of GIS

## The Six Parts of a GIS

The six parts of a GIS are: hardware, software, data, methods, people, and network. Previously, there were only five parts to a GIS. Network was later added with the advent in popularity of the Internet.

### Hardware

Hardware is perhaps the simplest part of a GIS to explain. The harder part of GIS refers to powerful computers, printers, output devices, such as monitors, input devices, and large amounts of disk space to store the, typically, large GIS data sets. A typical GIS computer workstation has a powerful CPU, large amounts of RAM, large monitors, fast network connectivity, and large amounts of disk space.

### Software

The software part of the GIS refers to the many free and commercial software packages that allow a GIS user to perform GIS functions such as the manipulation, storage, query, and analysis of spatial and non-spatial information.

### Data

The data part of a GIS refers to any and all spatial and non-spatial information stored in a digital format on the computer. A GIS can ingest any type of data in many different types of formats. The ability for a GIS to consume such a wide range of data does produce the effect of "data overload". With the wide proliferation of sensors, there is more data available to us today than ever before. With so much data, it is often difficult for a person to easily understand what is inside the data. A GIS can visualize the data as a map, graph, or other graphic, that helps humans understand and weed through all the massive amounts of data.

In fact, the problem of "too much data" does require a GIS user to spend quite a bit of time managing the data. In general, half of all the time you spend on a GIS project, will simply be working with and preparing the data for maintenance, storage, analysis, and output.

## Key Facts

A GIS can ingest any type of data including both spatial and non-spatial

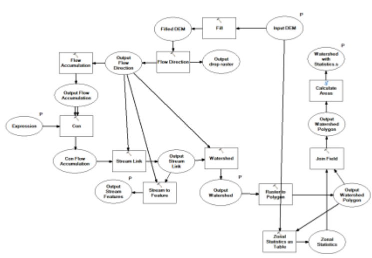
We are currently in "Data Overload." Visualizing data helps humans weed through it all.

Half of all time spent on a GIS projects will simply be working with data.

### Method

The methods part of a GIS refers to the formulas, statistics, analysis, and algorithms that are used to turn data into information, so that humans can turn that information into knowledge through interpretation.

## Key Facts




Methods

- Formulas
- Statistics
- Analysis
- Algorithms

### People

People, refers to anyone that uses a GIS. This includes the general public that may be using a printed map, or a digital globe, all the way up to skilled GIS professionals, that build new functionality and push the limits of what is possible in a GIS.

## Key Facts

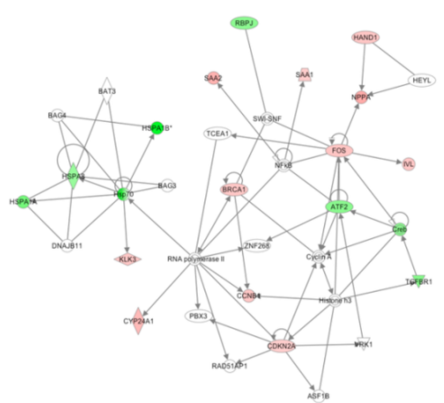

People

- Skilled GIS Professionals
- Skilled GIS Users
- Novice GIS Users
- Public Users

### Network

Network refers to both the computer, and social network. Both of these networks assist in the dissemination of data. Whether the dissemination of data is through transferring of data sets or collaboration, sharing the data from a GIS is a very common and useful operation. Additionally, these networks allow for the display of information in the form of web maps, web applications, or even paper maps using our social network.

## Key Facts



- Dissemination of data
  - Transfer datasets
  - Collaboration
- Display of information
  - Web maps
  - Web Applications

### How can a GIS be used?

Here are a few examples of a GIS in use.

- Cadastral Information
- Deliveries

- Military
- Farming
- Wildlife Mapping
- Disaster Management
- Infrastructure
- Decision Support
- 3D Analysis
- Mapping
- Redistricting
- Health
- Data Mining
- Natural Resources

Whether we are storing cadastral information to determine land valuation and taxes, tracking of our delivery vans commuting between our warehouses and customers, or managing our wildlife and natural resources, GIS can play a valuable role in assisting with these tasks.

To illustrate examples in which geospatial technology has been used, watch the Geospatial Revolution (http://youtu.be/poMGRbfgp38) which is produced by Penn State Public broadcasting.

# Thinking about Space

### Fundamental Concepts of Spatial Analysis: Space

There are three fundamental concepts of spatial analysis: space, location, and distance. Space can be measured in absolute, relatives, and cognitive terms.

#### Absolute Space

Absolute space can be thought of as mathematical space. Absolute space involves the precise measurement of location and space, such as an X, Y, and Z coordinate. Absolute space should provide an unambiguous description of space.

For example, all the information coded on the topographic map shown here, is stored and displayed as a representation of absolute location in space. All locations and features, are all described using precise measurements and calculations.

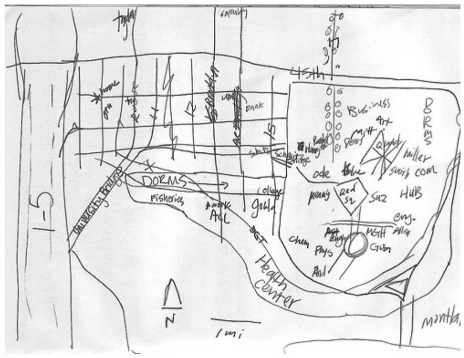
## Topological Space

Topological space can be thought of as relative space. Relative space deals with the definition of one location based on the location of another object. These topological relationships represent connectivity between features of the world. In topological space, precise measurement of space is not as important as the relative description of spatial features.

For example, this is a map of the Washington, DC Metrorail. This map shows all of the routes that the Metrorail takes. Notice that each stop is evenly spaced on this map.
Do you think that each stop is truly evenly spaced in the real world? No, of course not. As a rider of the Metrorail, are you concerned with how far apart the stops are, or how many stops before you exit the Metro? If you are a passenger on the Metrorail, topological space is much more important than absolute space as your only concern would be getting from point A to point B is how many times the doors open before you get off, not how many miles you travel before you exit the Metro.


## Cognitive Space

The third type of space is cognitive space. Cognitive space reflects people's beliefs, experiences, and perceptions about places. For example, this is a drawing of a university campus that my student drew from memory. A freshman student, may know the location of the dorms, the student union, and the dining hall, very well, but not know the rest of campus at all. Conversely, a senior, who lives off campus, may know the location of every parking lot, the locations of many classrooms, the location of the student union, but who may not be familiar with the location of the dorms.



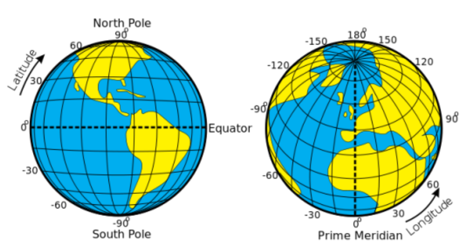

## Fundamental Concepts of Spatial Analysis: Location

There are three fundamental concepts of spatial analysis: space, location, and distance.  Location can be described in four ways: absolute, relative, cognitive, and nominal.

- Absolute - Latitude/longitude
- Relative – Site and situation
- Cognitive - Place name
- Nominal  - “Where were you when…?”

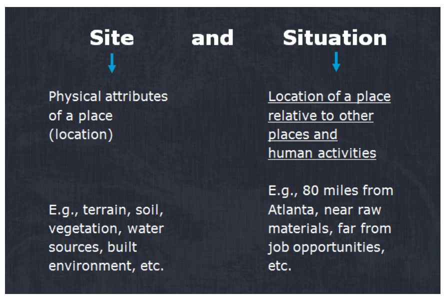
## Absolute Location

Let’s start with absolute location. Absolute location is considered an unambiguous descriptor of the location, typically expressed as a coordinate, such as latitude and longitude. An absolute location must not be confused with any other location on earth.



## Relative Location

Relative location is expressed as both the concept of the site, and a situation. Site refers to the physical attributes of the location, for example, the terrain, soil, vegetation, water source, built environment, and so on. The situation refers to the location of a place relative to other places and human activities. For instance, Walkula Springs is located 80 miles from Atlanta, near raw materials, far from job opportunities, etc.




## Cognitive Location

Cognitive location refers to the personal cognitive images of places and regions compiled from personal knowledge, experiences, and impressions. For instance, the same water body was considered an important fishing ground and source of food for the native inhabitants. Now, the same source of water is seen as a recreational area. So while this is the same physical location, it is perceived differently based on how it is experienced.


## Nominal Location

The last way in which we think of location is nominal location. A nominal location is a location that you remember being at when you heard about an event that took place at a different location. For instance, many people remember where they were when they heard the news of the 9/11 attacks. Even though they may not have been near the attacks, the location of the attacks is linked to the location the person was at when they heard the news. Therefore, the location where they heard the news is the nominal location that is linked to the distant location where the event took place.

**Key Fact**

**Systems of Measuring and Representing Location**
It is important to note that the systems of measuring and representing locations for instance absolute, relative, nominal are cultural systems which are embedded in different cultures and worldviews and can become inscribed in landscapes. Each culture will have influence on the way in which space is described and perceived.

## Fundamental Concepts of Spatial Analysis: Distance

There are three fundamental concepts of spatial analysis: space, location, and distance.  Distance is described in absolute, relative, and cognitive terms.

**Absolute**
Absolute distance is a physical unit of measure, for instance, number of miles between downtown Houston, and downtown Toronto.

**Relative**
Relative distance is calculated measuring distance, using metrics such as time, effort, or cost. For instance, the distance of two cities maybe 2000 miles apart, which is an absolute description of distance, becomes the distance of two cities measured in tanks of gas, or mileage charge.

**Cognitive**
Cognitive distance refers to an individual's perception of how far things are part. For instance, to some, driving 200 miles between Houston and San Antonio Texas is a reasonable drive. However, for others, a 200 mile drive may seem like a very, very far distance to travel if they are not used to traveling such a distance regularly.

## First Law of Geography

Lastly, you should be familiar with the first law of geography.

> "*Everything is related to everything else, but near things are more related to each other.*"
>
> Waldo Tobler
> (Professor Emeritus, UCSB, Dept. of Geography)

This is widely considered the first law of geography and speaks to the idea that space, and relative and cognitive relationships featured in the space have an influence on the development and interaction of those features.

# Lab 1: Introduction to ArcMap, ArcCatalog and Visualization



NOTE: Before beginning create a new folder called **c:\temp\GIS** that we will use to practice copying files to.

Download the dataset "lab1.zip" (zipfile format) from the class data website (see Syllabus) to your c:/temp/GIS directory and unzip the file.

**Part A**: Introduction to the ArcMap Document

**Objective:** Become familiar with the "look and feel" of the ArcGIS Desktop software interface by examining a pre-built Document.

**Part B**: ArcMap and ArcCatalog Basics Tutorial

**Objective:** Learn the basic functionality of ArcGIS Desktop software (ArcMap and ArcCatalog) by working hands-on

**Part C**: ***Lab Assignment***

**Objective:** Apply what you have learned

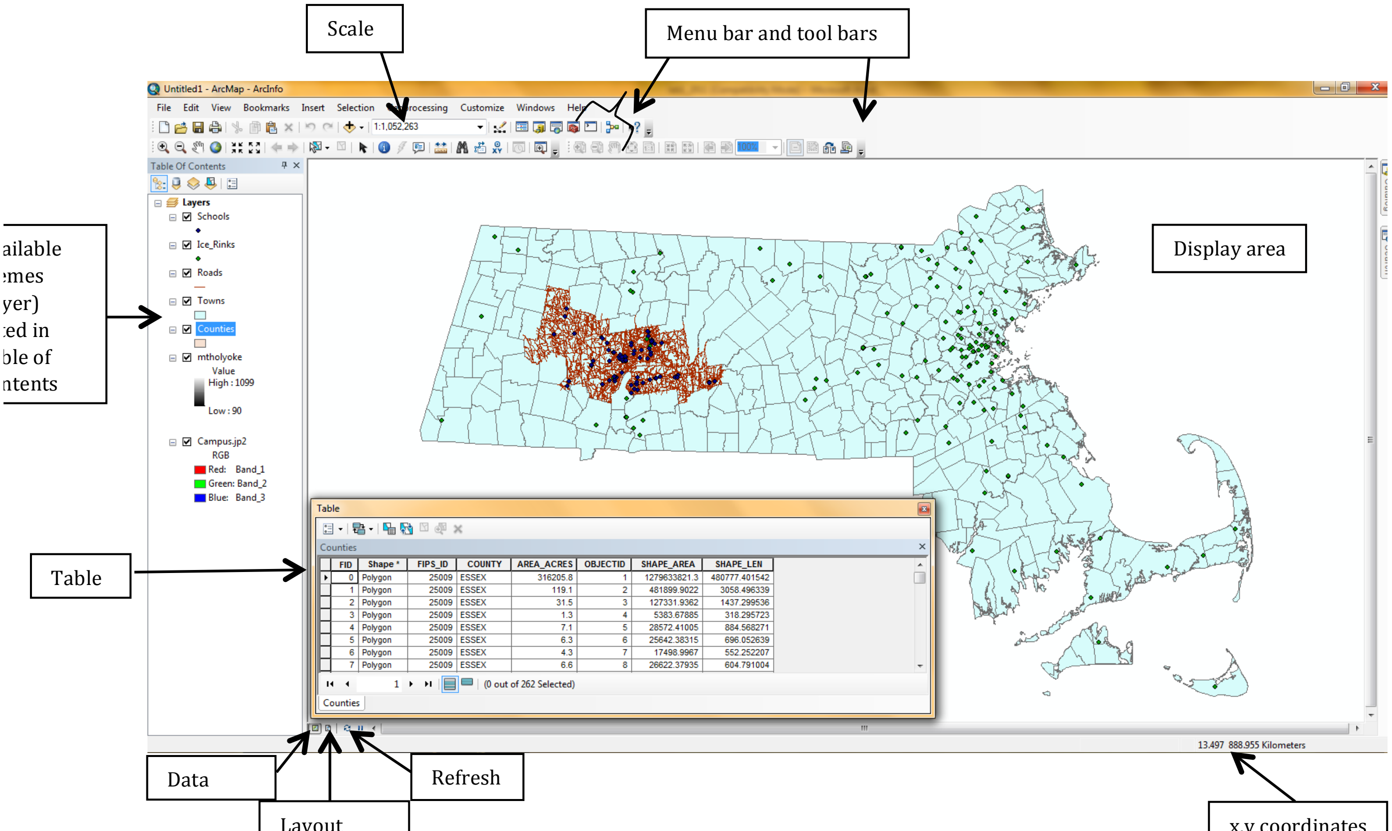
## Part A: Introduction to the ArcMap Document

1. General Characteristics

a. The ArcMap Document is a file that allows you to save the properties of the GIS you are working with so that you can open it later, retaining all the work you have done (added layers, color changes, query statements, etc.).

b. An ArcMap Document is saved with an *.mxd extension, for example, mymap.mxd.

The Anatomy of an ArcMap Document




c. An ArcMap Document is saved just like any other type of file: File → Save and give it a name (Make sure to save the document to your external hard drive!!). All your work customizing colors, text, labels, projections, added layers, etc. will be saved as is.

d. Note that our DATA files (the shapefiles, grids, and tables) are themselves not "IN" the Document. They are stored separately on the disk. The Document file simply "points" to the data files.
   - You can corrupt your Document file, but your DATA are still intact, so you can re-build your Document file (although the more complex it has become the less likely it is you will want to do this).
   - The reverse is not always true. If you corrupt one or more of your DATA files, you may or may not be able to still open your Document file; and you may have to re-create the corrupted data file(s) as well as the Document File.

**Part B: ArcMap and ArcCatalog Basics – Project Tutorial**

## A. The ArcCatalog Interface and practicing file management with ArcCatalog

a. Open the ArcCatalog Interface via Start -> All Programs -> ArcGIS -> ArcCatalog or 2x Click on the Icon on the Desktop

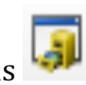
b. Note that ArcCatalog looks very similar to your normal Windows file manager (Windows Explorer)

c. The main difference is that ArcCatalog is designed to work specifically with GIS spatial data files. GIS data files have a unique structure that cannot be "read" by your normal file manager. Try to make a habit of using ArcCatalog instead of your windows file manager, as this will save you some trouble in the long-term.

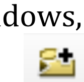
d. Copying **Shapefiles, Coverages, Grids, and Images** (Vector and Raster Data) must be done through ArcCatalog. Don't worry about these terms right now; you'll be learning more about them later. Just recognize that these are terms to describe spatial data layers.

e. You can also move and delete these spatial data files in ArcCatalog.

f. ArcCatalog also allows you to create new shapefiles, and specify coordinate systems for your spatial data which we will learn how to do in future labs.

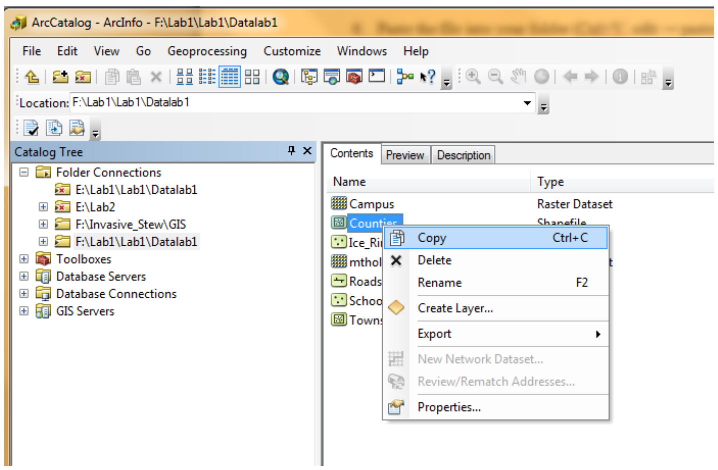
g. In the next section, ArcMap, you will see an icon in your toolbar which is  another way of opening the ArcCatalog interface.

h. To **copy** spatial data files the process is as follows:

   1. Make a folder somewhere to practice file copying to. This could be a new folder on c:\temp or a GIS\lab1 folder on your USB drive.
   2. Navigate to the directory where the shapefile, grid, or image is currently located in the catalog on the left. In this case, C:\temp\GIS\lab1.
   3. If you cannot see the C:\ drive even after it has been mapped in Windows, you need to "tell" ArcCatalog to connect to this folder using this icon → 
   4. Click on the folder (C:\temp\GIS\lab1) containing the shapefile, coverage, grid, or image in the left frame to open it.
   5. Select the file you wish to copy under the contents tab and copy the files (Edit → copy; or Ctrl+C; or right click → copy). In this case, choose counties.shp.

6. Navigate to the directory where you wish to place the copy in the left frame. This could be just be c:\temp for practice, or to a folder in your USB drive to practice.

7. Paste the file into your folder (Ctrl+V; edit → paste; or right click → paste).




*Using this method, copy the following six spatial data files into your new folder from your* **c:\temp\GIS\lab1** *folder (**note: this is for practice, but you will want to move your lab and project data into a central location whenever you work with GIS data**):*

1) Towns 2) Roads 3) Schools 4) Ice_Rinks

*Also copy the following grid:*

- mtholyoke

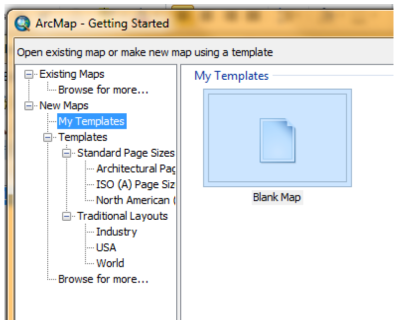
*And the following image:*

- campus

**8.** Finally, minimize the ArcCatalog window so it is off the desktop but still available if you need it.

## B. The ArcMap Interface

1. Open ArcMap by double clicking on the desktop icon or through the start menu  ArcMap will give the option to select a blank map or select from an existing map, select blank map.




2. Like other software programs, the work you do inside ArcMap can be saved and then re-opened. The terminology used to describe this saved file is an "ArcMap Document"

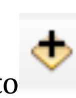
3. ArcMap Documents saved to your hard drive have an .mxd extension – for example, mymap.mxd.

4. For now, select "blank map", since you have not saved anything yet.

5. Once you click OK, you will have an empty document which you will proceed to "fill" with the "Layers" (aka "Themes") you copied previously using ArcCatalog.

6. A "Layer" may have been constructed in one of several formats. The four primary ones you will be using in the class are: 1) a shapefile, 3) a coverage 3) a grid, or 4) an image. All of these spatial data formats will become familiar to you over the next couple of weeks. For now, just recognize that the word layer is simply ArcMap's way of collectively referring to the numerous types spatial data file formats that it can open and manipulate.

Remember that Spatial Data Files are different from most types of digital files and thus, there are *specific methods* for copying them. Shapefiles and grids have *multiple files* associated with them. If you try to copy them manually you may lose or corrupt one of these files. Remember to always use ArcCatalog when moving, copying and deleting GIS Spatial Data Files.



7. Now that the data have been copied to your directory, you are ready to add them to your Document. To accomplish this, click the Add Data button in the Button Bar.

*You can also add layers by using the MENU BAR with File → Add Data. There are often two or three different methods for doing the same thing in ArcMap*

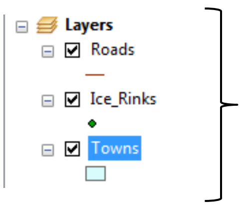
8. An 'Add Data' window will open. Navigate to <u>your folder</u>. Select the 9 layers (7 shapefiles, 1 grid, and 1 image) you have just copied and add them to the Document.

9. Notice that the names of the layers have now appeared in the bar on the left. This area is referred to as the Legend.

- You can add layers together or individually.  You can also add the same layer more than once.  To delete a layer from the project, right click on the name within the legend and choose remove.

- To change the layer order, click and drag the layer to your desired location. Change the layer order to match the legend shown below.

- Turn the layers on or off by clicking in the check box to the left of the layer's name.

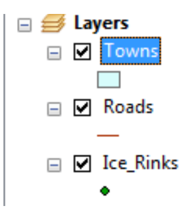



Legend

10.  Drawing Order

a. ArcMap draws layers from the bottom, to the top.  In the legend shown above, it is drawing Towns first, then Ice_Rinks, then Roads at the top.  This is important because there may be a situation when you have added a layer but you cannot see it.  In this case, one of the other layers is probably drawing over it.

b. To demonstrate this to yourself, drag the Towns layer to the top.  Your Legend should now look like this:




c. Note that, in the Document, you can no longer see Roads and Ice_Rinks because the features of the Towns Layer is covering them up.

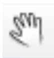
d. Drag Towns back to the bottom of the Legend.

11.  Exploring the **Tool Bar Icons**

a. The Zoom In and Zoom Out Tools allow you to zoom to particular areas of interest manually by clicking and dragging. You can also zoom to the extent of a particular layer by right clicking on that layer in the legend and selecting „Zoom to layer".

b. The Fixed Zoom In and Fixed Zoom Out Tools automatically zoom in and out from the center of the current view.

c. The Pan Tool allows you to move the center of the view around by clicking and dragging from anywhere in the scene.



d. The Full Extent Tool zooms to the full extent of all your active layers.

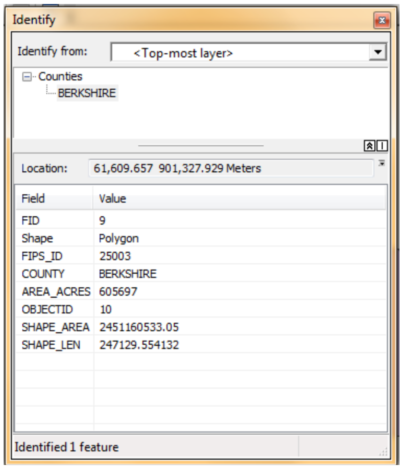
e. The Previous Extent Tool goes back to the previous extent of your view (this is particularly useful if you accidentally zoom somewhere you didn‟t want to go!).

f. The Select Features Tool allows you to click and drag over features active in the view and select ones of your choosing. (You will see why this is useful later when you view selected features within tables or export selected features into new shapefiles)

g. The Select Elements Tool allows you to select any text features or additional items (labels, legends) you have added to the view. You will use this tool later on since you have not added anything yet.

h. The Identify Tool is used to bring up a record about a geographic feature that you click on in the Document window. The default is the top-most layer in the legend, but you can select which layer you are interested in under the Layers drop down menu.

- Move the cursor into the Document window (the cursor should now have a little “I” symbol next to it) and over any feature and click once. An “Identify” window will appear. Within the results is all the known information about that feature.




- Experiment by changing the active layer in the drop down menu and identifying features of other layers you've added to the project.

i. The Find Tool allows you to select features within one of your files based on a keyword or property of the file. 

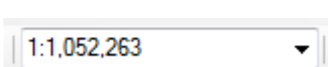
- Select the Find Tool. In the dialog box that appears, select the layer you wish to search (try *Counties*) and type in a keyword (try *Hampshire*). Hit find and features containing that word will appear at the bottom. By right clicking on the features, you can choose to zoom to that feature, select that feature, identify that feature etc.

j. The Measure Tool will measure distances on your view. 

- Click once on an ice rink, and then move the cursor (now a ruler) to another ice rink to measure the distance between.
- In the lower left hand corner of your ArcMap window you will see two measurements: segment and total length.
- Click once and move to another ice rink. Notice that total length is the cumulative distance you have measured, while segment length is just from the last click. Double click to delete this measurement and start over.
- Experiment with changing the units of measure which are displayed by right-clicking on the Layers → Properties → General Units → Display

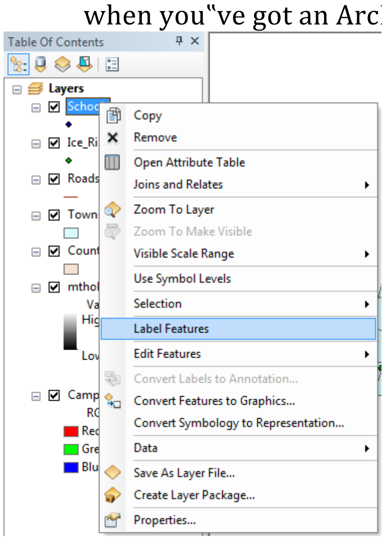
## 12. Other useful tools: **Scaling, Text, Labels, and Graphics**

a. The Scale Tool allows you to set the scale you choose, either via the drop down menu or by typing in your desired scaling. A scaling of 1:1 means that one unit on the screen (inch, centimeter, whatever) equals one unit on the ground. A scaling of 1:1000 means that one unit on the screen equals 1000 units on the ground. We will discuss more about scale in lecture. Experiment with different scaling to see which seems most appropriate for this view.

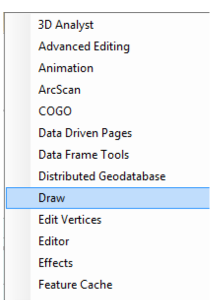


b. The Text Tool (under Insert → Text) lets you add free hand text to the view. Once you have added your desired text, use the Select Elements tool to move (click and drag) or edit (double click) the text. You can edit the text to specific sizes, colors, and rotation angles by double clicking the text and clicking Change Symbol. In the Change Symbol box, Properties will give you even more options.
c. The Label Tool (right click on the layer → Label Features) labels all of the features in that particular layer.

- The appearance of the labels can be changed by right-clicking the layer, then choose Properties.
- Once the Layer Properties window opens, find the Labels Tab. This is one of the places you can change the font, size, color and placement of labels.
- Getting labels to appear the way you'd like can be challenging – more reason to save often when you‟ve got an ArcMap Document looking the way you'd like!




d. The Draw (graphics) menu can be added to the view by right clicking on the menu bar and selecting Draw. The Draw menu will appear at the top of the screen. Using this you can add boxes, text, lines, shapes, call-outs etc. Experiment with the different options!





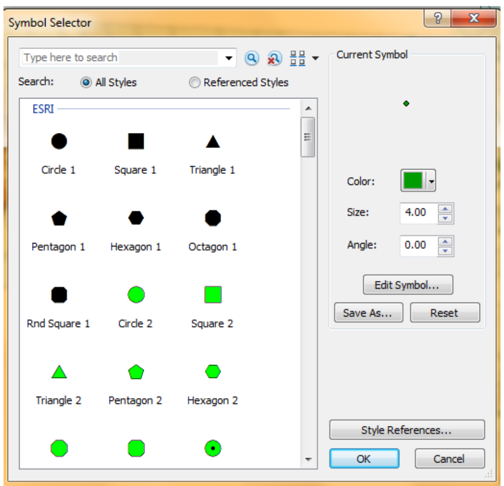
## 13. Changing how the **Layer Name** Appears

- You can change the name of a layer in the legend by selecting the layer in question, and then clicking it again (don't double click, that opens up another menu). You can edit the name directly in the legend.

- Note: <u>By doing this you are not changing the name of the file! You are simply changing what it is called in this Document</u>. When you save the project and return to it later, the new names will be maintained.
- Renaming files in the legend is a good idea, particularly when the files themselves aren't readily recognizable (What was bldgs.shp after all??).

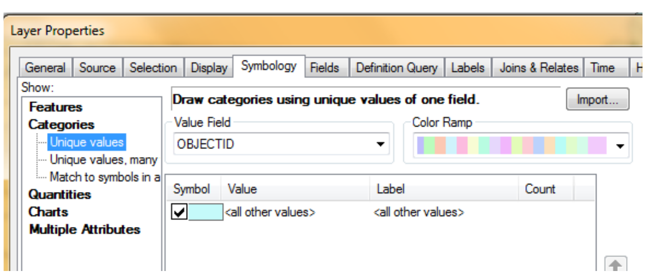
## 14. Changing the **Symbology** of a Layer

- You can change the color or shape of a feature within the layer by double clicking on the default symbol that appears below the Layer Name. Double-clicking will bring up a Symbol Selector menu. Experiment by bringing up the symbol selector for different types of data: Point (Ice_Rinks), line (Roads), and polygon (Towns).



Symbol Selector for a Point Shapefile

- Another way to change the way the layer is displayed is to select the Layer Name, then right-click. On the pop-up menu, choose Properties.
- Next, choose the Symbology Tab. Here, you can choose how to display this layer. Right now, it should be displayed as a single feature class. This means that all the Towns currently look the same




- You can make each Town appear a different color by:
  - Select Unique Values under the categories tab
  - Select *Town* from the Value Field drop down menu
  - Click the Add All Values button
  - Click OK
  - This will automatically assign a unique color to each Town

- Try displaying the Towns layer based on 2000 Population in graduated color, graduated symbols and dot density (Under the quantities tab).
- Are the default numbers of classes/dots appropriate for your data?

> *The ability of GIS Software to map geographic features (in this case Towns) based on an individual attribute or characteristic from a related database of information (in this case each Town's name or population) is one of the fundamental strengths of GIS. In addition, this ability is one way of distinguishing GIS from other software products such as Photoshop, which is simply a graphics editing program.*

15. If you haven't already, add the raster file mtholyoke to your Document.

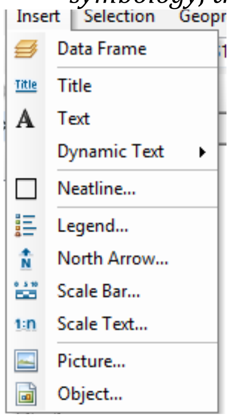
- This particular example of a raster file is referred to as a Digital Elevation Model (aka DEM) because it represents ground elevations (topography) of a portion of the Holyoke Range. We will discuss DEMs more in lecture.
- Make sure the file is at the top of your legend so you can see it! Double click on mtholyoke to open the layer properties window. Notice that the **Symbology** tab options are different. The reason is there is a great deal of difference between raster and vector data (towns, schools, roads). We'll also talk about this later in the class.
- For now, practice displaying your raster data using the stretched and classified options – is classified useful in this case? Next, try displaying it as unique values - is unique values useful in this case? When might you want to use unique values for a raster dataset?
- Finally, if you haven't already, add the raster file campus.jp2 to your view. This is an aerial photograph of part of Amherst (where are you now??). This is an example of another type of raster data that you might work with. Try using the identify tool to click on parts of the image. How is the identified result different from what you see when you click on a Town?

16. The Layout View: Adding Legends, Scale Bars, and other Map Goodies

In order to create a hardcopy map/poster of the work you have done, you need to change the ArcMap Document so it is displaying the **Layout View** (View → Layout View). Up until now, you have been displaying and working in the **Data View.**

a. Once you've changed to the **Layout View**, you'll notice the Document area looks different. Now ArcMap has confined you to a virtual piece of 8 ½ x 11 paper!
b. In addition, your Layers now appear inside (are contained in) a **Data Frame** on the paper.
c. And, a layout toolbar will now be present. The icons on this toolbar let you move around on the page.
d. You can resize the Data Frame by selecting it with your Select Elements Tool, and pulling on the corners of the Data Frame.
e. Under File -> Page Setup you have the options of changing the orientation of your paper to portrait or landscape, changing the page size, etc. For now, keep you pagesize at 8.5 x 11, but experiment with changing the page orientation so you can see how the Data Frame responds.
f. Also, check out the options in the Insert pull down menu. You can now add legends, scale bars, north arrows etc. Experiment with the insert options. *Note: A cool thing about GIS is*

*that all the graphic elements are linked to your original files. If you change the size, scale, or symbology, the elements will automatically update!*




g. After you have added legends other elements to your map, you can manually edit those items by right clicking on them and selecting 'Convert to Graphics'.

h. Once you have converted the element to a graphic, right-click the element again and click Ungroup.

i. You can then edit each individual component of the graphic element (such as re-wording text, or changing color) by double clicking the part you want to change. *Note: Once you've converted to graphics, the inserted items are no longer linked to the original files. Be careful about converting something to graphics and then changing the layers – you'll have to insert the item again in order to update it.*

j. You can also use the Layout View to add a grid to your map (i.e. coordinates or latitude/longitude marks around the outside). Do this by selecting the data frame, right clicking and selecting Properties (at the bottom). Select the Grids tab in the Data Frame Properties and add a new grid. Experiment with different types of grids.

**Part C**: **Lab 1** ***Assignment*** – "Mapping Some of UMass Infrastructure" (**20 points total**)

[Note to Charlie – test #3 below before the lab]

Scenario – You've just been hired by UMass to help create a base map of some physical infrastructure in and around our campus. UMass Facilities Planning (FP) is using GIS heavily to manage our campus infrastructure. Fortunately, FP has the spatial data you need already in digital format. The data are:

> **PVTA stops** = a point layer of PVTA bus stops on campus
> **Bike racks** = a point layer of bike rack locations on campus
> **Buildings** = The aerial footprint of campus buildings
> **Orthophoto** = a 2009 georeferenced color aerial photograph of the campus (a "webservice" – on a server hosted by campus planning)

1) Create a new directory for the **Lab1_HW** assignment on a USB drive or on your computer.
2) Download the "Lab1_UMass_data.zip" file from the course data repository (see Syllabus) your computer or USB stick. Unzip the file. Open the Lab1_UMass_data.zip folder.

3) Double-click on the "UMassphoto2009WebService.mxd" and ArcMap will invoke and if the UMass Facilities Planning data server is working, you should see the color orthophoto appear.

4) But if ArcMap opens up but you see a red "!" next to the orthophoto entry in the ArcMap table of contents and no aerial photo appears (just a white screen), then use the UMass_photo.zip file that downloaded for this assignment. Unzip that file and copy the contents to the Lab1_HW folder. Then click the "+" icon on the ArcMap menu (add data), use the "connect to folder" icon (the little folder with a + on it) to connect to your Lab1_HW folder, and navigate to find the "1_117902.tif" raster dataset. This is an older black and white aerial photo of the UMass campus. Select it in the list with a left click, and choose the Add button to add it to your ArcMap document. It may give you a "Geographic Coordinate System Warning". Just select "close." The black and white aerial photo image should appear.

4) Make a map using the data layers. While you have been given a certain amount of freedom to exercise your cartographic skills with color, font, symbology, etc., the UMass FP office has stipulated that the base map must at least contain the following information:

The map must include (**1 pt each**):

1) A title, a scale bar in *meters and a* north arrow

2) A legend including a color-coded building class (academic/research, residential, admin/support, etc.), PVTA stops and bike-rack locations.

3) Buildings must be displayed by Class (color coded)

4) PVTA stops should be designated with a yellow triangle symbol

5) Bike-rack locations should be identified with bright-green dots.
6) An aerial photograph must be displayed as a backdrop
7) The Campus Pond should be labeled
8) The Du Bois library (28 floors), Lederle Graduate Research Center (15 floors), and Thompson Hall (10 floors) should be labeled with their floor count numbers. (If you aren't familiar with the UMass Amherst campus, see http://www.umass.edu/visitorsctr/downloads/campusmap.pdf.
9) The scale of the View must be set to 1:9000 and with the Du Bois library in the middle, and the Lederle Grad Research Center visible in (roughly) the top right.
10) Add your name to the map: "Map author: your name"

Please answer the following questions should be provided at the bottom of the map (**2 pts each**):

11) What is the approximate E-W distance of campus from the Mullin's Center on the left of the image to the residential dorms on the right of the image (in feet)?
12) What is the approximate perimeter distance around the campus pond (in feet)?
13) What is the Stop_Name of the PVTA stop that is closest to the campus pond?
14) How many bike-racks are recorded in this database?
15) What is the approximate x,y coordinate location of the center of campus using this system's coordinate system?

**Lab Report Deliverable:**

Your lab report this week consists of

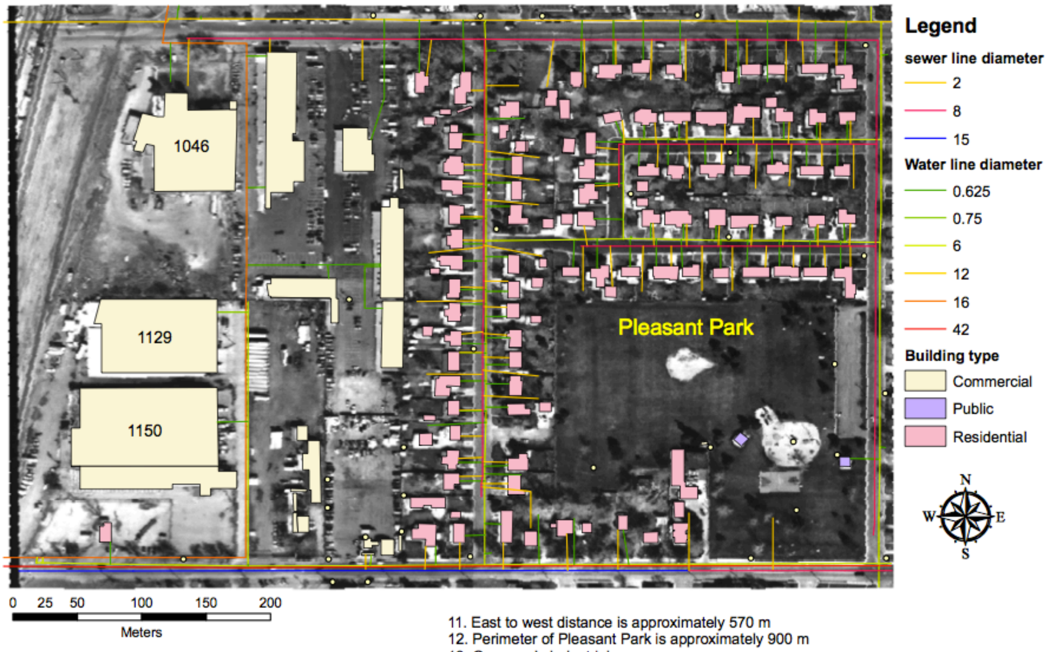
1) A digital image of the map you created inserted into a MS Word document.

2) The answers to questions 12 – 15 above below the map in this document.

Make sure to name the files with your name (e.g., **yourname_lab1.doc**), and upload it to the Lab 1 assignment on Moodle.

An example of the kind of output I am looking for is below. Note that this example has NOTHING to do with this assignment or UMass. I provide it to show you the kind of output we are looking for in this assignment.

## Maplewood (any similar title is fine)




11. East to west distance is approximately 570 m
12. Perimeter of Pleasant Park is approximately 900 m
13. Goorman's Industrial
14. 32 street lights
15. -84.89, 33.72

Extra. Higher numbers relate to brighter (whiter) areas on the aerial photo.

# Lab 2
# Vector-based Analysis in ArcMap: Selection and Data Querying

**2013 Bethany Bradley**



This week's lab focuses on working with and querying tables associated with spatial data. Everything you will learn this week is primarily relevant for shapefiles (vector data) which have attribute tables associated with them. These tables are similar to information you might find in Excel, except that each row is also linked to a specific spatial location.

## A. Getting Started

1. Create a new folder for Lab 2 on your USB drive
2. Unzip Lab2.zip into your Lab 2 folder. Open ArcMap and add the shapefile countries from \datalab2. This is a polygon shapefile of world countries as of 2010.

## B. Exploring the attribute table

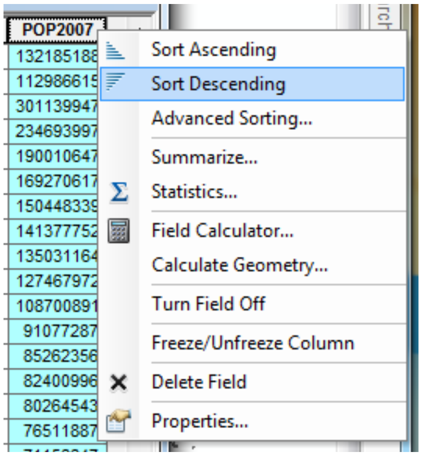
1. Right click on countries and select 'open attribute table'
2. The table associated with this shapefile will pop up. It currently contains information about the name of each country, the diplomatic status of the country, and the population of the country in 2007 (although there are some zeros representing unknown population counts).

Table

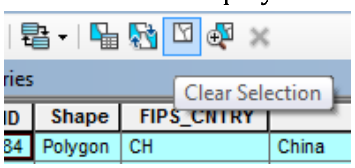
countries

| FID | Shape | FIPS_CNTRY | CNTRY_NAME | STATUS | POP2007 |
|---|---|---|---|---|---|
| 0 | Polygon | FK | Falkland Is. | Non-Self-Governing Territory of the UK | 0 |
| 1 | Polygon | FG | French Guiana | Oversea Department of France | 203321 |
| 2 | Polygon | GY | Guyana | UN Member State | 769095 |
| 3 | Polygon | PC | Pitcairn Is. | Non-Self-Governing Territory of the UK | 0 |
| 4 | Polygon | SX | South Georgia & the South Sandwich Is. | UK Territory | 0 |
| 5 | Polygon | SH | St. Helena | Non-Self-Governing Territory of the UK | 7543 |
| 6 | Polygon | NS | Suriname | UN Member State | 470784 |
| 7 | Polygon | TD | Trinidad & Tobago | UN Member State | 1056608 |
| 8 | Polygon | VE | Venezuela | UN Member State | 26084662 |
| 9 | Polygon | AQ | American Samoa | Non-Self-Governing Territory of the USA | 57663 |
| 10 | Polygon | CW | Cook Is. | Self-Governing in free association with New Zealand | 21750 |

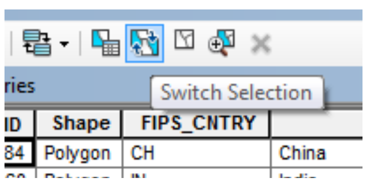
3. You can sort each column (as you might also do in excel) by right clicking on the column name and choosing 'sort ascending', 'sort descending', or 'advanced sorting'. For now, try sorting the table by population in descending order. What were the three countries with the highest population in 2007?



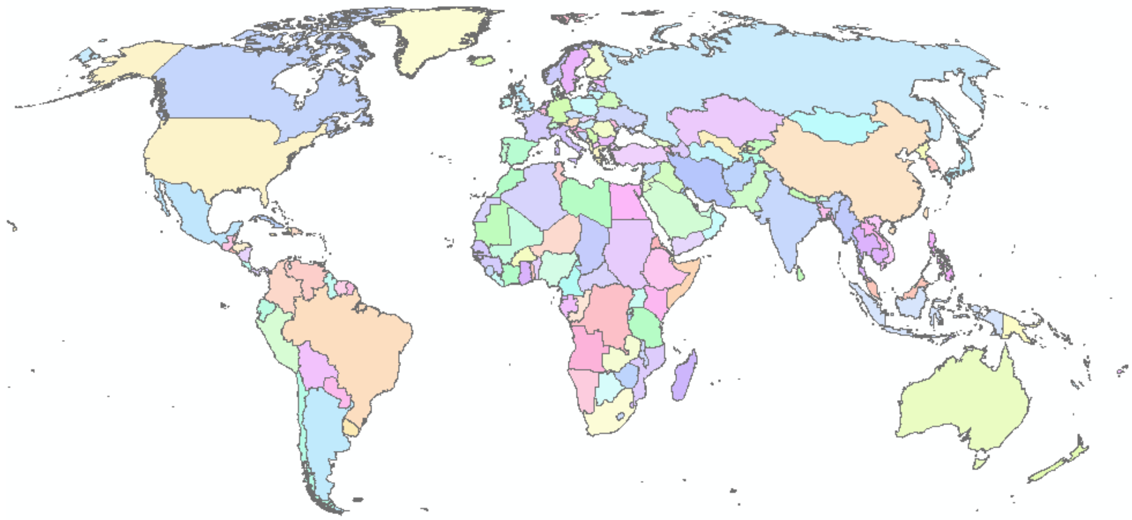

4. You can also use the table to select countries of interest. In the table, select the top 10 most populated countries by clicking on the grey squares at the left of the table. You can select multiples by <control> click, or by <shift> click.
5. You can clear a selection by clicking the 'clear selection' button in either the attribute table or in the display



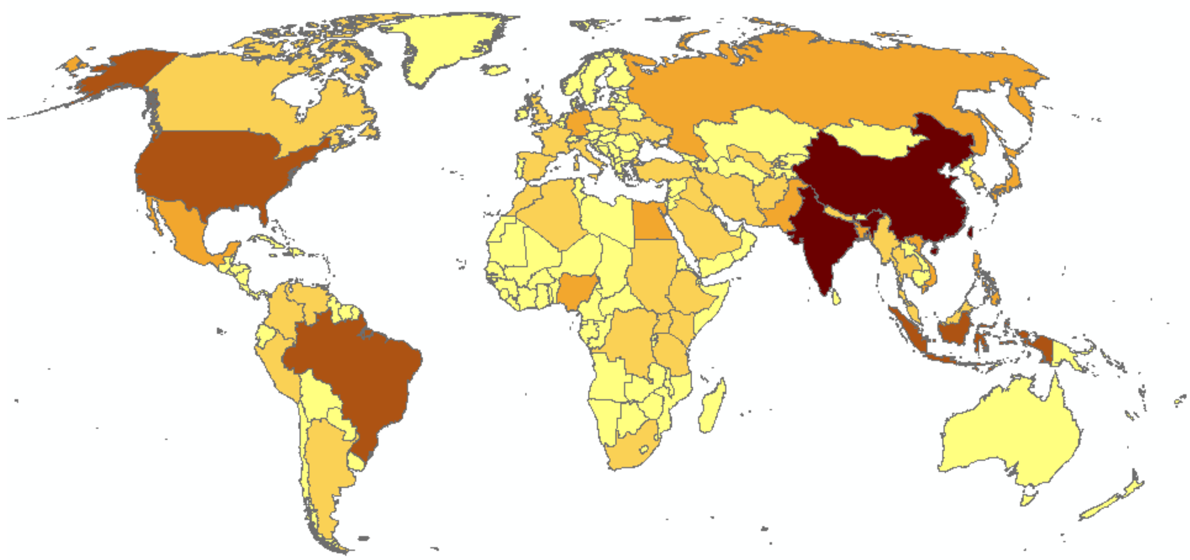

6. You can also switch the selection between selected and unselected features with the switch selection button in the attribute table.




7. Since we've got a nice range of populations in our different countries, let's use your skills from last week to make a map of population by country. Right click to open the layer properties of countries and change the color display of the shapefile. Last week you used a display by category. Try that for these data by adding all of the country names as unique values. Your map should looks something like this:



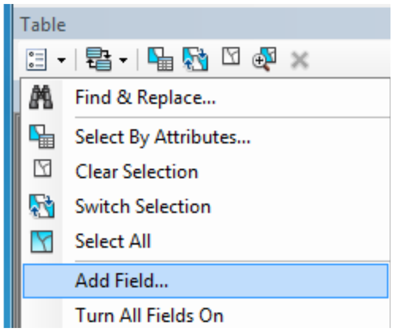
8. This map does a better job of emphasizing the different countries than just the borders alone.
9. Now try displaying population using graduated colors. Your new map should look something like this:



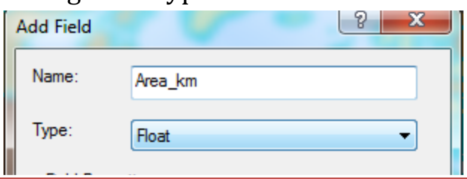
10. This new map still conveys the country borders, but also provides additional information about population.

**C. Table operations – calculating geometry and the field calculator**

1. Now, let's say that you want to create additional information in your shapefile. For example, we know population of each country. Let's use GIS to calculate population density of each country in persons/$km^2$
2. The first thing we need to calculate population density is country area. Because features in a GIS are geolocated, it is easy for ArcMap to calculate any geographic attributes.
3. First, you need to add a new column to your attribute table. Do this by opening the table options and clicking 'add field'.




4. Name your new field Area_km (***NOTE: GIS limits you to 10 characters in column titles. Use them wisely! You cannot use spaces or other special characters. If you are calculating something with units, I recommend trying to put the units in the title. For example, if you just titled the column 'Area', you wouldn't immediately know if it was square kilometers, square miles, acres, hectares etc.)*
5. Change the Type to float and hit OK




---

**What does Type mean?**

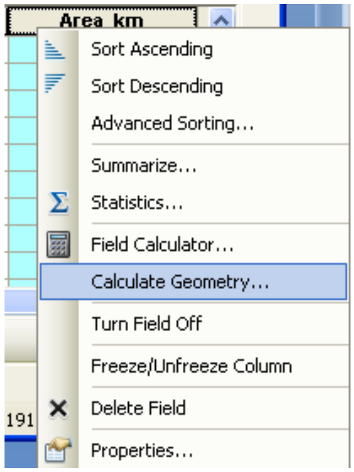
Whenever you add a field, ArcMap defaults to calling the data values "short integer". This may not seem like a big deal, but it is often a source of problems later on. "Type" refers to how the computer and software store information. An integer is a whole number (e.g., 1, 2, 3589). A "short" integer in computer terms ranges from -32768 to 32,768. Those weird numbers encompass $2^{16}$ numbers, which the computer stores as 16 bit (or 2 byte). If you go to "long" integer, you get $2^{24}$ integers to choose from (out to over 8 million).

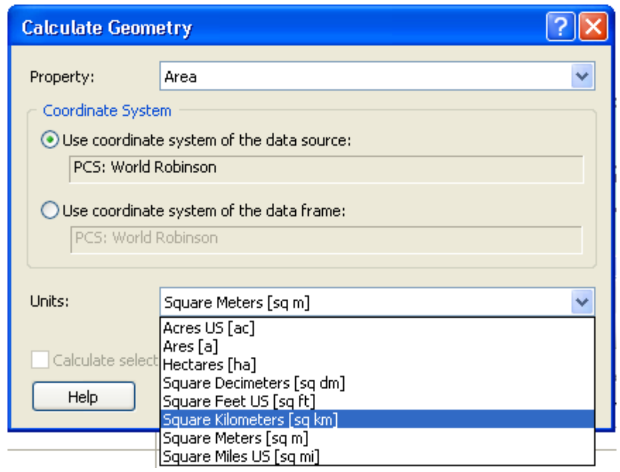
There are two other important "Type" categories that you'll probably use. The first is "float", which you just used for your new area column above. "Float" refers to floating point, which includes both integers non-integer numbers (e.g., 3.14, 3589.01). If you ever calculate a fraction, make sure that you have assigned a column value of floating point – otherwise ArcMap will round to whole numbers.

The second important "Type" category is "text", which is probably self evident. If you want the values of the column to be letters instead of numbers, make it text. As you might guess, you can't do any numeric calculations on text (try dividing Armenia/2 – doesn't work). Likewise, if you name a column type as text and put numeric values into it, ArcMap won't let you do any numeric operations because it sees the column as text.

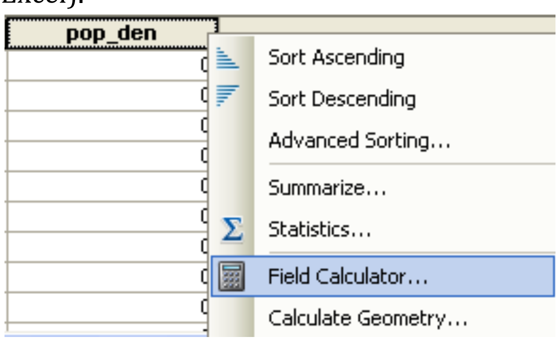
---

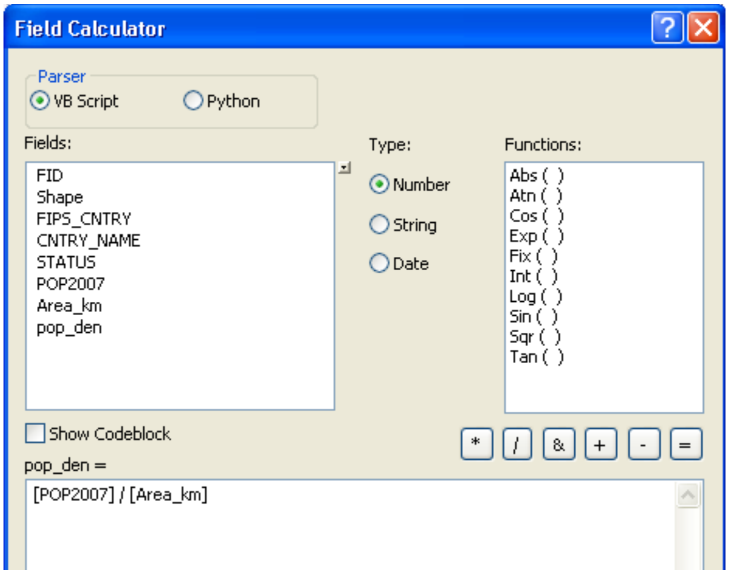
6. Right click on the new column title Area_km and go to calculate geometry. You may get a warning about calculations outside of an editing session – ignore this (click OK). You'll learn about Editor in a few weeks.



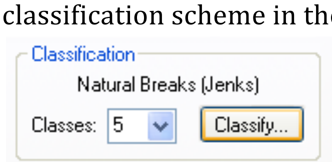

7. By changing the Property of calculate geometry, you can calculate area, perimeter, X centroid or Y centroid of a polygon shapefile. If this shapefile were a line file, you'd be able to calculate length (but not area).
8. Change the units to square kilometers and click OK. Now you've got a new column in your table of country land area in km$^2$.
9. Now you have the information you need to calculate population density. Add another new column, call it pop_den and make it floating point. Right click on pop_den and open the field calculator. Field calculator allows you to do any numeric manipulations on columns of data (similar to Excel, although it doesn't have quite as much functionality as Excel).



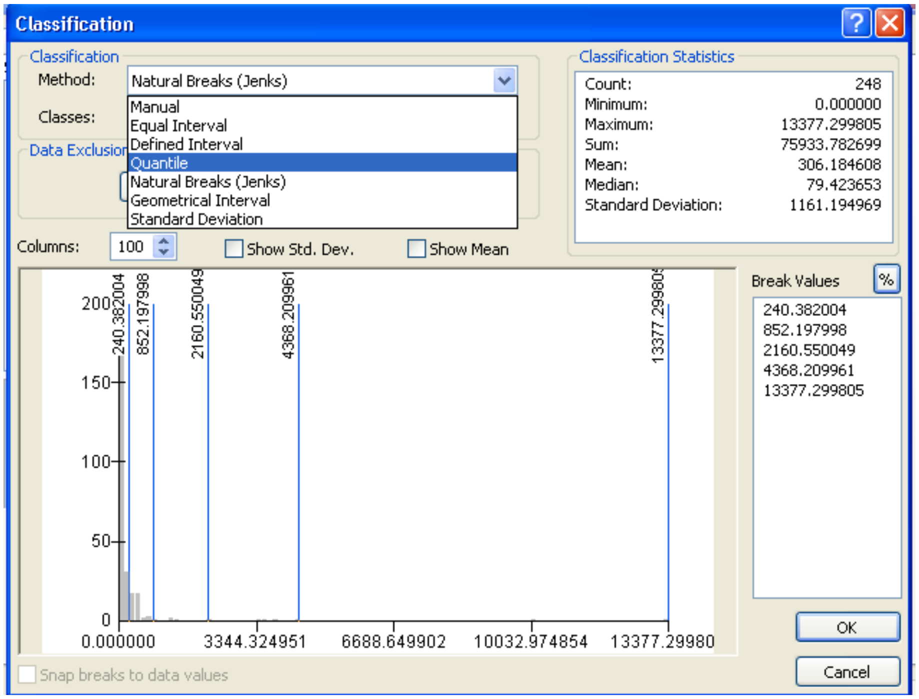



10. Use the field calculator to divide POP2007 by Area_km. The easiest way to do this (and least prone to error) is to double click on the field and the operations that you want. Click OK.
11. Now, make a map of graduated colors based on population density. The default map might look something like this:

12. This color scheme is not terribly useful because it divides the colors in the default "Natural breaks" between 0 and 13377. Make a better color stretch by adjusting the classification scheme in the symbology tab.



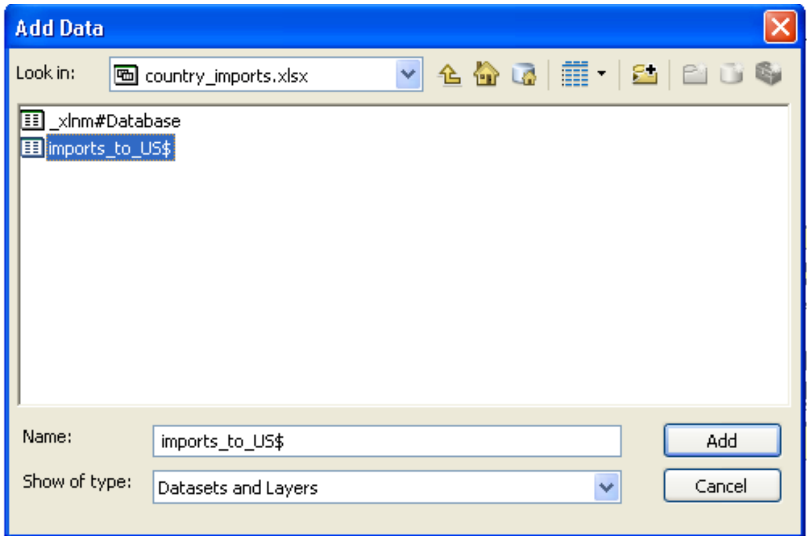

13. There are several methods of dividing the classes, but try using Quantile to see how that looks. If you have five quantiles, then the top 20% will be one color, second 20% another etc. If you choose ten quantiles, then the top 10% will be one color, second 10% another etc.



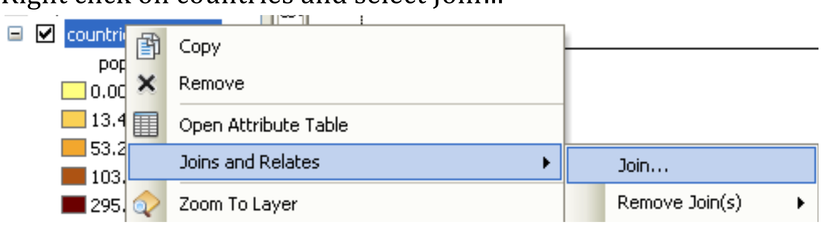

14. Now you've got a pretty new map that effectively displays global countries by population density.

## D. Joining tables to spatial data

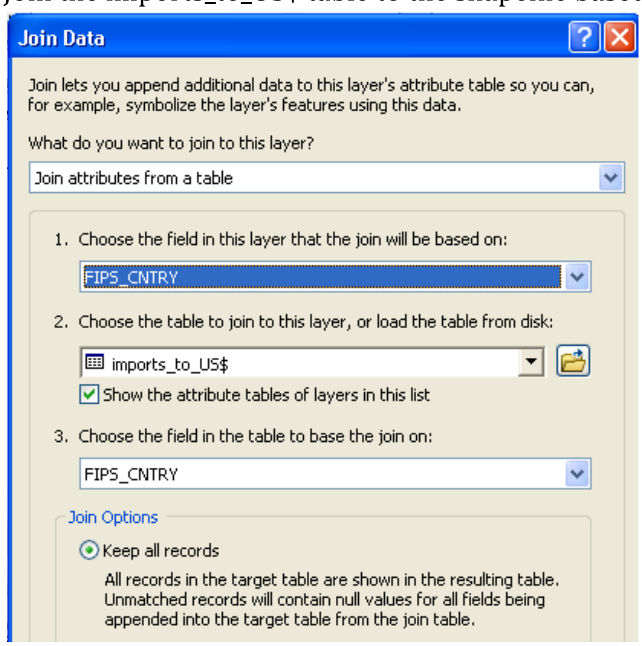
1. Sometimes spatial data already contain all the information you need. But, often you'll have some other information that needs to be connected to the spatial data. In this case, you've got a shapefile of countries of the world and information about their names and populations (and now population densities). Any additional information about countries (e.g., native language, GDP, percent forest cover) will need to be joined to the shapefile.
2. In your \datalab2 folder you will also find an excel spreadsheet named country_imports.xlsx. Open the spreadsheet in excel and check it out. This spreadsheet contains country codes, country names, and two new categories called all_imprt which is the value (in $US) of all imports to the U.S. coming from that country. Plnt_imprt is the value (in $US) of all live plant imports to the U.S. coming from that country. CLOSE EXCEL (if you try to open the same file in Excel and ArcMap at the same time, bad things can happen).
3. Add the Excel sheet to your ArcMap document




4. Right click on the file name to open it – you should see the same table that you just saw in Excel. Notice that the table values FIPS_CNTRY and CNTRY_NAME are identical in this table and in the countries shapefile. Joins will only work if the names are identical.
5. Right click on countries and select Join…




6. Join the imports_to_US$ table to the shapefile based on FIPS_CNTRY and click OK

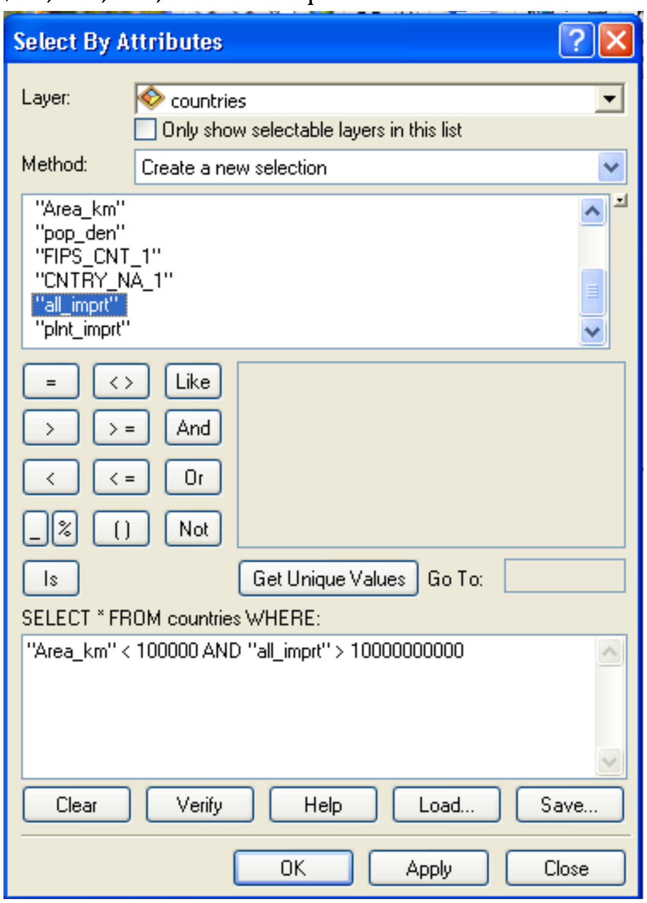



7. Now, open the attribute table of countries. If you scroll over to the right, you'll see that it now contains all of the columns from the imports_to_US table

8. If you wanted to, you could now perform new field calculator calculations with your new fields. For now, change the display to show each country by value of goods they export to the U.S. What country do we import the most from in terms of dollar value?
9. ***Note: Joins are not permanent! If you want one of the new imports columns permanently within your shapefile, you'd need to add a new field and set it equal to one of the imports columns using field calculator. Alternatively, you could save a new copy of the shapefile that contains all of the joined data.*

**E. Selecting by attributes**

1. Finding the highest or lowest value of something is easy to do by sorting the attribute table. But, sometimes you want to select features using more complicated queries. For example, what if you wanted to select all the small countries that do a lot of exporting to the U.S.?
2. Click the Selection menu at the top of ArcMap and choose Select by Attributes (you'll explore Select by Location in next week's lab).
3. Select by Attributes defaults to 'Create a new selection' (but you could also choose to add/remove or select from an existing selection).
4. This time, create a new selection of countries that have areas less than 100,000 km$^2$ (roughly the size of Nicaragua) AND total exports to the US greater than $10,000,000,000. The equation should look something like this:

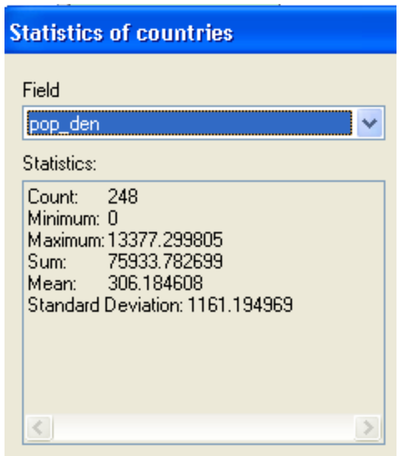



5. How many countries fit these criteria? (The answer should be 26 – make sure you can figure out how to find this information).

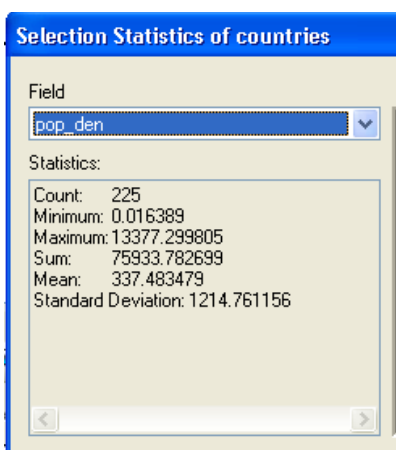
6. Now, see if you can find the small countries with areas of less than 100000 km$^2$ that export between $10,000,000 – $30,000,000 of plants to the U.S. Which countries fit these criteria?

**F. Summarizing and attribute table statistics**

1. Clear your selections and open the attribute table for countries again. What if you want some basic statistics about population density for all countries in the world? Right click on the pop_den column heading and select Statistics




2. Statistics gives you the min, max, mean of the column. But, there's a problem with the minimum and the mean here – recall that some countries didn't have data for population, so our calculated population values of 0 are not actually 0 people/km$^2$, instead they refer to "No Data". It would be better if our statistics didn't include the zeros.
3. Fortunately, you can also perform statistics on a selection within your table (in fact, statistics will always show you stats for the selection unless you have nothing selected). Use select by attributes to select all countries with pop_den greater than 0 and recalculate statistics.




4. Now you see that the numbers have changed, and 0 is no longer the smallest number. Can you figure out the mean value of total exports to the U.S. coming from each country? (Check – the number should be $37.5 billion)
5. Clear all selections

6. Finally, let's look at the Summarize tool in tables. This tool is extremely useful when you have multiple features that fall under the same category. For example, in last week's lab, Maplewood had commercial, residential, and public buildings. You could use calculate geometry to figure out the area of each building's footprint, and then use summarize to calculate the average area for each building type.
7. Open the countries attribute table and notice that several countries have the same "status" (e.g., UN Member Country). Right click on status and select Summarize. In this screen, you can rename the output table and specify where you want to keep it. This will be a temp file, so you can save it someplace you can delete later. Click OK and click yes to add the resulting table to the map.

Table

Sum_Output

| OID | STATUS | Count_STATUS |
|---|---|---|
| 0 | Australian Territory | 4 |
| 1 | British Crown Dependency | 3 |
| 2 | British Crown Territory | 1 |
| 3 | Commonwealth assocaited with the US | 1 |
| 4 | Commonwealth in poltical union with the US | 1 |
| 5 | Non-Self-Governing Territory | 1 |
| 6 | Non-Self-Governing Territory of France | 1 |

8. Open the table – because we didn't check any other columns, this summary just gives you a count of the number of features (countries) that have each status. How many countries are UN Member States? Can you think of another way that you could have found this information using Select by Attributes?
9. Let's try Summarize again, but this time we'll get a little more information out of it. Go back to the countries attribute table and right click Status to Summarize. This time, check 'Average' under pop_den and 'Sum' under all_imprt. Click OK and add the table to the map. Open the new table and check it out.

Table

Sum_Output_3

| OID | STATUS | Count_STATUS | Average_pop_den | Sum_all_imprt |
|---|---|---|---|---|
| 0 | Australian Territory | 4 | 0 | 117143900 |
| 1 | British Crown Dependency | 3 | 552.357 | 0 |
| 2 | British Crown Territory | 1 | 0 | 0 |
| 3 | Commonwealth assocaited with the US | 1 | 508.184 | 0 |
| 4 | Commonwealth in poltical union with the US | 1 | 425.149 | 0 |
| 5 | Non-Self-Governing Territory | 1 | 1.6244 | 0 |
| 6 | Non-Self-Governing Territory of France | 1 | 13.6932 | 0 |
| 7 | Non-Self-Governing Territory of New Zealand | 1 | 0 | 320019000 |
| 8 | Non-Self-Governing Territory of the UK | 10 | 805.5905 | 14679126000 |

10. With this new summary, you can now figure out (for example): What is the average population density of UN Member States? How does that compare to the Occupied territory? What is the total value of annual U.S. imports from Non-Self-Governing Territories of the UK?

## Analyzing Grizzly Bear Habitat

**Problem Statement:**

A century and a half ago, nearly one hundred thousand grizzly bears roamed the mountains and plains of the western United States. Today, only a few thousand remain. The grizzly bear, *Ursus arctos*, is listed as a threatened species under the U.S. Endangered Species Act, requiring that the federal government restore the bear's population to a level that removes them from the threat of extinction. To do this, the U.S. Fish and Wildlife Service designated five recovery zones in the Northern Rockies of Wyoming, Montana, Idaho, and northeastern Washington. A sixth is being considered in the central Idaho-Bitterroot area.

**Data:**

The data you need to answer these questions is in the \Grizzlies folder

- A shapefile called region.shp which is the study area
- A shapefile called grizone.shp which are the grizzly bear recovery zones
- An excel file called griznum.xlsx which contains a count of the number of bears observed in each recovery zone
- An image called Grizzly.jpeg (*note: this is an image, not a spatial file so you can't open it in the normal way in GIS*)

**Questions to be answered (10 pts):**

1) What are the names of the five grizzly bear recovery zones? (1 pt)
2) What is the total area of all recovery zones (*in square km*)? Briefly explain the steps you used to determine this information. (3 pts, 1 for answer, 2 for explanation)
3) What is the observed count of bears for each recovery zone? Briefly explain the steps you used to determine this information (3 pts, 1 for answer, 2 for explanation)
4) What is the density of bears in each recovery zone (*in bears per km$^2$*)? Briefly explain the steps you used to determine this information. (3 pts, 1 for answer, 2 for explanation)

**Map requirements: (10 pts – 1 pt each)**

1. Include two Data Frames in the Layout

   *Data Frame #1:*

   a. Includes the regional states for context
   b. Each of the grizzly bear recovery zones is symbolized with a unique color
   c. A legend is included showing the name of the recovery regions related to each color (Remove any unnecessary words or symbols from the legend)
   d. The map contains a scale bar (in km) and north arrow

   *Data Frame #2:*

   a. Includes the regional states for context
   b. Each of the grizzly bear recovery zones is symbolized by dot density relative to bear density
   c. A legend is included telling us dots/unit area
   d. The map contains a scale bar (in km) and north arrow
2. Include a labeled image of a grizzly bear (Insert → Picture)

Export your map as a jpeg and insert the jpeg into a Word doc with the answers to questions 1-4.

# Reading: Basic Geospatial Analysis Techniques

## INTRODUCTION

### LESSON OBJECTIVES

By the end of this lesson, you will be able to:

1. Identify the basic concepts of spatial analysis.

**LEARNING SEQUENCE**

| | |
|---|---|
| Required Reading | Read the following:<br><br>Basic Geospatial Analysis Techniques<br><br>• Data Analysis: Buffering and Dissolving<br>• Geospatial Data Analysis: Overlay Operations<br>• Geospatial Data Analysis: Table Operations |
| Resources | View the following:<br><br>1. Automatic Spatial Selection |

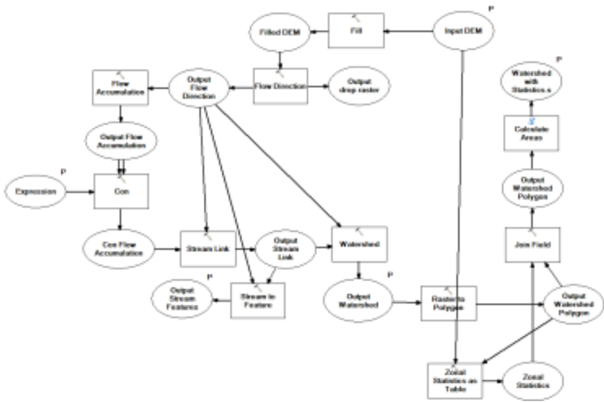
## INSTRUCTION

# Basic Geospatial Analysis Techniques

### What is Geospatial Data Analysis?

Geospatial data analysis is the application of operations to manipulate or calculate coordinates and/or related attribute data. Geospatial data analysis is applied to solve problems such as bus routes, determining flood zones, and in determining suitable sites for construction. Spatial data analysis

(SDA) uses spatial operations to manipulate or calculate. This is in contrast to analysis that does not use spatial operations to manipulate or calculate. As a note, SDA will be used to refer to spatial data analysis for the remainder of this lecture.




As an example of a SDA, this is an analysis flowchart for delineating watersheds using multiple input data layers, to produce an output of watersheds. While this model may look complex, you would be surprised at how each portion of the model is really not that difficult at all. The true power of SDA is the culmination of all of these different analysis methods.

## Important Considerations: SDA

For each SDA, you must consider the following:

- Input
- Operations
- Output

SDA typically involves using data from one or more layers to create an output. Therefore relationships can exist between the number of inputs, to the number of outputs: 1 to 1, one to many, many to one, and many to many. Speaking of the output it is important to note that the output may not always be spatial in nature. For instance, the output of an SDA may be a statistic, or report which is not spatial in nature.

**Key Facts**

SDA

- SDA typically involves using data from 1 or more layers to create output.
  - One to One
  - One to Many
  - Many to One
  - Many to Many
- Output may not always be spatial in nature. The output can be a:

- Statistic
- Report

## Basic Geospatial Analysis Techniques

There are seven basic geospatial analysis techniques that will be examined in this presentation.

- Selection
- Buffering
- Dissolve
- Overlay Operations
- Classification
- Table Operations
- Geocoding

## Selection

Selection is a basic type of geospatial data analysis that will probably be your most used operation when analyzing data within a GIS.

A selection operation is an operation that identifies features that meet certain conditions. The selection conditions may be spatial or aspatial. An aspatial selection condition selects objects based on attributes. A spatial selection condition selects objects based on position.

Whether you are performing an aspatial, or spatial selection, there are two types of selection methods available. The first is on-screen/manual selection, which is where you manually select the data with an input device. The second method is automatic where the user sets a condition, or writes a script, and lets the computer select the data that meets the conditions.

**Key Facts**

Selection

- Selection Operations identify features that meet conditions.
- Selection conditions may be spatial or aspatial
    - Select based on attributes – *aspatial*
    - Select based on position - *spatial*

### Aspatial Selection: Table Query

A table query is a very common GIS aspatial operation. A table query selects a subset of records based on values of specific attributes that meet certain criteria. Before you can do your first table query, however, you need to learn set algebra, and Boolean operators, as they are the building blocks of the table queries.

### Set Algebra

Let's start with set algebra. Set algebra uses operations to determine whether two values are equivalent or not. The four basic set algebra operations are less than, greater than, equal to, and not equal to.

#### Less than <

The less than operation checks to see whether the value on the left is less than the value on the right. The less than operation is represented by the left angle bracket symbol.

#### Greater than >

The greater than operation checks to see if the value on the left is greater than the value on the right. The greater than operation is represented by the right angle bracket symbol.

#### Equal to =

The equal to operation checks to see whether the values on both sides are equal to each other. The equal to operation uses the =.

#### Not Equal to < >

The not equal to operation checks to see if the values on both sides are different from each other, and is equivalent to the combination of less than and greater than. The not equal to operation is represented by both the left angle bracket and right angle bracket used together.

### Important Note

Less than <
Greater than >
Equal to =

The three operations listed above annotated can be applied alone, or in combination. So for instance, you can perform the test of whether the value on the left side is less than or equal to the value on the right side. The symbol that would represent this operation, would be both the left angle bracket followed by the =. The result of all of these operations is either the value of "true", or "false". As a quick challenge question, I am making the claim that greater than and less than may not be applied to nominal attributes. Why do you think this is the case? The answer is because nominal attributes are only descriptors, and it is illogical to compare them with respect to magnitude or rank. It is, however, logical to compare them using the equal to, or not equal to operators.

**Table Queries using Simple Set Algebra**

Provided are some examples of table queries using simple set algebra operations. Within the examples, we have the data of the lower 48 states of the United States of America. For each of the four maps, we will perform a simple set algebra operation to see which states are selected. Selected states are outlined in blue.

"Male Population">"Female Population" Map

The table query is used to see whether in each state, male population is greater than female population. The way this works behind the scenes, is that the table query goes row by row in the table, and compares the value on the left, which is the male populations value, the value on the right, which is the female population, and, if the male population is greater than the female population, then the result is true, and that row, and object, are selected. Running that query on the map selects five states.

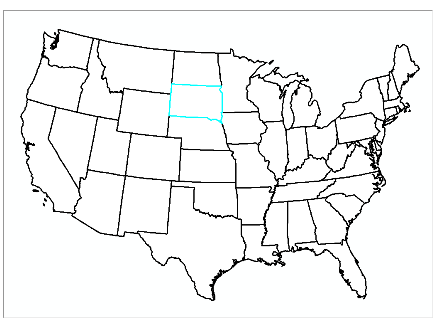
"Population" > 7,000,000 Map

This table query is comparing the population attribute with the number 7 million. So the states that are chosen are states where the population attribute is greater than the number 7 million.

“Region” <> ‘Mountain’ Map

This map selects all regions that are equal to “Mountain”. So again, it goes row by row in the table, and compares the region attribute to determine whether it is designated as a “Mountain” region. If the region attribute is equal to mountain, it is not selected. If the region attribute is not equal to “Mountain”, it is selected.

“State Name” = ‘South Dakota’ Map



This table query is state name equal South Dakota. If there is only one state with the name of South Dakota, only that state is selected from the table.

## Boolean Algebra

Now that we are familiar with set algebra; let’s move on to Boolean algebra. Boolean algebra is going to allow us to combine multiple set algebra operations into a more complex table query.

Boolean algebra has multiple conditional operators. The three operators we are going to focus on are the most common Boolean operators. They are: OR, AND, and NOT. The Boolean operators evaluate values on the left and right side of the operator, and then assign an outcome. The outcome is a

Boolean, or binary result such as true/false, 0/1, on/off, or any other dichotomy.

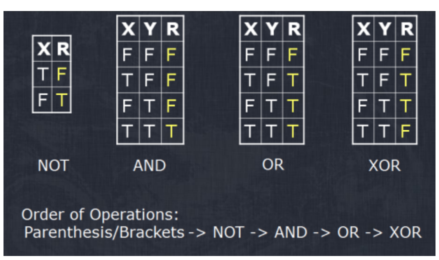
In Boolean algebra, the order of operations does matter; therefore it is not uncommon to use many sets of parentheses to force a particular order of operations. It is also important to note that Boolean operators are not distributable inside of parentheses.

**Key Facts**

**Boolean Algebra**

- Uses conditions *OR, AND,* and *NOT*
- Evaluated by assigning an outcome, true or false, to each condition
- Order of operations **do matter**
- Boolean operators are not distributable

## Boolean Operators




These are the truth tables for the three Boolean operators. A truth table allows us to see what the result of the operation, 'R' is in the table; we can apply the input of X or X and Y. Review each of the following Boolean Operators.

### Boolean Operator: NOT

Look at the truth table for the simplest Boolean operation: not. The "NOT Boolean operator" simply reverses the result of the input. So, if the input is true, then NOT true is false. If the input is false, then the result is true. The "AND Boolean operator", requires two inputs. One input is placed to the left of the AND Boolean operator, and one input is placed to the right of the AND Boolean operator.

### Boolean Operator: AND

The “AND Boolean operator” only assigns the value of true when the value to its left evaluates to true and the value to the right also, simultaneously, evaluates to true. In other words, in order for the “and Boolean operator” to be assigned a true value, both sides of the operation must have evaluated to true. Looking at the truth table, if the inputs are false AND false, the result is false. If the input is true AND false, the result is false. If the input is false AND true, the result is false. If the input is true AND true, the result is true. This makes the “AND operator” extremely selective.

### Boolean Operator: OR

Finally, let’s look at the “OR truth table”. With the “OR Boolean operator”, it only requires that one of the two inputs are true for the “OR operator” to return a true value. For instance, if the inputs are false OR false, the result is false, since neither of the inputs return be true. If the input is true or false, and the result is true if the input is false OR true, the result is true. If the input is true OR true, then the result is true. The order of operations of these Boolean operators are first, the innermost set of parentheses or brackets are evaluated first, followed by the Boolean operator, followed by the “AND Boolean operator”, followed by the “OR Boolean operator”.

## Practice Statements

Take a few moments, and consider each one of these statements. Determine whether each statement returns a value of true or false. To save space, false is represented by the letter F, and true is represented by the letter T.

**1. NOT (F)**

**2. T AND T**

**3. F OR NOT (T)**

**4. T OR (F OR (T AND T)**

**5. (T AND T) OR F AND (NOT (F) AND T)**

## Practice Statements: Answers

1. NOT (F) → True

The first one states NOT false. “NOT” simply reverses the truth of the statement, therefore, NOT false equal to true.

2. T AND T → True

The second statement is true AND true. If you remember, for the "AND Boolean operator", both sides must be true in order for the "AND operator" to return true, and it does in this case.

3. F OR NOT (T) → False

The next states false OR not true. For this one, we must remember our order of operations. Remember that not must be evaluated before the OR. Therefore, NOT true which leaves us with the statement of false OR false. The "OR operator" only returns true if one of the two sides returns true. Since a statement is false OR false, the result of the "OR operator" is false.

4. T OR (F OR (T AND T) → True

The next statement true OR open parenthesis false OR open parenthesis true AND close parenthesis followed by another close parenthesis. Remembering the order of operations, the innermost set of parentheses is evaluated first, which is the true AND true statement. True AND true evaluates to true which takes us to the next level of parentheses which is now false OR true. False OR true evaluates to true, because only one of the inputs needs to be true for the "OR operator" to return true. That leaves true OR true which evaluates to true.

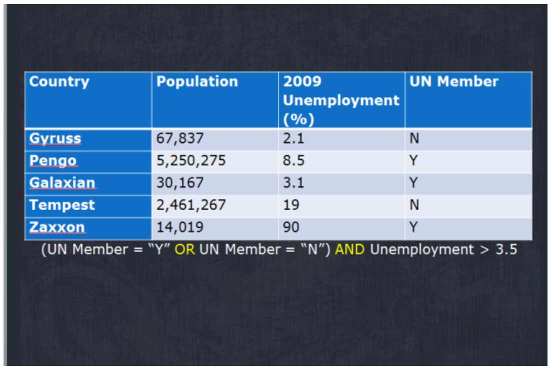
5. (T AND T) OR F AND (NOT (F) AND T) → True

The last statement reads open parenthesis true AND true close parenthesis OR false and open parenthesis not false AND true close parenthesis. Again, we need to go to the innermost set of parentheses, and in this case, we have two options. Let's start at the left and evaluate the true AND true. True AND true evaluates to true. Let's go to the other set of parentheses. Order of operations states that we must evaluate the NOT before we can evaluate AND. So, NOT false equals to true which leaves us with true AND true which evaluates to true. The statement now reads in total true OR false AND true. So, once again, order of operations states that we must evaluate the "AND" first so false AND true is equal to false. That now leaves the statement true OR false. True OR false equals true is the result of the statement.

## Combining Set and Boolean Operators

Now that you have been introduced to set and Boolean operators, let's combine them together. If you remember, the result of a set operation is a true OR false value, and Boolean operators evaluate the inputs that are been evaluated to either true OR false. Take a look at the following charts to see how it all fits together.

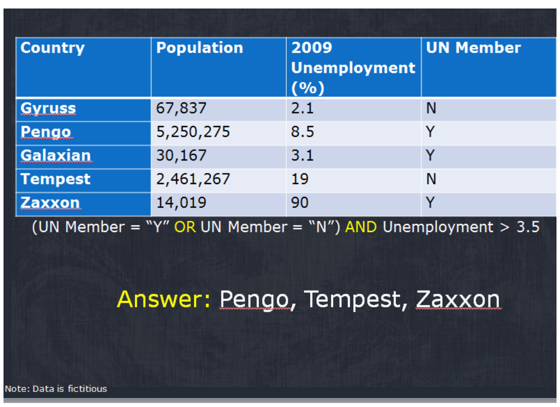
## Chart 1



| Country | Population | 2009 Unemployment (%) | UN Member |
|---|---|---|---|
| Gyruss | 67,837 | 2.1 | N |
| Pengo | 5,250,275 | 8.5 | Y |
| Galaxian | 30,167 | 3.1 | Y |
| Tempest | 2,461,267 | 19 | N |
| Zaxxon | 14,019 | 90 | Y |



Focusing on the chart, it is an attribute table that has four attributes of country, population, 2009 unemployment, and whether the country is a United Nations member. Below the table, we have the table query we wish to execute on the table. The table query reads open parenthesis UN member equals Y OR UN member equal N close parenthesis AND unemployment is greater than 3.5. Spend a few moments and go row by row through the attribute table to see if you can determine which countries are selected based on this query.

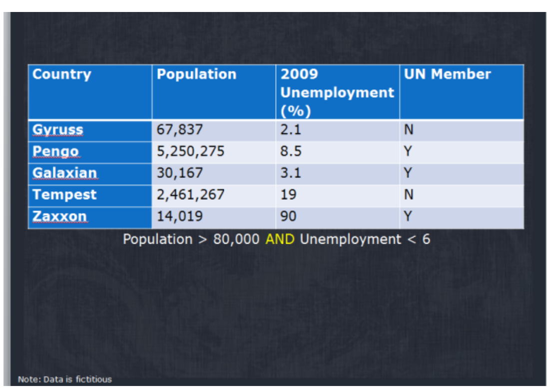


| Country | Population | 2009 Unemployment (%) | UN Member |
|---|---|---|---|
| Gyruss | 67,837 | 2.1 | N |
| Pengo | 5,250,275 | 8.5 | Y |
| Galaxian | 30,167 | 3.1 | Y |
| Tempest | 2,461,267 | 19 | N |
| Zaxxon | 14,019 | 90 | Y |



The query selects the countries of Pengo, Tempest, and Zaxxon.

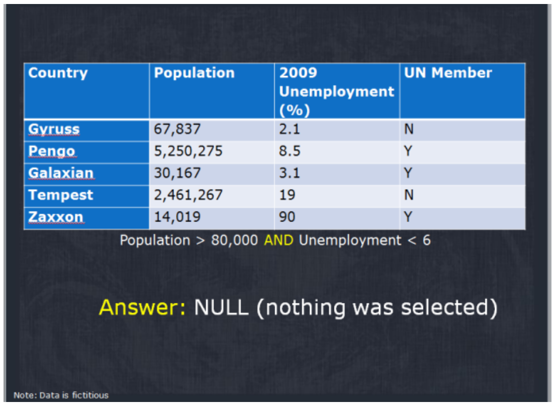
## Chart 2



| Country | Population | 2009 Unemployment (%) | UN Member |
|---|---|---|---|
| Gyruss | 67,837 | 2.1 | N |
| Pengo | 5,250,275 | 8.5 | Y |
| Galaxian | 30,167 | 3.1 | Y |
| Tempest | 2,461,267 | 19 | N |
| Zaxxon | 14,019 | 90 | Y |

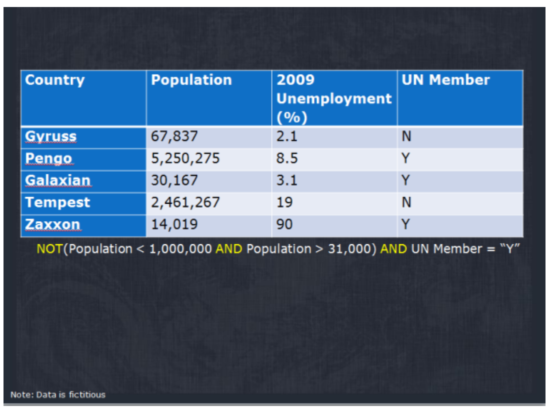


Now consider the table query of population greater than 80,000 and unemployment less than six. Spend a few moments to see if you can determine which rows were selected.

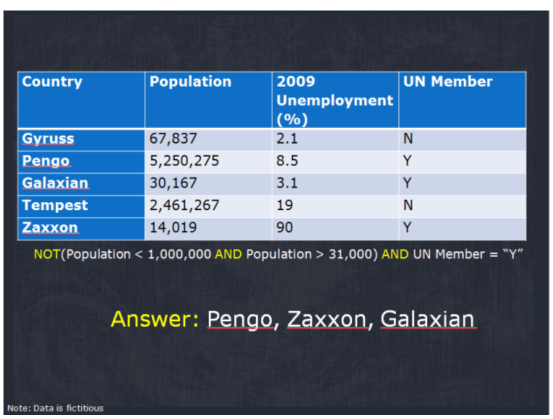



The answer in this case is NULL. The result of NULL means that no rows were selected using that query.

### Chart 3




Let's consider one last table query. The table query reads NOT open parenthesis population less than 1 million AND population greater than 31,000 close parenthesis AND UN member equals Y. Take a few moments to determine which rows are selected by this table query.

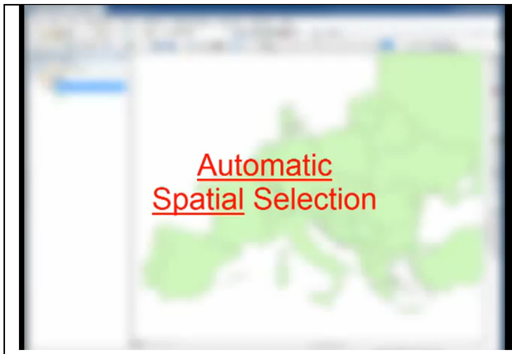



This table query selects the countries of Pengon, Zaxxon, and Galaxian.

## Spatial Selections

When a computer performs an automatic spatial selection, there are multiple requirements that you can set. As an example it is important to introduce the spatial idea of adjacency, intersect, containment, and distance.

In a spatial selection operation, adjacency identifies features that share a boundary segment with the input feature. The intersect operation identifies features that simply touch, or even overlap the input feature. Containment identifies features that are completely within, or completely contain the input feature. Distance identifies features that are within a set distance of an input feature.

<table>
<tr>
<td>
</td>
<td>Watch the video, Automatic Spatial Selection, https://www.youtube.com/watch?v=P_W734x7G8M (0:31). This video demonstrates both a manual and automatic spatial selection. The first selection made is a manual selection, where the user uses the mouse to select a country on screen. Next, the user performs an automatic spatial selection, by asking the computer to select all countries that touch the selected country. The computer then selects all countries that touch the boundary of the manually selected country.</td>
</tr>
</table>

**Key Facts**

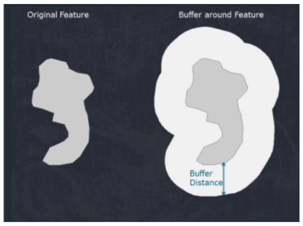
Spatial Selection Operations

- Adjacency identifies features that share a boundary segment.
- Intersect identifies features that "touch".
- Containment identifies features that are "contained by" or "contain" other features.
- Distance identifies features that are "within a set distance" of a feature.

# Data Analysis: Buffering and Dissolving

### What is a Buffer?

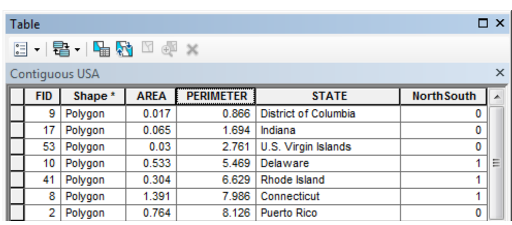
A buffer is a region that is less than or equal to a distance from one or more input features. Buffers can be created from point/line/polygon/raster geospatial data sets. Buffers are typically used to identify areas or objects "inside", or "outside" the threshold distance. Can you think of any examples where you might use a buffer? One example of the use of a buffer could be to determine how many homes are within 1 mile of the coastline. In this case, the coastline is the input; 1 mile is the threshold distance, which results, in a polygon that extends 1 mile out from the coastline. We can then perform spatial selection by selecting all houses that are inside the buffer.




To illustrate a buffer, the polygon to the left includes the original input features. After specifying a buffer distance, and running the buffer tool, a new polygon is created at the set buffer distance and surrounds and follows the outline of the original input feature.

## Dissolve Operation

Now let's discuss the dissolve operation. The dissolve operation combines similar features within a data layer based on a shared attribute. For example, the data set on the left has the attribute of north-south. If the state is a northern state, it has the value of one, and if the state is a southern state, it has a value of zero. If we use this data set, and dissolve on this field, all states that have the value of one, are dissolved into a single polygon. All states that have the value of zero for the north-south attribute are dissolved into a second single polygon. The dissolve field can have as many different values as you choose; however, only records with identical values for that attribute will be combined into a single feature.

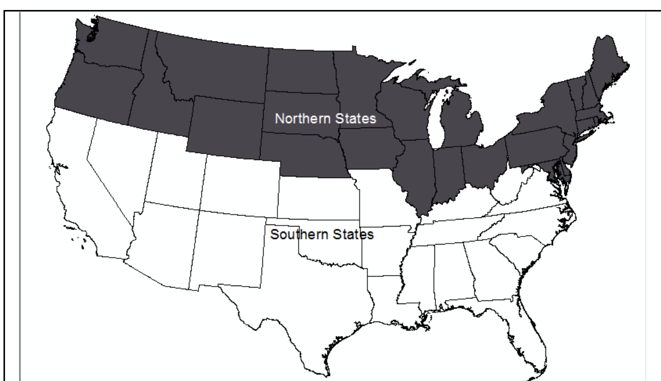


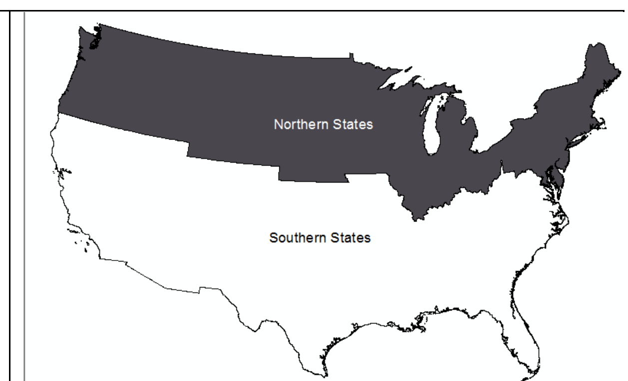
Table

Contiguous USA

| FID | Shape * | AREA | PERIMETER | STATE | NorthSouth |
|---|---|---|---|---|---|
| 9 | Polygon | 0.017 | 0.866 | District of Columbia | 0 |
| 17 | Polygon | 0.065 | 1.694 | Indiana | 0 |
| 53 | Polygon | 0.03 | 2.761 | U.S. Virgin Islands | 0 |
| 10 | Polygon | 0.533 | 5.469 | Delaware | 1 |
| 41 | Polygon | 0.304 | 6.629 | Rhode Island | 1 |
| 8 | Polygon | 1.391 | 7.986 | Connecticut | 1 |
| 2 | Polygon | 0.764 | 8.126 | Puerto Rico | 0 |

Dissolve combines similar features within a data layer based on an attribute.



# Geospatial Data Analysis: Overlay Operations

## Overlay Operations

Overlay operations involve combining spatial and attribute data from two or more spatial data layers. In other words, we are stacking the data, and looking for where the layers overlap. While the overlay operations may seem simple when used alone, they can actually be quite powerful operations when combined in a series. An example of an overlay operation would be the clip operation. For example, let's say I have a data set that contains all the streets in the United States of America. However, I only want to have the streets in Arkansas. I can use the clip operation to cut out all the streets that are only in Arkansas by providing the state outline of Arkansas as the clipping feature. One thing to note about overlay operations, are that overlays require that the data be in the same coordinate system. This is so that the software can perform the highest accuracy of overlay. Some software packages automatically re-project on-the-fly if the data is not in the same coordinate system, but this is not the preferred method.

**Key Facts**

Overlay Operations

- Overlay Operations involve combining spatial and attribute data from two or more spatial data layers. "Stacking data"
- Overlays require that data be in the same coordinate system.

## Vector Overlays

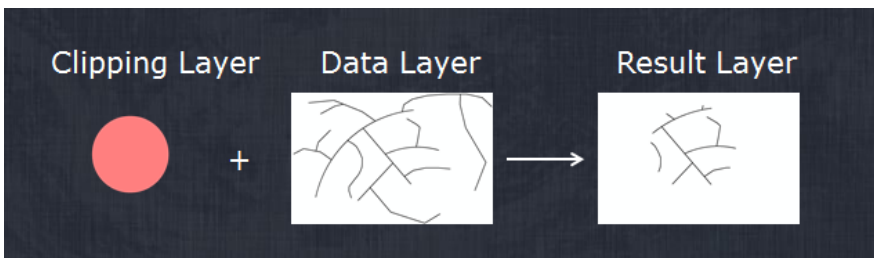
A vector overlay involves combining point, line, or polygon geometry and their associated attributes. All overlay operations create new geometry and a new output geospatial data set. You should be cautioned that with certain overlay operations, very large attribute tables may result if the overlay operations combine many layers, and each layer has a very large attribute table. Additionally, it might be possible that the combined attribute tables would cause duplicate attribute fields to exist. In these cases, you should consider reducing the number of transferred attributes to the minimum required, and rename duplicate fields so that there is no ambiguity.

## Basic Cases of Overlay

There are three basic cases of overlays: the clip function, intersection operation and the union overlay operation. Review each of the basic cases of overlay and the examples provided.

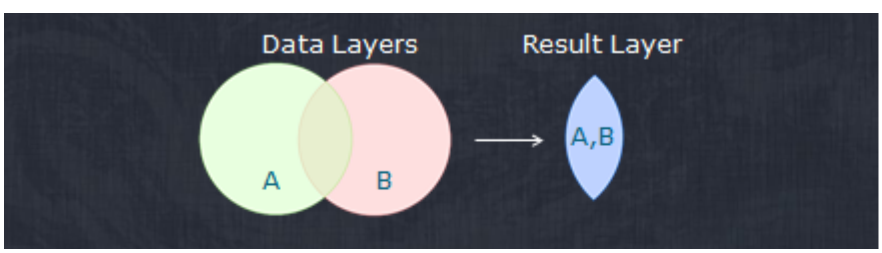
### Clip Function

The first basic case of overlays is the clip function. The clip function defines the area for which features will be output based on a "clipping" polygon. I like to think of it in terms of cookie dough, and a cookie-cutter. The cookie dough is the data layer that we want to reduce in size based on our cookie-cutter. The clipping layer is the cookie-cutter that will determine how we cut out the part of the cookie dough we are interested in. It is important to note that in the clip operation; only the geometry and attributes of the data layer are transferred to the result layer. Neither the geometry nor attributes of the clipping layer are included in the output, just as you would not make your cookies in the oven with the cookie-cutter still surrounding the cookie dough.



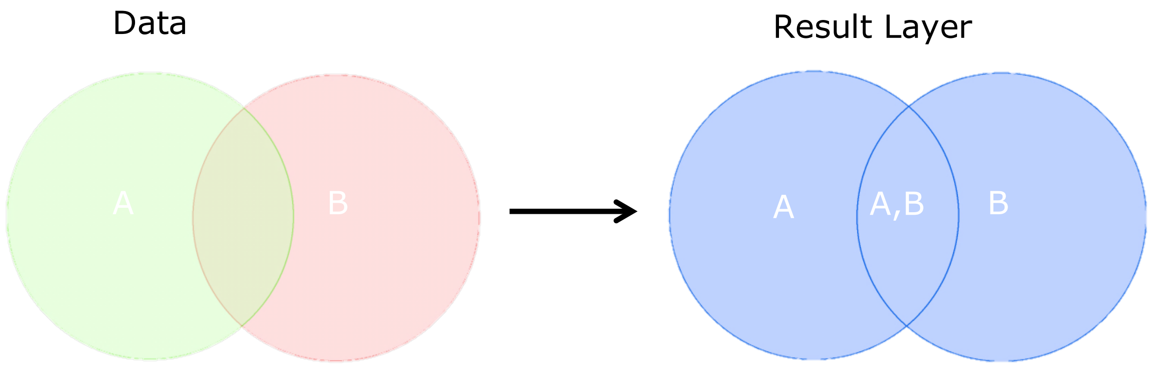

### Intersection Operation

The next basic case of overlay is the intersection operation. The intersection operation combines data from two or more input layers, but the combination only exists for the regions where all input layers contain data. This is similar to the clip function; however the attributes and geometry of all input data layers are transferred to the result layer.




### Union Overlay Operation

The third and final basic case of overlay is the union overlay operation. The union operation is an overlay that includes all data from all of the input data layers. Since the union operation combines all attributes, and all geometry, no geographic data is discarded in the union operation, and all corresponding attribute data is saved for all regions.



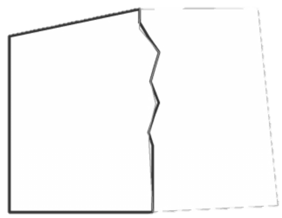

### Vector Overlay Problems

When performing vector overlays, there are a few problems that you should be aware of. In the case where the input features represent common features that are represented in both layers, they may have slightly different geometry. This creates sliver polygons when the overlay operations are performed. For instance, imagine we have one input data set the represents the national boundary of Canada created at a 1 to 5,000,000 scale. And we have a second input data set of the national boundary of the United States of America created at 12 1 million scale. If we stack these two layers on top of each other, it is probable that the boundary lines will not overlay exactly, as they were created at different scales, and therefore the Canadian boundary line has probably been simplified more than the United States boundary line. The areas where the two input data sets do not match, or result in very small, typically insignificant sliver polygons that need to be dealt with. There are several methods that exist to reduce the occurrence of sliver polygons, such as manually snapping the boundaries together, or introducing fuzziness to the edges of the overlay operation, or simply manually deleting the sliver polygons.

#### Example 1

To illustrate the idea of a sliver polygon, here are two different data sets, one outlined in a bold black line, and the other outlined in a thinner grey line. Where the two data sets overlay, they do not match exactly for whatever reason. If we perform an intersect, clip, or union operation on these two input data sets, we are going to create three sliver polygons.

### Example 2



The sliver polygons created between these two input data sets are highlighted in red. The sliver polygons are probably insignificant, and are the results of data sets being produced by different agencies, at different times, and probably at different scales. It is up to you to determine how you would like to resolve for the sliver polygons.

**Key Facts**

Vector Overlay Problems

- Input features
  - Common features represented in both layers, may have slightly different geometry
- Having slightly different geometry creates:
  - Sliver polygons when the overlay operations are performed
- Several methods exist to reduce the occurrence of sliver polygons:
  - Manually snapping boundaries together
  - Introducing fuzziness to edges of overlay
  - Manually deleting the sliver polygons

# Geospatial Data Analysis: Table Operations

### Table Operations

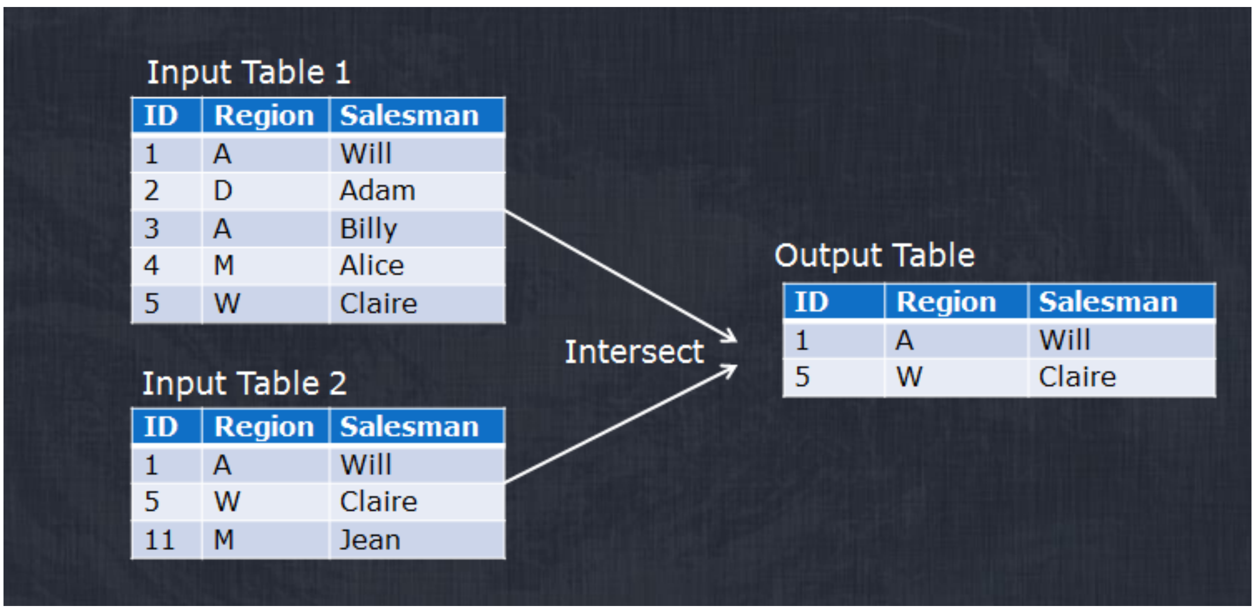
It is important to understand the concept of table operations and how they apply to attribute tables in a geospatial data set. Combining tables of attributes is a powerful analysis technique. There are three common types of table operations: intersect, union, and join.

Each one of these operations can be combined in tables in different ways, and for different purposes. It is important to note, that in order for multiple tables to be inputs to intersect, or a union table operation, the tables must have the same fields, and field types, which are commonly referred to as the schema. Let's take a look at each one of these three table operations in more detail.

### Intersect

The intersect table operation produces an output table that only contains records that all of the input tables had in common. That means that every attribute for the entire row was identical. As an example, we have two input tables that have the same schema, as all input tables must have the same schema for the intersect operation. The output table only includes the two records that were found in both input tables. In this case, the row where the ID was one, region was A, and salesman was Will, in the second row was where the ID was five, the region was W, and the salesman was Claire.

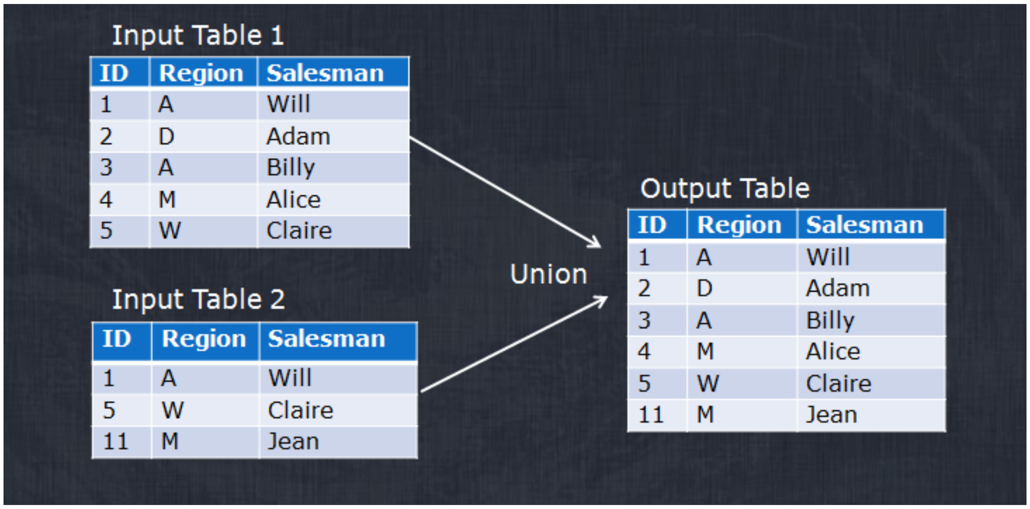
- An output table only contains records that all input tables have in common.




### Union

The union table operation produces an output table that contains all records that existed in all input tables. For example, we have the same two tables as before, and notice that the output table combines all the rows from both input tables. Again, note that both input tables have the same schema, as this is required for union.

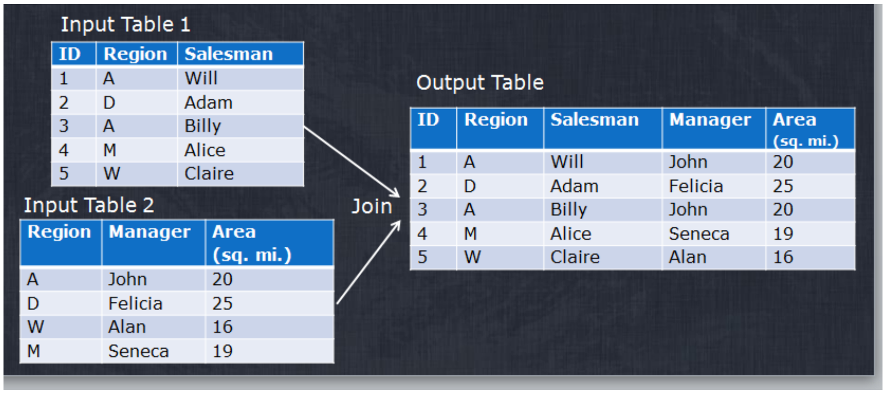
- An output table only contains all records that exist in all input tables




### Join

The third table operation is the join operation. The join operation brings together two tables based on a shared attribute. Based on the shared attribute, attributes of one input table is appended to the attributes of the other input table. For example, we have two input tables that share the common attribute of region. Based on the region attribute, the manager and area attributes of Input Table 2, are appended to the rows and input table 1. The region attribute, matches its value between the two tables, which determines which attributes of which row from Input Table 2 is copied over. For example, if we look at the output table, the row with ID 1 had a region A, which means that the manager John and area of 20 was appended. In row where the ID was 2, since the region was value of D, then the row with the same region of value D in the second input table is copied to the output table. Also note, that for the record were ID was 3, the region is A again, so once again, the manager John, and area 20 are copied to the output table for that row.

- Based on a shared attribute, attributes of one input table are appended to attributes of the other input table.



Input Table 1

| ID | Region | Salesman |
|---|---|---|
| 1 | A | Will |
| 2 | D | Adam |
| 3 | A | Billy |
| 4 | M | Alice |
| 5 | W | Claire |

Input Table 2

| Region | Manager | Area (sq. mi.) |
|---|---|---|
| A | John | 20 |
| D | Felicia | 25 |
| W | Alan | 16 |
| M | Seneca | 19 |

Join

Output Table

| ID | Region | Salesman | Manager | Area (sq. mi.) |
|---|---|---|---|---|
| 1 | A | Will | John | 20 |
| 2 | D | Adam | Felicia | 25 |
| 3 | A | Billy | John | 20 |
| 4 | M | Alice | Seneca | 19 |
| 5 | W | Claire | Alan | 16 |

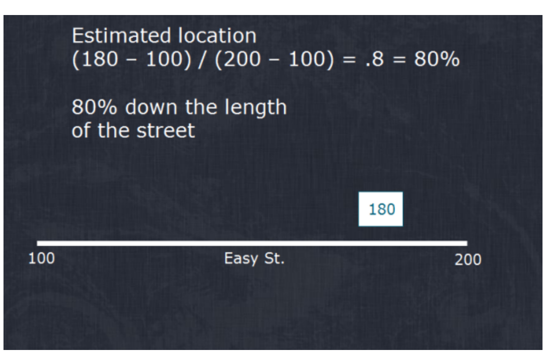
# Geospatial Data Analysis: Geocoding

## What is Geocoding?

The process of Geocoding involves spatially referencing features by address in order to determine their position on the face of the Earth. For example, the GPS unit in your car performs geocoding every time you enter a street address. The GPS unit takes the street address, and associates it with road data sets, to determine the coordinate that that addresses exists at. There are two typical types of geocoding: linear geocoding, and area geocoding. Linear geocoding assumes that addresses vary linearly along a feature which is typically a line. Area geocoding assigns the geocoded location to the entire area that matches the input address. Let’s look at an example of linear geocoding and area geocoding.

### Linear Geocoding

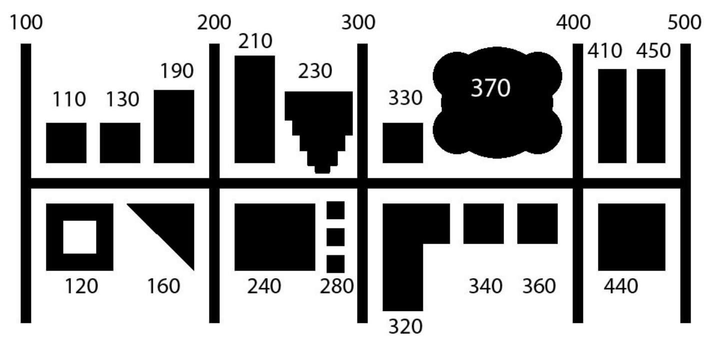
As an example of linear geocoding, imagine a geospatial data set that has the centerlines of roads. Between intersections, each road segment has a beginning address, and an ending address. If we geocode using that data set, since we do not know exactly where each address is on the block, we assume that the addresses are spaced evenly between the start and end range of addresses in the block, and therefore estimate where the addresses exists.




So for the example on the slide, this block of Easy Street starts at the address of 100, and ends at the address of 200. We wish to know where the address 180 Easy Street is located, we first identify the feature named Easy Street, then narrow down the results to the block where our input address exists. Next, we estimate how far down Easy Street that address is assuming equal spacing of addresses. The major downside to linear geocoding, is at the location may be misplaced if the addresses are linearly spaced. The major upside of linear geocoding is that the data sets are commonly, and often freely available and they are easy to construct.

### Area Geocoding

Now let's look at an example of area geocoding. This illustration shows four street segments, from addresses 100 to 500, and each block has the address range of 100. Here, the building footprints have been stored in a polygon geospatial data set, and each polygon has the address of that building. As each building has a slightly different shape and size, you can easily see that the addresses are not evenly distributed along the blocks. With area geocoding, we do not need to worry about whether the addresses are evenly spaced, as we can point to the exact area that the address exists. This is a much more accurate way to geocode than linear geocoding. The major downside to area geocoding is that the data sets are not commonly available, and are very expensive to construct.




**Acknowledgement**



The development of this document is funded by the United States Department of Labor in partnership with the Department of Education under the Trade Adjustment Assistance Community college and Career Training Grant Program (TAACCCT). 

UNITED STATES
DEPARTMENT OF LABOR

**Lab 3:**

# Select by Location & Vector Data Analysis



Open the hazmat folder in Lab 3 and add all of the shapefiles:

- Hazmat
- Schools
- Towns
- Counties
- Blackstone_watershed
- Blackstone_streams

These spatial layers surround the Blackstone River watershed, which provides drinking water to Rhode Island.

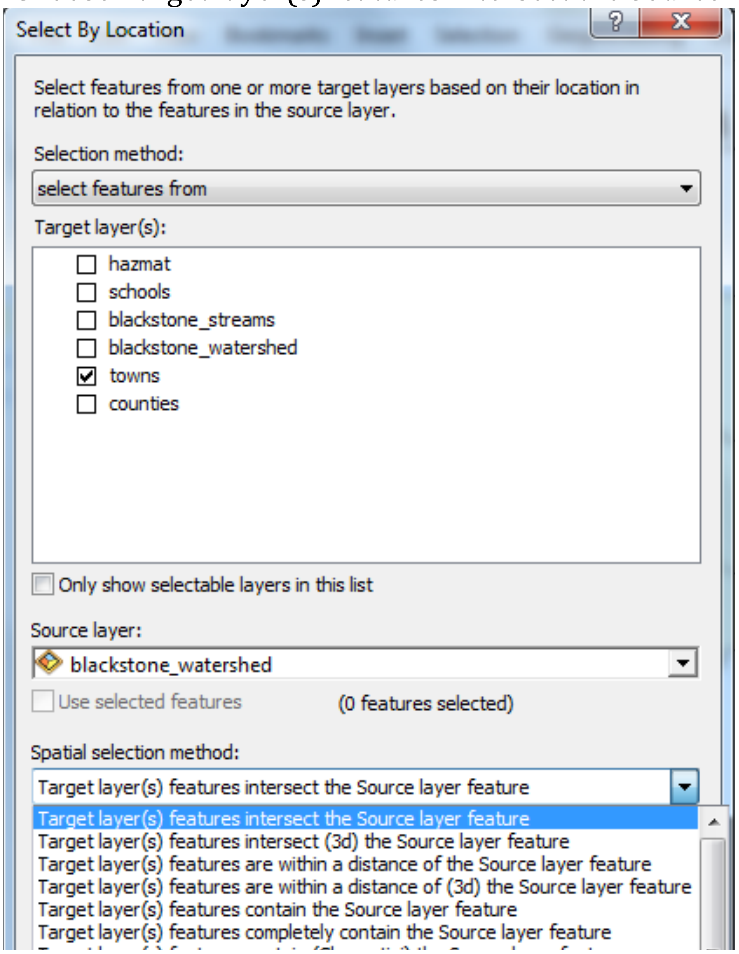
## Reminder exercise using concepts from last week

1. The hazmat folder also contains an excel spreadsheet with town population in 2000. Answer the following questions:
2. How many towns had a population of greater than 10,000 and an area of less than 50 $km^2$? (should be 11 – make sure you can do this)
3. Clear your selection.

## Select by Location

1. In addition to Select by Attributes, which you recall from last week, another option under the Selection menu is Select by Location. This tool allows you to select features from one shapefile that are within a specified distance of, or contained within another feature.
2. For example, let's say we want a count of all the Massachusetts towns that intersect the Blackstone River Watershed. Go to the Selection menu and choose Select by Location.
3. You want to select features from towns – this is the layer that will ultimately be selected.

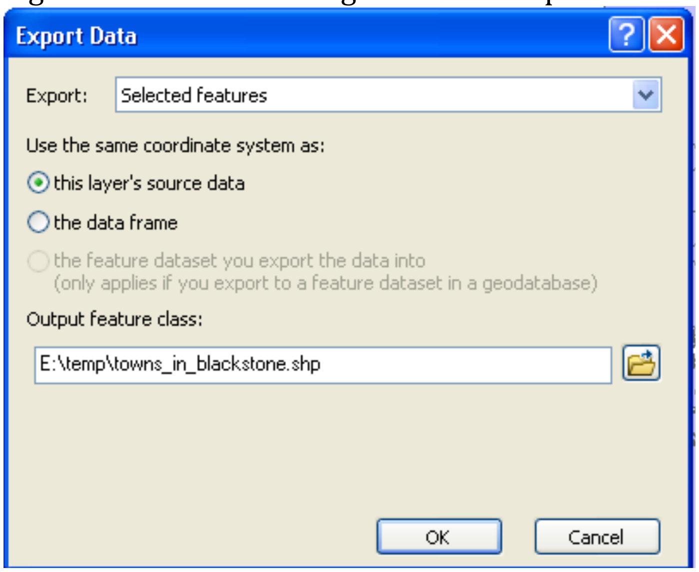
4. Choose Blackstone_watershed as your "source layer" – this is the layer that you'll be using to inform the selection.
5. Notice also that you have several options under Selection method – similar to Select by Attributes, you can create a new selection or add/remove from an existing selection. For example, by using select from an existing selection you could choose towns that intersect the Blackstone watershed AND contain a hazmat location.
6. Finally, notice that there are several spatial selection methods. The default is intersect. But, you can also select features within a specified distance of other features, or select features that only fall completely within other features. Check out the drop down menus to see your options.
7. Choose Target layer(s) features intersect the Source layer feature and click OK



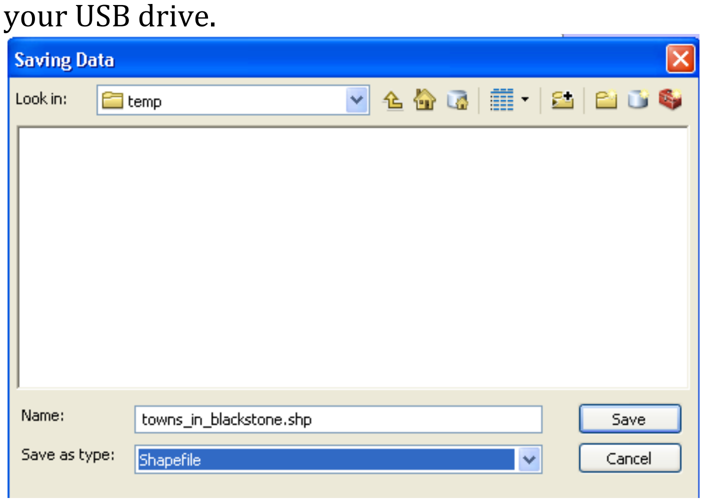

8. How many Massachusetts towns intersect the Blackstone watershed?
9. Using all of your selection skills – answer the following question: How many public schools (Type = 'PUB') are in the Blackstone watershed? (Answer should be 96 – make sure you know how to get that info)
10. How many hazmat sites are located within 100 m of one of the Blackstone streams? (Answer should be 33)

## Exporting a new shapefile

1. Let's say we're only interested in towns that intersect the Blackstone watershed. Why bother to keep all of these extra towns? They're just wasting space!
2. Use select by location to select all of the towns that intersect the Blackstone watershed (if they aren't already selected)
3. Right click on towns and go to Data > Export Data



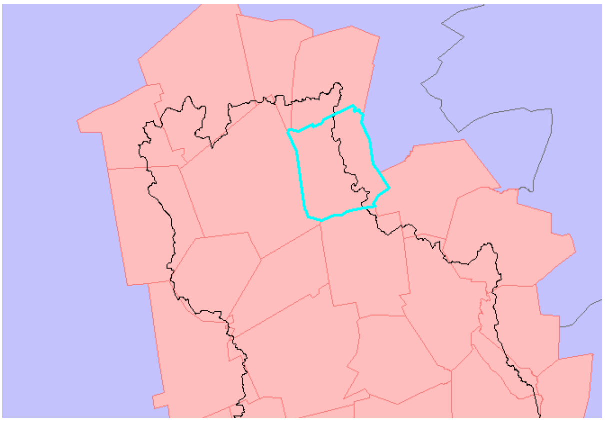

4. ***When you have data selected, ArcMap will default to perform the chosen operation on ONLY the selected data***. So, export will default to exporting the selected features. Field calculator and calculate geometry operations in the attribute tables will also default to the selected features (some of you noticed this last week).
5. Click the little folder button to save the output feature class to your Lab 3 folder on your USB drive.



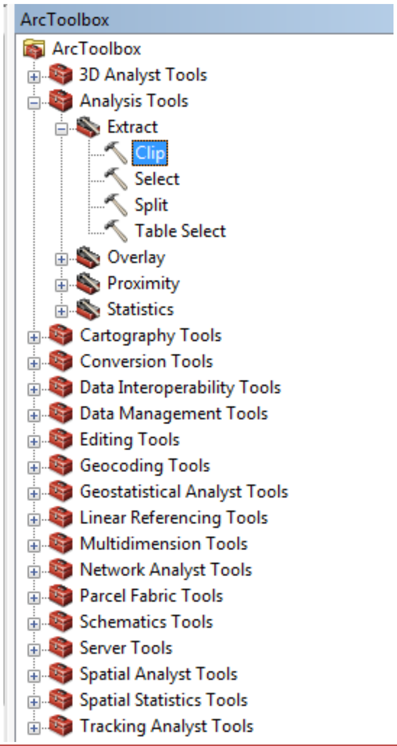

6. Click save and OK, and voila – a new shapefile that includes only the towns in the Blackstone watershed.

7. What is the area of all the towns that intersect the Blackstone watershed (should be 1381 $km^2$)

## Analysis tools - Clip

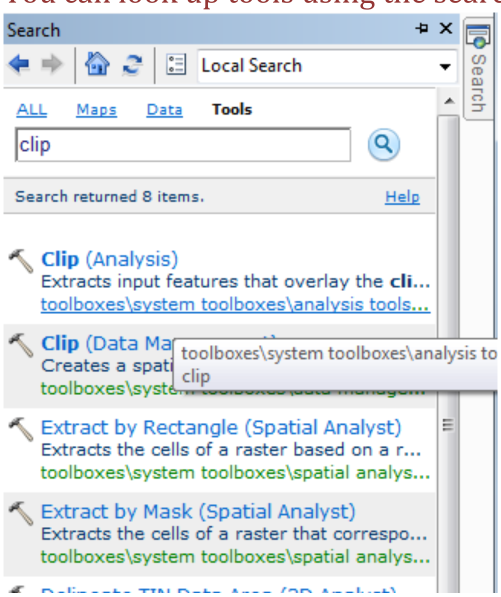
1. You just calculated the area of all the towns that intersect the Blackstone watershed. But, what if we don't want all the area that is outside of the watershed to be included? Take, for example, Shrewsbury. It is about half in the watershed and half out. How can we calculate the area of Shrewsbury that is inside the Blackstone watershed?

2. Now we're moving on to new and exciting ArcGIS frontiers. To do this calculation, you need to open ArcToolbox
3. Click the ArcToolbox icon to open ArcToolbox – you can click the top bar and drag it to dock it to one of the blue side arrows if it isn't already docked.

4. In ArcToolbox, you can expand the folders (toolboxes) to see all the different tools ArcGIS has to offer. They are numerous, and we will only use a fraction of them in this class.




## Note on finding tools in ArcToolbox

You do not have to remember where every tool is located in ArcToolbox. Unless you have a photographic memory, you will probably have to look them up each time. You can look up tools using the search tool.

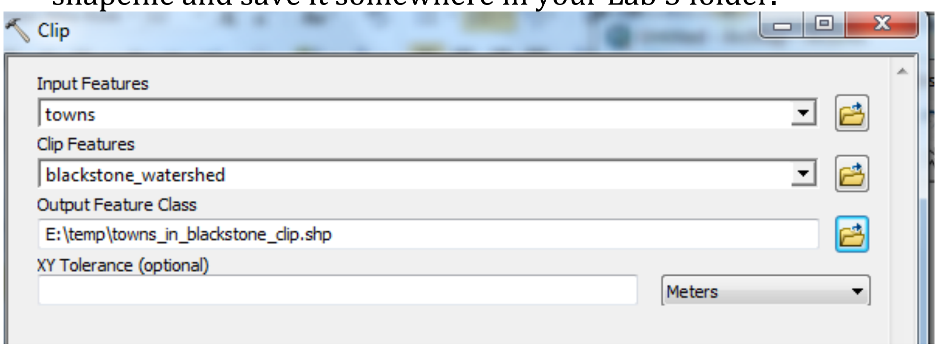



Clicking on the top link (Clip (Analysis)) will open the tool itself, clicking on the bottom link (toolboxes\...) will open the path in ArcToolbox.

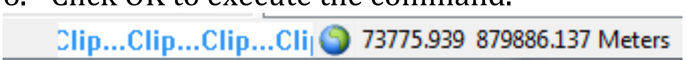
In order to search for tools in the Search menu, you must *already know what the tool is called*. If you don't know the name of the tool, I would recommend using Google rather than ArcMap help to try to find it.  Google has more flexibility in search terms – just make sure you include ArcMap or ArcGIS in addition to whatever you're searching for.  Or, better yet, ask one of us!

The more you use ArcGIS, the more familiar you'll become with the different tool names.  But, for now, consider keeping a notebook with some tool names that you can refer back to later.

5. Click to open the Clip tool from either the search window or ArcToolbox.  The 'input features' at the top is the layer you want to clip, the 'clip features' below it is the shape you want to use ('clip features' is like the cookie cutter).  Specify a new output shapefile and save it somewhere in your Lab 3 folder.




6. Click OK to execute the command.



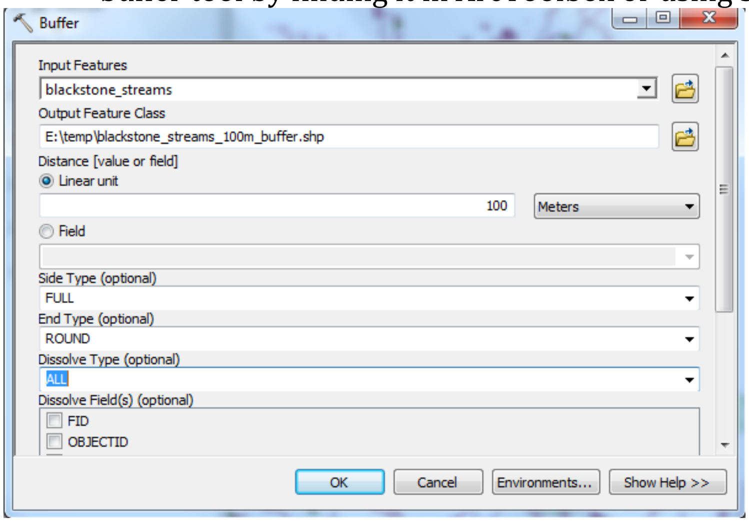

7. You'll know it's working if you see Clip in blue in the bottom right corner
8. Find Shrewsbury and see what it looks like now – it should only include the portion of the town that is inside the watershed.
9. What is the area of Shrewsbury inside of the Blackstone watershed?  (It should be 35.4 $km^2$)

**Warning: Calculate Area does not automatically update!!**

If you found the answers to the reminder question and #9 above (good job!), then you probably opened your attribute table to find the area of Shrewsbury and saw 56.27 in the Area column.  That number is the total area of all of Shrewsbury, not just the part in the Blackstone watershed.  Any time you clip a shapefile (or perform any other analysis tools), calculations like area or perimeter (for polygons) and length (for lines) will not automatically update.  They'll retain the original information in that field.  You will need to redo Calculate Geometry to find the correct area of Shrewsbury inside the Blackstone watershed.

## Analysis tools - Buffer

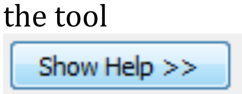
1. On to a new problem. Now, let's say that we want to find the area of land within 100 m of a Blackstone stream in the town of Shrewsbury. Blackstone streams is a line shapefile, so we'll need to create a new polygon before we can clip out the parts in Shrewsbury to calculate area. The tool to create a new polygon shapefile of area surrounding an existing point, line or polygon is called the buffer tool. Open the buffer tool by finding it in ArcToolbox or using search.



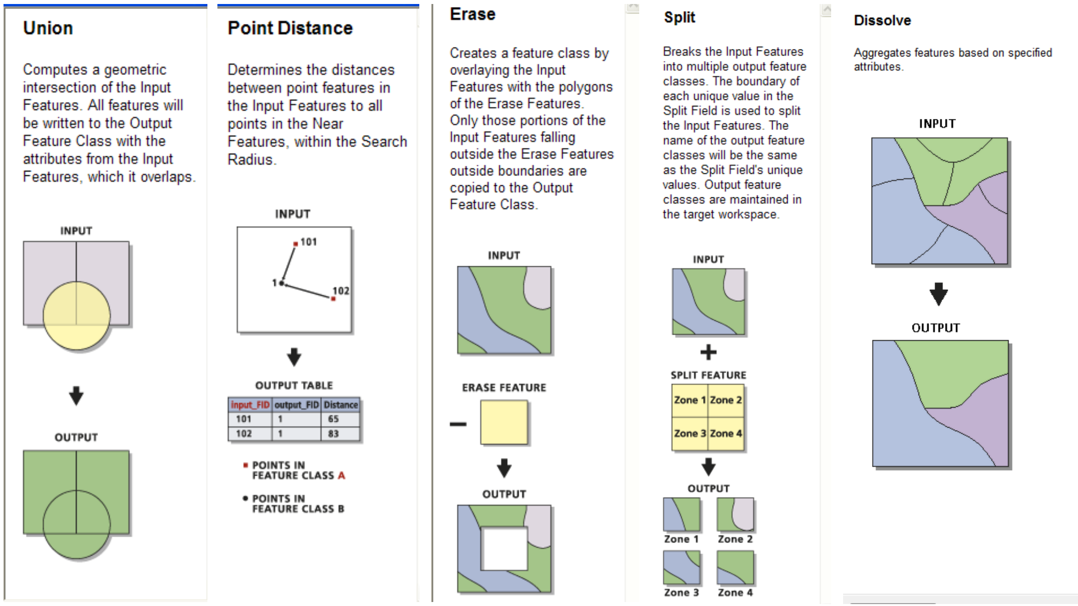

2. Input features should be blackstone_streams. Save a new output shapefile somewhere in your Lab 3 folder. In this case, we want the buffer to be 100 m. You should also change the dissolve type to ALL (the default is NONE).
3. With dissolve ALL, the output will be one single buffer feature surrounding all of the streams. If you use the default dissolve type NONE, you'll get a separate buffer around every individual stream feature. Buffers around each individual stream will include some overlapping areas, so the area calculation would be incorrect.
4. Click OK to execute the buffer
5. Now, clip your new stream buffer with the town of Shrewsbury (recall that ArcMap will default to perform calculations on the selected features if one or more are selected).
6. How much land area falls within 100 m of a Blackstone watershed stream in Shrewsbury? (Should be 14.4 $km^2$)

## Other Analysis Tools

1. There are lots of other analysis tools that you can explore in ArcToolbox. Below are a few examples. Chances are, if you can think of it, there's probably a tool to do it…it's just a matter of finding it!
2. You can read more about what tools do by clicking 'show help' once you've opened the tool



Example analysis tools:




3. Can you use your GIS skills to figure out how much land area falls between 100-200 m of a Blackstone stream in all the Massachusetts towns? (should be 347 km$^2$, 13.6 km$^2$ are in Shrewbury)

# To Spray Or Not To Spray:
# Problem Solving With Spatial Analysis

**Problem Statement**: The citizens and government in Middlesex County, Massachusetts are concerned by the recent discovery of a dead bird carrying West Nile Virus. The bird is believed to have lived in the local habitat, and the Centers for Disease Control (CDC) is considering whether to spray the surrounding are with a pesticide aimed at eliminating mosquitoes that may have contracted the disease. A representative from The CDC has arrived at your GIS lab and is asking you to help her evaluate whether or not CDCs current criteria for spraying a pesticide to eliminate the mosquitoes that carry the virus are likely to be effective.

**The Spraying Criteria**: According to the CDC's current criteria, spraying a pesticide by trucks <u>must</u> occur within a 2 kilometer radius around the site of any animal that has been found dead of the virus. However, it is known that pesticides sprayed by truck will only reach to 50 m on either side of roads. Further, the Environmental Protection Agency (EPA) has regulated that spraying must <u>not</u> occur within 100 m of any wetland because they don't want to harm wetland biota.

**Data:** To undertake the evaluation, you are provided with three layers in the cdcdata folder. The layers are:

1) *deadbird* containing the location of a bird that has been found to have died of the virus
2) *wetlands* representing wetlands as polygons
3) *roads* representing roads as lines

**Questions to be turned in:** Given this spatial data and the spraying criteria, answer the following questions (**10 pts**; **2 pts each**):

1) How much land area (units?) is available for pesticide spraying given the above wetland and road distance limitations?
2) What percentage of the recommended total spraying radius does this area represent?
3) In your opinion, how effective is this treatment likely to be?
4) List the GIS processes you used to arrive at this information.

**Map to be turned in (10 pts):**

5) Create an 8 ½ x 11 layout/poster showing the potential spray area, along with the location of the dead bird, roads, wetlands, and any other spatial data that would help illustrate your results. *Make sure your map has a legend, scale bar, and north arrow as well as a title.* When you are finished, export the map as jpeg and insert it in a word document with your answers to the above questions.
**Grading**: Potential spray area, final spray area, and dead bird = 2pts each; roads, wetlands, legend, title, scale bar, and North arrow) = 1pt. each

# Reading: Understanding Coordinate Systems and Map Projections

## INTRODUCTION

## LESSON OBJECTIVES

By the end of this lesson, you will be able to:

1. Select the appropriate coordinate system using provided metadata to determine spatial reference.

## LEARNING SEQUENCE

| | |
|---|---|
| Required Reading | Read the following:<br><br>Understanding Coordinate Systems and Map Projections<br><br>• Coordinate Systems and Map Projections<br>• Coordinates and Coordinate Systems<br>• Map Projections<br>• Developable Surfaces<br>• Map Projection Parameters<br>• Map Projections Properties<br>• Determining Deformation and its Distribution |
| Resources | View the following:<br><br>1. Map Projections |

## INSTRUCTION

# Coordinate Systems and Map Projections

### What is a datum?

A datum is a reference surface for measuring locations on the earth. A datum has two major components: the specification of an ellipsoid, which is an ellipsoid that has been surveyed and defines the origin and orientation of latitudes and longitude lines. We cannot assign any coordinates to a location without first specifying a datum and linking that datum to the shape of the earth through field measurements. There are three common datums that you will use in GIS: North American Datum of 1927 (NAD 27), North American Datum of 1983 (NAD 83), and The World Geodetic System of 1984 (WGS 84). It is very important that you read the documentation of the datasets that you using in your GIS to determine exactly which datum was used to determine locations.

### North American Datum of 1927 (NAD 27)

The first datum we will discuss is the North American datum of 1927 also referred to as NAD 27. The NAD 27 datum is based on the Clarke 1866 ellipsoid which holds a fixed latitude and longitude in Kansas. The locations were adjusted based on about 26,000 measurements across North America. Quite a bit of GIS data is still available in NAD 27, however, more recent data should use North American datum of 1983. Datums are not static, and often see updates and adjustments throughout time. In fact, both NAD 83 and WGS 84 have been updated multiple times.

### North American datum of 1983 (NAD 83) slide 3

The North American datum of 1983, or NAD 83, is the successor of NAD 27, and uses an earth centered reference ellipsoid rather than a fixed station in Kansas. Additionally, 250,000 points were measured to adjust the latitude and longitude locations.

It is worth mentioning that it is okay to use NAD 27 or NAD 83 data, however, when performing analysis, you should convert all the data into a single datum for analysis purposes.

### The World Geodetic System of 1984 (WGS 84)

The third common datum is the world geodetic system of 1984 commonly referred to as WGS 84. WGS 84 is based on satellite measurements and the WGS 84 ellipsoid which is similar to another ellipsoid named GRS 80. The major difference between the WGS 84 ellipsoid and the NAD 27 and NAD 83 datum is at the WGS 84 data has worldwide coverage, where NAD 27 and NAD 83 should only be used in North America. You should also note that the WGS 84 datum is used by the GPS system to report latitudes and longitudes.

### Vertical Datums for North America

Vertical datums also exist as a reference to specify heights and, like horizontal datums, such as NAD 27, they are used to establish elevation. There are two common vertical datums for North America: the national geodetic vertical datum of 1929 and the North American vertical datum of 1988. If you're working with three-dimensional information, in addition to reading about the horizontal datum of the data set, you should also read about the vertical datum to determine which data was used to derive the heights of the dataset.

**Key Facts**

**Vertical Datum**

- A vertical datum is a reference for specifying heights established through a set of surveyed control points and are used to establish elevation.

- Two most common vertical datums for North America
  - National Geodeteic Vertical Datum of 1929 (NGVD29)
  - North American Vertical Datum of 1988 (NAVD88)

# Coordinates and Coordinate Systems

### Representing Space

Representing space is very important in a GIS, as we will want to overlay information for visualization and analysis purposes. The question we must answer is how can we represent the space numerically, so that it can be stored, and manipulated by a computer? The answer is through the use of 2-D and 3-D coordinate systems. This section will discuss 2-D and 3-D coordinate systems commonly used in a Geospatial Interface System (GIS) in the United States.

### 3D Coordinate Systems

As it pertains to Earth, we use 3-D coordinate systems which represent a sphere such as the earth. The important thing to note about the 3-D coordinate systems that we are going to discuss is that it will not ignore the curvature of the earth, which makes it ideal for displaying locations and measuring across long distances where the curvature of the earth will become a factor. The 3-D coordinate system uses two angles of rotation commonly known as latitude and longitude, and a radius to specify a location. The angles of rotation will determine whether the location is north or south of the equator or east or west of the prime Meridian and the radius will specify how far from the center of the earth that location is.

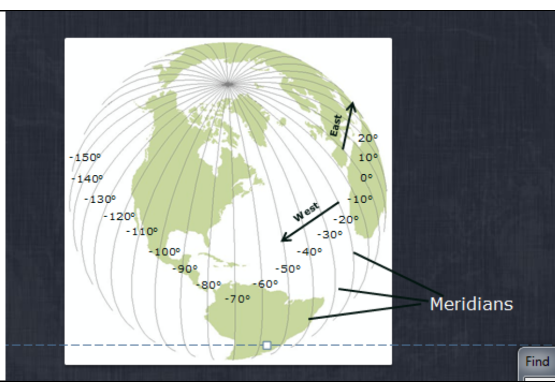
**Key Facts**

Spherical Coordinates

- 3D only
- do not ignore curvature of Earth
- uses two angles of rotation (latitude/longitude) and radius to specify locations

## Longitude

Longitude, also known as Meridian, is the angle of rotation measured East and West around the globe. What may be confusing is that the lines of longitude run north-south from the North Pole to the South Pole. Lines of longitude will vary from positive 180° East to -180° West measured relative to the line of longitude of 0°, which is known as the Prime Meridian. The Prime Meridian runs through Greenwich England. Lines of longitude west of Greenwich England up to and including 180° are represented as a negative number and/or as a Western longitude. Lines of longitude east of Greenwich England up to and including 180° is represented as a positive number and/or as an Eastern longitude.



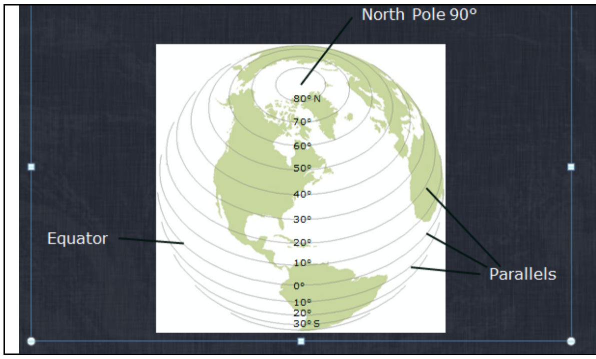

**Key Facts**

- Measures East-West
- Vary from +180° E to -180° W

## Latitude

The second angle of rotation is known as latitude, and is also referred to as a parallel. Parallels measure North to South on the globe, and the lines run in parallel to each other East and West from the North Pole to the South

Pole. The equator is the latitude of 0°. Lines of latitude measure from positive 90°North, which is located at the north pole, to -90° South, which is located at the south pole.




**Key Facts**

- Measures North-South
- Vary from +90° N to -90° S

## Displaying Latitude and Longitude Coordinates

There are two common formats that are used to display latitude and longitude coordinates: One format being degrees, minutes, and seconds, while the second format is decimal degrees. The degree, minute, second, format displays latitude and longitude broken down into the three separate measures of degrees, minutes, and seconds. The degrees number is represented by the degree symbol, the minutes number is followed by a single apostrophe, while the second's number is followed by two apostrophes. For example, degrees, minutes, seconds can be represented as -34° 23 minutes 45.23 seconds latitude, and positive 124° 12 minutes 45.32 seconds longitude.

**Key Facts**

Both latitude and longitude are typically represented in two ways:

- **Degrees, Minutes, Seconds** (DMS)

  -34°23'45.23" , +124°12'45.32"

- **Decimal Degrees** (DD) used by computers

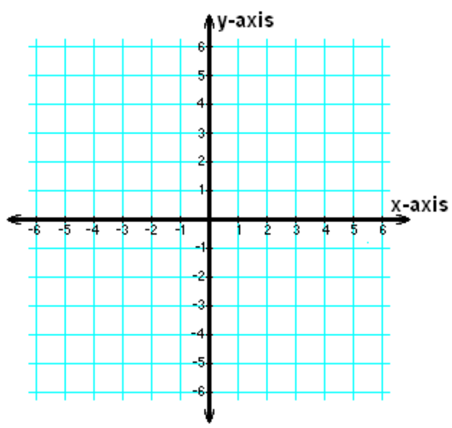
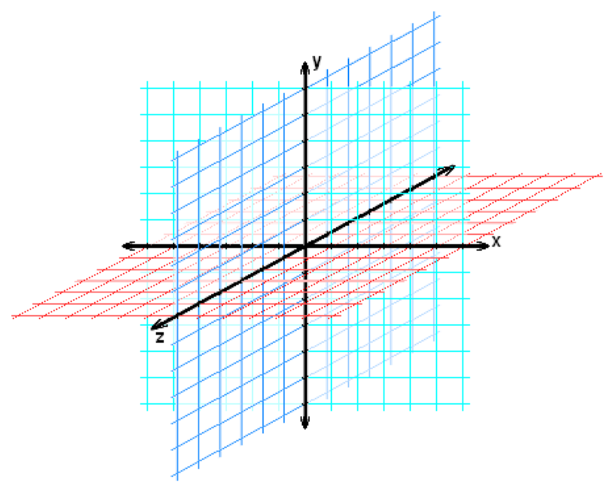
  -34.395897 , 124.212589

Another way latitude and longitude are represented is using decimal degrees. Decimal degrees are most commonly used by computers, as they can easily be stored in the float data type. Both latitude and longitude are simply represented as a decimal number in the decimal degrees format.

### 2-D Coordinate Systems

2-D coordinate systems are more commonly known as a Cartesian coordinate system. The Cartesian coordinate system defines spatial location and extent and can display data in two or three dimensions.

The difference between a 2-D and 3-D coordinate system, is the fact that the 2-D coordinate system ignores the curvature of the earth. Even if a 2-D coordinate system includes height, or elevation information, it is still assumed to be a height above a flat plane.

| 2D (x,y) | 3D (x,y,z) |
|---|---|
|  |  |





### 2-D Cartesian Coordinate System

GIS data typically uses the 2-D Cartesian coordinate system for simplicity. Additionally, for small areas, the curvature of the earth is typically not a factor for analysis or distance measurement, so we can safely ignore the curvature of the earth.

The 2-D Cartesian coordinate can represent many possible locations at many possible scales, and it is so flexible you can even create your own. However, there are two common representations that are widely used in North America, and the world. The first representation being the state plane coordinate system and the second being the Universal Transverse Mercator Coordinate System or UTM system.

**Key Facts**

- GIS data typically uses Cartesian system for simplicity.
- *Ignores curvature of Earth*
    - For small areas, usually acceptable
- Many possible representations
    - could even make your own
    - common representations
    - State Plane Coordinate System (SPCS)
    - Universal Transverse Mercator Coordinate System (UTM)

## State Plane Coordinate System

The state plane coordinate system is a set of 126 geographic zones that cover the United States of America. Each zone is designed specifically for the region of the United States of America that it covers, and is useful because it allows for simple calculations, and is reasonably accurate within each zone. In the state plane coordinate system, coordinates are always positive inside each zone. The state plane coordinate system can be based off of NAD 27 and NAD 83 datums, and the coordinates are represented and measured in feet.

## State Plane Zones

Each state may have multiple state plane zones, but this is not required. Each zone is strategically placed to minimize the amount of error within each zone. Additionally, each zone provides a common coordinate reference for horizontal coordinates over area such as counties while limiting error to specified maximums. Depending on the shape of the state, the state plane zone can be based on two types of map projections, the Lambert conformal conic, or the transverse Mercator.

This is an illustration of the state plane zones in the contiguous lower 48 United States of America. Notice that most states have multiple zones, and that zones typically either run north to south, or east to west. Also note that most of the smaller states, particularly in the New England area, only have one state plane zone covering the entire state, while larger states, such as California, and Texas, have multiple zones, so that the error can be minimized throughout the state. A notable exception to this is Montana, it is a large state, but only has a single zone.

State plane zones are typically designated with the name of the state, followed by a section indicator, such as north, central, or south, or combination of those, or, west, central, east. A notable exception is California, which specifies each zone using a different number.

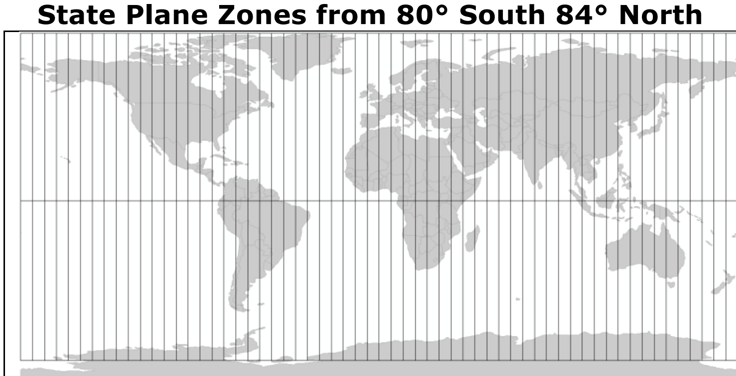
## Universal Transverse Mercator Coordinate System

The Universal Transverse Mercator Coordinate System is a worldwide 2-D coordinate system that splits the world into 60 zones. The Universal Transverse Mercator Coordinate System is useful because it provides for simple calculations, and manages error within each zone. Unlike the state plane coordinate system, which is measured in feet, and the Universal Transverse Mercator Coordinate System, is specified and measured in meters.

The Universal Transverse Mercator Coordinate System, commonly referred to as UTM, is a global coordinate system where each zone is 6° wide at the equator and extends from 80° South to 84° North. Each zone is numbered 1 to 60 heading east from 180° west, that is, on the opposite side of the earth from the prime Meridian. Those are also split in the north and south about the equator. So, zone three, for instance, will have both a zone three north, and zone three south designation.

UTM is common for data and study areas that cover large regions, or regions outside of the United States. It is important to note that coordinates are always positive, and specified in eastings, which is the vertical axis, and northings, which is the horizontal axis.

**State Plane Zones from 80° South 84° North**



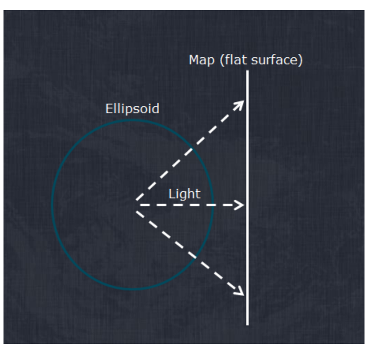
This is an illustration of the state plane zones covering the world from 80° South 84° North. Note the line running along the equator, which splits zones into North, and South.

# Map Projections

### What is a Map Projection?

A map projection is defined as a systematic rendering of locations from the curved earth surface onto a flat map surface. This allows us to flatten the curved surface onto a flat surface such as a piece of paper, or computer screen. The reason why we employ map projections is because globes are not very portable, or practical to use in some cases, therefore, we use map projections to flatten the earth into a map. For a broad overview of what a map projection is watch the video.

Map Projections
http://www.youtube.com/watch?v=2LcyMemJ3dE (1:00)




Basic Illustration of a Map Projection
This basic illustration displays the concept of a map projection. The map projection is broadly composed of three parts, the ellipsoid, which models the shape of the earth, a light source which is used to project features on the earth surface, and a developable surface, commonly a flat piece of paper, onto which the Earth's features are projected, and flattened, to be used as a map.

# Developable Surfaces

### What is a Developable Surface?

A developable surface is a geometric surface on which the curved surface of the earth is projected; the end result being what we know of as a map. Geometric forms that are commonly used as developable surfaces are planes, cylinders, cones, and mathematical surfaces.

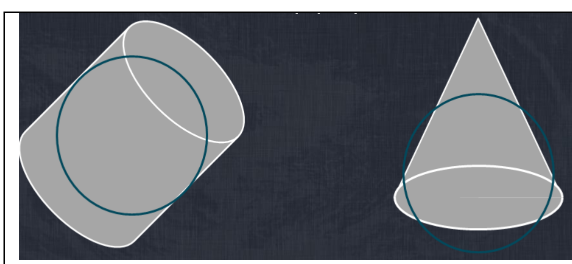
No matter which developable surface is used to create a map, the basic idea remains the same. The features of the curved Earth are projected onto one of the four geometric forms to produce a flat map.

Four common geometric forms used as developable surfaces.

Plane Cylinder Cone Mathematical

### Developable Surfaces: Interaction

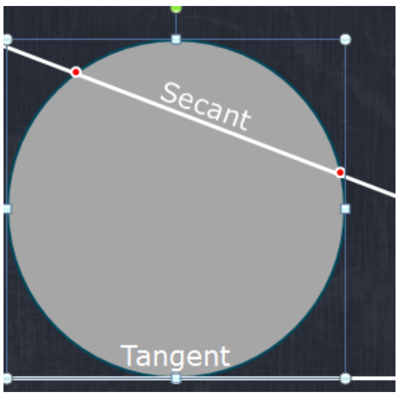
The way in which an ellipsoid and developable surface interact with each other is to place the ellipsoid in different locations and different rotations to gain a desired view of map properties. In this illustration, the blue circle is an ellipsoid representing the earth. The triangle represents a cone, although we are representing it in 2-D. The idea is that the developable surface can be placed on top of the ellipsoid as a hat, or pulled down through the surface of the earth, and even rotated side to side, or forwards and backwards, to create the desired view that we are looking for in a flat map.



Developable surfaces are placed relative to the spheroid/ellipsoid in different locations and at different rotations to gain the desired view and map properties.

### Developable Surfaces: Interacting with the Ellipsoid

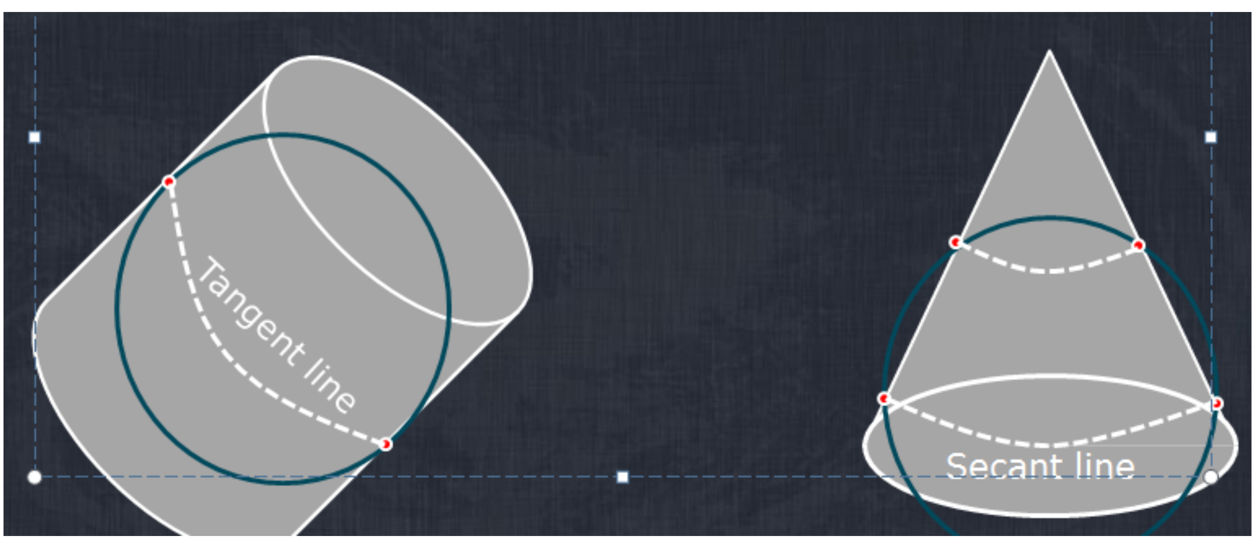
These developable surfaces will interact in a few different ways with the ellipsoid. Generally, the developable surfaces will touch the ellipsoid in either two places, which creates two secant lines, or at a single location, a single tangent line. In the illustration shown, the tangent line is touching the ellipsoid of the South Pole, which should give us a polar view of the earth. The secant intersection is intersecting the earth at two locations which provides us with an abnormal view of the northern portion of the ellipsoid. Both types of interactions are correct, as they depend on the purpose of the map, and the way in which we want to portray the earth on a flat map.




A developable surface touches a spheroid/ellipsoid in either two secant lines or one point or tangent line.

Example: Tangent and Secant Intersection

To further illustrate the idea of a tangent and secant intersection, two more illustrations are provided. For the tangent interaction, there is a line that can be drawn around the earth at its widest point that is perpendicular to the point tangents. This is known as the tangent line. For secant interactions, two lines are drawn following the earth's curvature between the secant intersection points.



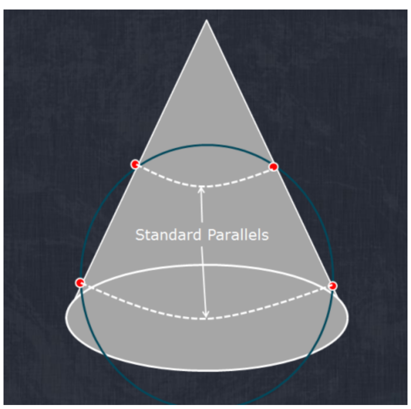
# Map Projection Parameters

## Types of Map Projection Parameters

There are five map projection parameters. They are: standard points and lines, projection aspect, central Meridian, latitude of origin, and light source location. Let’s start with a standard points and lines and work our way down.

## Standard Points and Lines

A standard point and line is defined as a point or line of intersection between the developable surface and the spheroid or ellipsoid. In the case of a secant intersection, there will be two standard lines that would define where the developable surface intersects with the spheroid. If the developable surface happens to intersect the spheroid along a line of latitude, it is known as a standard parallel. Additionally if a standard line falls on a line of longitude exactly, it is known as a standard Meridian. When defining a map projection, you must define the standard points and lines. It is not uncommon to have a map projection follow a standard parallel or standard Meridian.




In this illustration, the cone is in a normal aspect, and with the secant intersection, the secant intersecting lines follow parallels, therefore, they are both known as standard parallels.

So the question is why are standard points and lines important? The reason why standard points and lines are important is because the corresponding places on the map or along the standard point to lines will have no scale distortion. That means that where the developable surface intersects with the spheroid, there is little to no distortion on our flat map. The further away from the standard point or lines, the greater the distortion or deformation that will occur on the map. Secant intersection between the developable surface and the spheroid can help minimize distortion over a large area by providing more control than a tangent intersection at a single point or line.

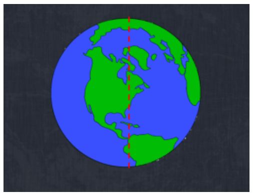
Therefore, placement of standard points and lines is one of the most important parameters to consider when defining a map projection.

**Why are Standard Points and Lines Important?**

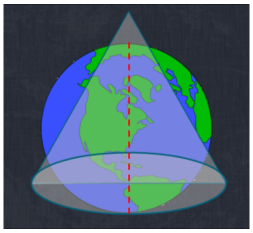
- Those corresponding places on the map will have no scale distortion.
- The father away from the standard point or line(s), the greater the distortion or deformation occurs.
- Secants can help minimize distortion over a large area by providing addition control.

## Projection Aspect

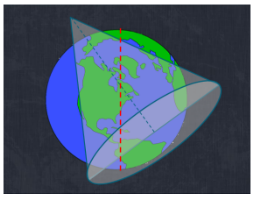
The next projection parameter to discuss is the projection aspect. A projection aspect is the position of the projected graticule relative to ordinary position of the geographic grid on earth. What that means is that if the developable surface's vertical axis coincides with the vertical axis of the earth, then this defines a normal aspect. Should the vertical axis of the developable surface differ from the vertical axis of the earth, then this would be an abnormal aspect.



This illustration displays the normal axis of the globe, which runs from the North Pole through the center of the Earth to the South Pole.



This illustration shows a cone developable surface at a secant intersection with the earth. As the vertical axis of the cone, which is displayed as a blue – line, coincides with the vertical axis of the earth, which is displayed as a red dotted line; this is considered a normal aspect. Remember normal aspect.



This illustration shows a non-aspect. Note that the vertical axis of the cone does not coincide with the vertical axis of the earth thus creating a non-aspect.

### Central Meridian

The next projection parameter to discuss is the central Meridian. The central Meridian defines the center point of the projection. That means that this is the Meridian, or longitude line, that displays in the center of the map. Essentially, this allows you to rotate the earth about the vertical axis to determine what portion of the earth you want to have in the center of the map.

> **Definition**
>
> **Central Meridian**
> The meridian that defines the center of the projection.

### Latitude of Origin

Related to the central Meridian is latitude of origin. Latitude of origin is latitude that defines the center of the projection. That means that this is the latitude that will be in the center of the map projection. Changing the latitude of origin moves the projection about the horizontal axis to determine which portion of the earth will be shown in the middle the map.

> **Definition**
>
> **Latitude of Origin**
> The Latitude that defines the center of the projection.

### Light Source Location

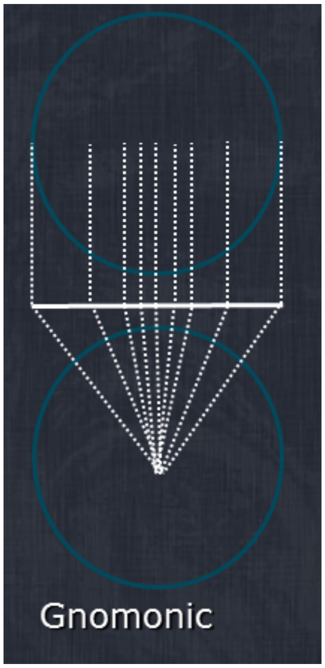
The light source location map projection parameter is the location of the hypothetical light source in reference to the globe being projected. Remember that there are three parts of a map projection: the ellipsoid, the light source, and the developable surface. The light source is what projects the surface of the earth onto the developable surface. There are three primary positions in which we can place the light source.

> **Definition**
>
> **Light Source Location**
> The location of the hypothetical light source in reference to the globe being projected.

### Three Primary Positions of Light Sources

The three primary positions of my sources are gnomonic, stereographic, and orthographic.

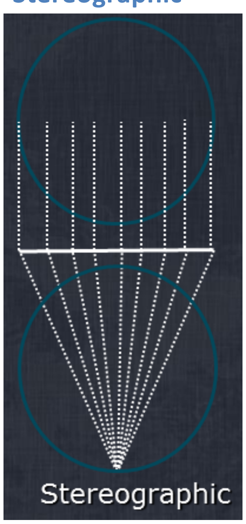
### Gnomonic



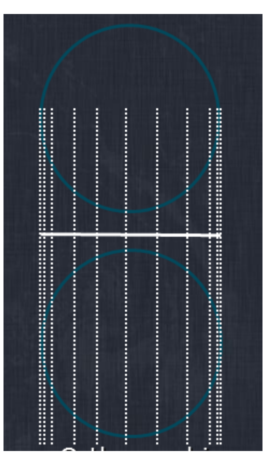

In the gnomonic light source position, the light source is placed at the center, or core, of the earth. The light is then projected through the Earth's surface and projects the landmasses onto the developable surface. In the illustration, the earth is the bottom circle, the lines are the light source, and the solid white line is the developable surface, in this case a plane. The top circle and the dotted lines represent where the earth will be compressed, or stretched, based on the position of the light source.

In the case of the gnomonic light source, looking at a polar projection, locations at the center of the earth are held closer to true, however, the locations towards the extremities, are elongated as they are stretched out to meet the developable surface.

### Stereographic




In the stereographic projection, the light sources placed at the opposite side of the earth from where the developable surface has its secant or tangent intersection. In this case, we see a less severe differential between where the earth is compressed and elongated, but no location is clearly free from distortion.



### Orthographic

In the orthographic position, the light is placed at a theoretical infinite distance from Earth opposite from the point of intersection or tangency. This allows formidable distortion in the center the projection, however significant compression at the extremities of the map.

Together, these five map projection properties allow us to selectively display, and distort the earth to create a map that is suitable for our needs.

# Map Projection Properties

### What is Map Projection Property?

A map projection property is defined as an alteration of area, shape, distance, and direction on a map projection. Map projection properties exist because of the required conversion from a three dimensional object, for example the Earth, to a two-dimensional representation, such as a flat paper map, require the deformation of the three-dimensional object to fit onto a flat map. The three-dimensional spherical surface is torn, sheared, or compressed in order to level it into a flat developable surface.

**Key Facts**

**Map Projection Property**

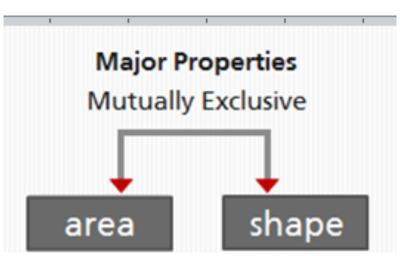
**Definition**
Alterations of area, shape, distance, and direction on map projections.

**Why?**
All maps contain error because of the 3D -> 2D transformation process.

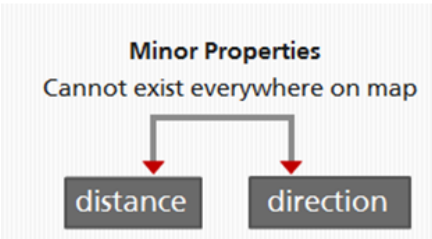
**How?**
Rendering a spherical surface on a plane causes tearing, shearing, or compression of the surface.

### Four Map Projection Properties

There are four map projection properties: area, shape, distance, and direction. These four map projection properties describe four facets of a map projection that can either be held true, or be distorted.



Area and shape are considered major properties and are mutually exclusive. This means, that if area is held to its true form on a map, shape must be distorted, and vice versa.




Distance and direction, on the other hand, are minor properties, and can coexist with any of the other projection properties. However, distance and direction cannot be true everywhere on a map as will be discussed soon.

### Map Distortion

Anytime we create a flat map of a three-dimensional object, we must distort the three-dimensional object. Distortions are unavoidable when making flat maps of the earth. Distortion is not constant across the map, as distortion may take different forms in different parts of the map. There are few points where distortions are going to be zero, however, distortion is usually less near the points or lines of intersection where the developable surface intersects the globe. By determining where the standard points and lines are placed will directly affect where the map will have the least and most amount of distortion.

**Key Facts**

**Map Distortion**

A map can show one or more but never all of the following at the same time:

- True directions
- True distances
- True areas
- True shapes

### Equal Area Map Projection

The equal area map projection, also known as the equivalent map projection, aims to preserve the area relationships of all parts of the globe. You can easily identify most equal area map projections by noting that the meridians and parallels are not at right angles to each other. Additionally, distance distortion is often present on an equal area map projection, and, shape is often skewed.

Even with the distortion of distance and shape, equal area map projection is useful for general quantitative thematic maps when it is desirable to retain area properties. This is especially useful for choropleth maps, when the attribute is normalized by area. Holding areal properties to be true, allows for an apples to apples comparison of density between different enumeration units, such as counties.

**Key Facts**

**Equal Area Map Projection**

**Also Known As (AKA)**
Equivalent Map Projection

**Goal**

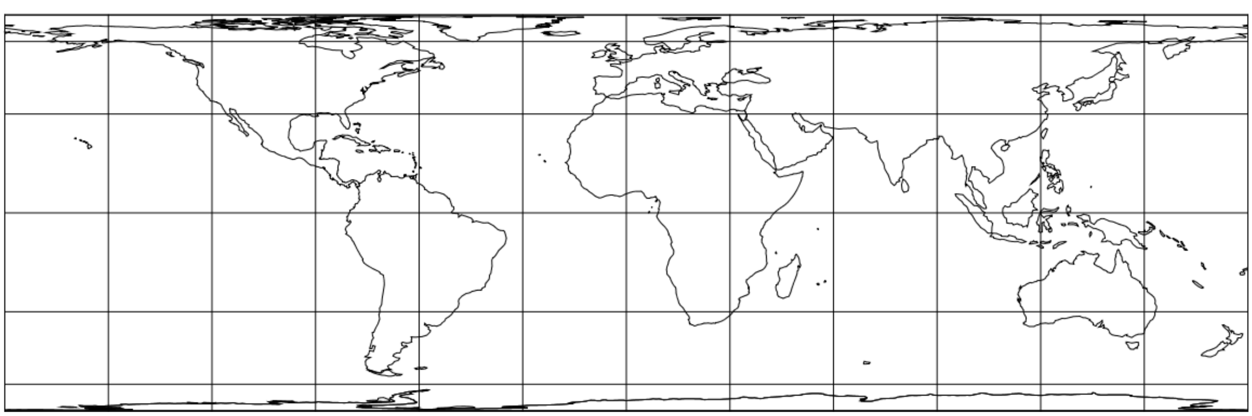
- Preserve area relationships of all parts of globe

**Identifying Marks**

- Meridians and parallels are not at right angles.
- Distance distortion is often present.
- Shape is often skewed.

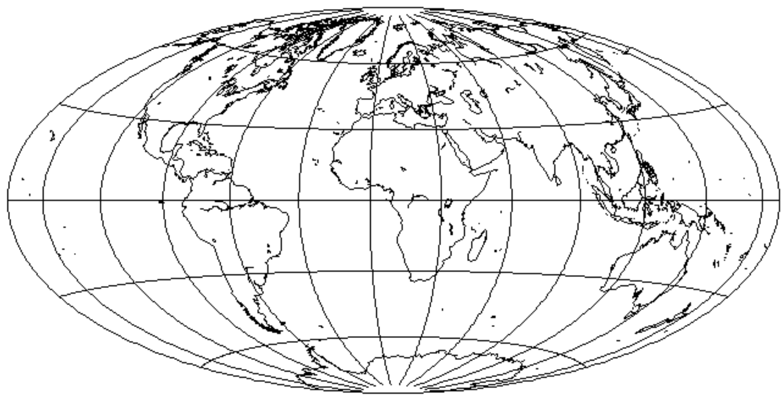
### Cylindrical Equal Area Map Projection

The cylindrical equal area map projection is an example of an equal area, or equivalent map projection, which aims to keep the areal relationships of all parts of the globe correct.



### Hammer-Aitoff Map Projection

A second example of an equal area projection is the Hammer-Aitoff map projection. Again, like the cylindrical equal area projection, this map projection aims to hold areas true. Also note that on this map projection, the parallels and meridians do not intersect at 90° angles, which is a hint that lets us know that this may be an equal area projection.



### Conformal Map Projections

Conformal map projections, also known as orthomorphic map projection, preserves angles around points, and shape of small areas. Additionally, it allows for the same scale in all directions to or from a single point on the map. Conformal map projections can usually be identified by the fact that meridians intersect parallels at right angles, areas are distorted significantly, its small scales, and shapes of large regions may be severely distorted.

Even with the potential for large shape distortion. Conformal map projections are useful for large scale mapping and phenomenon with circular radial patterns such as radio broadcasts for average wind directions.

**Key Facts**

**Conformal Map Projection**

**Also Known As (AKA)**
Orthomorphic Map Projection

**Goal**

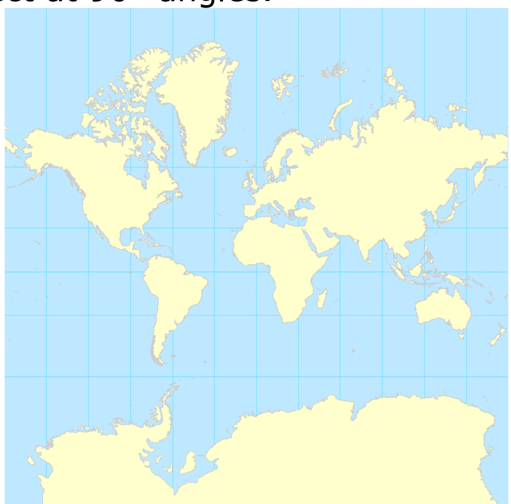
- Preserves angles around points and shape of small areas
- Same scale in all directions from or to a point.

**Identifying Marks**

- Meridians intersect parallels at right angles.
- Areas distorted significantly at small scales.
- Shapes of large regions may be severely distorted.

### Mercator Map Projection

The Mercator projection, perhaps the most famous of all map projections, is a conformal map projection that preserves shape. However, notice the massive amount of distortion in the lower latitudes towards the South Pole, and the northern latitudes, near the North Pole. Also note that the parallels and meridians intersect at 90° angles.



### Equidistant Map Projection

The third map projection family is the equidistant map projection which aims to preserve great circle distances. That means a distance can be held true from one point to all other points, or from a few select points, to others, but not from all points to all other points. It is also important to note the scale is uniform along these lines a true distance from the select points on the map. Identifying marks of the equidistant map projection are that they are neither conformal nor equal area, and look less distorted.

Equidistant map projections are useful for general purpose maps and Atlas maps.



An example of an equidistant map projection is the equidistant cylindrical map projection. Notice that compared to the conformal map projection, there is less distortion at the North and South Pole, and also notices that the shapes do not look overly distorted.

**Key Facts**

**Equidistant Map Projection**

**Goal**
- Preserves great circle distances.
- Distance can be held true from one to all other points or from a few points to others, but not from all points to all other points.
- Scale is uniform along lines of true distance.

**Identifying Marks**
- Neither conformal nor equal area; looks less distorted.

## Azimuthal Map Projection

The azimuthal map projection, also known as the true direction map projection, preserves direction from one point to all other points in the map. It is important to note that direction is not true between non-central points. Direction is only true when measured to or from the specific points chosen. Azimuthal map projection is most useful for preserving direction two or from one point, often used for navigation.

The azimuthal equidistant map projection is an example of a true direction map projection that also holds distance to be true. While not all azimuthal map projections look like this, this particular map projection allows you to measure across the poles, and around the world, to determine true distance and direction from a single point.

**Key Facts**



**Azimuthal Equidistant Map Projection**

**Also Known As (AKA)**
True Direction

**Goal**
Preserve true direction from one point to all other points. Direction not true between non-central points.

**Useful for:**
Preserving direction from one point.

## Combining Map Projection Properties

As seen on a few example map projections previously, we can combine map projection properties onto a single projection. For example, an equal area map projection can also combine parts of it azimuthal map projection. Conformal can combine with azimuthal, equidistant can combine with azimuthal, and azimuthal can combine with equal area, conformal, and/or equidistant.

| | Equal Area | Conformal | Equidistant | Azimuthal |
|---|---|---|---|---|
| **Equal Area** | -- | No | No | Yes |
| **Conformal** | No | -- | No | Yes |
| **Equidistant** | No | No | -- | Yes |
| **Azimuthal** | Yes | Yes | Yes | -- |

**Yes** denotes they can be combined.
**No** denotes they cannot be combined.

## Minimum Error or Compromise Map Projection

There is another map projection family that does not try to hold a single map projection property true. This map projection family is known as the minimum error, or compromise, map projection. The goal of the compromise map projection is to simultaneously minimize all four map projection properties, but may not hold any of the four map projection properties as true. The compromise map projection is useful for general geographic cartography.

**Key Facts**

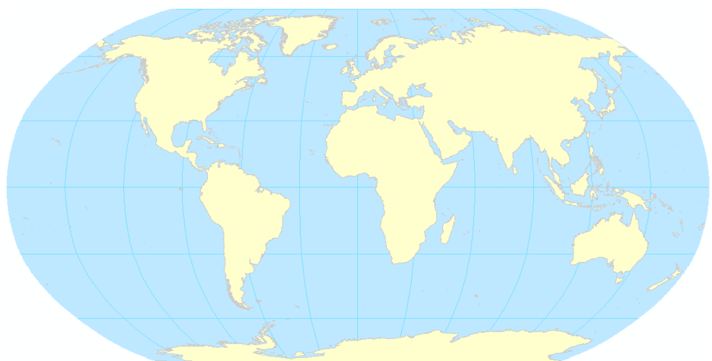
**Minimum Error Map Projection**

**Also Known As (AKA)**
Compromise Map Projection

| |
|---|
| **Goal**<br>Simultaneously minimize all four map projection properties<br><br>**Useful for:**<br>General Geographic Cartography |

### Robinson Map Projection

An example of a compromise map projection is the Robinson map projection. It does not greatly distort any of the four map projection properties nor does it hold any of the four properties true. However what is nice about the Robinson map projection is that it does a reasonable job of showing the true shape, distance, direction, and size of the features of the earth.



# Determining Deformation and its Distribution over the Projection

### Tissot's Indicatrix

A common method to determine deformation on a map projection is called Tissot's indicatrix. As we know, all flat maps distort shape, area, direction, or length when displaying features of a three-dimensional object, such as the earth. Tissot's indicatrix helps to quantify the distortion and projection properties shown on the map projection. Tissot's indicatrix is composed of immeasurably small circles centered at points on the earth. The earth is then projected onto a map using a map projection, and we consider the shape of the circles after projecting the map to determine the deformation and distribution of error throughout the map.

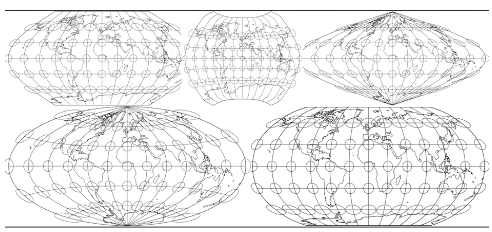
**Key Facts**

**Tissot's Indicatrix:**

- helps to quantify distortion and projection properties
- is composed of infinitesimally small circles centered at points on the Earth
- considers the shape of the circle after projecting the map

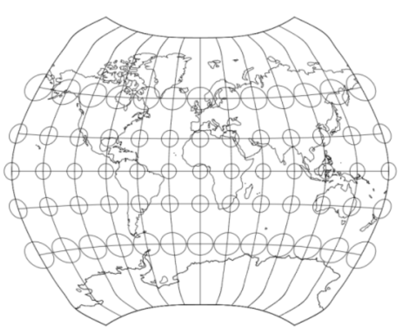
## Interpreting Tissot's Indicatrix

As a quick illustration, if you look at the five map projections, remember that the circles were all originally the same size and were perfect circles. We will interpret Tissot's indicatrix by seeing what happens to the circles when the earth is projected.



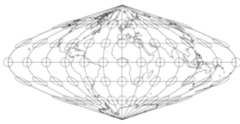
Let's start with considering how the circles change on an equal area and conformal map projection. On equal area map projection, the circles will be transformed into ellipses, but the area of the ellipses, will be the same as the area of the original circles. That is, although they will change shape, it will not change size. On a conformal map projection, the circles will continue to be perfect circles, but their size will vary over the map.

## Tissot's Indicatrix: Interpreting Map Projections

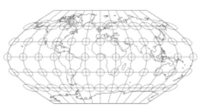


If we look at this map projection, we must consider how the circles look compared to the original size and shape of the circles. On this map projection, even if the circles are varying in size, they are all still generally perfect circles. That tells us that this is a conformal map projection, as it preserves shape.

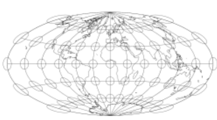
The size of the circles lets us know where deformation is greater. So for instance, on this map projection, we can see that as we move towards the north or south pole, the circles are growing greater in size very quickly, which tells us that this is where we are seeing more distortion.



Let us consider Tissot's indicatrix on this map projection. As we move north and south, or east or west, from the Prime Meridian and the equator intersection, the circles are deforming and shape however, the size of the ellipses and circles tend to stay the same. This lets us know that this is an equal area map projection, as area is preserved throughout the map projection, even though shape is not. Again, by considering how the circle is changing throughout the map projection, we can note that there is severe shape distortion points at the North and South Pole and towards the east and west extremities.



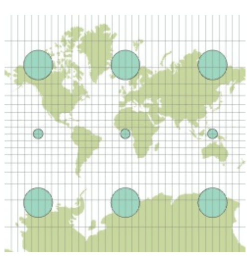
Now let us consider this map projection. It is immediately apparent that the size of the circles is not staying the same, which would indicate that we do not have an equal area projection. However, additionally, the shape of the circles is also being deformed near the extremities of the map. This lets us know this map is neither conformal nor equal area, however, it does still indicate the pattern and distribution of deformation over this map projection.



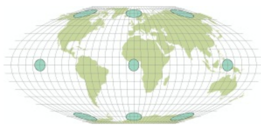
This map projection also does not hold shape or size to be true based on the interpretation of Tissot's indicatrix. On this map projection, it is clear that shape, and area, become distorted as you near the extremities of the map, and are only held fairly well when you intersect the equator and the prime Meridian.

### Conformal Projection Property

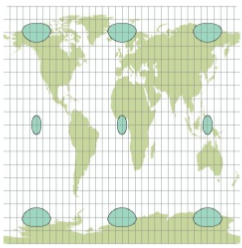
Now that we have a little bit of practice interpreting Tissot's indicatrix, let's look at a few common map projections and see how Tissot's indicatrix performs.



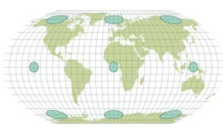
On the Mercator map projection, which is a conformal map projection, the perfect circles would continue to remain circular however they will be larger or smaller than the original circle size. This because, again, the conformal map projection property keeps shape true, but must distort size, as shape and size are two major map projection properties, and are mutually exclusive.



Now let’s consider the flat polar quartic projection, and how Tissot’s indicatrix performs on it. As the flat polar quartic projection is an equal area projection, circles will not have the same shape as a perfect circle; however, they will have the same area as the original circle. As we see on this map projection, the circles roughly have the same size; however we are still seeing some areal distortion towards the North and South Poles.



Now let us look at how Tissot’s indicatrix performs on other projection properties. On an equidistant map projection, such as the equidistant cylindrical projection, where distances are true between two points, the circles are formed into different shapes throughout the map projection. This tells us that this map projection distorts both shape, and size, however, we are not able to immediately tell that this is an equidistant map projection based solely on the size and shape of the circles.



If we look at an aphylactic map projection, which is another way of referring to a compromise map projection, note that we do not have perfect circles anywhere on the map, however, the distortion is not severe anywhere on the map. Again, this is because the aphylactic map projection distorts all four properties, but no one property distorted much greater than the other.

### Map Projection Reference Websites

These are three recommended map projection reference websites. The USGS Map Projections poster provides illustrations and information about many common map projections, and useful matrices to show when the use of a particular map projection is appropriate. The Radical Cartography Projection Reference also shows helpful illustrations, and information about when to use certain map projections. Flex projector is a free piece of software that allows you to create your own map projections, and export them into projection files for use elsewhere.

# Lab 4:
# Coordinate Systems and Map Projections



**Objective**: Understand how ArcGIS software works with coordinate systems and map projections. Learn how to change between coordinate systems and map projections. Understand how images are referenced in GIS so that other spatial data can be utilized with them.

**You will hand in the answers to the <u>numbered</u> questions scattered through this lab. Make sure you answer them as you are going through it.**

## Background/Introduction

Without thinking about it, you use a type of coordinate system every day as you go about your daily routine. This coordinate system is a "map" in your head of locations - for example, where your 9:00 am class is, where your 10:30 class is, and how to get from one building to another. You typically reference these locations by names (e.g. Holdsworth, Orchard Hill, Trader Joe's) and directions (go south on University, take a right just before rte. 9).

All of these locations and directions depend on some spatial knowledge of distance and direction between objects. At the heart of a GIS is the ability of the software to use a coordinate system – to keep a "map" of how locations relate to each other in space. However, in a GIS, the coordinate system references locations as a pair of numbers rather than names and directions.

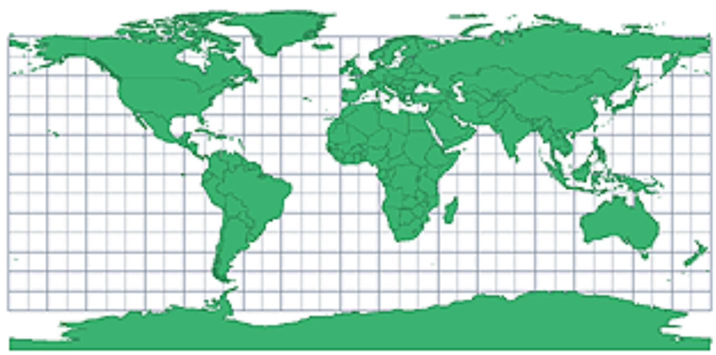
One example: in a GIS, "Holdsworth Hall" could be represented as "72.53˚W, 42.39˚N". Those numbers are probably familiar to you – they're in decimal degrees, with lines of latitude moving north from the equator and lines of longitude moving west from the prime meridian.

You probably aren't as familiar with another possible representation of the location of Holdsworth "115120, 905220 meters"

Open ArcMap and add the file 'mtholyoke' from your datalab4 folder. Zoom in on part of the grid and take a look at the x,y coordinates in the lower right corner of the display – they'll look something like this:

115291.903 904562.003 Meters

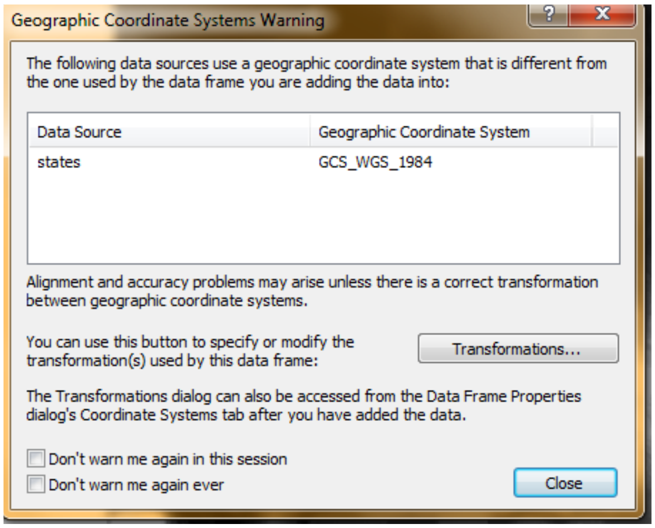
Move the cursor around the map and notice that the numbers change. As you move east to west, the values on the left change, as you move north to south, the values on the right change. What's up with these weird numbers? Why don't we just use lat/lon for everything?

 ←What's wrong with this map?

As you have learned, it is a challenge to represent a spheroid like the earth on a flat plane. No matter what you do, you always have some level of distortion. In the case of a "lat/lon", also known as a "geographic" projection, you get quite a bit of area distortion as you move away from the equator because lines of longitude get closer and closer together. Antarctica is actually about the same size as Europe…doesn't look that way on this map, does it?

Turns out, mtholyoke is projected using a **conic** projection designed to best represent the state of Massachusetts with as little distortion as possible. The units are in meters, and the origin (0,0 point) is located somewhere in the Atlantic off the coast of North Carolina. This projection is the standard one that you'd get when downloading data from MassGIS – it's also known as "Massachusetts State Plane – Mainland NAD83". Here's what the northern hemisphere looks like using a conic projection.

Let's check this out in ArcMap. Add the states shapefile from your datalab4 folder to the project. You'll notice that a warning pops up




This warning is telling us that the layer we're adding has a different projection from the current data frame. In this case, the data frame is projected as MA State Plane, and "states" has a geographic (lat/lon) coordinate system. Fortunately, ArcMap is capable of displaying 'on the fly' and can display both of these layers despite two (or more) different coordinate systems.

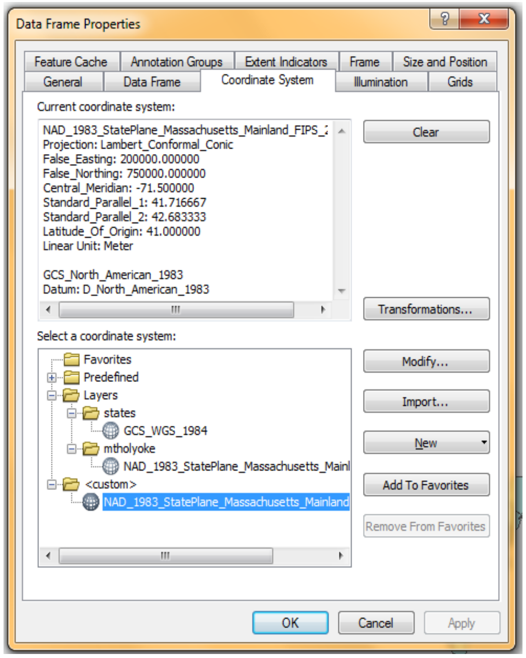
Click 'close'.

**Changing the look of the map display by using different projections**

ArcMap data frame projections default to the coordinate system of whichever layer (shapefile, grid) you add first. If the first thing you add does not have a defined projection (we'll talk about this later) then the data frame will also not have a defined projection.

Zoom to the full extent of states and check out how it looks in the display – it should be reminiscent of the conic projection image on the previous page.

In the Table of Contents, right click on **Layers >** Properties, and select the 'Coordinate System' Tab.



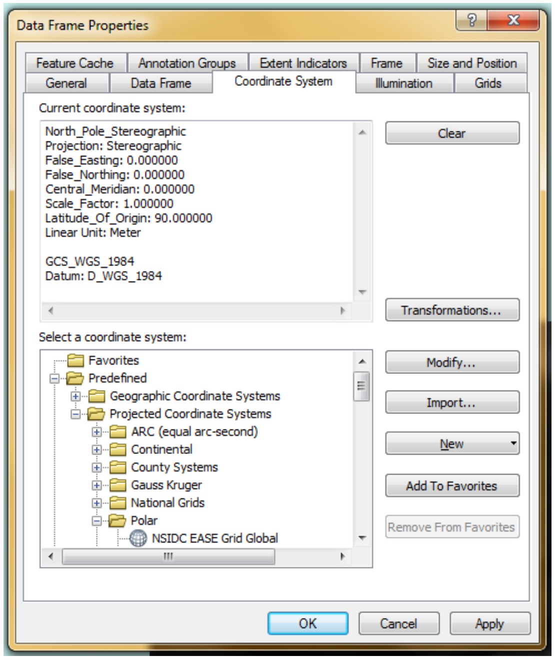

This shows you your current coordinate system based on mtholyoke – if you expand Layers like the image above, you can see the two different projections of your layers. Select GCS_WGS_84, the native projection for 'states' and click OK. You'll get another warning message telling you that your layers are in two different projections (but you already know that!) so click OK.

Now zoom to the layer extent of 'states'. Looks different, right? Now it's projected like the lat/lon map from page 1.

Now zoom to the layer extent of mtholyoke. Zoom in really close on the north edge of the range. Check out the pixels. Are they square?

**Question 1: Why aren't the pixels square?**

Now let's try setting the map display to a projection that is not native to one of the current layers. Go back to the Coordinate System tab in Data Frame Properties. Expand 'Predefined' and navigate to projected coordinate systems > polar > North Pole Stereographic




Click 'OK' and zoom back to the Mt Holyoke topographic map. Is this projection appropriate for displaying mtholyoke? Which way is north?

*Quick Exercise: Set the data frame projection to UTM, WGS84 (Datum), Zone 19N. This is a commonly used projection for areas in New England.*

## Reprojecting Spatial Layers

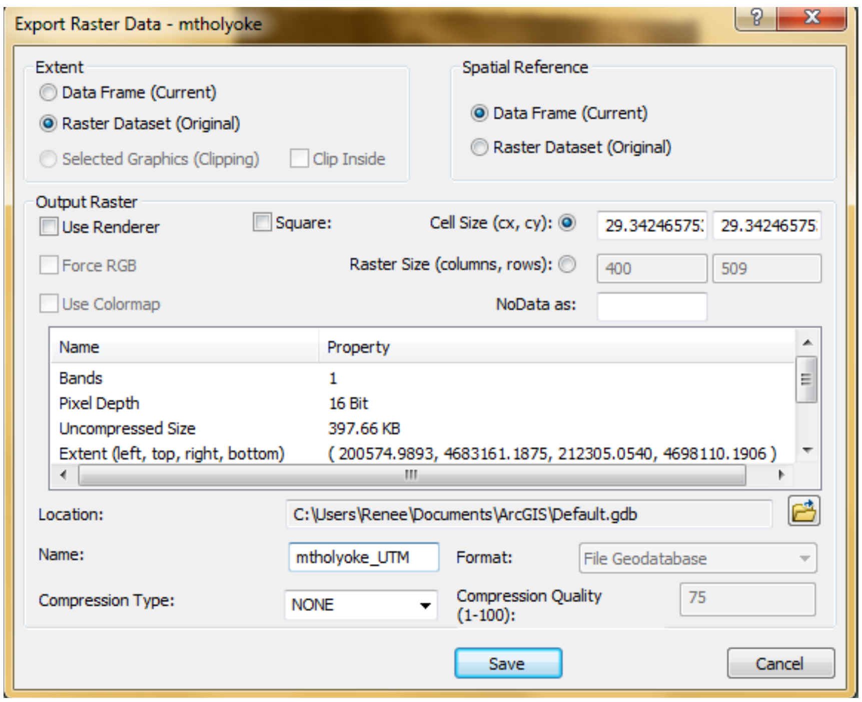
1) In the section above, you have applied different Coordinate Systems containing different map projections to a Data Frame containing layers representing geographic features.

2) However, the coordinate system of the Data Frame is not necessarily the same as the **native/defined** coordinate system of the individual layers you have added to the data frame
   a. You know this because the two layers you've been working with (states and mtholyoke) have two different native projections (Geographic and MA State Plane respectively).
   b. As you know, you can use the Data Frame properties to project your map in the same projection as the layers, or a totally different projection depending on how you want to represent your data.

3) In some circumstances, particularly if you are using certain spatial data frequently and performing multiple spatial processes on your data, you might want to keep all of your data layers in the same projection. This requires *reprojecting* your data. This is different from

just changing the projection of the display.  Now, we want to change the native projection of the data layer.

4) The easiest way to do this is through ArcMap.  Let's reproject mtholyoke so that its native projection is UTM rather than MA State Plane.
5) You should have mtholyoke loaded in your Data Frame and the Data Frame projection set to UTM Zone 19N (double check that this is the case).
6) Reprojecting a shapefile or grid is as simple as exporting a new file with the projection of the Data Frame.
7) Right click on mtholyoke and select Data > Export Data

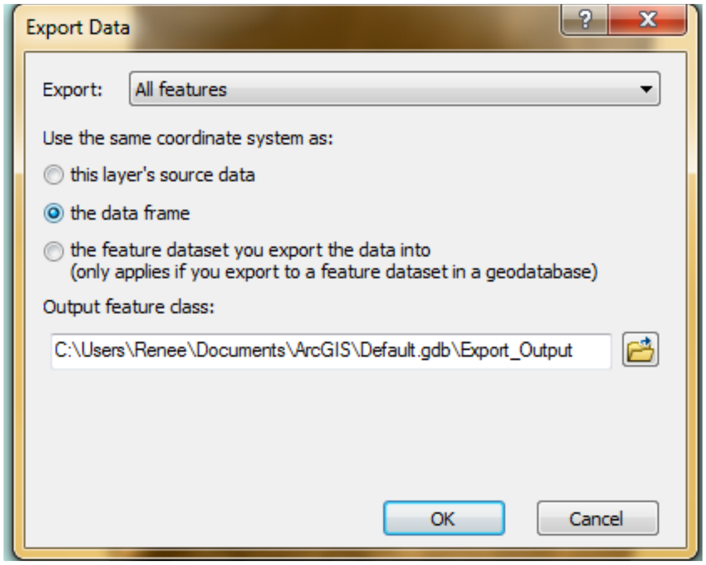



8) Name this output Mtholyoke_UTM and choose an output location in your personal datalab4 folder.
9) In the upper right, select the box in Spatial Reference marked 'Data Frame'.  Note how the Spatial Reference information changes in the box in the center.  Click OK.
10) Add the new file to your Data Frame.  Navigate back to the Coordinate System tab in Data Frame Properties and notice how the two Mtholyoke files now have two different native projections under the Layers tab.
11) Zoom in on the north edge of the range again and compare the two mtholyoke layers by turning one on and off in the display.  Are the pixels exactly the same?

**Question 2: Why aren't the pixels of the two Mtholyoke layers exactly the same?**

Now let's do the same for a shapefile.  You'll notice when working more with ArcGIS that processes performed for RASTER data are often similar, but rarely identical to processes performed for VECTOR data.

1) Set the Data Frame to UTM WGS84 Zone 19N if it isn't already set to that projection.
2) Right click on states, choose Data -> Export Data
3) Select an output location in your datalab4 folder and name the file in a manner appropriate to the content and new projection.
4) Select 'Use the same Coordinate System as the data frame' and add the new layer to the data frame.




5) Zoom in on the Boston area and compare the location of the vertices in the two *states* shapefiles. Are they different?

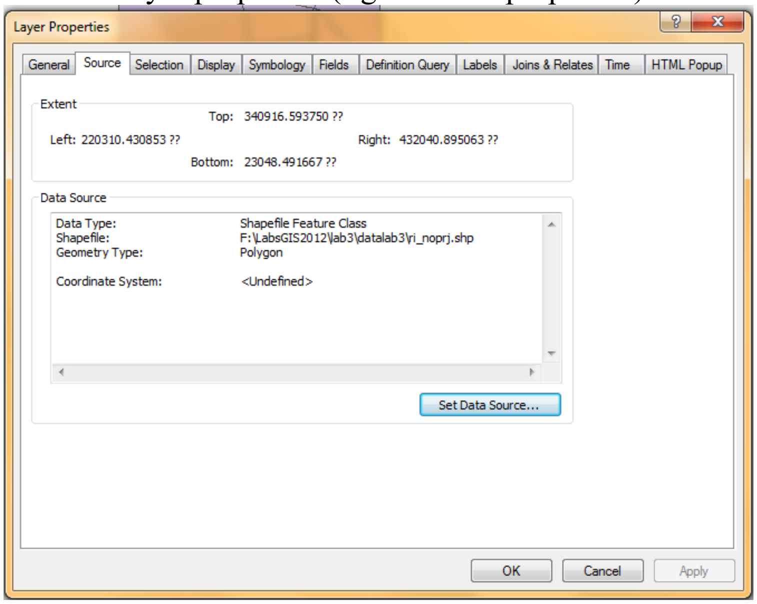
**Question 3: Why do the vertices of this shapefile look the same in the two different projections, while the grid squares in mtholyoke looked different?**

**Defining a Projection/Coordinate System when it is 'undefined'**

1) Sometimes when you're working with spatial data you may come across a data file that doesn't appear to have a coordinate system. In these instances, ArcMap may present a Warning box indicating that the coordinate system is unknown.
2) Another indication of this is that the Coordinate System of the Data Frame says *"Undefined"*
3) This is because ArcMap *requires* a specific file to be associated with the data from which it "reads" the coordinate system. This file is the "projection file" (aka the ***.prj*** file) for vector files and grids; or for images, the "world file" (aka the **.hdr, .jpw, .tfw** or **.sdw** depending on the image format).
4) Without a projection file, ArcMap has no idea what the projection is…and so, we run into trouble.
5) Start a new ArcMap project and add the shapefile 'ri_noprj' to the data frame

**Question 4: Is the native projection of this shapefile Geographic? Why or why not?**

7) You can also tell that this shapefile has no projection information by going to the Source tab in its layer properties (right click > properties)




8) Want to see why undefined projections can be a problem? Add *states* again from your datalab4 folder. Where is states? Zoom to each layer file. You can see that the undefined projection means that Rhode Island is not in its correct geographic location relative to the rest of the U.S.

9) How do we figure out what the native projection is for this layer? *This can be a challenge!* The best way is to look for metadata that you might have downloaded along with the layer. There is often some spatial information in a text file or elsewhere. Another thing to consider is the source of your data. Did you download it from MassGIS? In that case, it's likely to be in MA State Plane projection. Did you download other data along with this one? It might have the same projection. Sometimes this just requires guess work…UTM projections are very commonly used, as are geographic projections. Those are good places to start if you ever run into this problem (I hope you do not!).

10) For now, I'll tell you that the native projection of this unprojected file is RI State Plane, NAD83 (Feet).

11) Use what you learned above (with reprojecting data) to define the correct projection for this file. [Hint: Make sure to adjust the data frame to the correct projection first – you'll know it's correct because Rhode Island will now appear in the correct geographic location relative to states]

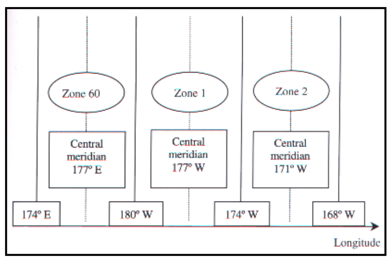
**Learning about a frequently used Coordinate System – UTM**

1. UTM is short notation for "Universal Transverse Mercator". We're going to spend a little time playing with this one so you better understand this commonly used coordinate system.
2. Insert a new Data Frame and add the layers **cntry04.shp** and **utmzone.shp** from the datalab3 folder. Put cntry04 on top and make it hollow so you can see the utm zones through it.
3. The UTM Coordinate System is widely used by all countries because, like the Geographic Coordinate System, it is a worldwide reference for xy coordinates, as opposed to a "local reference system" (which you'll examine next).

   a. Because the UTM Coordinate System is a worldwide reference system satellite imagery (remote sensing) is often provided in this format.
4. In this system, geographic features in a layer are projected using the transverse mercator projection algorithm.
5. In addition to the projection, the entire globe has been divided up into sixty "zones" of longitude and one of these zones is another defining parameter of the UTM coordinate system for any given spatial data file.

**Question 5: What is the native coordinate system of the utmzone.shp file?**

6. Display utmzones such that each 'Zone' is a different color (Layer Properties > Symbology). You'll see that there are quite a few different zones depending on where on earth your study area is. There are a total of 60 zones starting at 180° and increasing to the east.




7. Change the projection of the Data Frame to UTM WGS_84 Zone 19N – the zone commonly used for projecting Massachusetts.

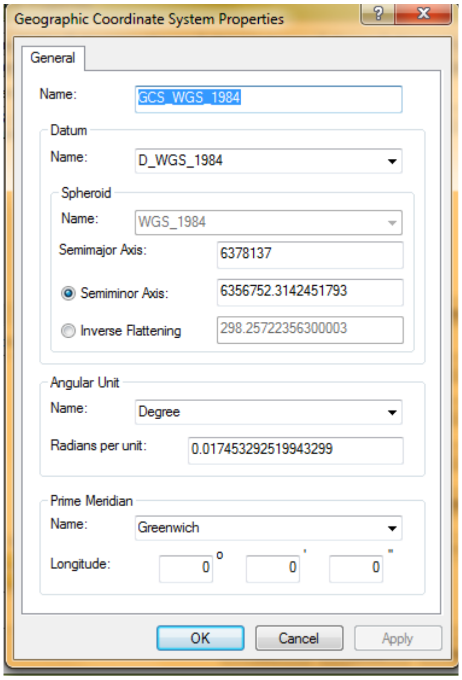
**Question 6: What other Zone might you want to use for projecting Massachusetts? Under what circumstances might you choose a different Zone?**

**Question 7: Where is the origin of UTM Zone 19N? [Hint: The origin is the point from which northing and easting meters are measured]**

8. Find an appropriate UTM projection for some GIS research in the country of Lesotho.

**Datums - What's up with this WGS84 and NAD27 stuff?**
Wikipedia does a better job of explaining this one than I can – so check out the article on *reference datums* at http://en.wikipedia.org/wiki/Datum_(geodesy)

If you ever want additional information on a coordinate system and projection than what is provided in the Coordinate System Tab, you can find it by clicking on 'Modify'.
9. Change the Data Frame projection back to Geographic Coordinate System using World Geodetic System (WGS84).
10. Click 'Modify' once you've selected this coordinate system and you'll see a new window.




11. Here, you're presented with additional information on the projection, including the axis (6378137 is the radius of the earth in meters as defined by this Datum), and the units (Degrees in this case). If necessary, you could edit this information to create a new coordinate system and/or datum…but you'll hopefully never need to do this.

**Question 8: What is the radius of the earth defined as if you choose North American Datum 1927?**

**Assignment: Turn in the answers to questions 1-8 (above in the lab), 9-11 (below) along with the map associated with the following exercise:**

You are the GIS Analyst for the Natural Resource Conservation Service in Montana. You've been asked by a local farmer to determine the number of different soil polygons and the names of the soils in each of his three fields so that he can better determine his fertilization requirements. *(Hint: this lab is all about projections, so expect some problems with the definition of the projection for your spatial data that you will need to fix before you can complete the analysis)*

You have two spatial data files which are in a directory called Montana Soils.

1) fields = which the farmer created by using his GPS unit
2) Montana_soils_utm = the soils in the vicinity of his farm
   **Hint: You may find that using states.shp and utmzone.shp are helpful.

You will also need the names of soil types listed in MT_soil_names.xls

**Question 9: Name all of the soil types in West Williams Field**

**Question 10: Kobase Silty Clay Loam (on 0-4% slopes) is the best soil for farming. How much land area (in square meters) of this soil type is available in each of the three fields?**

**Question 11: Explain how you found the answers to Questions 9 & 10.**

Create a map for the farmer showing his fields and the different soil types in his fields. Include a legend, north arrow and scale bar. Export your map as a jpeg and insert it into a Word doc containing your answers to questions 1-11.

**NOTE: Grading rubric for this lab**

Questions:

1-8 = 1 pt. each

9-11 = 2 pts. each

Map requirements:

Fields, legend, scale bar, north arrow = 1 pt. each

Fields displayed by different soil types = 2pts.

# Project-related lab

**(There is NO deliverable for this lab)**

# Searching for Data – MassGIS
# Importing MassGIS Data into ArcMap:
# MassGIS Highway Roads and a USGS Topographic map example



The reason we start with MassGIS in this class is because often people affiliated with our institution, UMass, want to work on projects *IN* Massachusetts. For that reason, MassGIS is a terrific resource for GIS-relevant data. For anyone in the class interested in doing analysis at other states or internationally, MassGIS still provides some learning that is generic for other GIS-relevant Internet data gathering work.

*In short, this lab walks you through the steps of downloading some data from MassGIS and bringing it in ArcCatalog and ArcMap*. If you'd rather follow this but download data more relevant to your project – GO FOR IT.

Note: MassGIS continually changes, and this lab was written for ArcGIS version 10. For those reasons, directions in this lab may not be totally correct. If you have problems, contact the class instructor.

In this lab we will:

1. Learn how to find a scanned and georeferenced topomap from Mass GIS for the UMass campus (MrSID format) and download it
2. Download the Massachusetts highway roads file
3. Read their metadata in the MassGIS dbase file
4. Define their projections in ArcCatalog – THIS IS A CRUCIAL STEP YOU WILL USE OVER AND OVER!
5. Display them in ArcMap

NOTE: STEP 7 BELOW is particularly important to remember. It shows you how to define a new layer's projection in ArcCatalog.

1. Create a directory in c:\temp\gisdata if one does not already exist. Erase old files in it if it does exist.

2. Go to the MassGIS site: [http://www.mass.gov/mgis](http://www.mass.gov/mgis)

**Review the MassGIS Data Overview page**. http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/overview.html

Obvious (but important) question: **What datum, projection and units are all the MassGIS data in?**

> The datum for the MassGIS database is North American Datum 1983 (NAD83). The data are registered to the Massachusetts State Plane Coordinate System, Mainland Zone (Fipszone 2001). Units are meters.

3. Go back to the main MassGIS site. http://www.mass.gov/mgis/

   In this exercise we will download two layers: the **USGS topographic map for the town of Amherst and the corresponding MassHighway roads layer**.

4. Let's first download a 1:25,000 scale USGS topographic map grid

   Note: if you are unfamiliar with some of the available USGS maps at different spatial (map) scales, see http://erg.usgs.gov/isb/pubs/booklets/usgsmaps/usgsmaps.html

   > See if you can figure these questions on your own (the answers are provided at the end of this document).
   >
   > Question 1: Where on MassGIS can you find USGS topographic maps at 1:25,000 map scale?
   >
   > Question 2: Which quad map number(s) covers the town of Amherst? What is a MassGIS "tile"?
   >
   > Question 3: Can you successfully find and download the Amherst "tile" #117902 as a MrSid image?
   >
   > The answers to these questions are at the end of this lab.

5. Now let's download the corresponding Mass Highways roads layer for this same area. See if you can figure this out on your own.

   > Question 4: Where can you download this data on MassGIS?
   >
   > Download the roads layer for the "Town Tile" for Amherst. These are shapefiles that are stored in an executable zip file (.exe extension).
   >
   > Using the Firefox browser, you can right click on the Amherst shape file (eotroads_8.exe) and choose "save as." Save to your c:/temp/gisdata folder.
   >
   > Why should you be careful about downloading .exe files? – because they are EXECUTABLE files (programs) and can contain viruses. But MassGIS is a "safe website" so it is OK to do.
   >
   > Once you have stored it on your desktop computer, you need to unzip the data. To "Unzip" the roads shape file in executable format (.exe) you just downloaded, you simply

find the file in Windows Explorer and double click on it. You should see several files appear that are all related to the ESRI shape file format. Notice the types of files produced.

What projection is this layer in? How would we find out? See the next step.

6. Understanding MassGIS metadata

   a. First we need to understand how to work with the metadata.

Question 5. Does anyone have an idea of what file of the ones downloaded store the metadata for this layer?

Question 6. How would you read the metadata? What projection is this data in?

The projection is **State Plane NAD 1983**, the units are in meters. The same is true for the topomap.

Now let's pull it into ArcMap. (NOTE: **ALL MASSGIS DATA IS IN THIS PROJECTION!)**

**7. Define the shape file's coordinate system to ArcCatalog *** NOTE! THIS IS IMPORTANT AND SOMETHING YOU WILL HAVE TO DO ANYTIME YOU HAVE A NEW LAYER!!! *****

**Also note – the steps below might have changed slightly in ArcGIS v 10.1… so see if you can figure it out if it has changed.**

1. Start ArcCatalog
2. Navigate to C:temp/gisdata
3. Right-click on the shapefile whose coordinate system you want to define (etoroads_8.shp).
4. Click Properties.
5. The spatial information should have been carried down and should be there. This is because a .prj (projection) file was downloaded.
6. If the projection didn't come down (which may be the case in some MassGIS data, you'd have to define it).
7. You'd click on the X/Y Coordinate system tab.
8. Click the Select button and then click the "projected coordinate system." (What is "geographic" versus "projected" coordinates telling you?)

Select the State Plane, NAD1983 (NOT FEET! – METERS!), and find the

"NAD 1983 StatePlane Massachusetts Mainland FIPS 2001.prj" projection

9. Click OK in the Spatial Reference Properties dialog box.

   In the Shapefile Properties dialog box, the name of the coordinate system appears next to the Spatial Reference

10. Close the window.

8. What about the Mr. Sid topo image?

We have to do a similar process to see if the projection info is known or unknown.

Right click on the Sid image.

Choose properties.

Scroll down to “Spatial Reference” – see if the projection information is listed.

The projection **is** known at this point, and this is because the “MrSid world file” was downloaded with the raster layer that had the projection information within it. If it wasn’t, we’d have to do a similar process (don’t do the below unless the projection information is unknown):

Edit the spatial reference information.

Select.

Projected coordinate systems.

State plane.

NAD 1983

Massachusetts Mainland. Click Add. You should see projection info appear.

Click OK.

9. Start ArcMap. Let’s look at the roads layer draped on top of the MrSid topomap.
    a. Drag the roads shape file layer – etoroads_8.shp from the ArcCatalog window to the TOC in ArcMap.
    b. Drag the Mr.sid image over and place BENEATH the ETOroads in the TOC.
    c. You may have to move your Arc Catalog window to see the topomap
    d. Use the zoom in magnifying glass to zoom in to the Umass campus topo.

This exercise is completed!

**Lab Deliverable:**

**Use this lab opportunity – while this is fresh in your mind – to gather some “basemap” data for your own project (if you are doing something in MA).**

**Answers to Questions above:**

Question 1. Where on MassGIS can you find USGS topographic maps at 1:25,000 scale?

Mass.gov/mgis, Datalayers, MassGIS Datalayers, USGS Topographic Quadrangle Images, Download these images. The direct URL is: http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/imquad.html

Question 2. What is a MASSGIS "tile" # for the town of Amherst?

"The Scanned USGS Topographic Quad Images are available on MassGIS in TIFF and MrSID (a compressed file) format. The images are tiled as 4 kilometer squares, to match an Orthophoto Index, rather than by the original 7.5-minute sheets used by the U.S. Geological Survey for its paper map versions."

The pdf map that has this information on MassGIS is at http://www.mass.gov/anf/docs/itd/services/massgis/quad-imgs.pdf

The page that has the above link is http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/usgs-topographic-quadrangle-images.html

The town of Amherst's orthophoto tile number is 117902.

Question 3. To download  a USGS Topographic Quad tile for Amherst, #117902, in Mr Sid format:

Note: MrSID is an acronym for Multi-resolution Seamless Image Database, a powerful wavelet-based image compressor, viewer and file format for massive raster images that enables instantaneous viewing and manipulation of images locally and over networks while maintaining maximum image quality. The reason we will use MrSid is because it is a smaller file to download, compared to a TIFF file (the other available format).

Download area on this page: http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/usgs-topographic-quadrangle-images.html

MRSID Format" Download page. It is better if you try and navigate and find this page again on your own, but if you can't find it, it is here: http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/usgs-topographic-quadrangle-images-mrsid-format.html

The files are provided in a .zip (compressed format). Find the zip file for 117902. (Helpful Hint: **in your browser, do a Edit, Find for the number**):

- q117902.zip is what you want to download. This compressed file needs to be UNZIPPED using an unzip software. Your computer might be able to recognize this, but in case it can't, you need to find an Unzip program on your computer. One that is freely available is 7-zip (7-zip.org). Within this file are three files that, together, make up the raster topomap for the town of Amherst:

- q117902.sdw – the georeferencing information (Mass State Plane)
- q117902.sid – the image "grid" (this is the scanned pixels of color of the topographic map)
- q117902.aux – some additional information

**To download, using Firefox, right click on each file and "save link as"** to folder c:\temp\gisdata. **Note**: MassGIS historically gets overloaded and times out. If it doesn't work the first time, you may have to try several times. Eventually, a list of MrSid tile files should appear. (Different browsers act differently – but somehow using your browser you should be able to download the zip file to your computer).

Question 4. Where can you download the Mass Highways data for Amherst on MassGIS?

MassGIS Datalayers, Infrastructure, MassDOT Roads. http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/eotroads.html

Question 5. There is an XML file (etoroads_8.shp.XML).that was downloaded with the other shape file data. What is XML? eXtensible Markup Language. It uses various "tags" to label parts of data. This is quite handy in part so that computers can talk to each other and share data using these tags.

Question 6. How would you read the metadata? What projection is this data in?

Open up ArcCatalog and navigate to the layer you downloaded. View the metadata via the Metadata tab. State Plane, NAD83

**Project Related Lab-**

**Searching for GIS Data on the Internet –**
**The "quick list" of US and Global GIS data websites**

June 2013 version

**2013 Charlie Schweik**



Note: These sites keep changing so it is hard for me to keep this completely up-to-date. If you notice any problems with any of these links, please notify me (cschweik@pubpol.umass.edu).

This page provides a list of some of the resources out on the Internet that provide access to GIS-related data.

**LAB INSTRUCTIONS AND DELIVERABLE:**

The objective of this lab is to give you some experience gathering and using GIS taken from the Internet. Ideally, this data might be something you might use for the project later in this class (but not necessary).

For this lab see if you can successfully locate and download data of interest for you for your project from one or more of below websites. Specically:

1. Download a layer from each for a certain area of interest, find information on the coordinate or projection information, unzip the data (if needed);
2. Bring them into ArcCatalog and make sure their coordinate system or projection is defined;
3. Make an ArcMap layout of the layers. If they are for the same geographic area, overlay them. If they are not, make two different map layouts.

There is **NO DELIVERABLE** (NOTHING TO HAND IN) for this lab but ideally – what you do here moves your semester GIS project forward!

**INTERNET SITES:**

General information on searching the Internet. The UC Berkeley librarians have provided a very nice overview summarizing a broader Internet search approach that helps you do a more thorough search than simply "going to Google."

http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html

Federal Geographic Data Commission – www.fgdc.gov

US Geospatial data websites

1. http://www.Data.gov – Try out the geodata catalog and the tool catalog
2. www.Geodata.gov – Federal portal to lots of GIS data
3. www.usgs.gov\ngpo (national geospatial program)
4. www.Geographynetwork.com – ESRI run data portal
5. http://nationalmap.gov/viewer.html – seamless data (meaning no map edges) for download
6. National Historical GIS – https://www.nhgis.org/
7. http://earthexplorer.usgs.gov - USGS Earth explorer – lots of data for download or purchase – especially satellite imagery, aerial photos, etc. This was around before Geodata.gov and data.gov. NOTE to use this to get data you have to register (free) and login. I *encourage you* to try this one. Lots of good data. Note: This site doesn't work well using Safari. If it doesn't appear to work with your browser, try Google Chrome. It works well with that.
8. http://www.census.gov/ - Look at Geography, TIGER – this is where Census data are (e.g., Blocks or Block Group – note that .shp is a defacto standard…
9. http://www.communitycommons.org/ - useful for Census information
10. http://www.nationalatlas.gov/ - Note the "mapping professionals" – here you can get "raw" GIS data including some national layers related to census info
11. Topoquest – USGS Topomaps (GeoTiff) http://www.topoquest.com/
12. Fish and Wildlife service list of New England GIS sites -- http://www.fws.gov/northeast/gis/stategis.html
13. Open Topography (US NSF funded) website - http://opentopo.sdsc.edu/gridsphere/gridsphere?cid=datasets
14. NOAA: Land cover change data for the Northeast (NOAA) - http://www.csc.noaa.gov/crs/lca/northeast.html. Also other data – see http://www.csc.noaa.gov/data/index.html and http://www.csc.noaa.gov/digitalcoast/
15. USDA Natural Resource Conservation Service Data Gateway http://datagateway.nrcs.usda.gov/GDGHome.aspx

NonSpatial Data

1. http://www.fedstats.gov
2. New England City Data. http://www.bostonfed.org/commdev/data-resources/city-data/index.htm?wt.source=email-gen

A few international sites

1. German Earth Observation Center - http://eoweb.dlr.de:8080/index.html
   Here you can get, among other things, Space Shuttle Topography data for most of the entire terrestrial earth.

2. Global spatial data infrastructure association - www.gsdi.org and its http://www.gsdi.org/SDILinks.php page (lots of links to other international data sites)
3. Free GIS, Remote sensing and hydrology blog - http://free-gis-data.blogspot.com/
4. www.fao.org/geonetwork - Food and Agriculture Org site
5. http://geodata.grid.unep.ch/ - United Nations Environmental Program's geodata portal

6. http://www.gbif.org – Global Biodiversity Information Facility (Denmark)
7. http://www.naturalearthdata.com/ -- Natural Earth – free global data

**Lab 5:**
**Internet Data Search**

Charles M. Schweik



Assignment:

Your task in this lab is to find a georeferenced 1:24,000 scale USGS topomap of the area around the town of *Manistee, Michigan*. Hint – I found it looking for a "Digital Raster Graphics" or DRG file – which is a common name for scanned topomap files.

Specifically – and this is important -- locate a "***GeoTif" (TIF file)*** for use in ArcGIS. (note: In a USGS site, I found a version in "GeoPDF" which is not ArcGIS compatible). There are other websites that provide scans of topomaps (e.g., .jpg) but these are not GIS ready (they'd have to be georeferenced).

What you want to find is a download location that has:

- a Tif file;
- and an associated ".tfw" file. This is a "Tiff world file" which has the projection information in it.
- There will be a third file as well, most likely.

To do this, go back to some of the "Searching the Internet for GIS Data – The Quick List of GIS data websites"). At least one of the websites listed has the data you want on it and it can be downloaded, unzipped and viewed in ArcGIS.

Note that topomaps can be referred to as: (1) 7.5 Minute Quads or (2) Digital Raster Graphics.

Once you find one that looks usable, make sure you check the website about what the datum and coordinate system it uses.

Make sure the dataset coordinate system is set using ArcCatalog.

**Deliverable:**

1) Make a map layout for this topomap, and save it to a jpg or other image format. Paste it in a Word document.
2) Name the Word document "yourname_lab5.doc"
3) Upload it to Moodle for the Lab 5 assignment.

If you can't find a file that would work after no more than 2-3 hours of searching. **STOP**. It shouldn't take more than that. In the case of "not found," write a short paragraph in your Word document explaining a little on what you tried in your search. We'll provide you the answer in the future.

Note: In case you find the site I found, FYI, a "".tgz" or a ".tar" file is another type of compressed file (like a .zip file). You can unzip it the way you can a .zip file.

Note 2: you can find this data without having to pay anything for it.

Good luck!

**Grading rubric for this lab**:

Lab is worth 20 points. Proof that you found the data or that you gave it a good effort will receive full credit.

# Optional Lab:
# Model Builder Exercise.

## "Commercial Areas of Interest in Hadley, MA"

Charles M. Schweik



### Model Buider

The model builder is a modelling environment available in all Arc GIS products. It is a more intuitive way of doing geoprocessing than using scripts or comands. In the language of Arc GIS help, "it is a powerful and efficient tool which can streamline workflows. One of the benefits of using the model builder is self-documentation. The models can be saved and run again, whereas if you simply use a series of manual operations, it may be difficult to recreate the process exactly. Models can also be placed within geodatabases or saved as files for distribution. You could bundle a number of data sources along with a model to allow a collaborator to validate your methods."

The best way to learn Model Builder, is to "play" with it. This exercise is intended to show you just a little of how to find your way in this environment. You are encouraged to try this same model with other data/geographical location, modify the model to do something else, or even construct your own model from scratch.

### This exercise

Imagine that you work for a company that is trying to set up a retail location for their products in the Hadley area. They ask you to produce a map with potential areas for their consideration. The areas of interest should be flat (less than 3 degrees slope), close to roads (less than 150 m from any road) and should fall in the zones allocated to business/industrial areas in the town of Hadley.

### Data

Navidate to the folder … Open the project "Modelbuider.mxd". Look in the table of contents: there are three data sets: "ROADS", "zn117p1", and "hadleyDEM". This is where the data comes from:

"**ROADS**": [http://www.mass.gov/mgis/eotroads.htm](http://www.mass.gov/mgis/eotroads.htm) (you are already familiar with Mass GIS roads layer tiled by town).

“**zn117p1**”: http://www.mass.gov/mgis/zn.htm (this is the “zoning” layer for the town of Hadley)

Here you have some relevant documentation about this data obtained from the Mass GIS site:http://www.mass.gov/mgis/zn.jpg

“The MassGIS zoning datalayer represents the boundaries of municipal zoning districts. Because zoning is established at the town level, there is no standard district classification across the state. While districts in different towns may have similar or even identical names, their definitions are often quite different. Generalized codes have been added to make these data useful for regional display. A related table contains detailed information about the districts such as setbacks or text descriptions from each town’s zoning bylaws”.

This is an important issue that we should consider for the proposed problem: if this were “real life”, we should obtain the most updated zoning map directly form the town (but for the purpose of this exercise, we will go ahead and use this layer):

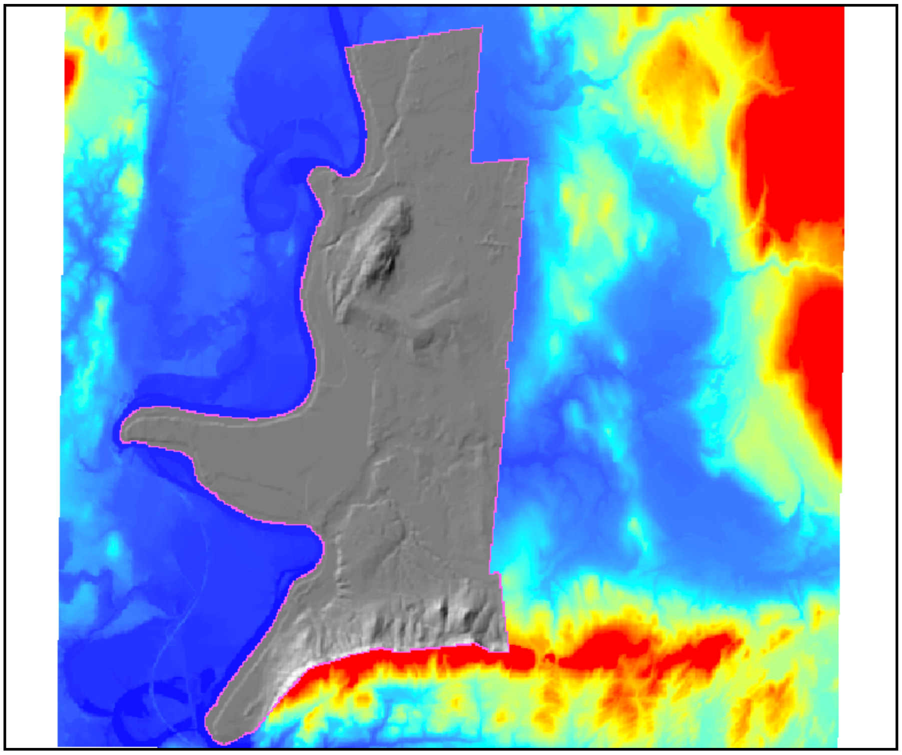
“Zoning district boundaries change frequently and MassGIS currently has no formal process in place to regularly update these coverages. These data should therefore be used for regional analysis only and not as official zoning maps. The town’s own official zoning map and current copy of the by-law should be considered as the final word on zoning boundary questions or issues”.

**“HadleyDEM”:** http://seamless.usgs.gov/

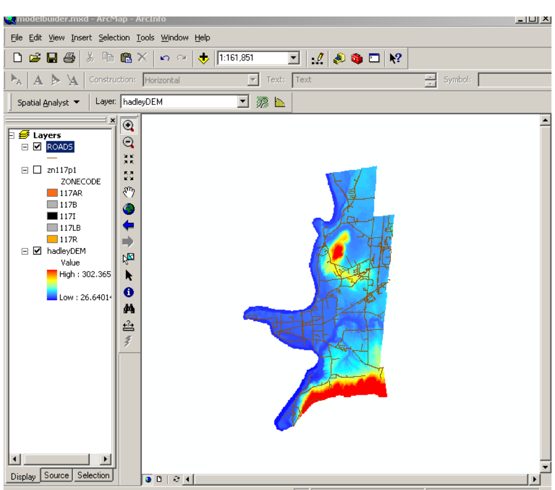
In this website, we zoomed to the Hadley area and obtained a 30 m DEM in geographical coordinates NAD 83. We reprojected to Mass State Plane Nad 83, and “clipped” it using the boundary of Hadley as a mask. (If you ever need to perform this operation, here you will find detailed instructions:

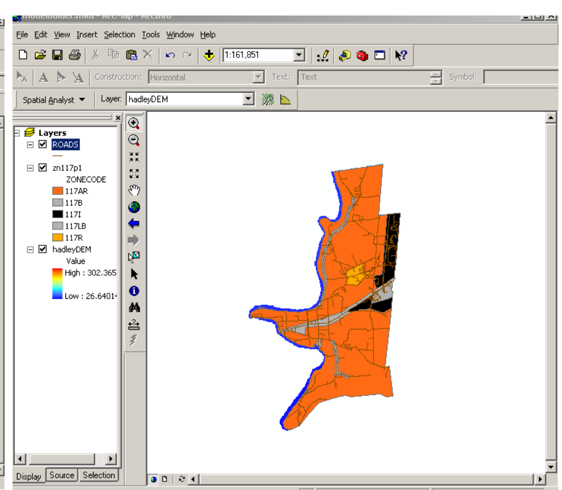
How to clip a multiband image using ArcGIS and Spatial Analyst:

http://support.esri.com/index.cfm?fa=knowledgebase.techarticles.articleShow&d=22526



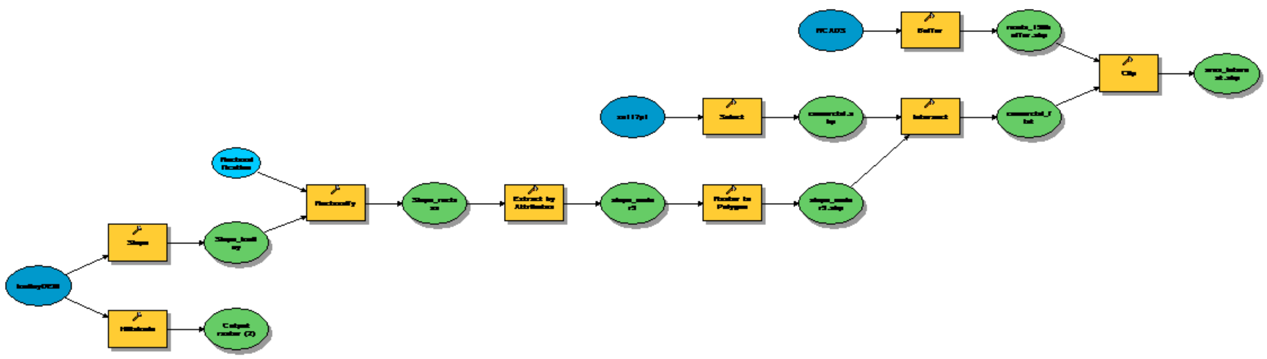
These are the layers that we will use in this exercise:



### *Overview of the process*

The logic that we will follow is to intersect a layer of the flat zones (slope < 3 degrees) with the polygons that represent commercial/industrial areas in the town of Hadley. We will clip the intersection result with the 150 m buffered roads. The idea is pretty simple but there are lots of steps in the process. This is how the final model looks for you to have a general idea of what we are about to do:

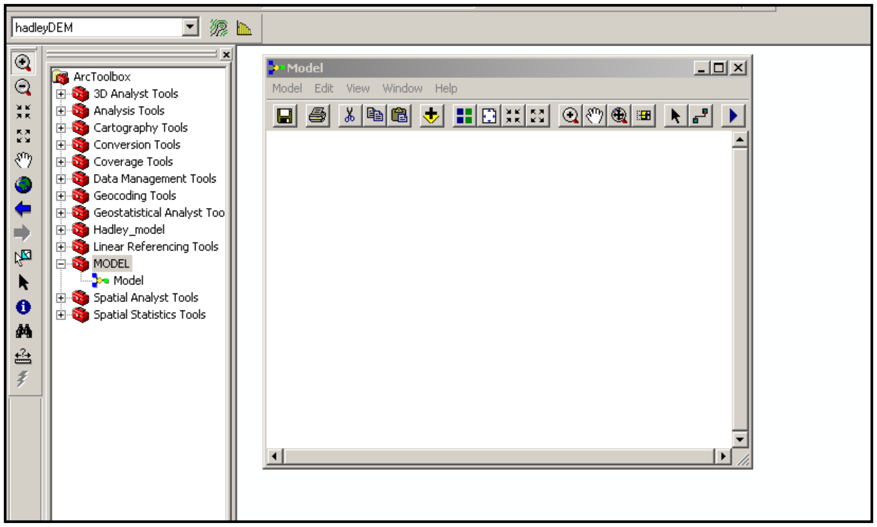



## <u>1. Obtaining a polygon map with "flat areas"</u>

### Step 1. Create a model in toolbox:

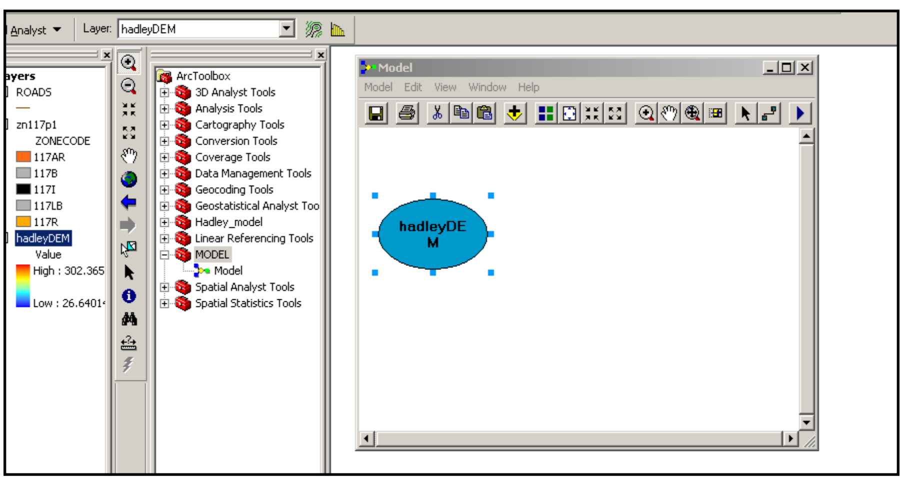
Open Arc Toolbox, right click and go to "new toolbox". Call it "model". Right click on it, new>model:


An empty window called "model" appears. This is the environment where we will build the model.




**Step 2. Import data into the model:**

We will import the DEM into the model, and add analytical operators in a “work flow” to obtain the areas under 3 degrees of slope. Drag “HadleyDEM” from the table of contents. (You could also drag it from ArcCatalog or use the “Add data” button).

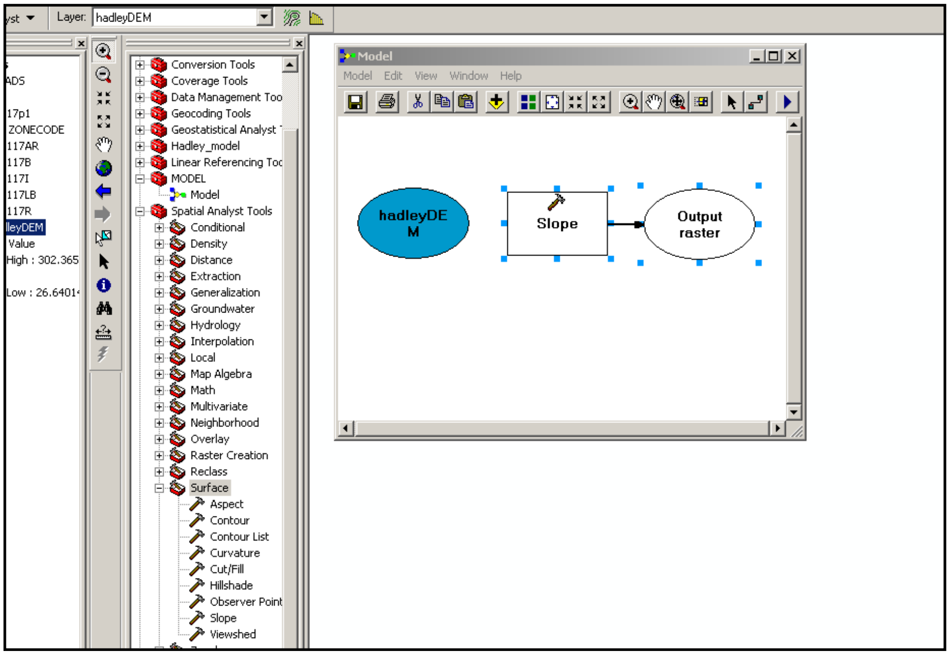



As you can see, data is represented with a blue oval in Model Builder.

**Step 3: Add tools from ArcToolbox .**

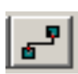
Activate the extension “Spatial Analyst”: “Tools>Extensions>check spatial analyst”

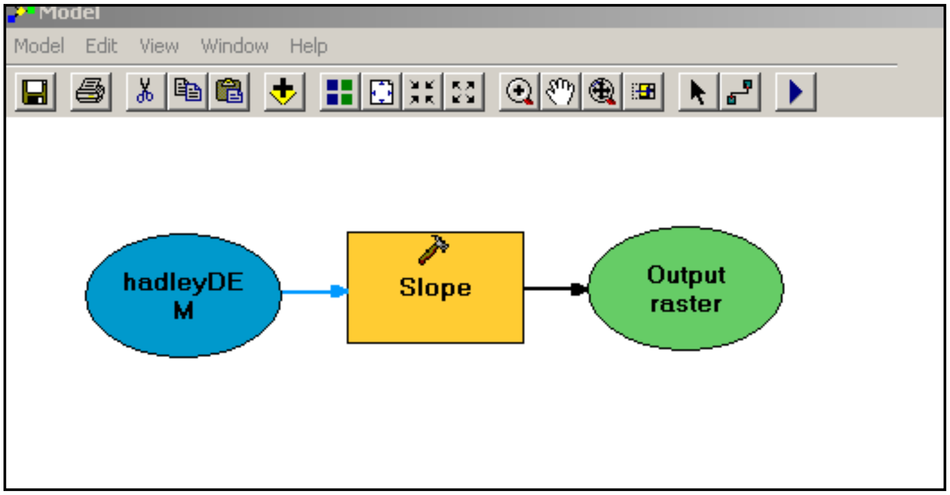
Go to the ArcToolbox tree, and navigate to “Spatial Analysis tools>Surface”, grab the tool “Slope” and drop it inside the model:




The slope tool comes up connected to an “output data set”.

**Step 4. Adding connections:**

Now, grab the “connection” button  in the model toolbar and connect “hadleyDEM” with “slope”.




The tool and the output get "activated". In model builder, tools are represented with yellow rectangles, and "outputs" or "intermediate data sets" with green ovals.

**Step 5. Adjust/enter parameters:**

Grab the selection button  and double click in the "slope" tool to set the parameters:

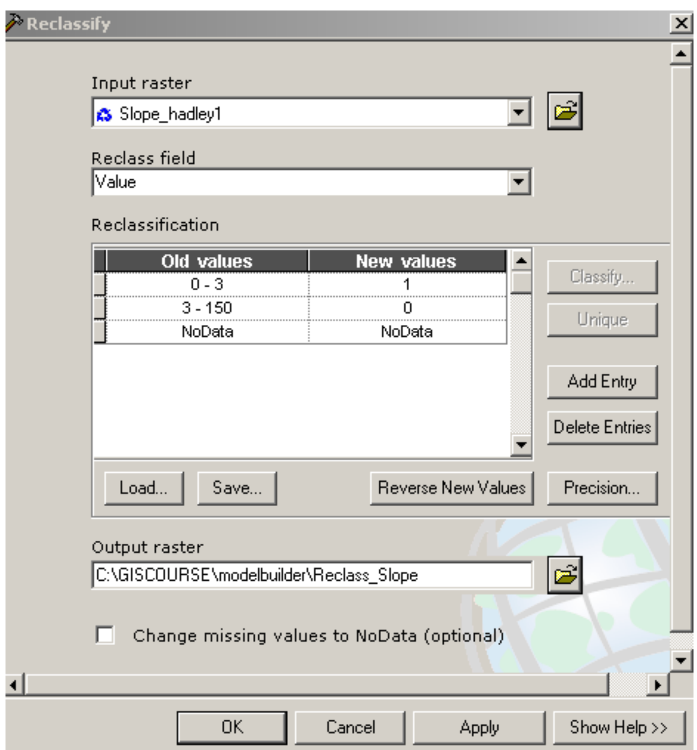

You can rename any output (green oval) right clicking: "Slope_hadley1".

**Step 6: Continue adding more tools/processes to the model:**

Now we will **reclassify** the raster dragging into the model the reclassifying tool:

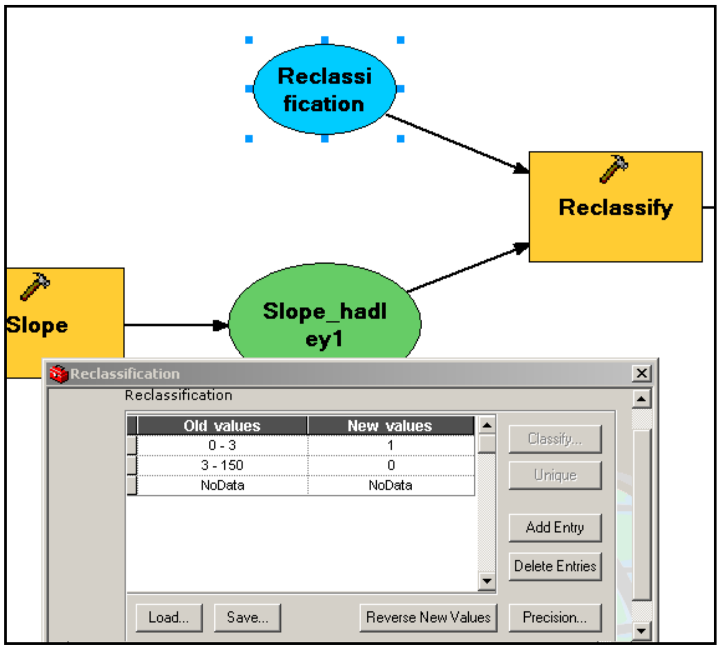
(Spatial Analysis tools> Reclass> Reclassify). Connect the tool to Slope_hadley1.


Set the values for the reclassify parameter:  double click in "reclassify" tool, and a window with parameters will appear. Modify the values ("Add Entry") to obtain two classes: class 1 for slope < 3, class 0 for all other values. Rename the output to SlopeReclass.



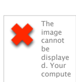

**Step 7. Create a variable.**

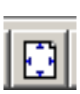
We could create a variable that represents the reclass table that we just created. This will be a light blue oval: a “Value parameter” that references non geographical data. Right click in “Reclassify”, and go to “make variable>from parameter>reclassification”. Double click in the variable blue oval, you could save the current table as “reclass3”, but also modify it and save it with another name. You can also “load” any reclass table that you may have created. This allows to run a model several times with different parameters in no time!




### Step 8. Reorganize the model appearance

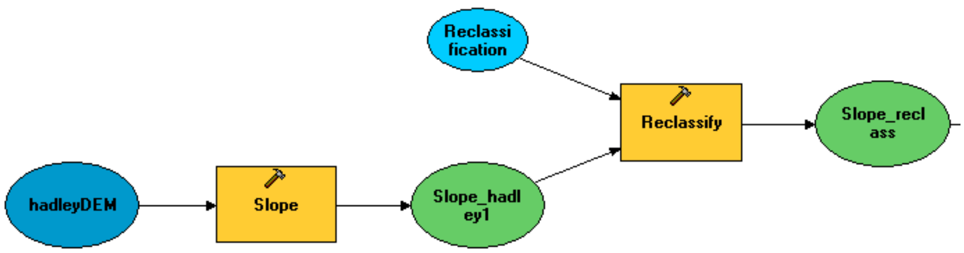
At any time you can use the buttons of the model window toolbar:

will reorganize the apearance of the model.

 will fit all the model in the window

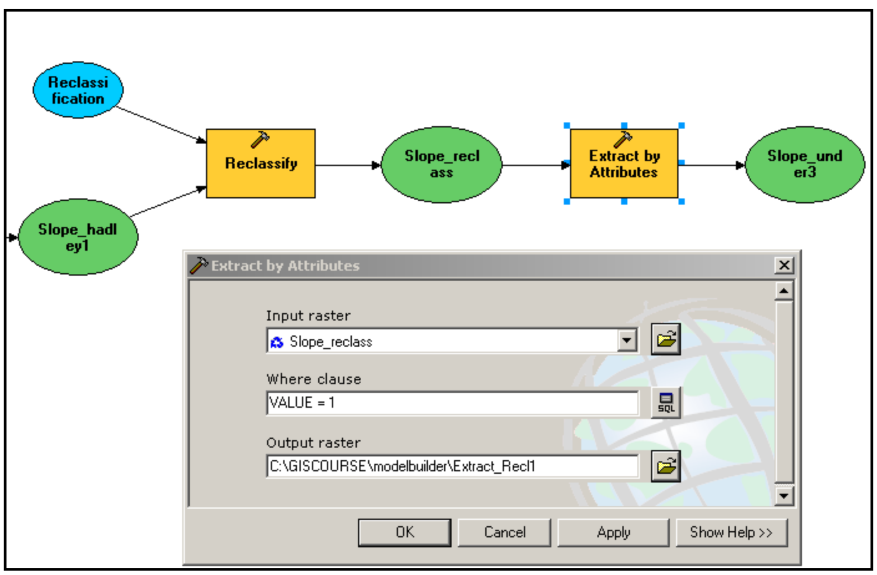
 will help you “find your way” in the model layout.

Your model should look like this so far:

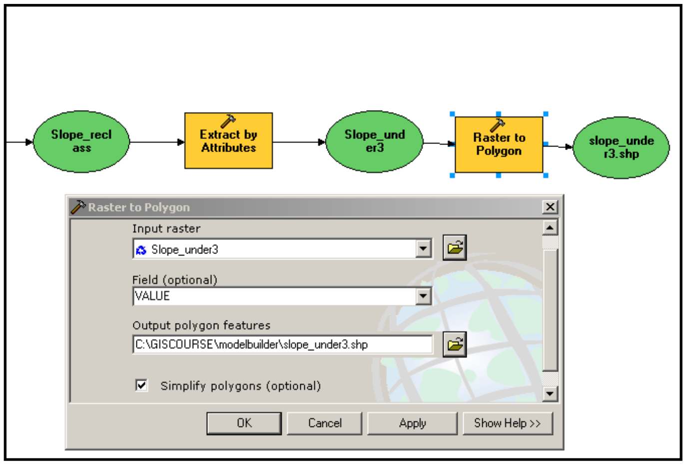



We will go a little faster from now on, as you know the basics of constructing a model. Drag the tool"Spatial Analysis Tools>Extraction>Extract by attributes". Connect to "Slope Reclass".

 Double click in "Extrac by Attributes" tool and modify the "where clause" to extract the "values = 1", as we know those are the raster cells where slope is under 3 degrees. Rename the output to Slope_under3.




Now drag the tool "Arc Toolbox>Conversion Tools>From Raster>Raster to Polygon". Connect the tool to "Slope_under3".  Double click and adjust the parameters, rename the output "slope_under3.shp"



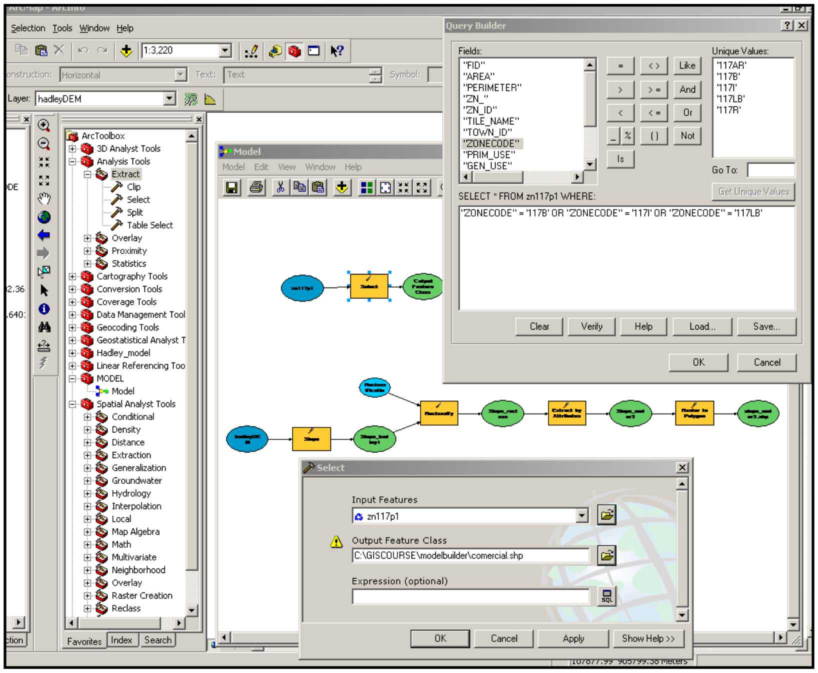

We got the slope polygon map! We will intersect this dataset with the zoning dataset, but first we should select the zoning areas for commercial/industrial purposes.

## 2. Select commercial/industrial areas in the town of Hadley

**Step 1.** Drag the shapefile "zn117p1.shp" from the Table of Contents into the model.

**Step 2.** Add the tool "select" (Analysis Tools>Extract>Select).

**Step 3.** Connect those two.  Double click in "select" and set up the parameters:




This is the expression that you should type in the query builder to select business and industrial zones:

"ZONECODE" = '117B' OR "ZONECODE" = '117I' OR "ZONECODE" = '117LB'

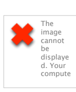
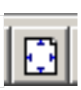
(a look at the metadata for the layer, the attibute table and other aditional .dbf tables that came with the data would take you to the same conclusion).

Rename the output “Commercial.shp”

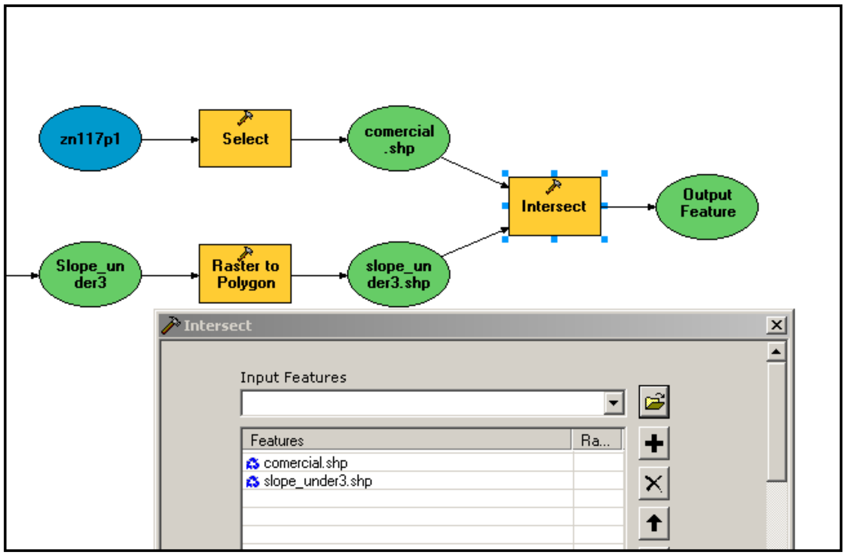
## 3. Intersect “Slope_under3.shp” and “Commercial.shp”

**Step 1**. Add the tool “Intersect” (Analysis Tools>Overlay>Intersect).

**Step 2.** Connect the tool to the two shapefiles we want to intersect.

**Step 3.**   Reorganize the appearance of the model.

**Step 4.**  Double click in the tool and set parameters: Select the two shapefiles as input features, call the output “commercial_flat.shp” and go ahead with the other defaults.



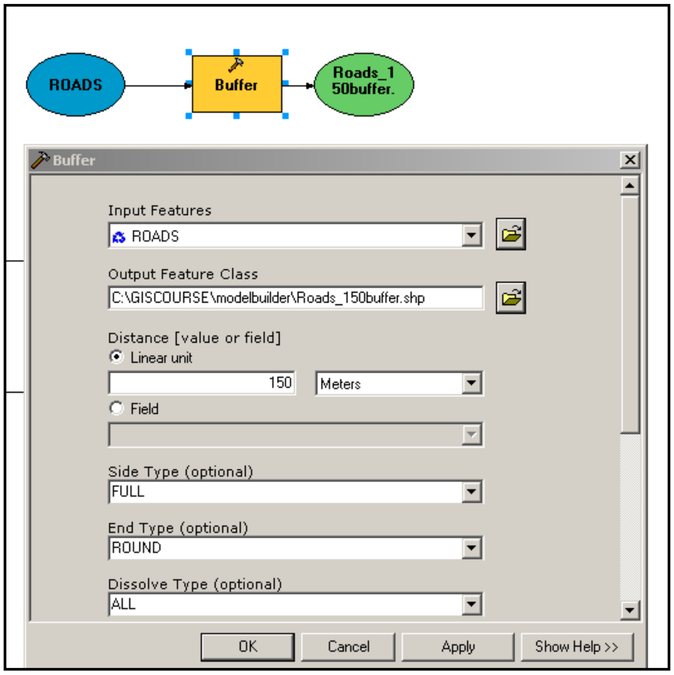

## 4. Create Road Buffers and clip "commercial_flat.shp".

**Step1.** Drag "Roads" layer from table of contents.

**Step 2.** Drag "buffer" tool into the model (Analysis tools>proximity>buffer)

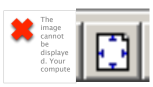
**Step 3.** Connect both.  Double click buffer and set parameters.

Call the output: Roads_150buffer.shp, define a buffer of 150m and select "Dissolve Type": ALL

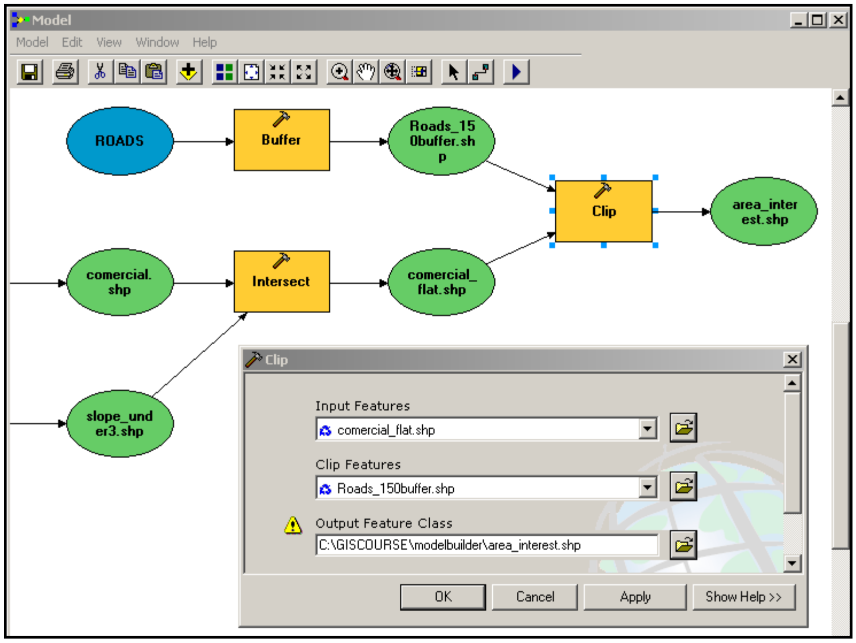



**Step 4**. Drag the tool "clip" (Analysis tools>Extract>Clip).

**Step 5.** Connect it to Roads_150buffer.shp and Commercial_flat.shp

**Step 6.**  Reorganize the appearance of the model.

**Step 7.**  Double click in the "clip" tool and set parameters.

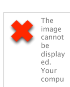



Set commercial_flat.shp as input, roads_150buffer.shp as clip features, and name the output "area_interes.shp".

## 5. Set Model Parameters

You can set up some of the parameters as Model's Parameters. The user will have to define those parameters when running the model from the Arct Toolbox window. If you want to see how this work, you can right click "Reclassification" light blue oval, and check "Model parameter".

## 6. Save the model

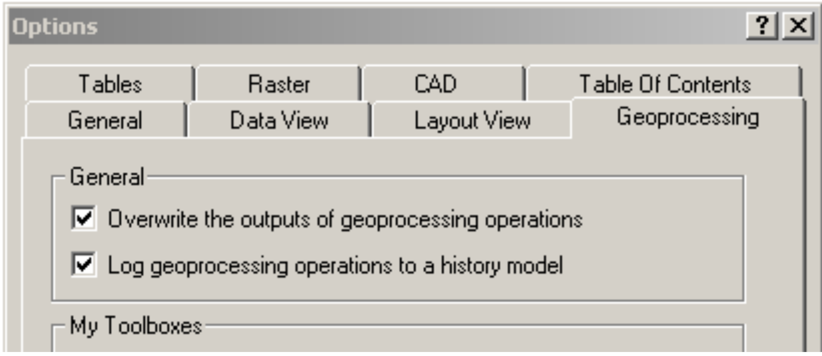


☺

## 7. Bugs/Disclaimers/Problem solving.

There have been some bugs in the past because Model Buider is not able to "overwrite"the outputs of the model, so it is not possible to run a model twice unless you delete the intermediate products in ArcCatalog, or set the green ovals to "intermediate" data (right click>intermediate) and tell Model Builder to delete them: clicking in the Model Menu "Delete Intermediate Data".

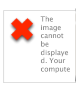
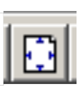
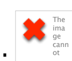
You could also set the program to allow the overwritting of geoprocessing operations: Arc Map menu: Tools>Options>Geoprocessing Tab>Check "Overright the outputs of Geoprocessing.."




IMPORTANT! Even by choosing this option and deleting intermediate data, Model Builder doesn't refresh its memory sometimes, if you try to run a model two times in a row, it may create an error message, and to enable the second run you may have to close Arc Map. Patience here.

**8. Run the model**

We are almost done!

.   Reorganize the apearance of the model. Go to Model>Run, or click the button: 

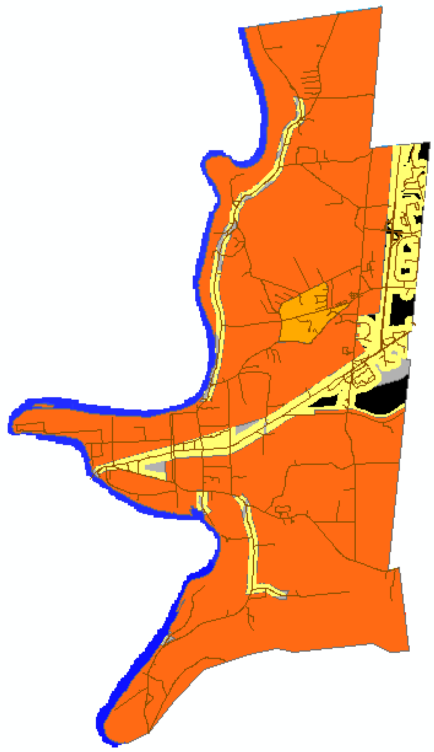
Relax, sit back and observe the model running.

If all the parameters are correct, the model should run to completion. Otherwise, it will stop where there are errors, and you will have a chance to fix any problem. This always takes a little bit of "debugging". You can run the model all at once, or run any tool you want (Right click the tool> Run), and then the model will run only to produce any inputs that the tool requires. Many of the "green ovals" are intermediate data sets, but if some of them are of interest, you could right click them and select "Add to display", so you could examine the intermediate outputs.

If you didn't have time to get here, open the tool "Hadley_model" in the Arc Toolbox tree: there is a model ready to go so you can explore it. Right click "model">edit, and run the model from the window tool.

**9. Display results**

Add “Area_interest.shp” to the table of contents and look at the results. Yellow (Clear) polygons represent the “flat-commercial-close to roads” areas that we were looking for.



# Reading: Creating Geospatial Data

## INTRODUCTION

### LESSON OBJECTIVES

By the end of this unit, you will be able to:

1. Create a new Geodatabase to store digitized features.

### LEARNING SEQUENCE

| | |
|---|---|
| Required Reading | Read the following:<br><br>Creating Geospatial Data<br><br>• Geospatial Data<br>• Coordinate Transformation<br>• Digitizing<br>• Manual Digitizing<br>• Automatic Digitizing<br>• Feature Generalization |
| Resources | View the following:<br><br>1. Manual Digitizing |

## INSTRUCTION

# Geospatial Data

### Creating Geospatial Data

There are many reasons why you might want to create your own geospatial data. The main reason is that the data you need has not been created by anyone else. Even though GIS and digital data have been around for a long time, there is still a large need for the creation of new data covering many different facets of our world at many different scales. A couple of additional reasons that you may want to create geospatial data are that the data you need is on a paper map and needs to be converted to digital format so it can be used in your GIS software, and the data available may be a different scale than you need. For example, let's say you want to show a map of where all the state parks are in Virginia. However, all of the park data sets are extremely detailed, and look very poor at a statewide scale because of the intricate detail of the park boundaries. In this case, you would need to create a new park dataset suitable for display at a smaller scale, or, simplify the existing dataset so that it can display better at the smaller scale. Another reason to create new geospatial data is the possibility that the data available is too old for your purposes and you need an updated version of the data. There are many other reasons why you would want to create geospatial data, but these are the most common reasons that you should be aware of.

**Why Create your own Geospatial Data?**

- Data you need has not been created
- Data you need is on a paper map and needs to be converted to a digital format
- Data available is at a different scale than you need
- Data available is too old for your purposes

### Two Types of Data Sources

There are two types of data sources that you can derive digital geospatial data from: hardcopy and digital.

Each of these data sources has their own positive and negative aspects as a way to store data. Let's look at these aspects, starting with hardcopy data.

### Hardcopy Data

Hardcopy data, such as data stored on paper as maps, was the most common medium of storage for GIS until the 1980's. The positive aspects of hardcopy storage is that it is a stable medium which means that it cannot get a virus, run out of battery power, or crash, like a computer. Hardcopy data is also cheap to produce, is easily portable, and does not require any specialized hardware or software to use. This provides a very low barrier for use, and makes a paper map easy to use and understand by a very wide range of users. Hardcopy data is semi-permanent so long it must be properly stored. The negative aspects of hardcopy data are that it is not reusable because it cannot be easily updated. A hardcopy map, for instance, is a static representation, and can only be updated by printing over the existing map, which is often not desirable.

**Key Facts**

Hardcopy data is…

- stable
- cheap to produce
- does not require hardware/software to use
- semi-Permanent
- portable
- familiar to many users/audiences
- reusable
- not easily updateable

Hardcopy data, such as data stored on paper as maps, was the most common medium of storage for GIS until the 1980's. The positive aspects of hardcopy storage is that it is a stable medium which means that it cannot get a virus, run out of battery power, or crash, like a computer. Hardcopy data is also cheap to produce, is easily portable, and does not require any specialized hardware or software to use. This provides a very low barrier for use, and makes a paper map easy to use and understand by a very wide range of users. Hardcopy data is semi-permanent so long it must be properly stored. The negative aspects of hardcopy data are that it is not reusable because it cannot be easily updated. A hardcopy map, for instance, is a static representation, and can only be updated by printing over the existing map, which is often not desirable.

**Key Facts**

Hardcopy data is…

- stable
- cheap to produce
- does not require hardware/software to use
- semi-Permanent
- portable
- familiar to many users/audiences
- reusable
- not easily updateable

### Digital Data

Digital data has become a common storage medium for GIS data since the 1980s as computers became more widely used and storage capacities grew larger and prices dropped significantly.

| **Key Facts** |
| --- |
| Digital data is… <br><br>• stable<br>• cheap to produce<br>• semi-permanent<br>• flexible<br>• easily updateable<br>• not portable<br>• requires hardware and software to use<br>• may be tied to specific systems that may deprecate |

### Advantages of Digital Data

The positive aspects of digital data are that it is stable, so long as it is properly maintained, backed up, and occasionally transferred to a modern storage medium. Digital storage is now extremely cheap and can store massive amounts of information in a very small space. Digital data is semi-permanent. What is meant by that, is that the data structures must be set up to allow for archiving of historical values information, however, this is often not the case, therefore previous data is often overridden when values are updated and there is no way to see the previous version or value. Digital data is extremely flexible. Digital data can easily be updated and can be used for many different purposes.

### Disadvantages of Digital Data

Negative aspects of digital data are that it is not portable without specialized hardware and software. This significantly raises the barrier of entry for users, and the hardware and software may not be easily understood by the users either. Additionally, the data may be tied to specific systems that may deprecate over time, so there is a need to continuously update the data into newer hardware and software systems.

# Coordinate Transformation

## What is Coordinate Transformation?

Coordinate Transformation is the process used to bring spatial data into an earth-based map coordinate system using a series of control points.

For instance, imagine you are in an airplane and took a bird's eye view of Earth. You can copy that picture from the camera to the computer and attempt to place it in your GIS software. What would happen is that the image will show up not in the right place on Earth, but will, instead, most likely be placed at the intersection of the equator and the Prime Meridian. This is because the aerial photo you took is not referenced to any location on Earth, in fact, the picture has its own arbitrary coordinate system which starts from 0, 0 point, typically in the lower left-hand corner, and increases in a positive direction along the X and Y axis.

That is why, when you add a picture to a map, it places it at the intersection of the equator and Prime Meridian, because that is the 0, 0 point in the latitude and longitude system. In order to make the aerial photo show up in the right position on Earth, we must perform a coordinate transformation. The idea is that you identify locations that are in your picture on the ground, and get the coordinates for those locations. You then match up the known coordinates of the positions on the ground with where it is located on your picture. Once you have chosen enough control points, the computer software can perform a transformation, by placing the picture where it belongs on Earth. During the transformation, the picture is often stretched, rotated, and deformed, to match the location of the control points on the earth and on the photo.


### Performing Coordinate Transformation

In order to perform the coordinate transformation, you must use considerable care when choosing which control points to use. The control points are used to transform the data from the source coordinate system, to the map coordinate system. The source or digitizer coordinate system in the case of our example is the coordinate system used in the digital camera. The map coordinate system is the coordinate system that you wish to transfer to, such as the state plane coordinate system, or latitude and longitude. Therefore, in order to perform the coordinate transformation, we needed to identify control points. One set of control points from the source, or digitizer and one set of control points from the map.

### Control Point Criteria

When choosing control points on both the source and map, you should choose control points that have the highest feasible coordinate accuracy. This means, that if you're looking for an accuracy of 1 foot, you control point should be no larger than 1 square foot in size. Additionally, for the control point, the accuracy of the control point's location should be at least as good as the desired overall positional accuracy. That means that if you want centimeter level accuracy, then you should not measure the location of your control point to the nearest 5 meters. And last, your control points should be evenly distributed throughout the area. By distributing the control points evenly throughout the area, you are providing sufficient coverage to provide a good transformation across the entire area. This in contrast to only choosing control points in one corner of the data set, which in that case, only the corner of the data set will be accurate, and the rest of the data set will contain significant amounts of error.

**Control Point Criteria**

1. Should provide highest feasible coordinate accuracy
2. Accuracy should be at least as good as desired overall positional accuracy
3. Should be evenly distributed through the data area

### Sources of Control Points

There are two common sources of control points. The first source is a land survey performed by a licensed land surveyor. Land surveyors are experts at measurement, and can provide very accurate and precise control points based on existing or new benchmarks that are placed around the nation. The second data source is an existing map that covers the same location as the source data set. If the existing map is already registered, then you can pick the coordinates from the map as your control points. When choosing control points off of an existing map, good choices are permanent features that are easy to see from the air, such as concrete corners, bases of buildings, or aerial photo targets which are those white crosses you often see painted on the highway. Poor choices will be items that move, or are difficult to see from the air such as cars, trees, or street paint.

- Land Survey – base control points on benchmarks or set new control points
- Existing Maps – base control points on existing maps

Land Survey
Bases control points on benchmarks or set new control points

Existing Maps
bases control points on existing maps

| Good Choices | Poor Choices |
|---|---|
| • Concrete corners<br>• Base of buildings<br>• Aerial photo targets | • Cars<br>• Trees<br>• Street Paint |

# Digitizing

### What is Digitizing?

Digitizing is the process of converting coordinates of features from a data source, such as a paper map, into a digital format. Digitizing can be thought of as tracing an object. The general idea is that we will use an input data source, such as an aerial photograph rectified to earth-based coordinates, and then use an input device to trace and record the features from the map as a vector feature. There are two methods of digitizing: manual digitizing and automatic digitizing. Manual digitizing is when a person digitizes the features from the source data manually. This process is often time-consuming, tedious, and prone to human error, however, it allows the human to intelligently determine where to digitize, and what to digitize. The second method of digitizing, automatic digitizing is where a computer program is trained to identify features on the input data source, and then automatically traces those features. Automatic digitizing can save a significant amount time; however, the features need to be easily identifiable on the input data source. Training the software to identify the features may take quite a bit of time, and a human may still need to clean up errors in the automatic digitizing results.

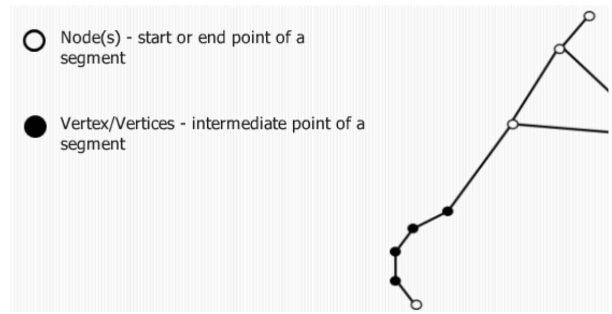
### Anatomy of a Digitized Object

When an object is digitized into the computer from the source data set, it is composed of two parts: a node, and a vertex.

A node is a start or end point of a line segment, or a point.

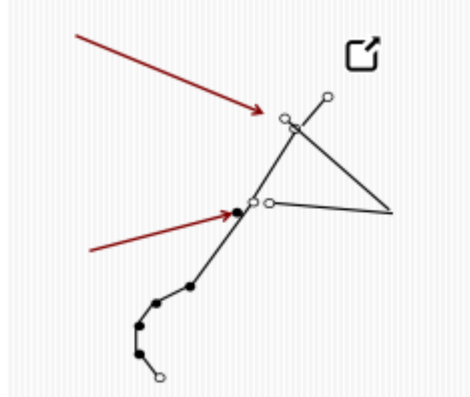
A vertex, is an intermediate point of a segment and must exist between two nodes.

As an example, if we look at the digitized object on the slide, we can see that there are four nodes, and five vertices.  As we know that nodes determine the end and beginning of a segment, and vertices are continuations of the same line segment, we can identify four separate digitized line segments in this figure. Starting at the bottom, we have a node, followed by four vertices, that terminates at the base of the triangle. That is one line segment. The second line segment is between the two nodes on the small leg of the triangle. The third line segment, are the two nodes on the short leg of the triangle, that go through the vertex at the other end of the triangle. The fourth and final segment is the short line that offshoots from the top node of the triangle to the end of the short line.




## Common Digitizing Errors

When digitizing manually, there are two common digitizing errors. The first common digitizing error is known as an overshoot. An overshoot is where lines cross over existing lines and nodes where they should've actually connected. For instance, imagine if we are digitizing pipelines off of a map and we have two pipes that connect in a T intersection. However, when digitizing, the pipe that is forming the stem of the T does not connect exactly with the other pipeline; instead, it passes over and through it in the digital data making a cross instead. This is an example of an overshoot. An undershoot is when nodes do not quite reach the intended line or node. Going back to the pipeline example, if the two pipelines do not intersect, because the second pipeline was terminated just short of the pipeline was supposed connect to, this would be considered an undershoot.

| | |
|---|---|
| *Overshoot*<br>lines that cross over existing lines or nodes where they should have connected<br><br>*Undershoot*<br>nodes that do not quite reach the intended line or another node |  |

### Digitizing Process

When performing digitizing, you should follow these five steps.

Step 1: Create or use an existing data set to store the digitized data. You should select the geometry type of the data set.
Step 2: Load the source data into the GIS program.
Step 3: Register the source to Earth-based coordinates if it is not already.
Step 4: Manually, or automatically, digitize the features into the digital data set.
Step 5: Save the newly digitized features.

### Manual Digitizing

Manual digitizing is the human guided capture of features from a map image or source. There are two methods of manual digitizing: on-screen digitizing, and hardcopy digitizing.

On-screen digitizing is probably the most common form of digitizing today and is where a source is scanned, or downloaded into the computer's memory and loaded into digitizing software. The mouse and keyboard are used to digitize features while the sources feed on the screen.

Hardcopy digitizing is where the original source is taped to a digitizing table which is connected to a computer. The digitizing puck, which is similar to a mouse, is used to digitize features from the hardcopy source into the computer. The purpose of the digitizing puck is to feed coordinates into the computer, based on its location on the digitizing table, which would be the source's coordinate system.

Watch, [Manual Digitizing](http://www.youtube.com/v/t9NnO_T2fuw?version=3&hl=en_US&rel=0) (http://www.youtube.com/v/t9NnO_T2fuw?version=3&hl=en_US&rel=0) (1:46), a demonstration of on-screen manual digitization.

Video Transcript: Manual Digitizing

This video is an example of on-screen manual digitizing. Note that as we are digitizing the centerlines of the roads, it is being stored into a line vector data set. Additionally, after the road is digitized, it provides an attribute with the name of the road. As we digitize the houses, it is being stored into a separate polygon vector data set. It is also being attributed after the polygon is completely digitized. At the end of the digitizing process, do not forget to save your work, otherwise all of your digitizing will be lost.

### Manual Digitizing: Positives and Negatives

**Positive Aspects of Manual Digitizing**
Manual digitizing is typically sufficiently accurate as humans can interpret maps easier than computers. As digitizing is not typically a job that requires highly skilled operators, the training period is often short and the labor is cheap.

**Negative Aspects of Manual Digitizing**
There are a couple of negative aspects to using manual digitizing. For instance, the map scale impacts the accuracy. Different operators may work better or worse at different scales, as some maps may have such fine detail that the operator has a hard time accurately digitizing the intended feature. The hardware that the operators are using may also impact accuracy, if the operators are not provided with capable hardware and software. Finally, different operators will digitize at different qualities and may interpret what is being digitized differently. This may lead to inconsistent digitizing results.

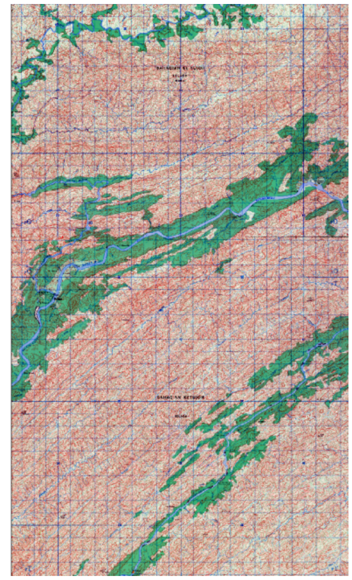
# Automatic Digitizing

### What is Automatic Digitizing?

Automatic digitizing is the computer guided capture from a map image or source. This is different from manual digitizing, where it is a human guided capture. The way in which the computer performs automatic digitizing is that it uses algorithms to identify and digitize the features. While these algorithms are often quite powerful, it often requires an initial training period that may take time. Some of the algorithms available in automatic digitizing software can learn how to identify the features and are supposed to trace based on full color scans, instead of requiring an initial binary classification.

### Example: Automatic Digitizing

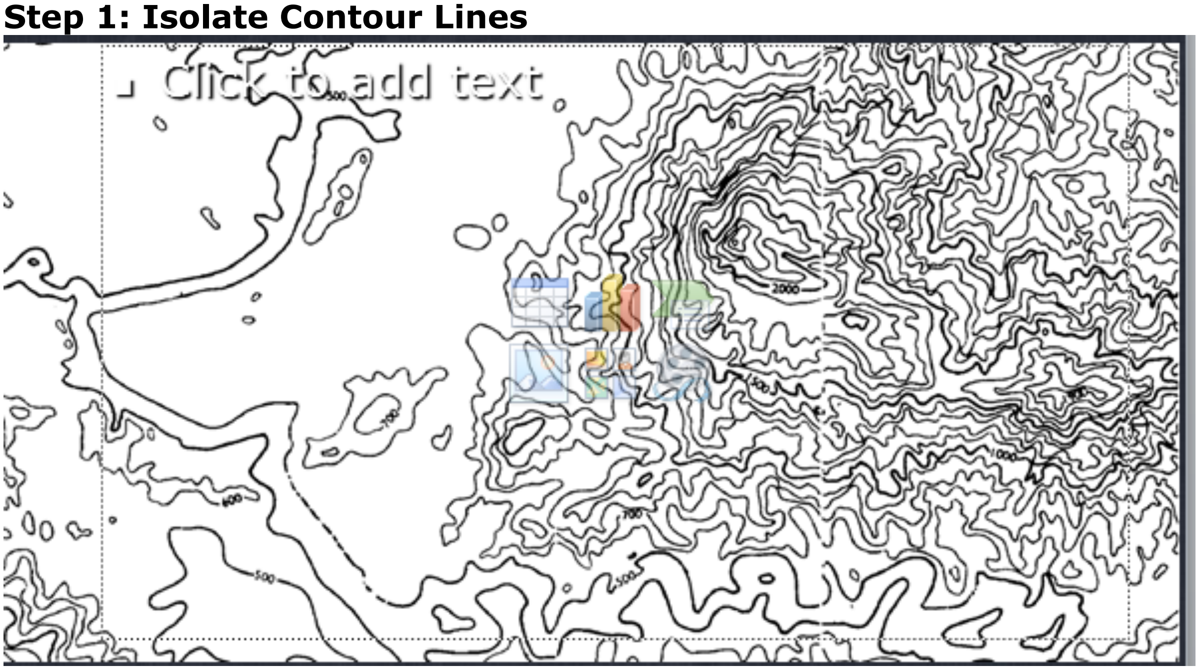
Here is an actual example of automatically- digitizing contour lines from a contour map. What you are seeing on the slide, is actually three contour maps that have been mosaicked, or pasted, together. As you can see, there are quite a few red contour lines on this map, which would take quite a long time for someone to manually digitize. In an effort to save time, we will automatically digitize the contour lines from this map. This map is a good candidate for automatic digitizing because the contour lines are drawn in a contrasting red color, and no other features on the map are the color red. This helps the computer easily identify, and isolate the contour lines of the map.



#### Step 1

In order to assist the computer in isolating the contour lines, the digitized contour map has been classified into a binary image. Wherever there was red on the map, is now black, and where the color red did not exist on the original map, it is now white. This initial step in the automatic digitizing process and is critical for ensuring that the computer can easily distinguish features from the map. It is important to note, that other digitizing software does not need the user to isolate the features for it, as it can work with selected colors.

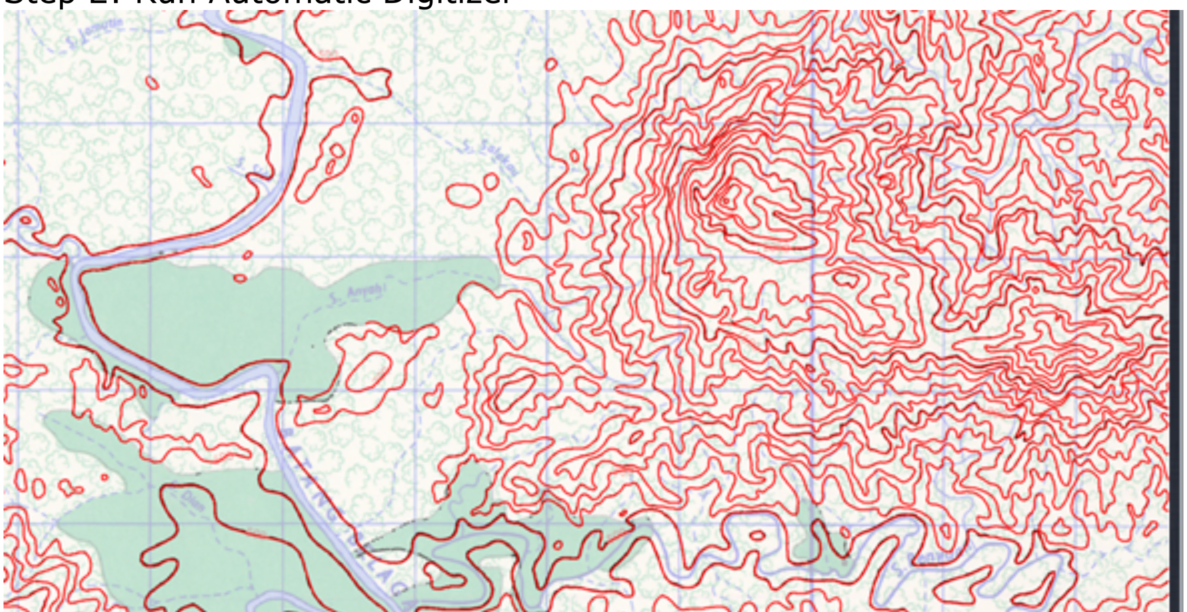
**Step 1: Isolate Contour Lines**



### Step 2

The second step of automatic digitizing is to run the automatic digitizing algorithm. The red lines now shown on the map are the actual digitized vector lines. We cannot see the original raster representation of the contour lines, as they are underneath the digitized lines; this means that the automatic digitizing algorithm did a good job of tracing the contour lines.

Step 2: Run Automatic Digitizer

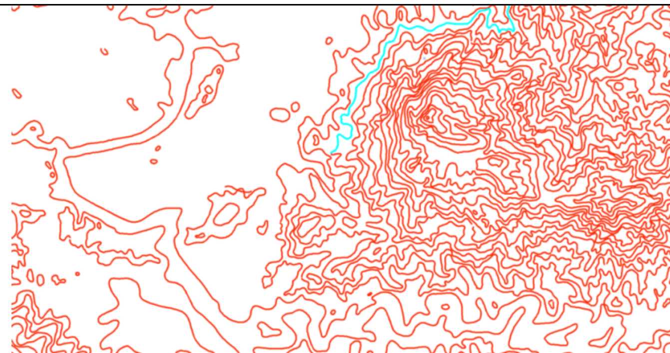
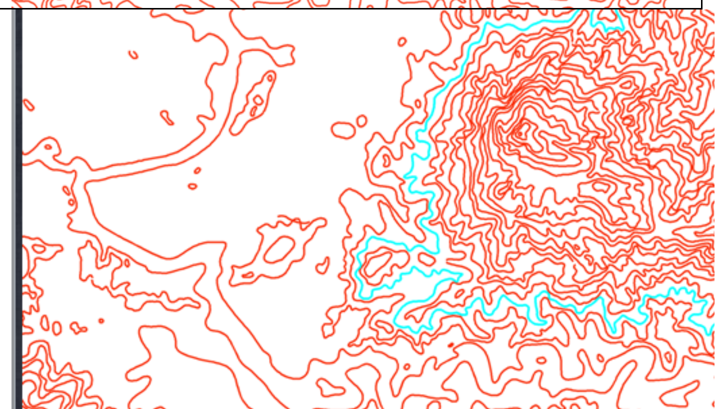


Initial Results of Automatic Digitization

### Step 3

The third step in automatic digitizing is to clean the results of step two. On this particular map, if the contour lines make sharp turns, the contour lines become disconnected. For instance, the blue contour line should connect with the red contour line that it touches toward the center of the screen. However, as this is a significant change of angle, the automatic digitizer considers this to be separate contour lines. Therefore, it is up to the user to either: modify digitizing parameters so that this may be caught, and run step two again, or, manually connect the contour lines.

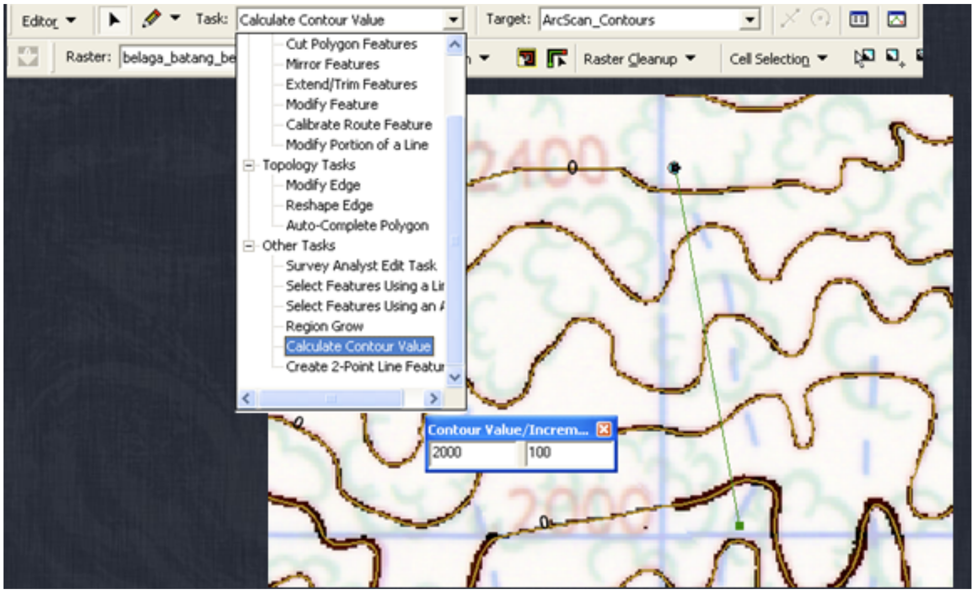
Cleanup Task Connecting Disconnected Contours

| | |
|---|---|
| Example of Cleanup Task Connecting Disconnected Contours |  |
| Example Cleanup Task Connecting Disconnected Contours<br><br>This is what the corrected contour line looks like. In this case, the user opted to manually connect the contour lines. Now that that error has been corrected, the user would continue to review the automatic digitizing results to find the other errors, and correct those. |  |

### Step 4

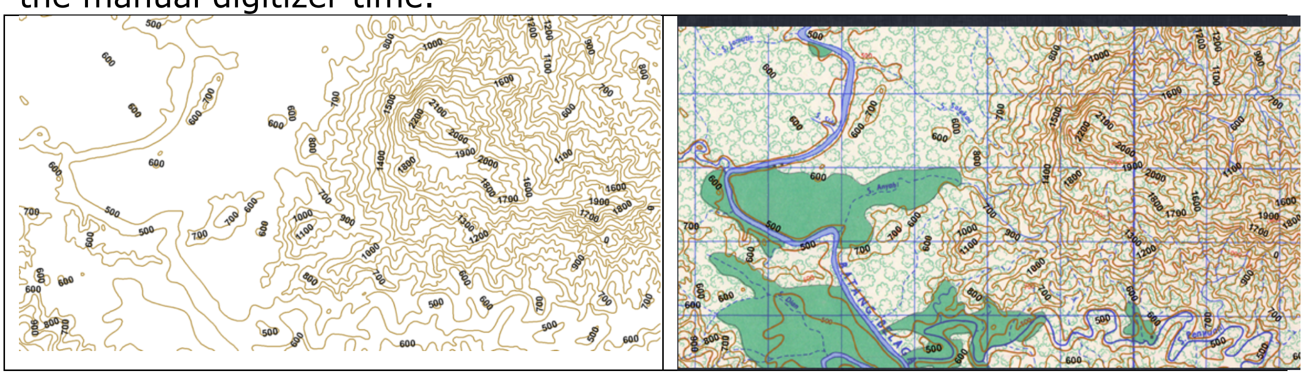
The fourth step is to assign the contour line values into the attribute table of the newly digitized contour lines. In this case, a tool was downloaded that makes it easy to assign multiple contour lines at once, by drawing a line across multiple contour lines, and then specifying the first contour line it crosses contour value, and how much it should increment each additional contour line that passes over. So, in this case, the contour line at the bottom would be assigned a value of 2000, the line above it 2100, line above that in the line above that 2220 and 2300 respectively, and finally, 2400 as the last, top contour line is crossed.

**Step 4: Assign Contour Line Values**



This is the result of the automatically digitized contour lines with the contour values labeled.

**Completed Contour Lines** SLIDE 10
Finally, this shows the completed digitized contour lines overlaid on top of the original contour map. As you can see, the digitized contour lines follow the original contour lines extremely close and it does a good job of saving the manual digitizer time.



## Automatic Digitizing: Positives and Negatives

### Positive Aspects of Automatic Digitizing

Automatic digitizing is good for large projects that would take a large amount of staff and time to complete. There is no initial operator error but the results will require manual editing, as we just saw. Automatic digitizing can work well with a large number of elements on a map, so long as those elements are easily distinguishable by the computer.

**Negative Aspects of Automatic Digitizing**
There are a few negative aspects of using automatic digitizing. One negative aspect of automatic digitizing is that the software can be very expensive; however, if the project is sufficiently large, the expensive software may ultimately be cheaper than the labor that would be required for manual digitizing. Automatic digitizing can also only translate and trace, but cannot interpret what it is actually looking at. Automatic digitizing may be more susceptible to miss digitizing areas in the map where the scanner messed up or had a smudge.

For most of your digitizing work, if it is sufficiently small in scope, manual digitizing will be the preferred method. However, if the digitizing work covers an extremely large area and the source has easily distinguishable features, automatic digitizing can provide a large time and money savings.

# Feature Generalization

## What is Feature Generalization?

We should first step back and remind ourselves that maps are simply abstractions of reality. That is, they are someone's viewpoint of how the world is organized. While we would like to store infinite detail about reality in geospatial data sets, it is simply not possible. Constraints on an infinite representation are the size of the map medium, insufficient resolution of sensors, finite disk space, scale at which we are representing reality, and so on. Additionally, you may not need to represent all aspects of reality for your mapping purposes.

Our inability to store infinite detail of reality and geospatial data sets means that we must generalize the features that we are storing. Feature generalization is the approximation and simplification of real features represented on a map. The main purpose of feature generalization is to aid in map legibility by reducing the complexity of the features being shown. What is most important to remember about all feature generalization, is that a properly generalized feature, should still maintain the same character of the original objects. That means, for example, if we reduce the number of curves in a stream, we should still be able to identify that feature as a simplified version of that stream.

## Four Methods of Feature Generalization

In order to discuss the four methods of feature generalization, let us first see what the source data was. The source data for this digitizing project was a digital orthophoto quadrangle quad. Houses could be easily identified on these photos, and the user wanted to digitize the footprints of the houses and roads. The original source data is shown on the left with the digitized house footprints laid on top. On the right, we see the digitized data without the source imagery behind it. We will now focus on only the digitized objects and the ways in which we can generalize those objects to aid in map legibility when we display this information on maps.


### Feature Generalization Methods

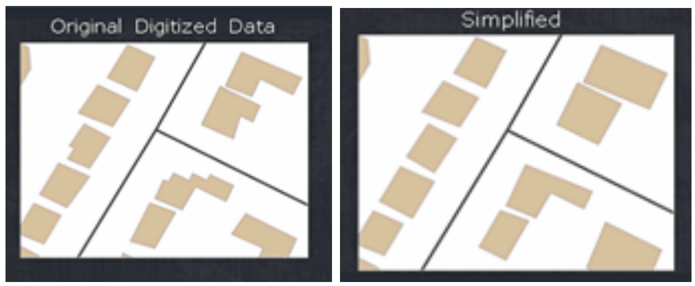
View the following examples of feature generalization methods: fused, simplification, displaced, omitted and exaggeration. It is important to reference, the original digitized data image in order to gain an understanding of how each method affects the image.

#### Fused

The process of fusing is to combine the geometry of adjacent features that are close together. In this example, we have fused together eight houses so that the fused data now only has nine building structures, instead of the original 13.

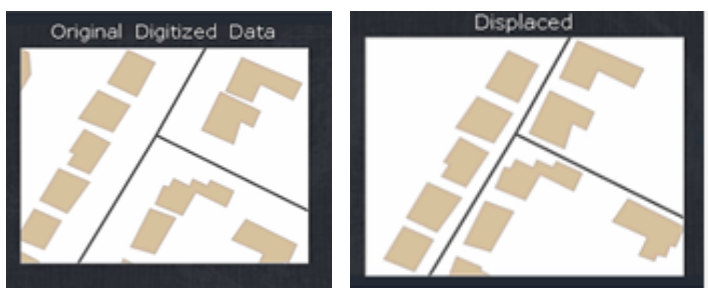
### Simplification

To simplify features, you remove vertices where the feature shows a reasonable amount of complexity.  For instance, looking at the house south of the intersection, we see a sort of stair-step pattern in its construction. If we remove the stair-step's, and square off that corner, we still maintain the character of that house, as the relative location and shape are still intact.



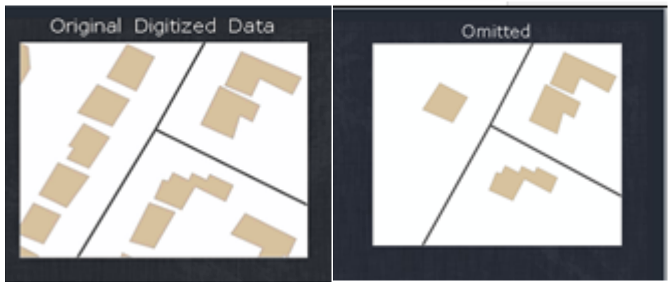

### Displaced

Here, the houses are moved closer to the streets. The houses could also have been moved to be further away from the streets. In either case, what is important, is that the relative positioning of all the features involved still be maintained.



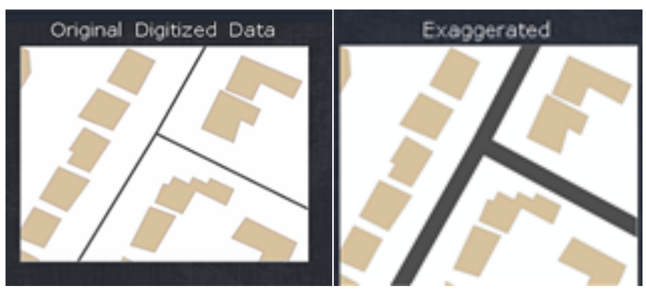

### Omitted

Here, features are simply removed from the data set. You may want to do this for lakes, for instance, where a large lake is surrounded by multiple small, insignificant lakes. If we remove the small insignificant lakes, keep the large lake, and zoom out, the map reader can still understand that there is a water body in that location.




### Exaggeration

Exaggeration is, perhaps, the most common method of feature generalization. For instance, on a roadmap for state, the roads are greatly exaggerated so that they can be seen on the map. If the roads were drawn to scale, they would probably be so thin, that the printer cannot print the map, and/or the reader could not read the thin line. Therefore, by exaggerating the road, the original characteristic of the road is still maintained, and a map reader can easily see the road features of importance on the map.




## Broad Directives for Generalization

Now that you are familiar with the methods for feature generalization, it is important to review some broad directives.

1. Observe the map purpose at all times.
   Do not include pointless details or features and vice versa. Additionally, the degree of simplification must be related to the purpose. For instance, if you are creating a map of a river for a kayaking club, you would not want to overly simplify the river, as all of the intricate bends and features of the river are what the kayakers really want to see.
2. Scale has a bearing on the need for simplification.

The scale that the map is being produced at should have a bearing on the need for simplification. If, at the current map scale, you can easily see all the features on the map, then there is no need for feature generalization. Again, you should maintain the essential character of the feature and area. Even if the features have been fused, simplified, displaced, omitted, or exaggerated, the map reader should still identify those features.

3. Be consistent in treatment.
   If you omit a large number of one type of feature layers on a map, but do not omit any of another type of feature layer on the map, the user may assume that the map layer with all of the omitted features is just as accurate as the detailed layer.

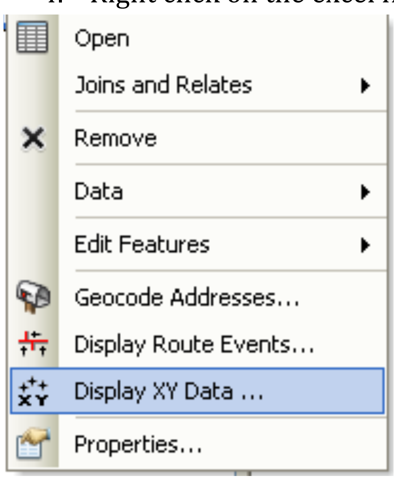
# Lab 6:
# Mapping GPS points, Creating and Editing Vector Shapefiles and Adding Attributes

**2013 Bethany Bradley**

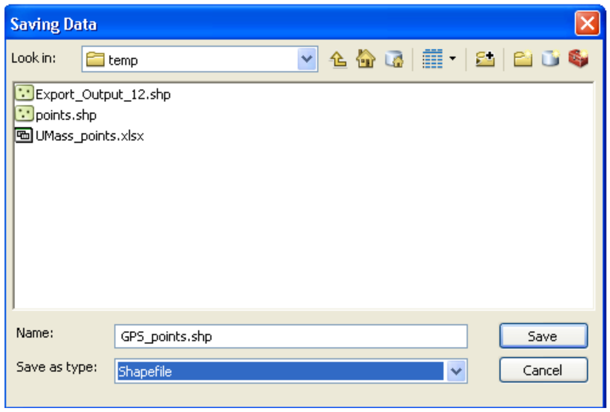


## A. Importing GPS coordinates as a point shapefile

1. Create an excel file with three columns called Lat, Lon, and Descript. Type in the coordinates for the locations you identified in your scavenger hunt. In the third column, type a description of the location.
2. **You will need to change your longitude coordinates to negative values to signify 'west' rather than east!
3. In ArcMap add the excel file to your data frame – you'll probably have to select 'sheet1$'.
4. Right click on the excel file in the data frame and select Display XY Data



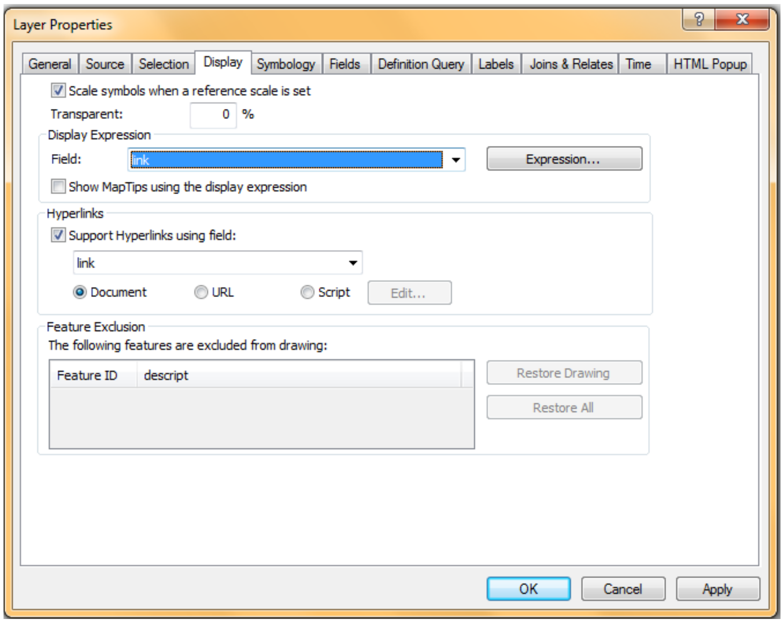

5. This will bring up a screen that allows you to specify LON as the X field and LAT as the Y field. Specify the coordinate system as GCS with the datum WGS 84.
6. **How do you know the coordinate system is GCS? How do you know the datum is WGS 84?**
7. Hit OK and a temporary shapefile will appear called sheet1$ events.
8. Save this shapefile as permanent by right clicking and under Data → Export Data. Click 'yes' to add the new shapefile to your data frame.



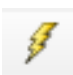

9. Add the grid UMASS.tif **Note – the data frame may default to an undefined coordinate system. Set the data frame coordinate system to be the same as UMASS.tif to solve this problem.
10. Are your point locations where you expect them to be relative to the UMass campus? If so, congratulations! That's all it takes to import GPS point locations that you collected into ArcGIS.

## B. Adding a hyperlink

1. Did you remember to take a picture at the location you thought had a nice view? If so, get that file and add it to your Lab 5 folder. If not, download a pic from the internet of a view you'd like to see from UMass.
2. Create a new column called 'link' in the attribute table of the UMass campus points, make sure the type is set to text.
3. Add the editor toolbar to document (if it isn't already there), and select **start editing** for the UMass campus points shapefile (don't worry about the error about different coordinate systems).
4. In the attribute table, type in the full path name (or URL) of the image you want to link to in the 'link' column. For example, E:\Bethany\Lab5\view.jpeg
5. In the editor toolbar, save your edits and stop editing
6. Turn on hyperlinks in shapefile properties > Display tab, and select 'link' as the field supporting hyperlinks.




7. Now, in the regular tools menu, you should see a new highlighted symbol of a lightning bolt

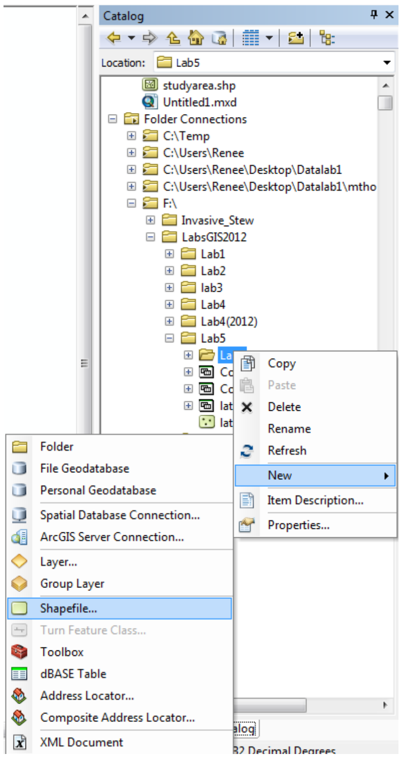


8. Select that symbol, and any points with hyperlinks will be highlighted on your map, click them to bring up your picture.

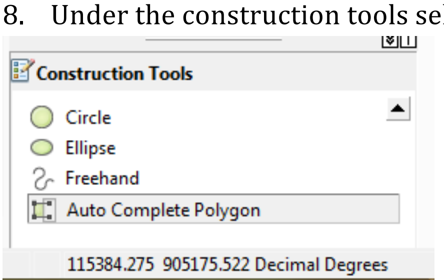
9. Hyperlinks in ArcGIS are a nice way to embed photos, videos, or other information into a more dynamic map.

## C. Creating a New Shapefile

1. Often times you'll want to create a new shapefile and draw your own features.
2. Open ArcCatalog and navigate to your Lab 5 directory.
3. Create a new shapefile, right click in catalog window → New → Shapefile

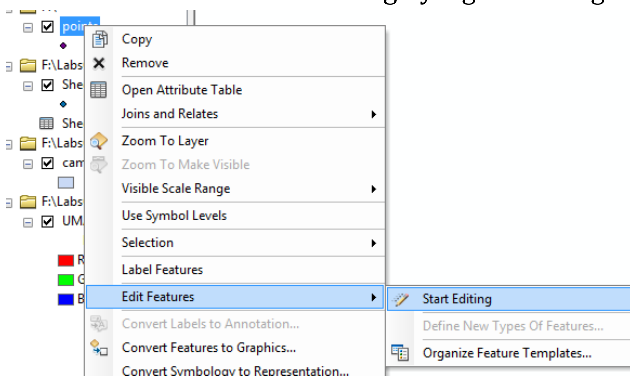



4. Create a new **polygon** shapefile and name it campus_buildings. Set the projection be the same as UMASS.tif (Mass State Plane)
5. Add campus_buildings to ArcMap. Note that the shapefile is empty. Nothing appears because you haven't yet created any polygons!
6. Zoom in on Holdsworth on UMASS.tif.
7. In order to enable the tools on this toolbar, click on the Editor drop-down menu and choose **Start Editing**. Select campus_buildings to edit.
8. Under the construction tools select polygon.

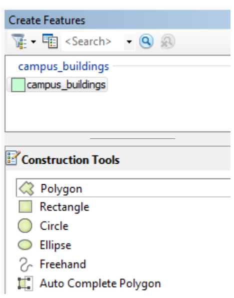


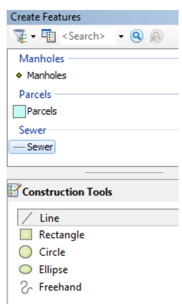

9. Create a shapefile of Holdsworth by clicking once to create a new point at each corner, and 2x clicking to complete the polygon.
10. Save your edits.

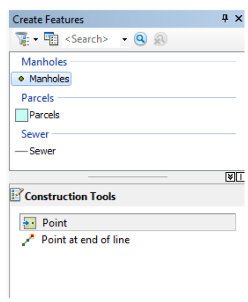
11. You can change the target to your gps points shapefile, under the create features box, and add new points as well. Since you're in editing mode, you can also select points (using the black arrow in the editor toolbar), move them around, and delete them.
12. Note - You can also start editing by right clicking on the layer -> edit features-> start editing




13. Notice depending on which layer type you are clicked on in the create features box, the construction tools change to be the same type of shapes. Select your polygon layer to create polygons or rectangles, your line layer to create lines, and points layer for points.




13. Bring in your UMASS.tiff file and play with these different construction tools. Create a point over the pond. Draw a polygon around the pond, first using the polygon tool and then using the freehand tool. Which tool works best for this task?

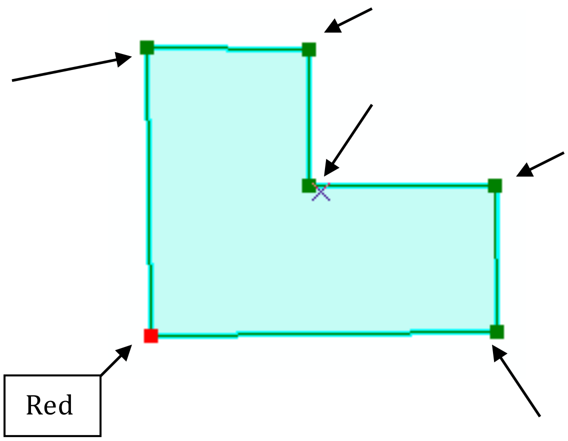
## D. Editing an existing shapefile

1. Copy the file EdDataArcMap.exe from the datalab5 folder to your own folder.
2. 2x click this file to extract the data. This will create another folder called EdData in your folder.
3. Open the Spatial_Editing.mxd file that is located in this folder.
4. The project contains three spatial data files in shapefile format from a municipal GIS, including parcels, sewer lines, and sewer manholes.
5. When you start an edit session in ArcMap, every feature in the target layer (every polygon, line or point) is represented by a **sketch**. A sketch consists of points in a points shapefile, or vertices and segments for a line or polygon shapefile.
6. When you edit a feature in ArcMap, you are working with its sketch, not the original features. Any changes you make are not saved until you choose to save them via Editor -> Save Edits or Editor -> Stop Editing.
7. See how this works by the following:

a. Click the Edit Button in the Editor toolbar

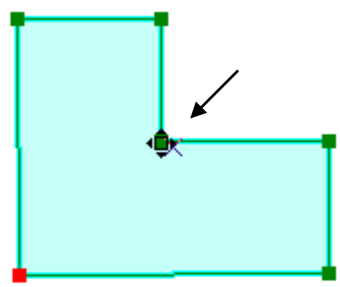
b. Then move your cursor into the display area and 2x click the L-shaped parcel in the display.

c. The feature's sketch is now activated and its vertices revealed.

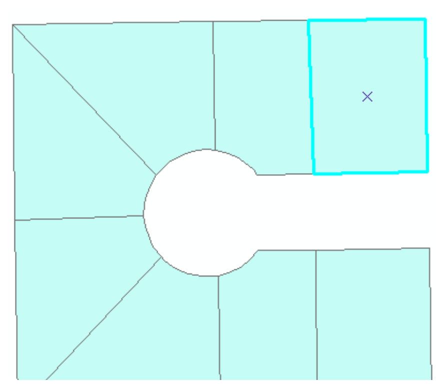



In your display, the lower left vertex will be red, while all the others will be green. The red one indicates that last vertex which was placed to close this polygon when it was originally digitized.

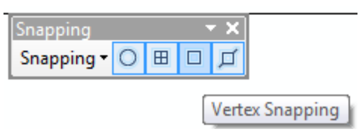
8. Now, place your cursor directly over the vertex located on the inside corner of the L, as shown.



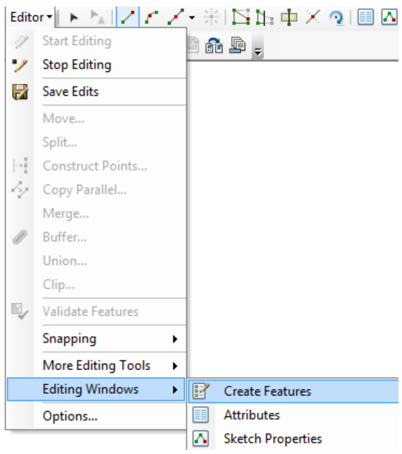
9. Right-click and from the popup menu that appears choose Delete Vertex.
10. Notice the **sketch** reflects your change, but the underlying feature is still unchanged. The feature's geometry will not permanently change until you finish the sketch and save your changes.
11. Right click once more and from the popup menu select Finish Sketch to make the change. It is still not permanent yet because you haven't saved it.
12. You can undo this change under edit > undo (or CTRL Z)
13. You can also move a vertex by clicking and dragging it around.
14. If you need the shape to be less rectangular, you can add new vertices. Right-click anywhere on the green sketch of the polygon feature and from the popup menu choose Insert Vertex.
15. By moving, inserting, deleting vertices in this manner you can modify a polygon to take on any shape that is needed to adequately represent the geographic feature in question (in this case it is a house lot, but it could be a lake, a town boundary, a building footprint, etc.)



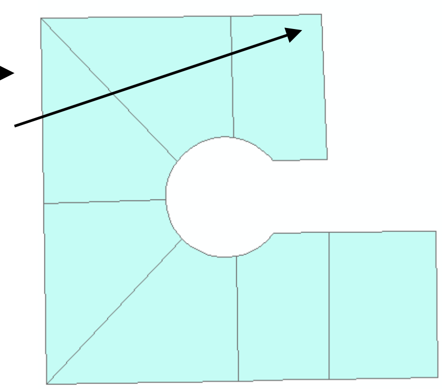
16. You can move and rotate features as well:
    a. Use the edit tool and click once to select the parcel in the northeast corner
    b. Now simply hold the mouse button down and drag the parcel to the right to move it.
    c. You can move it anywhere you want in the display
    d. Notice the small "x" in the center. This is called the "selection anchor" and this is the basis for a rotation.
    e. On the Editor toolbar, click the Rotate button in the toolbar 
    f. Move your cursor back onto the parcel you have selected and rotate it!
    g. You can change the "focus" of the rotation by moving the selection anchor
    h. With the rotation tool still active, hold the cursor directly over the anchor. After a moment, the tool will change in appearance. When it does, you can drag the anchor anywhere it is needed so that when you go back to the rotate tool, the feature will rotate using the new location of the anchor as the pivot point
17. To delete a feature, select it and hit the delete key.
18. **Snapping** is a process in which GIS automatically sets nearby points to the same coordinates, so that polygon or line edges and vertices join perfectly.
    a. From the Editor menu, choose Snapping.
    b. A snapping tool bar will appear within the ArcMap session




    c. Select vertex snapping – the square symbol
    d. Select polygon under construction tools. Click the vertices between two polygon shapes, 2x click to complete, and a polygon will be "snapped" in between the two polygons
    e. Next, from the snapping tool bar , choose **Options**.
    f. In the popup window, find the Snapping Tolerance text box, enter 100**.**
    g. Based on the snapping environment and snapping tolerance you just set; new line vertices will automatically snap to the vertices in the Parcels layer if they are placed within 100 pixels of the existing ones.
    h. To create a new feature, under the editor drop down -> editing windows -> create features.




i. Move the cursor into the display area. It will look like a big plus sign.
j. Drag the cursor around without clicking. Note how as soon as you get within 100 pixels of a vertex on an existing parcel that the blue circle "jumps" (snaps) to that vertex.
k. Once your plus sign has snapped to an existing vertex, click once with the mouse to secure the intial vertex
l. Then proceed to digitize the other corners of the parcel (these can go anywhere you wish in this exercise)
m. After you make the last vertex, right click and select "Finish Sketch"
n. Voila, you have a new polygon shape record representing a parcel!
o. Note that you can decrease the snapping tolerance if you need to create any parcel sides that are less than 100 pixels from a vertex
p. If you make a mistake, you can
    i. Right click -> Delete Sketch
    ii. Right click -> Finish Sketch -> Edit -> Undo Create
    iii. Stop Editing -> No to save changes
q. You can also create a feature with specific dimensions:
    i. Delete any parcels you just made so that you have this configuration once again



    ii. Click once to create a vertex for a new parcel at the upper right corner
    iii. Drag the cursor to the right but before you click again to create another vertex, right click to obtain the popup menu
    iv. In the popup menu choose Direction/Length
    v. For direction enter 0; for length enter 92. Hit the return key on the keyboard.
    vi. A new vertex will be placed exactly 92 feet from the first vertex and at the directly east

vii. For the next vertex, again right click to bring up the popup menu and Direction/Length. Enter 270 for the direction (directly south) and 120 for the length. Hit Return

viii. Allow the cursor to snap to the last vertex, and then right click -> Finish Sketch.

ix. In this manner you can be quite precise about the dimensions/directions of features you create.

r. It is also possible to snap features in one layer to features in another layer (as long as the features are in the same coordinate system!). In this step you will add a new sewer manhole at the eastern end of the existing sewer line.

i. From the Bookmarks menu click Municipal.

ii. In the Snapping tool bar select End snapping (the four small squares symbol), this is used for lines so that snapping only occurs at the ends of the line, not somewhere in the middle of the line.

iii. Under create features select manholes, and point under construction tools

iv. Place your cursor over the eastern end of the existing sewer line (it will snap there!), and click once to add the new manhole.

19. During the edit session you can add attribute information as you go

a. For now, first stop editing (Editor -> Stop Editing)

b. Don't bother to save any changes at this point

c. Right click your Parcels Layer and open it's attribute table

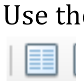
d. This table is currently missing a lot of information that we might need

e. Add a field called Owner

i. Make this a text field that is 50 characters in length

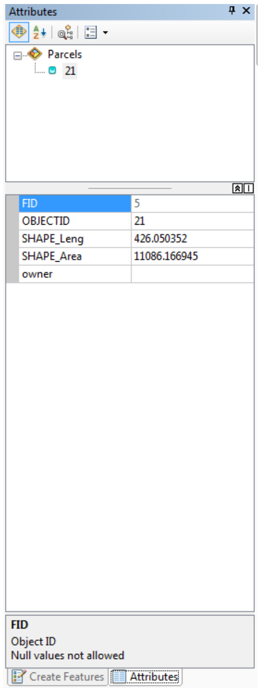
ii. Close the table and restart an editing session

iii. Use the Edit Tool to select a parcel



iv. From the Editor Toolbar, choose the Attributes icon

v. Once you do this, a popup window will appear with the information for the selected parcel




vi. Click to the right of the attribute Owner and you will be able to enter an owner's name

vii. In this manner you can add a feature, and edit the tabular data simultaneously

**Exercise: Digitizing Information on the UMass Campus**

A new student who is unfamiliar with the campus wants to take a tour beginning at **Orchard Hill**. She would like to visit the following locations:

1. **Whitmore administration building**
2. **The integrated science building**
3. **Holdsworth**
4. **The recreation center**

And, she would like to find a **nice spot to eat lunch** along the way. Since you're guiding her tour, also take her past a location with a **nice view**, and a **labeled tree** (she's into the environment).

Create points of the different locations that she will visit, and create a line file showing a recommended route. Make sure that you guide her along walkable paths (i.e., the route may not be a straight line).

Hand in:

A map of the UMass campus showing the 8 locations above (in bold). All of the locations should be labeled. The map should also show a recommended walking route around campus that includes those 8 points. Include a picture of the nice view so that the student will be able to orient herself (if you didn't take a picture, download a picture from the internet of a view you wish we had from UMass). Be sure to include a scale bar and north arrow on your map.

Also include answers to the following questions:

1. What is the total length the new student will walk to get to all these locations (in m)?

2. What is the area of the campus pond (in $m^2$)?

**Grading rubric:**

1 pt for: each of the 8 labeled locations, North arrow, and scale bar. (10 points total)

2 pts for: campus photo as background, walking trail, nice view picture, question 1, and question 2. (10 points total)

# Lab 7:
# Spatial Analyst – working with Raster Grids



## Introduction

The Spatial Analyst extension to ArcMap allows you to work with *raster data* as opposed to vector data. The raster data format (known as a GRID in ArcGIS nomenclature) represents spatially explicit continuous surfaces. Examples include: digital elevation models (DEMs), aerial photography, bathymetry, remotely sensed imagery, and countless others. Any given point, or "pixel" in a raster dataset has a defined value, thus they are spatially continuous. Pixels are not assigned attributes like vector files. Instead, they have a single measurable value at any given location.

Each *pixel* in a Grid contains a single value. This number could be an actual measured value (e.g., elevation in meters), or it could be a unique code (e.g., a numeric code of 110 could be designated to mean the pixel is silty loam). Non-numeric cell values are also possible, so instead of 110, there could be a text string "silty loam" for a pixel.

The *cell size* is an important piece of information about any raster grid. Cell size tells you how much land area a single pixel represents. For example, a 30 m pixel represents the average value within a 30 m x 30 m (900 $m^2$) area.

Like vector data, raster grids can be queried and overlaid with other raster or vector layers to analyze their spatial relationships. However, since the "attribute" is a numeric value, the analysis commands involve **mathematical statements** that can be very simple (e.g., elevation > 20) to more complex depending on the application.

## Types of rater grids

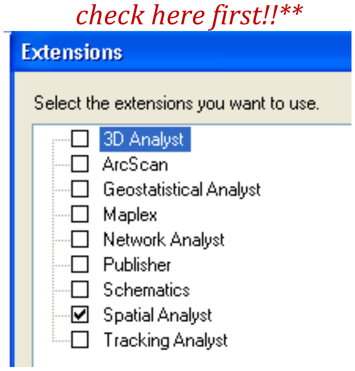
Raster grids are typically either *discrete* or *continuous*. **Discrete** grids have a limited number of values and are usually derived from other spatial data. For example, habitat quality could be represented as a discrete grid (1 = high, 2 = medium, 3 = low). **Continuous** grids contain data that could (theoretically) have infinite values. For example, topography on Earth could be represented as a continuous grid from 0 (sea level) to 29,000 ft (Everest). You can always reclassify a continuous grid to create a discrete grid (e.g., elevations less than 1000 ft = 1, elevations greater than 1000 ft = 0 would define all low elevation areas as 1).

Pixel values (in either discrete or continuous grids) can range from simple binary to complex text strings. The more complex your data values, the more memory the file will take to store and process.

| | Data Format | Example Values | Use |
|---|---|---|---|
| | Binary | 0, 1 | Represents true/false; discrete classes |
| More memory intensive ↓ | Integer | 1, 2, 3, 4, 178 | Could be discrete (classes) or continuous (values) |
| | Float | 2.9, 3.14, 1111.111 | Usually continuous values |
| | Text | Flat, steep | Represents categories; discrete classes |

## A. The Basics of Working with Rasters (aka Grids)

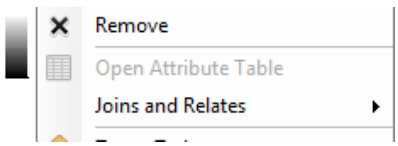
1) Check to make sure that the Spatial Analyst Extension is enabled via Customize -> Extensions. If there is a check mark next to Spatial Analyst it is already enabled; if there is not, select the spatial analyst box. ***If you ever get an error message when using a tool in spatial analyst, check here first!!***




2) From your lab 6 directory, add the grid **dem250k** to your ArcMap document.
3) Zoom in to a small enough area so that you can see that the grid is comprised of individual pixels.
4) Using the Identify tool, click on a single pixel. The attribute *Pixel value* contains the measured value within that pixel. In this case, it represents elevation in meters.

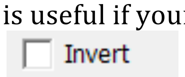
5) **Quick Exercise: Is this grid continuous or discrete? What is the data format (binary, integer, floating point, text)?**

6) Close the identify tool and right click on the file in the layers window. Notice that the 'open attribute table' is grayed out. Because the raster grid is continuous, there are too many unique values for ArcGIS to build an attribute table.

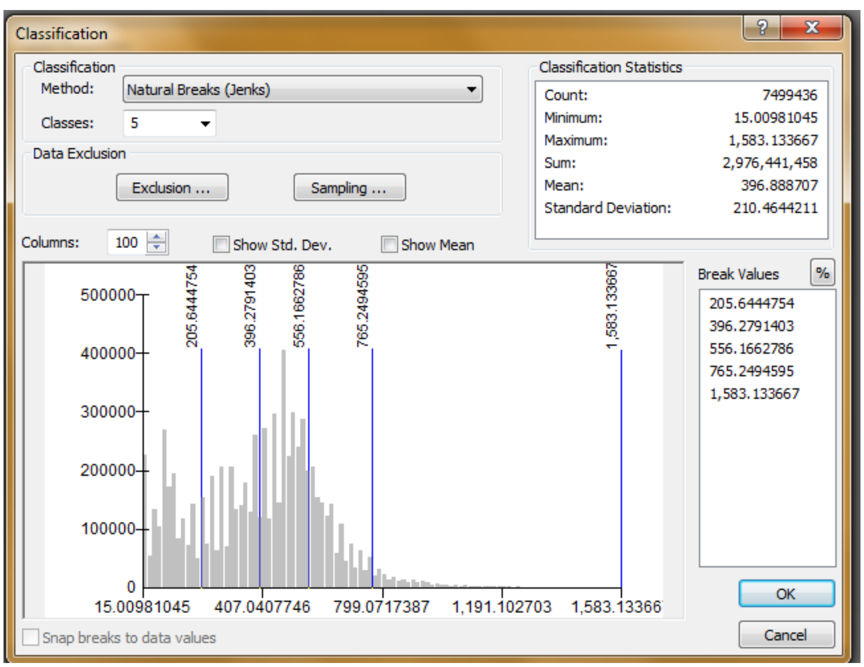



7) Select the layer properties for the grid and navigate to the symbology tab.

8) The default color scheme is stretched grayscale. Change the color ramp to something else to make it more colorful. Note that there is also an invert button to invert the color ramp. This is useful if your data are better represented by warm to cool rather than cool to warm.



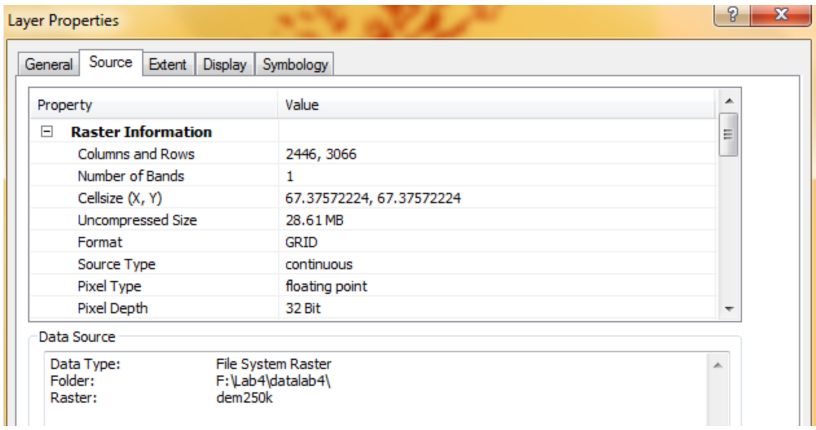

9) Click on classified instead of stretched. A window will pop up asking if you want to build histograms – click 'yes', building histograms is necessary to define a classified stretch. The default is 5 categories, but you can select anything from 1 to 32. Click the '**classify'** button to edit the ranges. You can change the ranges either by typing in values to the right, or by dragging the blue bars in the histogram window.




10) Try changing the classification method to 'Quantile'. Click okay and apply this to the map. Although the raster grid is *continuous*, you have now displayed it as if it were 5 *discrete* elevation bands.

11) **Quick Exercise: An ecologist is interested in visualizing habitat for the Pine Marten in the Adirondacks. She thinks this animal prefers elevations between 500-600 m. Change the classified color scheme to emphasize this elevation range.**

12) Before we leave layer properties, let's get some more information on this grid by clicking the Source tab. In the source tab, you'll find some important information about this grid. For example, at the top you can see the columns and rows in the grid (2446 pixels x 3066 pixels), as well as the pixel size (67.37). If you scroll down, you can find the projection of the grid (under spatial reference) and some statistics.

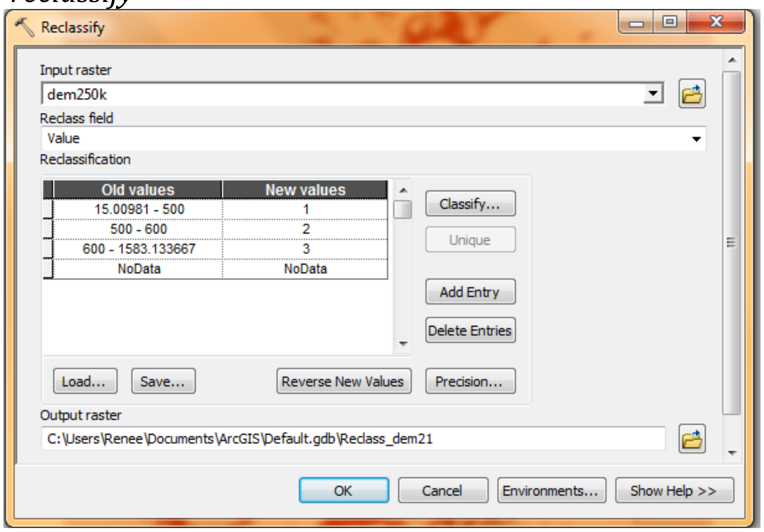



13) **Quick Question: What is the unit of the 67.37 pixel size (inches, feet, meters, km, degrees??)? How do you know?**

## B. Reclassifying a grid

1) In the exercise above, we changed the color scheme of the grid to visualize potential habitat for the Pine Marten. This is useful for producing a reference map (if all we were interested in were elevation). But, if we were constructing a habitat model, we'd need to create a discrete classification of elevation related to habitat preference. In this case, we want to create a new raster grid where "suitable habitat" (between 500-600 m) = 1, and "unsuitable habitat" (everything else) = 0. Recall that in raster processing, we often use 1 = true and 0 = false.

2) In ArcToolbox, select “Spatial Analyst Tools” > “Reclass” > “Reclassify” (double click “Reclassify”). ***Note: If you forget where the reclassify tool is located, you can always search for it in the search for it in the search tool. But, you have to know that it is called ‘reclassify’*




3) Two things to notice. First, the ‘old’ values that ArcGIS chooses match what you’ve defined previously in the symbology tab. If you haven’t defined anything, ArcGIS will use an arbitrary default. You can edit these old value ranges by clicking on the classify button. This will bring up the same screen you saw previously in the symbology tab.
4) Second, the default output raster of the reclassify tool is not likely where you want it saved. To ensure that you know where to find the reclassified file you must select the folder icon and save it to your flash drive in the lab 6 folder, name it Reclass_dem250.
5) For this example, create a grid with values of ‘1’ (= ‘true’) for elevation ranges between 500-600 and values of ‘0’ (= ‘false’) for all other elevation ranges.

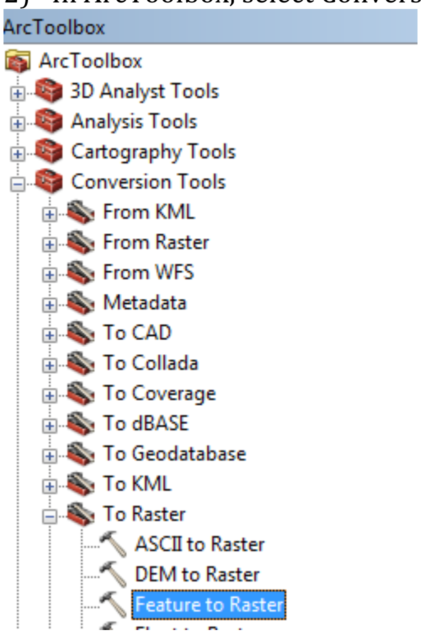
6) The grid will appear in your layer window. Use the identify tool to click around the grid, and notice that all the values are either 1 or 0 (as you defined in the reclassify tool). Now that you have a discrete grid rather than a continuous grid, you can open the attribute table.

7) **Quick Question: How many pixels are between 500-600m? Now that you know both the number of pixels and the pixel size, what is the total land area between 500-600m?**

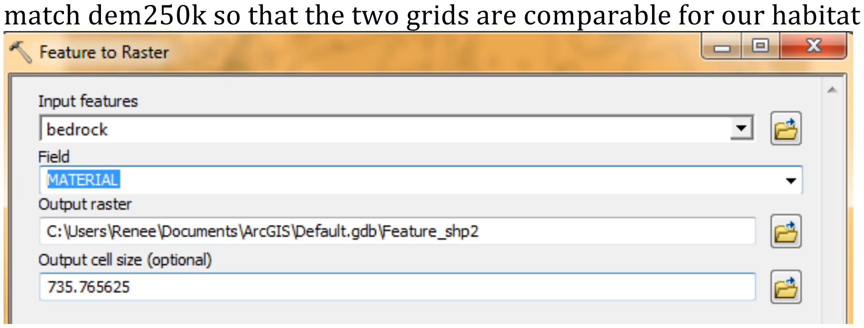
## C. Converting a polygon shapefile to a grid

1) Add the layer bedrock.shp to your ArcMap document. Open the attribute table and notice that the geologic units are recorded under the heading 'Material'. Let's continue with our example of Pine Marten habitat. The animal prefers to forage in soils derived from Mu and MuG geologic parent material. We need to find a way to get that information from polygon to grid format so we can add it to our existing topographic analysis.

2) In ArcToolbox, select Conversion Tools → To Raster → Feature to Raster.

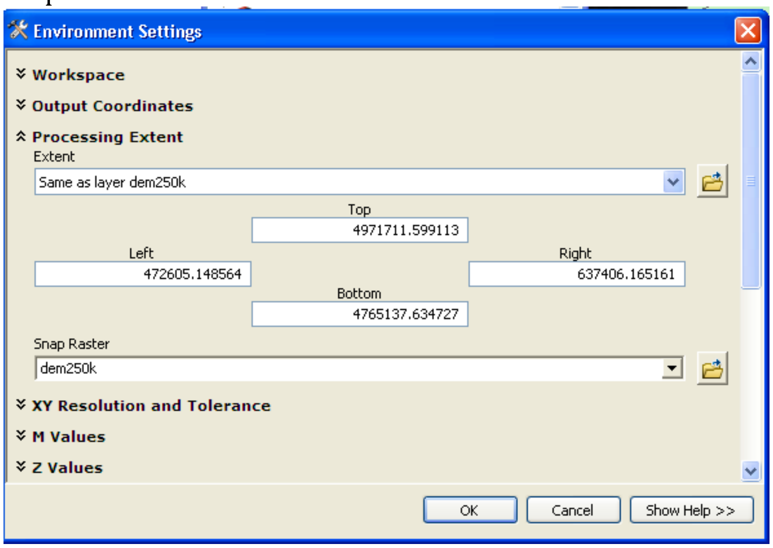



3) Select bedrock from the Input features dropdown, and Material as the field we want to convert into a raster grid. Name the output raster "bedrock1" in your lab6 folder.
4) **BEFORE YOU CLICK OK**, Notice that the output cell size is an arbitrary number that ArcGIS has pulled out of thin air. Any time you are working with raster grids, it is a good idea to make sure they're all the same cell size and extents. In this case, you want your new bedrock grid to match dem250k so that the two grids are comparable for our habitat analysis.

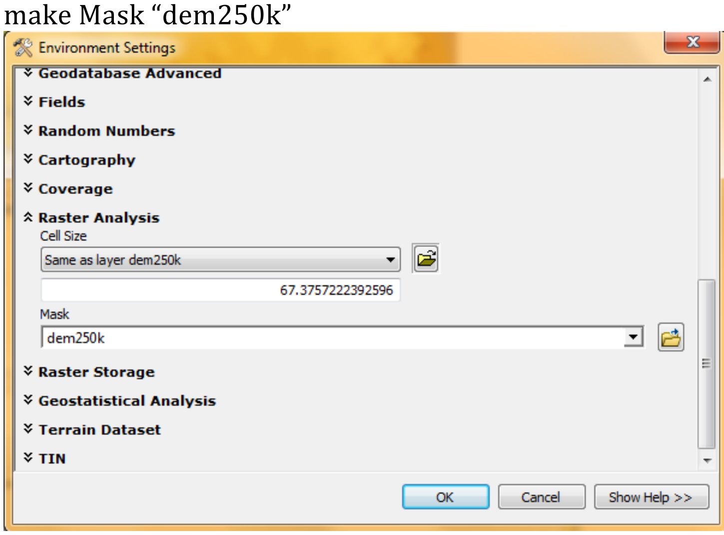



5) Fix the cell size by clicking 'Environments' at the bottom of the feature to raster tool (note, Environments is an option in all spatial analyst tools).

6) Select "Processing Extent", under Extent choose "same as layer dem250k" and "dem250k" for Snap Raster.

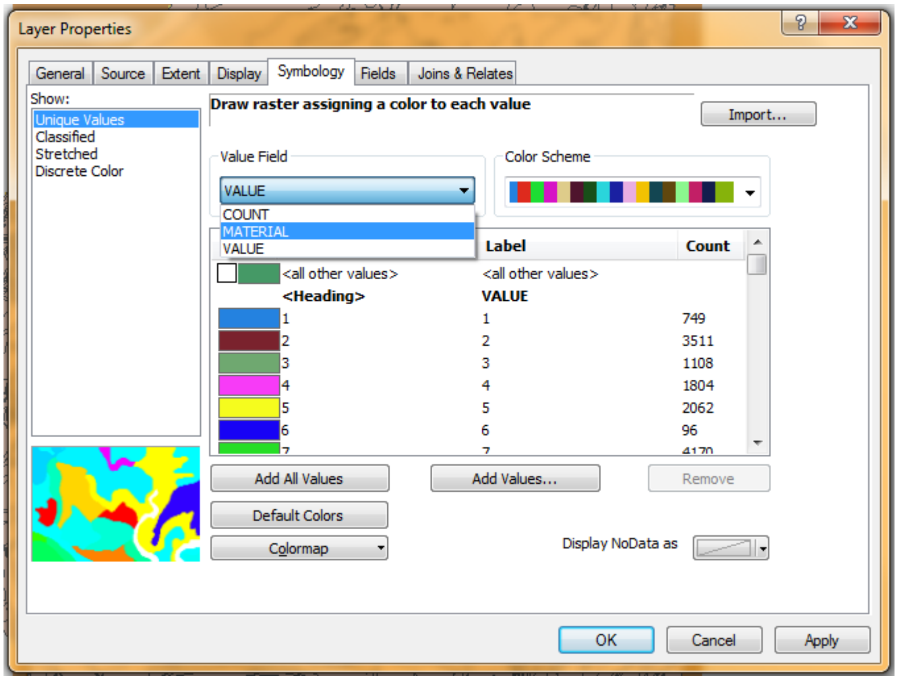



7) Next, scroll down and select Raster Analysis, make Cell Size "same as layer dem250k" and make Mask "dem250k"



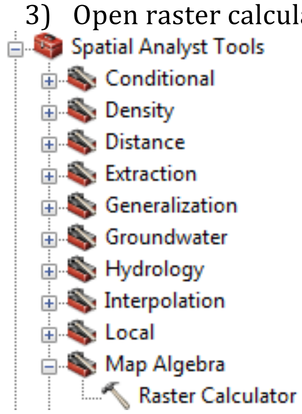

8) Now, the output will be exactly the same pixel size and extents as dem250k. Click OK on environments and OK to execute feature to raster.

9) 'bedrock1' will appear in your layer window and the default symbology will be a set of numbers. Open the attribute table to figure out what these numbers are – they correspond to Material, but grouping them 1-7, 8-14 etc. doesn't make much sense here. Instead, what we want to display is each geological unit uniquely based on parent material. Do this in the symbology tab by clicking on 'unique values' and changing the value field to 'Material'




## D. Raster Calculator

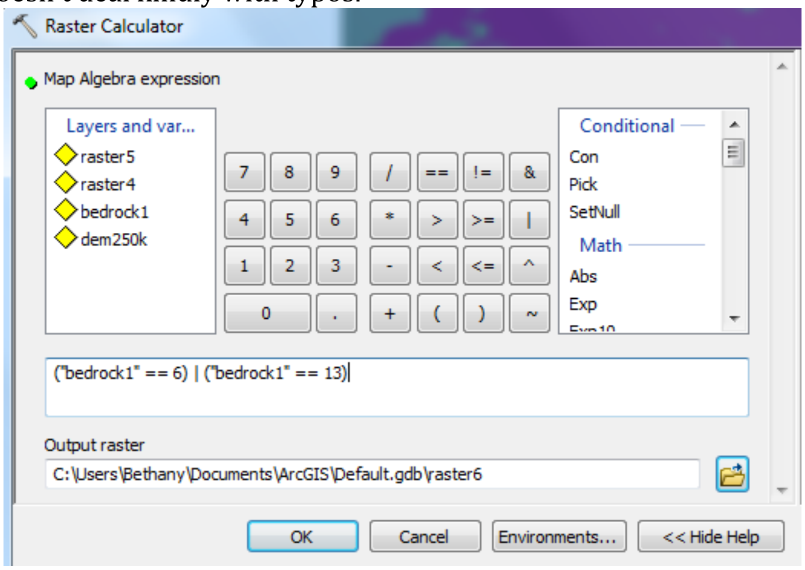
1) Back to our original goal: to classify Mu and MuG parent material as suitable habitat. There are multiple ways to do this through spatial analyst. The first you've learned already. You could use the Reclassify tool to set every material to '0' except for Mu and MuG. But, like just about anything in GIS, there are multiple ways to skin a cat. Another tool that you will learn to love in spatial analyst is Raster Calculator (**See page 9 of this lab for more info on Raster Calculator and where to look to find help**).
2) Open the attribute table for bedrock1 and notice that the grid has assigned a Value for all of the parent materials, and the values for Mu and MuG are 6 and 13. Those values are important to notice because raster calculator doesn't deal with text.
3) Open raster calculator. ArcToolbox > Spatial Analyst Tools > Map Algebra > Raster Calculator




4) Our aim is to create a grid where 1 = true and 0 = false. 1 (true) should be assigned to any pixel with the Value of 6 OR 13. In Raster calculator, this statement is written as:

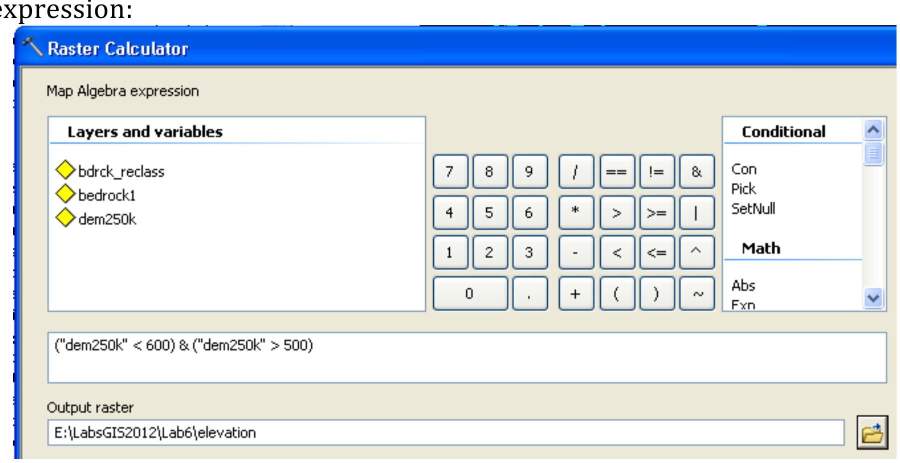
('bedrock1'== 6) | ('bedrock1 == 13)

5) Remember, it's usually better to click the operators rather than type them yourself – ArcMap doesn't deal kindly with typos.




6) Raster calculator will output a calculation, which has values '6' and '13' = 1 and all other values = 0. Notice that the names of the parent material have been lost from the attribute table. If you wanted to preserve this information, you'd have to add them back in manually in a new field.

7) Now, let's use Raster calculator to recreate your classification of areas between 500-600 m elevation. Recall that previously we did this using reclassify. Both approaches will produce the same result. Open Raster calculator and use the buttons to input or type in the following expression:

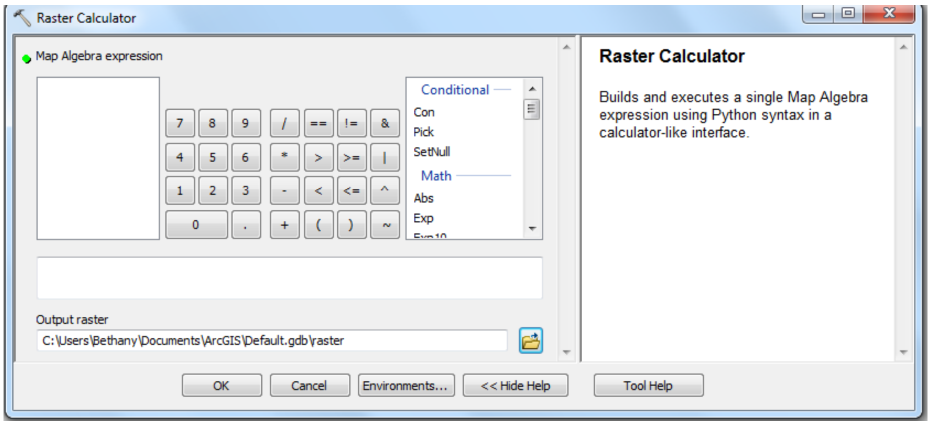



8) Make sure the Output raster is your flash drive, name this file suitable_elev

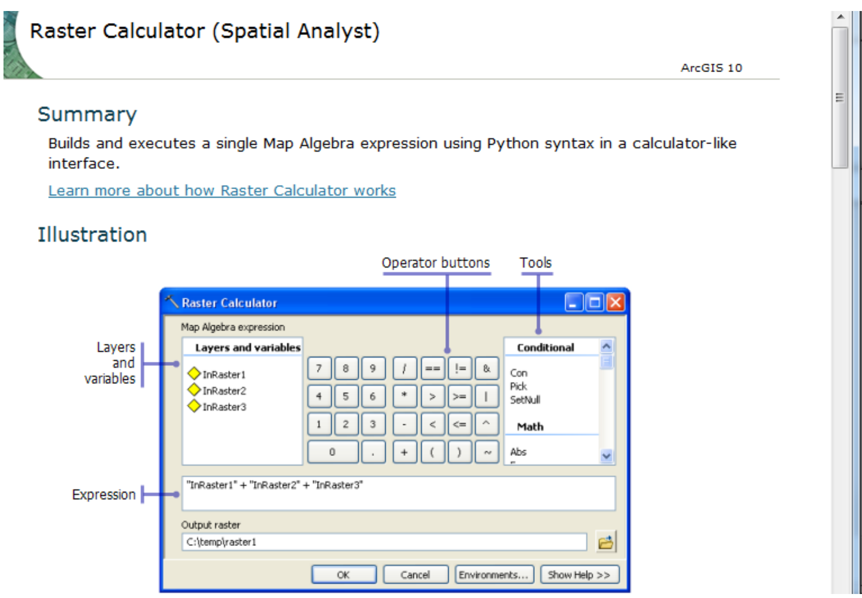
9) This expression tells ArcGIS that we want it to produce a grid that is "true" (value = 1) where dem250k is less than 600 AND dem250k is greater than 500. All other values will be assigned "false" (value=0). Click OK to execute the command. Now, you've recreated the classification of elevations between 500-600. This should be exactly the same as the grid you created previously using reclassify.

## How to find more help with Raster Calculator

If you run into trouble with raster calculator expressions (trouble is raster calculator's middle name), expand "show help" and click the "tool help" button. This is a rare occasion where ArcMap's help menu is actually somewhat helpful.




Below is what you get when you click "tool help" – if you scroll through this help page, you'll get some useful information…



For example, here's some info on those weird operator buttons and what they actually mean…

he operators in the Raster Calculator tool dialog box are:

| | | | |
|---|---|---|---|
| / (Division) | == (Equal To) | != (Not Equal) | & (Boolean And) |
| * (Multiplication) | > (Greater Than) | >= (Greater Than or Equal to) | \| (Boolean Or) |
| - (Subtraction) (Negate) | < (Less Than) | <= (Less Than or Equal to) | ^ (Boolean XOr) |
| + (Addition) | | | ~ (Boolean Not) |

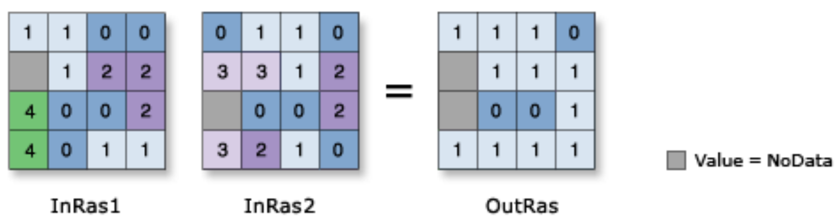
*Spatial Analyst operators*

Click on the | (Boolean or) symbol to get more information about what | means in raster calculator:

## Summary

Performs a Boolean Or operation on the cell values of two input rasters.

## Illustration



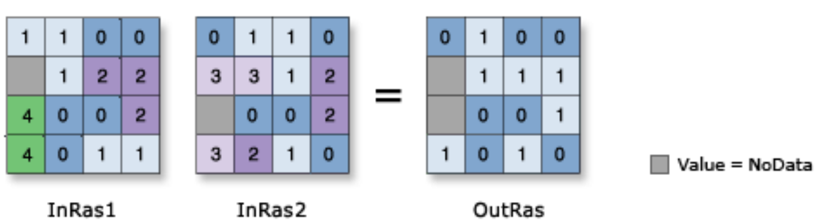

*OutRas = Raster("InRas1") | Raster("InRas2")*

Now click on the & (Boolean and) symbol to get more information about what | means in raster calculator:

## Summary

Performs a Boolean And operation on the cell values of two input rasters.

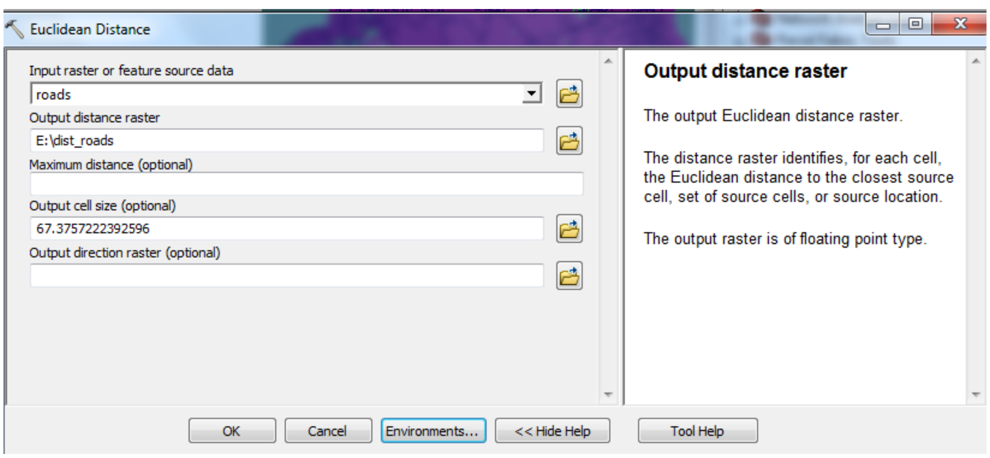
## Illustration



Compare these two images to see how the | (OR) operation differs from the & (AND) operation. In the | (OR) illustration, you can see that if there is a value in InRas1 **or** InRas2 the output is a 1, and the rest are 0. In the & (AND) illustration, you see there needs to be a value in InRas1 **and** InRas2 for the output to be 1, otherwise the output is 0. Notice also that all NoData values stay as NoData in the output regardless of the operation you perform.

## E. Creating Distance Grids

1) One final thing that's important for Pine Marten habitat is being far away from cars, and therefore far away from roads. The safest habitat occurs greater than 1 km (1000 m) from a road, while areas less than 1 km from a road are dangerous and therefore poor habitat.
2) Add the line shapefile roads.shp to your ArcMap document.
3) Under spatial analyst tools select distance -> Euclidean distance. Select roads under Input raster and save to your lab6 folder.
4) **BEFORE YOU CLICK OK**, make sure that your environments are set to be the same extents and cell size as dem250k

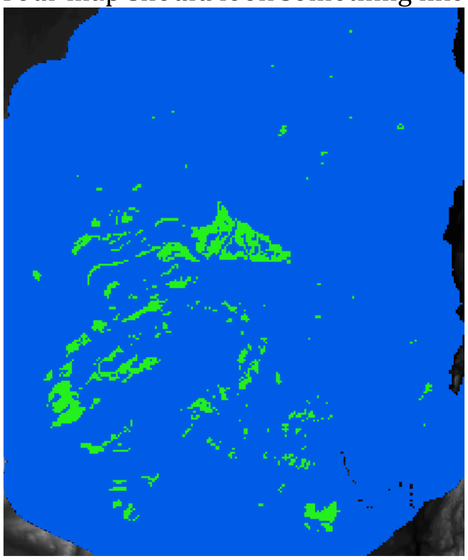


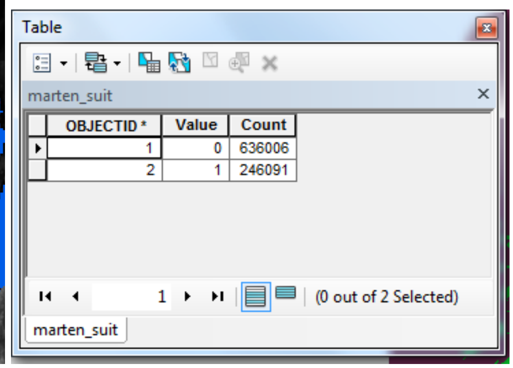

5) Click OK to create a new *continuous* grid of distance to any road. Just like dem250k, you can't view the attribute table because there are too many values for ArcGIS to calculate.
6) Use *either* the reclassify tool or raster calculator to reclassify all lands greater than 1 km from roads as "1" (meaning best habitat), and lands less than 1 km from roads as "0" (meaning poor habitat).
7) Now you have three layers of Pine Marten habitat suitability: Suitable soils (1 = yes), Suitable elevation (1 = yes), and suitable distance from roads (1 = yes). You want to combine these layers to create a single grid of habitat suitability.
8) You've just run into a research question. Does the Pine Marten require all three of those suitability criteria to be met? Or, is it okay for the Pine Marten if any of those conditions are met (e.g., either 500-600 m OR Mu OR MuG bedrock).
9) Let's start by assuming all three conditions must be met. This is known as a conditional "AND" statement. A & B & C works just like the 'intersect' tool does for shapefiles. It identifies

areas where all three overlap. Open Raster Calculator and click to create the following statement:

“Suitable distance from roads” & “Suitable elevation” & “Suitable soils”

10) Your map should look something like this:







| OBJECTID * | Value | Count |
|---|---|---|
| 1 | 0 | 636006 |
| 2 | 1 | 246091 |



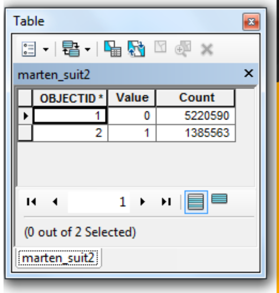
11) There are a total of 246091 pixels on suitable soils, within the suitable elevation range, and more than 1 km from a road.

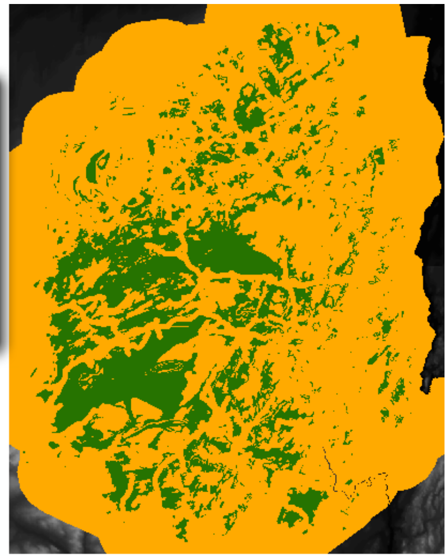
12) Now, let’s assume that we want to define suitable habitat as land that must be greater than 1 km from a road, but can be either on suitable soils OR within the suitable elevation range.

13) Now use raster calculator to identify areas that are suitable distance to road and either suitable soils or suitable elevation range. In raster calculator, you have to be careful with the expression and where you put parentheses. Here’s what the expression should look like:

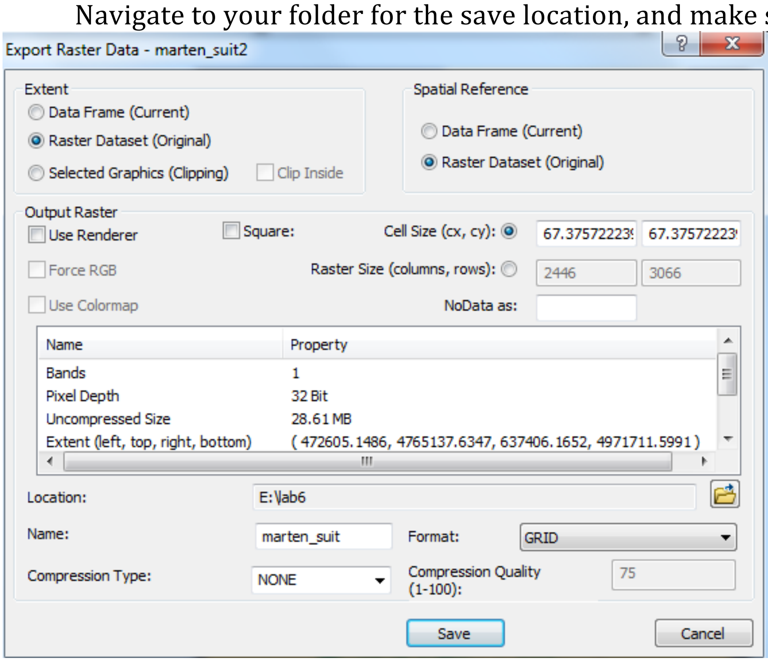
“Suitable distance to roads” & (“Suitable elevation” | “Suitable soils”)

14) Using this definition of suitable habitat for the Pine Marten, we get a much larger land area. There are a total of 1385563 pixels that are more than 1 km from a road and on *either* suitable soil *or* at suitable elevation. How you select your criteria for modeling suitability really affects your end result.



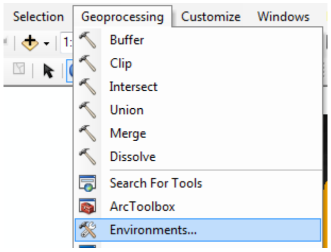
15) We've now got a model result that we like. But, the calculation is probably saved in a temporary file or named something inscrutable like raster8. To rename the grid and save it in your lab6 folder, export a copy by right clicking the file and selecting Data > Export Data. Navigate to your folder for the save location, and make sure the format is GRID.




## Setting Environments for all Raster Processing

1. In the exercises above, you had to twice set the Environments tab in raster calculator so that all of your output grids were the same pixel size and extent. That's okay if you do it once or twice. But, if you're doing a lot of raster processing, repeating that step is annoying.
2. Fortunately, you can save Environments for the whole ArcMap document, so every time you open that document and do some raster processing the output grids will be the same.

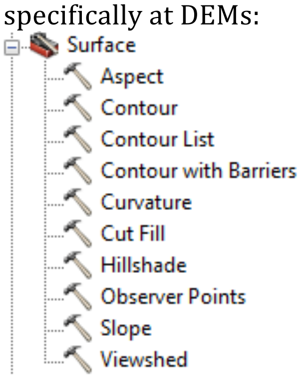
3. Find the environments tab under Geoprocessing > Environments. Notice that the options are identical to what you found in raster calculator and feature to raster. Choose the same options here (Processing extent and Raster analysis set to match dem250k).



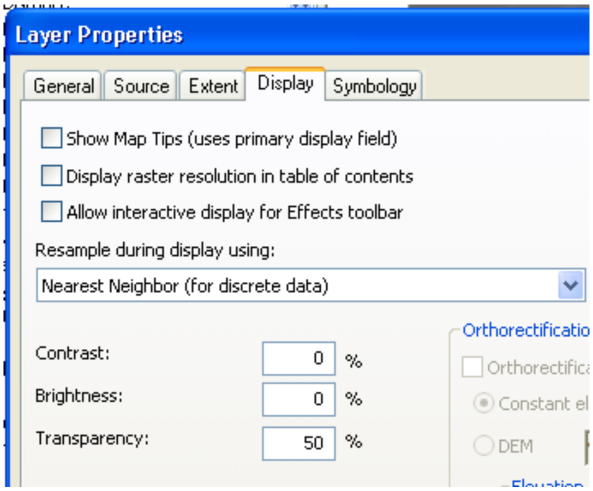

**4.** Open Raster Calculator and notice that now the environments are set to what you specified for the ArcMap document.

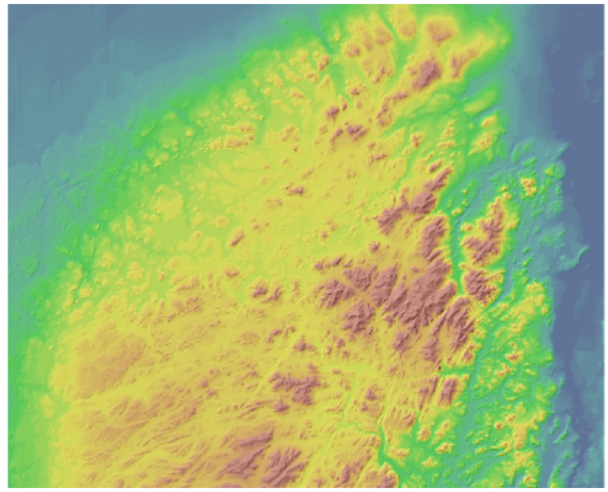
## F. Topographic analysis

1) Topographic data (also known as a digital elevation model 'DEM' or digital terrain model 'DTM') are one of the most common forms of raster grid that you'll run into. ArcGIS has a couple of special processing tools specific to topographic data
2) Remove any grids and shapefiles from your data frame and add back dem250k.
3) In ArcToolbox > spatial analyst tools > surface, you'll see a number of options directed specifically at DEMs:




4) *Aspect*: creates a new grid of aspect (direction of slope where 0 = N, 90 = E, 180 = S, 270 = W) based on the topographic grid
5) *Contour*: creates a line shapefile of contours based on the topographic grid at intervals that you specify
6) *Slope*: creates a new grid of slope (in either degree or percent) based on the topographic grid
7) *Hillshade*: creates a '3D' type visualization of topographic depth based on a sun direction and angle that you specify (used to make pretty maps only, not for analysis)
8) *Viewshed*: This requires a point shapefile to represent 'where you're standing' and creates a new grid of where you can see from that point location.
9) Create a hillshade grid of dem250k using the default values. Your output file should look like a shaded relief map. Now for our next visualization trick, move the hillshade below dem250k and turn both layers on.
10) In dem250k layer properties, find the 'Display' tab and change transparency to 50%. Click OK and check out your new map of elevation with shaded relief. For an even fancier display, change dem250k symbology to a color (rather than grayscale) color ramp.


Layer Properties
General
Source
Extent
Display
Symbology
Show Map Tips (uses primary display field)
Display raster resolution in table of contents
Allow interactive display for Effects toolbar
Resample during display using:
Nearest Neighbor (for discrete data)
Contrast:
0
%
Brightness:
0
%
Transparency:
50
%
DEM



## Lab Report Exercise: Adirondack Park

Scenario: The Adirondack Park Agency is concerned about some cases of small-scale illegal logging within the park. Park managers think that local land owners have been cutting down trees for fire wood. The Adirondack Park Agency has asked for your help to identify locations likely to be targeted this summer so that they can position law enforcement to prevent loss of trees. Here are the criteria that the park managers believe create elevated risk for illegal logging:

1) Sites within 300 m of roads are at **high risk**
2) Sites with slopes less than 10 degrees are at **high risk**
3) Sites with hardwood trees (land cover value = 1) are at **high risk**, and sites with mixed, but predominantly hardwood trees (land cover value = 2) are at **moderate risk**
4) The park agency is only interested in lands within park boundaries

Use the files in the Lab 6 folder to identify high risk (meets all the high risk criteria and within park boundaries), and moderate risk (meets high risk criteria #1 & #2, moderate risk criteria #3, and within park boundaries).

Rule: Since this is the Spatial Analyst lab, use Raster grids to identify risk. Also, the Park Agency would like the final map in the same spatial extents and pixel resolution as dem250k.

**Deliverables**:

1) Create a map for the Adirondack Park Agency highlighting areas at high and moderate risk. For visualization, make low risk areas no color and overlay the map on a shaded relief background. Also include the outlines of the towns that intersect the park so the managers can figure out jurisdiction. Include a legend, scale bar, and title on your map.
2) Describe the steps you took to create this risk analysis.
3) What is the total land area (include units) of high risk within the park boundaries? What is the total land area of moderate risk?

**Grading rubric:**

2 pts for: high risk, moderate risk, shaded relief background, park boundaries, legend, scale bar, title and question 3 (total area) (16 points total)

4 pts for: question 2 (steps)

# Lab 8:
# Surfaces and Geostatistics

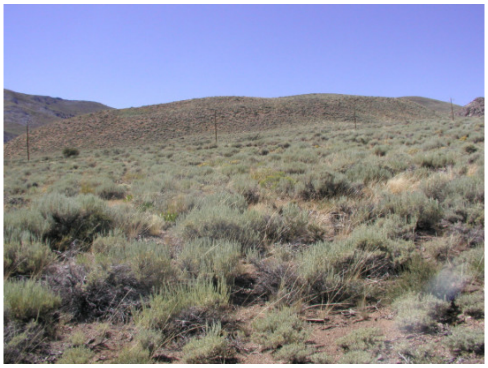


This lab will explore the geostatistical analyst toolbox in ArcGIS, as well as a few other tools related to understanding surfaces and spatial distributions. Previously, you have worked with vector data analysis, and raster data analysis. This lab will look at a few more tools that **combine** vector and raster data analysis.

Before you begin, Make sure the spatial analyst and geostatistical analyst extension are active (customize -> extensions).

The data that you'll be working with for the lab portion are related to the distribution of an invasive grass called cheatgrass, which has invaded over 100,000 square kilometers of the western U.S. (for comparison, the state of Massachusetts is about 20,000 square kilometers). Here's what an invaded landscape (left) looks like compared to a native landscape (right):

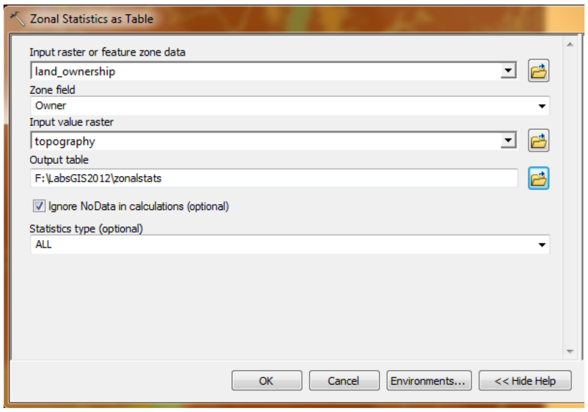
vs.



**A. Getting info from a Grid based on a Polygon: Zonal Statistics**

1. This command works with a polygon shapefile and a raster grid. By applying zonal statistics, you will calculate the statistics of the raster grid for each class specified within the polygon shapefile.
2. For this example, our goal is to compile some statistics about topography (from a grid) within different land classes (in a polygon shapefile).
3. Add topography (if a message pops up about building pyramids choose no) and land_ownership to your ArcMap document.

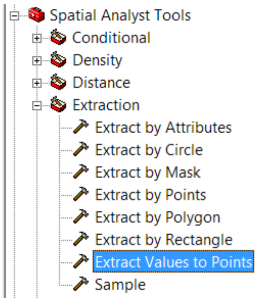
4. Open the attribute table of land_ownership and notice that the column ‘owner’ lists different land owners. This study site is in Nevada, so a lot of the land is public.
5. Change the symbology of the land_ownership layer to classify each owner in a different color. What agency owns/manages the bulk of land in this study area?
6. Under Spatial Analyst Tools in the toolbox, select zonal, then zonal statistics as table. Set the feature zone (polygon) data to land_ownership, the zone field to owner, and the value raster to topography. This command will calculate basic statistics about topography (elevation) for each type of land ownership. Remember to choose your lab folder for output raster and name the file "zonalstats”



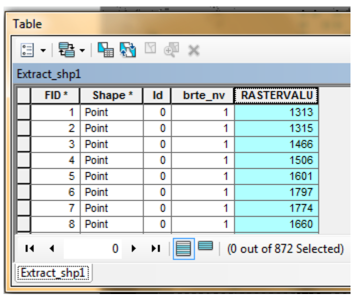

7. The nice thing about this command is that you can have multiple polygons of a particular class (e.g., BLM owned) and zonal statistics will consider them all when calculating statistics.
8. The mean elevation of BLM lands within our study site is 1605 m. What is the mean elevation of USFS (Forest Service) lands within our study site? Why do you think they are so different?
9. The output stats file is a .dbf, which you can open in MS excel (although the latest excel won’t output a .dbf table), or another spreadsheet program to create charts (at least, better charts than ArcMap can make) or tables to compare the elevation ranges of the different land ownership classes.

**Extract Values to Points**

1. Now, let’s say we have a point distribution of a species that we’re interested in. In this case, we have point distribution data for cheatgrass within the study area. Add the shapefile cheatgrass_present to the ArcMap document. We want to know whether cheatgrass is more likely within particular ranges of elevation. But, the elevation (topography) data are a raster grid. We can extract the values of the raster grid into our point shapefile using the ‘extract values to points’ command, which you can find under Spatial Analyst tools.




2. Select cheatgrass_present as your feature layer, and topography as your raster layer. ArcMap will output a new shapefile for this process, so specify a name and location for that shapefile.
3. Open the attribute table of the new shapefile you just created, and notice that it has a new column called 'Rastervalu', which in this case is the elevation (in meters) of each point location.



Table

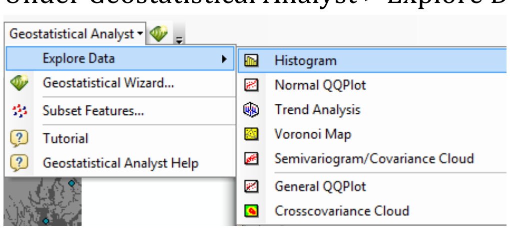
Extract_shp1

| FID * | Shape * | Id | brte_nv | RASTERVALU |
|---|---|---|---|---|
| 1 | Point | 0 | 1 | 1313 |
| 2 | Point | 0 | 1 | 1315 |
| 3 | Point | 0 | 1 | 1466 |
| 4 | Point | 0 | 1 | 1506 |
| 5 | Point | 0 | 1 | 1601 |
| 6 | Point | 0 | 1 | 1797 |
| 7 | Point | 0 | 1 | 1774 |
| 8 | Point | 0 | 1 | 1660 |

0 (0 out of 872 Selected)

Extract_shp1

4. You could open the .dbf file in excel to figure out basic statistics of these 872 points, or, you can calculate mean and standard deviation by right clicking RASTERVALU > statistics. But, let's explore one more tool for analyzing this column of data.
5. Under Geostatistical Analyst > Explore Data choose 'histogram'

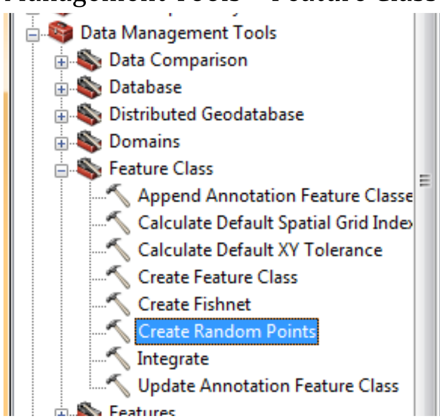



6. In the histogram, make sure the layer is the shapefile you just created, and the attribute is set to 'Rastervalu', the elevation values that you just extracted. In the upper right, you can see the summary statistics for these elevations. The mean elevation is 1448.7 m, and the standard deviation is 154 m.
7. You can change the number of bins in your histogram, hide or show those summary statistics, and perform a transformation on your data if they were skewed. (You can see the value of data skewness in the statistics box in the upper right – the larger the absolute value, the more skewed your data) In this case, the distribution is close to normal so there's no need for a transformation, but you can check out what happens if you transform the histogram anyway.
8. Click to select the bins at the low elevation end of the histogram, and notice that the points corresponding to these elevations are also selected in the data view. This way, you can quickly identify and highlight any outliers from your distribution data.

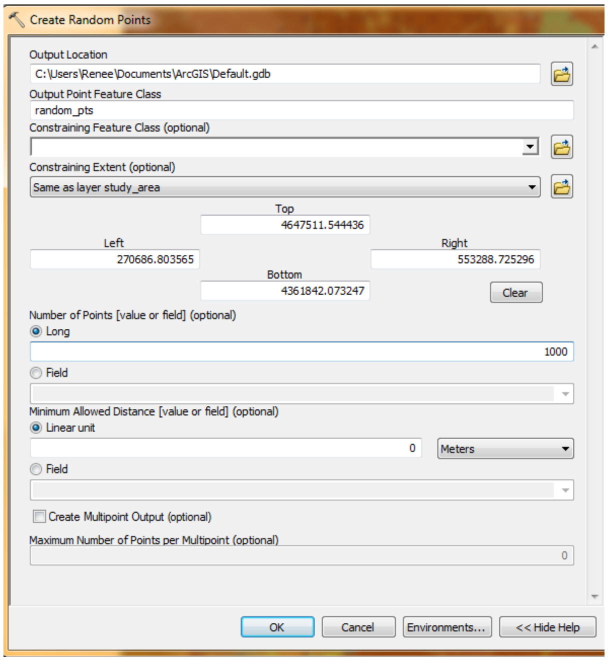
9. If you want to put this histogram into a map, you can add it to your layout and include it in a presentation of your data later on.
10. We've figured out that our cheatgrass presence points have a mean elevation of 1448.7 m. But, is that elevation significantly different from the available land area? To calculate summary statistics for the study area, we can use zonal statistics. But, first we need a polygon shapefile of the study area
11. Use what you learned in Lab 5 to create a new polygon shapefile outlining the study area (approximate extents are okay). Make sure to set the projection to be the same as topography.
12. Use the zonal statistics as table tool to calculate the mean elevation of the study area. The number should be around 1580 with a standard deviation of 300.
13. Based on these numbers, we can see that cheatgrass tends to invade a slightly lower elevation than the mean of the study area.

**Create Random Points for Comparison**

1. If you want more information than just the mean topography value for comparison to the values for cheatgrass distribution (e.g., you want to compare two histograms), then you might want to create a set of random points to compare to the species distribution.
2. Create random points in ArcToolbox is in a very non-intuitive location, under Data Management Tools > Feature Class > Create Random Points.



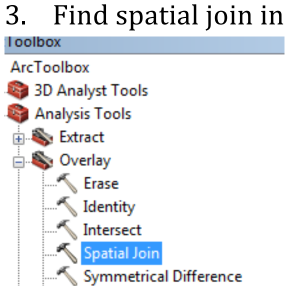

3. In the dialogue box, choose your output folder, and name your file. You'll also want to specify output extents for your points in the 'constraining extent' box. Set this to the study area polygon that you created earlier. Also specify an output of 1000 points (the more points you make, the longer it will take to process).



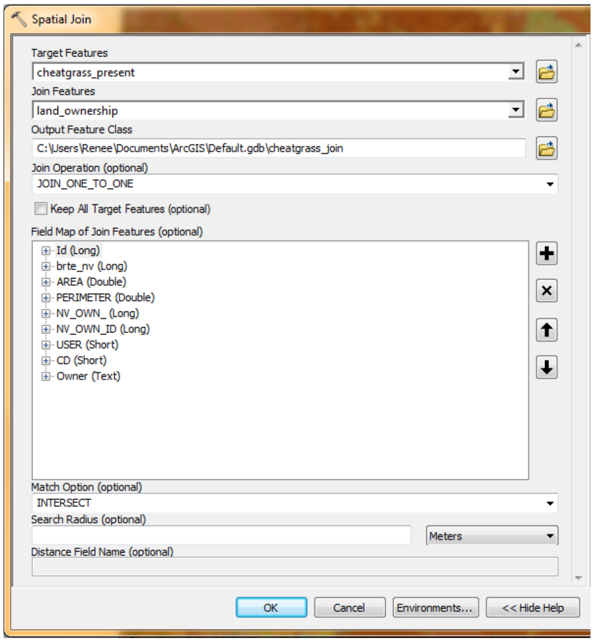

4. Click OK to create your random point distribution within the study area extents.
5. Now, like you just did with the cheatgrass points shapefile, use extract values to points to pull the topography values from the grid into the random points shapefile.
6. Add the histograms from the cheatgrass points and the random points to your layout view so that you can compare them side by side. The random points histogram tells you about topography in the whole study area, while the cheatgrass points histogram tells you about topography where cheatgrass is present.
7. *Note*: (In an ideal world, for analytical purposes we'd prefer to compare the environmental characteristics of locations where a species is *present* to locations where a species is *absent*. But, if we don't have absence data for our target species, we can instead compare presence to total available land area.)

## Spatial Join – getting info from one shapefile into another

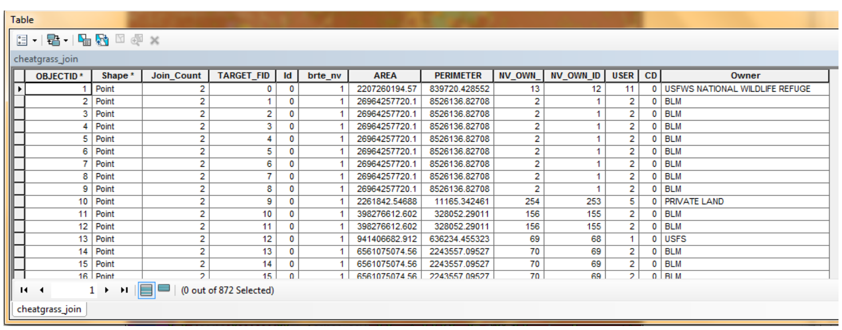
1. You've calculated the distribution of cheatgrass present relative to topography. Now, let's look at its distribution relative to land ownership. If you deleted it previously, add land_ownership back to your ArcMap document.
2. In this case, we can't use the extract values to points command because land ownership is a polygon shapefile, not a raster grid. Instead, we need to use what's called a spatial join, which combines the attribute tables of two shapefiles based on their location.
3. Find spatial join in the Analysis Tools -> Overlay



4. The target features should be the point shapefile, to which you want to join land_ownership, in this case cheatgrass_present. Choose a logical output shapefile name and location and click OK.




5. Your new output should contain the attributes from land ownership, including ‘owner’.




| OBJECTID * | Shape * | Join_Count | TARGET_FID | Id | brte_nv | AREA | PERIMETER | NV_OWN_ | NV_OWN_ID | USER | CD | Owner |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Point | 2 | 0 | 0 | 1 | 2207260194.57 | 839720.428552 | 13 | 12 | 11 | 0 | USFWS NATIONAL WILDLIFE REFUGE |
| 2 | Point | 2 | 1 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 3 | Point | 2 | 2 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 4 | Point | 2 | 3 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 5 | Point | 2 | 4 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 6 | Point | 2 | 5 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 7 | Point | 2 | 6 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 8 | Point | 2 | 7 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 9 | Point | 2 | 8 | 0 | 1 | 26964257720.1 | 8526136.82708 | 2 | 1 | 2 | 0 | BLM |
| 10 | Point | 2 | 9 | 0 | 1 | 2261842.54688 | 11165.342461 | 254 | 253 | 5 | 0 | PRIVATE LAND |
| 11 | Point | 2 | 10 | 0 | 1 | 398276612.602 | 328052.29011 | 156 | 155 | 2 | 0 | BLM |
| 12 | Point | 2 | 11 | 0 | 1 | 398276612.602 | 328052.29011 | 156 | 155 | 2 | 0 | BLM |
| 13 | Point | 2 | 12 | 0 | 1 | 941406682.912 | 636234.455323 | 69 | 68 | 1 | 0 | USFS |
| 14 | Point | 2 | 13 | 0 | 1 | 6561075074.56 | 2243557.09527 | 70 | 69 | 2 | 0 | BLM |
| 15 | Point | 2 | 14 | 0 | 1 | 6561075074.56 | 2243557.09527 | 70 | 69 | 2 | 0 | BLM |
| 16 | Point | 2 | 15 | 0 | 1 | 6561075074.56 | 2243557.09527 | 70 | 69 | 2 | 0 | BLM |

(0 out of 872 Selected)

cheatgrass_join

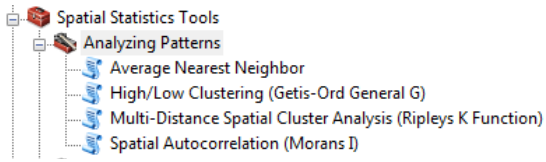
6. Summarize this chart from the owner column to calculate a count of points in each land ownership category. What percentage of cheatgrass points are on BLM lands?
7. Now, follow the same procedure using your random points. What percentage of random points are on BLM lands? Does that seem very different from the cheatgrass points? Do you think that cheatgrass is preferentially invading BLM lands?

## Spatial Clustering 1

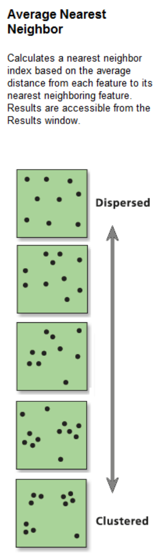
1. Clustering is relevant for point, polygon, or line distributions. A cluster analysis could help you determine whether points that you collected are spatially biased, or whether species presence points are clumped vs. randomly distributed.
2. If presence points are clumped, that suggests that something is affecting their distribution (e.g., a species only grows on south facing slopes, or on sandy soils, or near roads). If

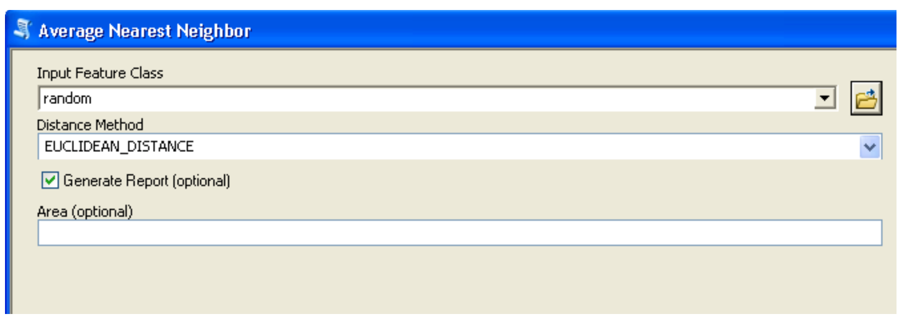
something is affecting a species distribution, then chances are good that we can try to predict what that is and better understand why that species is where it is (the essence of biogeography).

3. The ArcMap tools used to measure clustering can be found in ArcToolbox under spatial statistics tools > analyzing patterns



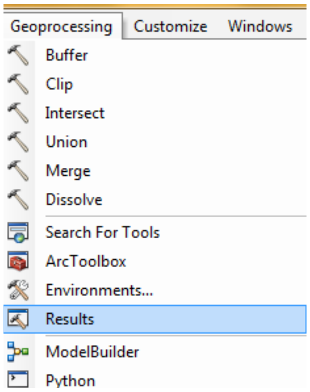

4. Double click Average Nearest Neighbor, which compares the locations of points to determine whether they are clustered, random, or evenly distributed. Select the random points file you created earlier. Make sure you click 'generate report' as an option so that ArcMap produces an html output. Click OK.



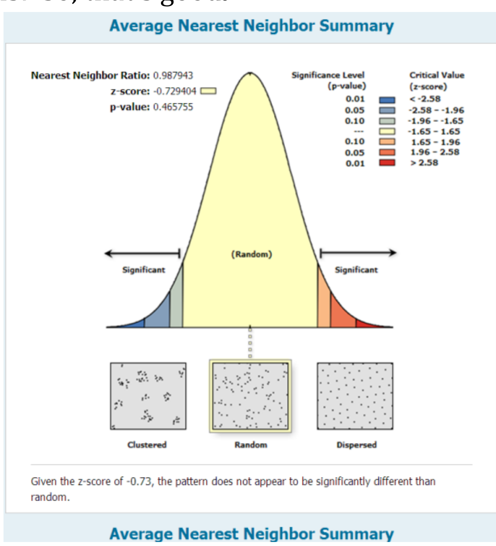




5. Once this process finishes, it will look like nothing has happened. That's because ArcMap conveniently stores the .html report in the folder where your default geodatabase is located. You can either navigate to it through Windows explorer, or find it in the ArcMap results window located in the Geoprocessing tab.

6. Note – the results window is a helpful window to know about because when processing don't work properly, the results window includes some description of what went wrong (not always, but occasionally useful)



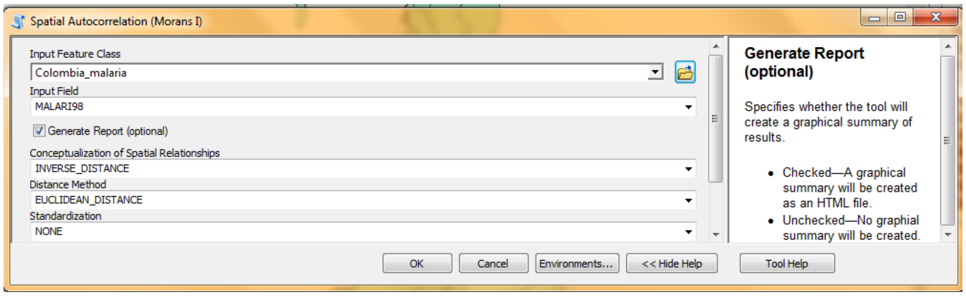

7. The output .html document should tell you that your point distribution is random…which it is! So, that's good.



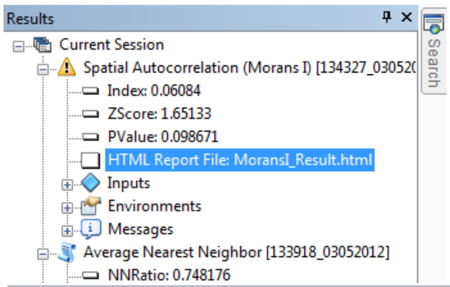

8. Use the editor tool to delete some of your random points and see if you can make a more clustered distribution. Rerun average nearest neighbor to see if you succeeded.

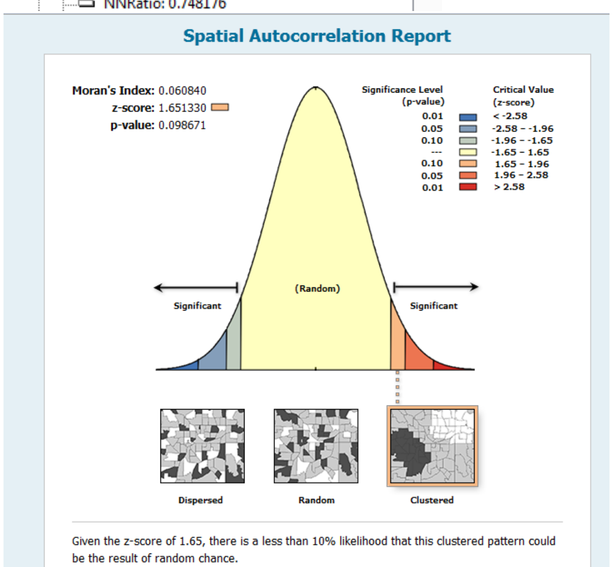
**Spatial Clustering 2**

1. Another pattern analysis tool is Moran's I, which analyzes how spatially autocorrelated (clustered) the data in an attribute table are relative to the distribution of the shapefile.
2. Add the layer *Colombia_malaria* and zoom to the layer extents. Open the attribute table and notice that, along with lots of other information, there is a column of malaria incidents in 1998.
3. From the ArcToolbox spatial statistics toolbox > Analyzing patterns, select 'spatial autocorrelation'.
4. Use Colombia_malaria as the input shapefile, and Malari98 as the input field. Make sure to select Generate Report.




5. Click okay. ArcMap will now calculate whether there is a spatial pattern of number of incidents of Malaria within the Colombian states. To view the results, find the .html document in your default geodatabase folder, or go to the results window and double click on the HTML Report File under Spatial Autocorrelation.



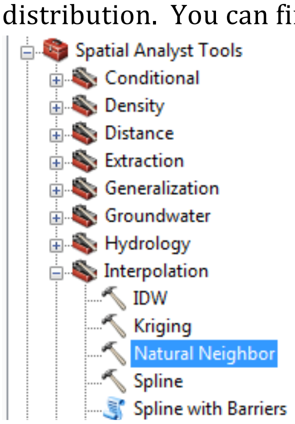

6. The Moran's I output shows that the data are clustered rather than randomly distributed. The Moran's I score (0.06) also gives you a confidence range. 0.06 means that if we created 100 random distributions, only 6 of them (0.06 * 100) would be as clustered or more so than this one. The other 94 would be less clustered. So, that gives us some confidence that the clustering we're measuring here is real and not just a random occurrence.
7. Change the symbology of Colombia_malaria to display by graduated color, and make ten classes based on quantile. Does the map display also suggest some clustering? Why do you think malaria outbreaks would be spatially clustered in Colombia?
8. Why would you bother with any of this stuff? Only if you have data *or attributes of data* that you think might be clustered. The previous two analyses provide spatial statistics (nearest

neighbor and Moran's I) to support any assertion on your part that your data are indeed clustered (or, that your data are most likely randomly distributed).

9. If your GIS project involves some clustering analysis, you can also learn about and experiment with the other spatial statistics tools related to patterns.

**Interpolation**

1. Let's say you have a bunch of points, and you want to estimate values in between those points. Here's a case where you need interpolation to create a raster surface from your point data.
2. Open a new ArcMap document and add the layer Baltimore_house_sales_1978
3. Note that this shapefile *does not* have a projection. The projection information has been lost, and the data do not use a standard projection so they can't be easily projected (the units aren't decimal degrees, but they distances are pretty small to be in meters). For this example, we'll just ignore that problem and assume that these points represent locations in the city of Baltimore and surrounding suburbs in 1978. However, if these were your data, you would want to make sure to define the projection before performing an interpolation because the interpolated surfaces will also be without a projection.
4. Change the colors of the points to a graduated color scheme showing low to high prices. Generally, where are the low priced houses and where are the high priced houses? If we assume that the center of this distribution of points is the city center, does this make intuitive sense to you?
5. Start with a natural neighbors interpolation. This method assumes that every value is weighted equally, and it does not interpolate beyond the boundaries of the point distribution. You can find it in the spatial analyst toolbox > interpolation




6. Make sure to select PRICE as your Z value field, and keep the default cell size. Click OK to create the interpolated grid. The output defaults to an 'earth tone' color ramp, but you can change this to help you visualize the patterns better.
7. Move the point location shapefile to the top of your display so that you can visualize the locations of the points relative to the interpolation.
8. Run the IDW and Spline interpolations from the same toolbox, also using PRICE as your Z value field, but keeping all the other values as their defaults.
9. Change the color ramps of the three interpolations to be the same, and compare the resulting interpolations. Recall that IDW weights closer points much more heavily than far

away points (weighting default is 1/distance$^2$), which can create a 'bullseye' effect. Also recall that Splines are a mathematical best fit that work well along regular gradients, but can over and/or under predict interpolated values relative to your data values.

10. What is the lowest pixel value for the spline interpolation? Does that seem reasonable to you? Can you explain why that happened?
11. If you want to get fancy with interpolations, you would also want to check out kriging, which is a much more mathematically complex method that also has numerous options you could play with. We won't do that here, but if your GIS project requires some interpolation, kriging might be a method you'll want to look into further.

# Lab Exercise: Suitability Analysis of White Top in Idaho

White top (*Lepidium draba*) is an invasive plant that has become more and more problematic in western states.

White Top invasion in Wyoming

Concerned about this invasion, the Idaho state department of agriculture (ISDA) has collected point locations of White Top invasion within the state using GPS. They have asked you to figure out if the data that have been collected so far are representative of the state as a whole.

ISDA has provided you with a list of lat/lon coordinates where White Top has been found in the file *white_top_idaho.xlsx*. Assume that the Datum used by the GPS receivers was NAD27. Convert this table to a shapefile.

You also have the following spatial datasets available to you:

*Idaho_major_roads* (a line shapefile of major roads in the state)
*Idaho* (a polygon outline of the state)
*Idaho_land_ownership* (a polygon shapefile of land ownership in the state)
*Id_ann_ppt* (a raster grid of interpolated annual precipitation in mm)

First, since we're only interested in Idaho, delete all of the points that are outside the state boundaries. Second, set the geoprocessing environments processing extent to be the same as the Idaho layer.

Answer the following questions:

Distribution relative to roads

1. What is the average distance to a major road of the collected White Top points?
2. What is the average distance to a major road within the state of Idaho?
3. Explain how you calculated 1&2.
4. Are the White Top points skewed relative to distance to road? Explain how you know this. Give one hypothesis why the point location data might (or might not) be skewed relative to distance to major roads.

Distribution relative to land ownership

5. Calculate the area (in $km^2$) of each land ownership class in Idaho (you don't have to report it here). Explain how you did this.

6. Only BLM, Private & State lands are available for future sampling. These lands represent a total land area of 126,123 $km^2$. Relative to the total land area available for future sampling, what percentage of land area is owned by:
    a. BLM
    b. Private
    c. State
7. Calculate the percentage of White Top points on land owned by:
    a. BLM
    b. Private
    c. State
8. Of the three land ownership classes, which have been oversampled? Which have been undersampled?

Distribution relative to annual rainfall

9. Are the White Top points found in areas of lower or higher annual rainfall than the state as a whole?
10. Create a histogram of the White Top points relative to annual rainfall. Create a histogram to compare this to for a random set of points in Idaho (*note: if you wanted to make the histograms look nicer and have identical bins, you could create them in excel based on the extracted point values. But, for our purposes, the ArcMap output is fine)*.

Hand in: Answers to the 9 questions above along with the two histograms of white top and random points relative to annual rainfall in Idaho.

**Grading rubric:**

Questions 1-10 are all 2 pts each.

# Reading: Understanding Remote Sensing and Aerial Photography

### INTRODUCTION

### LEARNING SEQUENCE

| | |
|---|---|
| Required Reading | Read the following:<br><br>Understanding Remote Sensing and Aerial Photography<br><br>• Remote Sensing and Aerial Photogrammetry<br>• Basic Principles of Remote Sensing<br>• Aerial Images<br>• Quality of Images<br>• Uses of Aerial Imagery<br>• Stereo Images<br>• Satellite Images |
| Resources | View the following:<br><br>1. Introduction to the Electromagnetic Spectrum<br>2. Visible Light Waves<br>3. Infrared Waves |

### INSTRUCTION

# Remote Sensing and Aerial Photogrammetry

### Remote Sensed Data

Let's first define the idea of remotely sensed data. Broadly speaking, within the context of GIS, there are two common types of remotely sensed data: aerial imagery, and satellite imagery. Aerial and satellite imagery are both remotely sensed data, because it records reflected energy from a distance. This remotely sensed information is useful for many reasons, such as: its large area coverage and single capture, its ability to look outside the visible spectrum of the electromagnetic spectrum, its geometric accuracy once it is rectified to Earth-based coordinate systems, and the fact that it is a permanent record of a single snapshot in time. Let's take a look at a few examples of remotely sensed data to get an idea of what we're talking about.

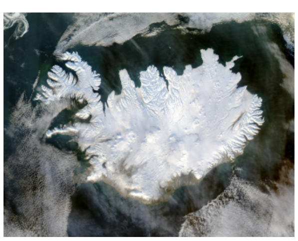
**Key Facts**

Remotely-Sensed Data

- Two Types
    - Aerial Imagery
    - Satellite Imagery

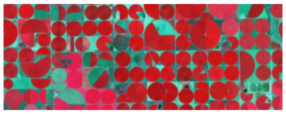
- Useful for many reasons:
    - Large area coverage
    - Extended spectral range
    - Geometric accuracy
    - Permanent record

This is a satellite image of the Middle East, the Horn of Africa, and India. This image clearly illustrates what is covering the land on this portion of the earth. The brown represents sand and dirt, while the green represents vegetation. In one single image, so much information is stored about that portion of the earth at that particular time.

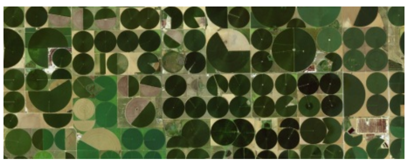


This is a satellite image of Iceland. Again, based on the colors and textures that are shown on this image, we can derive quite a bit of information about the environment in that part of the earth. For instance, if we look at Iceland, along the top edge, we can see the texture that represents mountainous terrain.

This next image is an aerial photograph of the port of Corpus Christi, Texas. In this image, we see great detail about structures and ships in the port.



This image looks quite a bit different than the previous images. For one, you are probably wondering what all the red represents. In this image, the red represents healthy vegetation. The healthy vegetation is red, instead of green, because this image is looking outside the visible spectrum. This is useful, because, as you will find out in a later section, objects reflect different portions of the electromagnetic spectrum differently, and can tell us useful and interesting things, that we would not be able to see by viewing the visible portion of electromagnetic spectrum.



This is the true color image of the same infrared photography we just saw. The circles are irrigation circles of crops. It is clear to see that where the crops are irrigated, they have very healthy vegetation.

# Basic Principles of Remote Sensing

### Remote Sensing

The main idea behind remote sensing is the storing of reflected electromagnetic energy by a sensor that is distant from the object being sensed. The reflected electromagnetic energy is commonly referred to as light. Light energy is characterized by a wavelength and each wavelength is defined by a distinctive color. Examples of colors would be red, orange, and yellow. As humans, we see only the visible portion of the electromagnetic spectrum, which is composed of smaller and longer wavelengths that surround the visible section. Examples of waves that are longer and shorter than the visible portion of the spectrum are x-rays, gamma rays, and radio waves. The full range of wavelengths is called the electromagnetic spectrum.

**Key Facts**

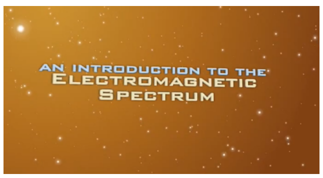
Basic Principles of Remote Sensing

- Remote sensing is based on reflected electromagnetic energy.
- Light energy is characterized by *wavelength.*
  - Each "color" of light has a distinctive wavelength.
  - Full range of wavelengths is called *electromagnetic spectrum.*

- Our eyes sense light in the visible portion of spectrum.

## Electromagnetic Spectrum

To provide a more thorough introduction to the electromagnetic spectrum, you will now watch three videos created by NASA. The first video will introduce you to the entire electromagnetic spectrum. The second video will discuss the visible portion of the electromagnetic spectrum which contains the colors that we see. The third video discusses infrared waves and their uses in remote sensing and science.



Introduction to the Electromagnetic Spectrum (5:03)

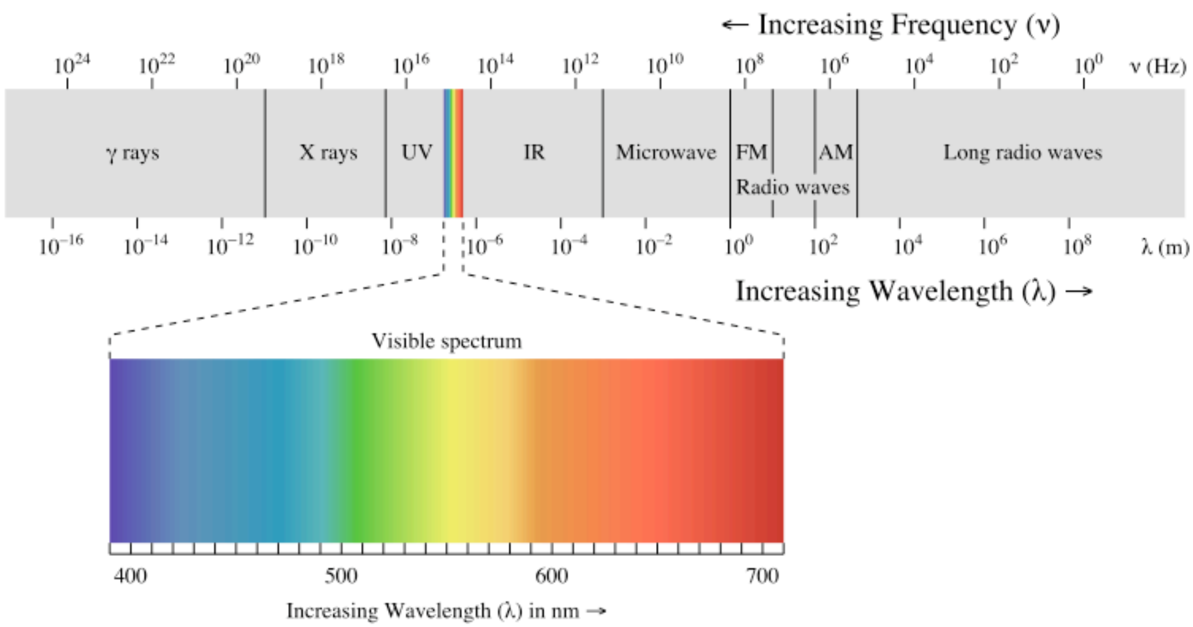
Visible Light Waves (4:50)

Infrared Waves (5:23)

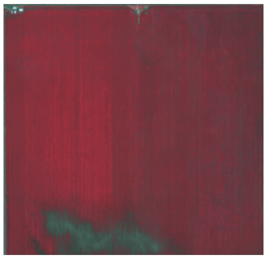
## Electromagnetic Spectrum

Let’s briefly reinforce the concept of the electromagnetic spectrum shown in the previous three videos. The electromagnetic spectrum is composed of energy at different wavelengths. The portion of the spectrum that we can see with our eyes is the visible spectrum which is roughly between 400 and 700 nm. The infrared portion of the spectrum contains longer wavelengths than the color red and is quite useful when studying environmental features such as vegetation. Generally, in GIS and remote sensing the visible and infrared portions of the electromagnetic spectrum are the most widely used portions of the spectrum.

We just spent all that time discussing the electromagnetic spectrum because it is the reflected energy, and where it lies within the electromagnetic spectrum, that we are interested in recording. Electromagnetic energy is all around us and it is constantly striking objects. The energy is then reflected, absorbed, or transmitted from those objects back to our recording sensor, such as a camera. The energy that is reflected is what it appears as in the picture. In the visible spectrum, if an object reflects the color green, and absorbs all other colors in the visible spectrum, we will record the green wavelength for that object in our sensor. In addition to the energy being reflected from the visible portion of the electromagnetic spectrum, it is also simultaneously reflecting, absorbing, or transmitting electromagnetic energy outside the visible spectrum. We may or may not, see the electromagnetic energy outside the visible spectrum based on which portion of the spectrum the sensor is calibrated to detect.




We will now look at an example of an image taken of a farm in South Texas. This image represents the visible portion of the electromagnetic spectrum and therefore, the healthy vegetation shows as green. Based on what our eye sees this looks like a perfectly healthy crop. However, let's now take a look at the same crop but focus on the infrared portion of the electromagnetic spectrum.

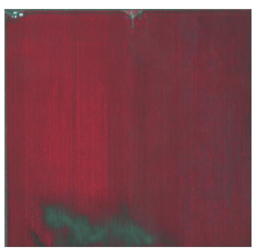


This image shows a different story of the crop area in the electromagnetic spectrum, healthy vegetation reflects a large amount of infrared energy. That means, that the brighter red areas are where the crop is healthier. Let's focus on the area at the bottom of the image. If you remember, on the image of the visible spectrum, this was all green and looked normal. However, the energy reflected in the electromagnetic spectrum tells us that these are unhealthy plants because of unusually high salt content in that portion of the farm. So while the plant looks healthy on the outside, inside it is an unhealthy plant.

## Two Types of Remote Sensing

There are two types of remote sensing systems: passive and active.

### Passive Remote Sensing Systems



These two example images taken of the farm in South Texas represent passive remote sensing systems. This is because; the camera did not transmit any energy down to the earth to be reflected back and recorded. Instead, the camera relied on an external source of energy, in this case, the sun. Relying on an external source of energy has its positives and negatives. The positive, is that the sensor does not require a large amount of energy to use. The negative, is that it can only record reflected energy when the external source is available and may be affected by atmospheric conditions and do not penetrate clouds. Therefore, when sensing with passive remote sensing systems, it is important to plan to record when there is ample energy and little to no obstructions.

**Key Facts**

Passive Remote Sensing Systems

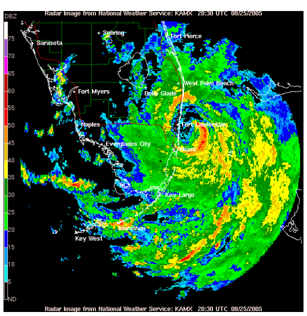
- Passive systems may be affected by atmospheric conditions
- Most systems rely on sun's energy
- Do not penetrate clouds

### Active Remote Sensing Systems

An active remote sensing system generates its own energy signal and that energy is sent out and returned. A major positive for active remote sensing systems is that it can collect data day or night and can typically penetrate clouds using wavelengths that are cloud penetrating. The negative, is that it can require a large energy source to operate. As a quick challenge, can you name a well-known active remote sensing system?

Radar is a well-known active remote sensing system. Radar stands for radio detection and ranging and it works by transmitting radio waves, and recording the intensity of the radio waves that are returned. Radar is the most common active remote sensing system in the world, except for cameras with a flash. Radar systems are typically restricted to one wavelength, so they can identify the radio waves that are returned as its own. Because of this, radar only records the intensity of the return signal, therefore, produces a monochromatic or black and white, image.

This is a picture of a radar system at Heathrow Airport. The radar tower rotates 360° transmitting a single radio frequency and measures the intensity of the returns, which represent airplanes. The air-traffic controllers can then use the resulting image to direct traffic.



This is the resulting image of a weather radar showing a hurricane covering the Southeast Coast of Florida. While this image has multiple colors, the colors are false and simply represent the intensity at which the radio wave has been returned to the radar sensor.

**Key Facts**

Active Remote Sensing Systems

- Generate an energy signal and detect the energy returned
- Active systems can collect day or night and typically penetrate clouds
- Radar is the most common active remote sensing system.

Another active remote sensing system is Lidar. Lidar stands for light detection and ranging and is becoming an increasingly popular active remote sensing system. Lidar works by scattering laser pulses to find the range of a distant target. Because lasers produce higher frequencies and lower wavelengths, they are much smaller than radio waves and can provide a much higher resolution image. To further explain Lidar, watch this short video.

Lidar (2:50)

This is another example of a Lidar image; this is Lidar data rendered in three dimensions of the Ground Zero site from the terrorist attacks of September 11, 2001. In the lower center of the image, you can see where the two towers used to stand, and how tall the rubble pile is. This image shows how detailed a Lidar image can be from the air.

**Key Facts**

LiDAR

- Stands for light detection and ranging
- Increasingly popular
- Scatters laser pulses to find range of a distant target
- Much higher resolution than radar
    - Uses more focused beams

# Aerial Images

## Aerial Imagery

Aerial imagery has been a primary source of geographic data for quite a long time. Aerial imagery started almost as soon as portable cameras were invented and became practical with the invention of the airplane. As the camera was used more and more for aerial imagery, the science and profession of photogrammetry was defined. Photogrammetry is the science of measuring geometry from images. What that means, is that if we take a bird's eye view of the downtown area and we know the flying height, focal length, and other variables of how the photo was taken, we can accurately measure features from the photo, such as length of roads, area of land or even elevation differences if we have overlapping photos.

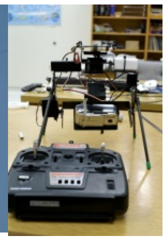
This is a picture of an aerial camera being installed in the nose of an airplane. This aerial camera is looking up the side of the nose; therefore, it is not providing a bird's eye view, but a perspective view. Perspective views are useful to capture images of areas you can fly directly over, want to show more information in a single photo frame, without flying higher. Notice how large the aerial photo camera is. Old film aerial photo cameras used very large film formats. For instance, in the typical consumer film camera, film size is 35 mm. In an aerial photo camera, a film size is 240 mm, quite a difference in size, but also, quite a difference in quality.

This is a perspective image from a camera like the one shown in the previous image. While this photo is useful for identifying features, there is a significant amount of geometric distortion which it makes it more difficult to accurately measure distances from this photo. Because of this, the most common perspective of aerial photos used in GIS, are the top down, or bird's eye view, which will allow us the maximum geometric accuracy.

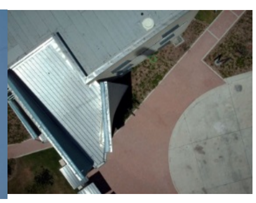
This is an example of a pointer installed on a helicopter. This camera can take both perspective, and bird's eye view photos.

In more recent history, cameras are being outfitted on unmanned aerial vehicles. The unmanned aerial vehicle can fly itself along a predefined path, or be flown by a pilot looking at a video screen from afar. These unmanned aerial vehicles can take perspective, or birds eye views of the earth, much like an aerial photo camera on an airplane.

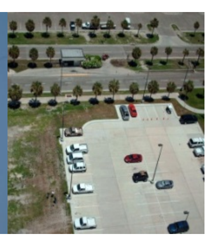
This is an example of aerial photo retrieved from an airplane that had an aerial photo camera fixed at a bird's eye view. After the photogrammetry expert corrects this photo for geometric error, we can make meaningful measurements of the size of the objects in the photo.

Not all aerial photo cameras need to be large or expensive,
Here is a project where a consumer digital camera was connected to an apparatus that can be radio controlled to take a picture. The apparatus was then attached on a kite line and the kite was launched into the air. The camera was remotely controlled from the ground and when the kite landed, pictures like this were retrieved.

Here's one image from the kite aerial photo showing the top and side of the building from a bird's eye view.

Using the radio control, the camera can also be pointed into a perspective view to take a picture like this.

Even though these photos are not from as high of a height, they can quickly be launched, are cheap to create and maintain, and can help you view your surroundings from a different, higher perspective.

## Aerial Image Attributes

Now that we are familiar with the concept of an aerial photograph, it is time to discuss how an aerial image is described. An aerial image has multiple attributes that are important to us as users: image scale, extent, and image resolution.

### Image Scale

The image scale is the relative distance on an image to the distance on the ground. For instance, we would specify that a particular aerial image has a scale of 1 inch measure of the image representing 100 feet on the ground. With respect to scale of distances, images contain distortions typically not found on maps. Additionally, scale is not uniform across the image and therefore, the scale is usually specified as an average scale. What that means, is that due to the way in which an aerial camera is designed and the distortion inherent in collecting an image from a distance, we cannot say with certainty that 1 inch measure of the image represents 100 feet on the ground anywhere on the image.

**Definition**

**Image scale**
relative distance on image to distance on ground

**Key Facts**

**Image Scale**

- Images contain distortions typically not found on maps
- Non-uniform scale across image
    - Usually specify an average scale

### Extent

The second aerial image attribute is the extent. The extent is the area covered by the image. The extent depends on the physical size of the recording media, the lens system installed in the camera, and the flying height of the platform the camera is installed on. Extents can be specified in units such as square miles, or degrees of latitude and longitude, among others.

**Definition**

**Extent**
area covered by image

**Key Facts**

**Extent depends on…**

- Physical size of recording media
- Lens system
- Flying height

### Image Resolution

**Definition**

**Image Resolution**
smallest object that can be reliably detected on an image

The image resolution attribute represents the smallest object that can reliably be detected on an aerial photo. Typically, resolution is measured between two highly contrasting colors of adjacent objects that can give clues as to what the resolution is of the images. Resolution will depend on the size of the films grain in analog cameras, and the sensor density in digital cameras. When testing an aerial camera to determine the image resolution, oftentimes, an alternating pattern of black and white lines are painted on the ground. The aerial camera is flown at a specific height, with specific camera settings, and then based on how clear the resulting images, the image resolution can be determined. Image resolution is defined as how many square units of the ground are being covered in a single pixel in an aerial photograph. For example, a high resolution aerial photo camera, can take a picture at a resolution of 6 inches squared.

Below is an example of a photo test pattern painted accurately on the top of the building. The image on the left is from an aerial camera with a high resolution. On that image, the alternating lines are very clear. On the picture to the right was taken with a low resolution aerial camera and the alternating patterns are fuzzier and therefore, at a lower image resolution.

**Photo Test Pattern**

Higher Resolution Lower Resolution

**Key Facts**

Image Resolution

- Typically resolution is measured between two highly contrasting colors.
- Resolution depends on film grain in analog cameras and sensor density in digital cameras.
- Resolution is often tested using alternating patterns of black and white lines.

## Types of Film

Aerial cameras that still use film are still quite common. In order to record different parts of the electromagnetic spectrum, the cameras use different types of film. There are three widely used types of film: panchromatic, true color, and infrared.

### Panchromatic Film

The panchromatic film records the intensity of the returned energy and creates a black and white photo. The positive take of the panchromatic image is that it is inexpensive, and has a wide exposure range.

### True Color Film

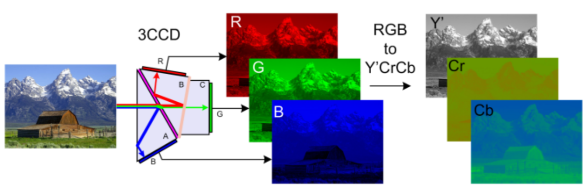
The true color film is sensitive to light across the visible spectrum. This requires three emulsion layers, one for each primary color of red, green, and blue. This makes the film expensive, and does not have as wide exposure range as the panchromatic film.

### Infrared Film

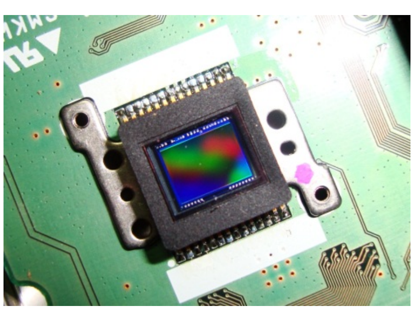
The third widely used type of film is infrared. Infrared film is sensitive to visible and infrared wavelengths. It is sensitive to difference in brightness, which makes it a high contrast image. Infrared film is widely used for vegetation mapping.

## Digital Aerial Cameras

Digital aerial cameras are quickly replacing film cameras as the sensor of choice. One major positive for a digital aerial camera is that it eliminates converting the film into a digital image through scanning. Scanners can introduce error in the photo. Instead of using film, digital aerial cameras use charge coupled devices, more commonly known as CCD's, or a similar element depending on which technology is being used by the camera. The CCD is a rectangular array of pixels that respond to light and record the intensity of the range of the electromagnetic spectrum that it is calibrated to record. It is possible to split multiple CCD's across multiple lenses, or use multiple CCD's with a single lens to simultaneously record different portions of the electromagnetic spectrum.




This image shows how reflected energy from the visible portion of the electromagnetic spectrum reacts with a three CCD system. In a three CCD system, there is one CCD for each primary color. A complex set of mirrors and filters redirect the primary colors onto its respective CCD. This creates three separate images, each reflecting the intensity of returned energy for one of the three primary colors. When those three images are laid on top of each other, they create a color photograph as we know it.



The picture on the bottom shows a CCD on a circuit board. The CCD is placed, typically, behind the lens of the camera, so it can record the reflected electromagnetic energy. Include content on topic.

# Quality of Images

### Geometric Distortion

All aerial images contain geometric distortion. It is impossible to take an aerial image of an area without introducing some sort of error and distortion. Therefore, you should never use aerial imagery that has not corrected these geometric distortions. What you should use instead is an orthographic photo. An orthographic photo is a photo that has been corrected for geometric distortion by a trained photogrammetrist and now shows positions of objects after being projected onto a common plane. This is now equivalent to looking vertically downward onto a scene from an infinite height. Viewing a location from an infinite height is the only way we can improve all geometric distortion. Since we cannot view from such a distance, photogrammetry employs algorithms, and equations to emulate the view from an infinite height.

**Key Facts**

Geometric Quality of Aerial Images

- All aerial images contain geometric distortion.
- Orthographic photo: plots position of objects after being projected onto a common plane.
    - Equivalent to looking vertically downward onto a scene from an infinite height

### Perspective View

Contrary to popular belief, most aerial photographs do not provide an orthographic or top-down view. Instead, they provide a perspective view which is a geometrically distorted image of the earth surface. This is caused for multiple reasons, such as: not being able to reach an infinite height, distortions in the camera’s lens system, or undesirable movement of the aerial platform while taking a photo.

**Definition**

**Perspective View**
a geometrically distorted image of the Earth's surface.

This image shows an example of a perspective view of a building. We know this is a perspective view because we can see the side of the building. If this were a top-down, or orthographic photo, we would only see the top of the building and the tops of all of the surrounding features.

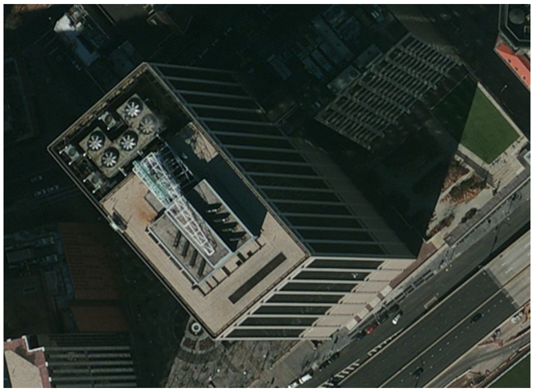
## Primary Sources of Aerial Image Distortion

There are six primary sources of aerial image distortion: terrain, camera tilts, film deformation, camera lens, atmospheric bending, and other camera errors. The terrain, and camera tilts sources of error are considered to be the major sources that contribute to the greatest amount of error. We will now discuss each one of these six primary sources of error in more detail.

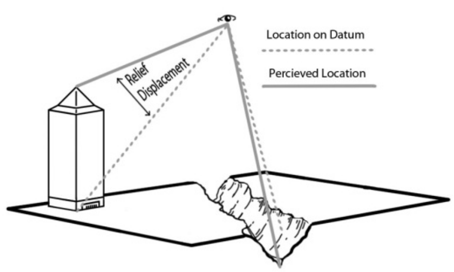
| Source | Amount of Error |
|---|---|
| **Terrain** | **Major** |
| **Camera Tilt** | **Major** |
| Film Deformation | Minor |
| Camera Lens | Minor |
| Atmospheric Bending | Minor |
| Other Camera Errors | Minor |

### Terrain Distortion

Terrain distortion, a major source of error, is often the largest source of geometric distortion. Trained distortion is caused when the terrain greatly varies in elevation. That means, the position of features on the image, are displaced from its original position and therefore, its position cannot be accurately measured.



Returning to the image of the building, this is an example relief displacement. The location at the top of tall objects is displaced away from the center of the photograph that means that the true location of the building corner, measured at the top of the building, is displaced outward from its actual location.




When we take an aerial photo, we are recording reflected energy onto a flat sensor, which is represented as a conceptual flat surface on the ground for the cameras pointed. Objects that exceed the elevation of the datum are displaced outward from the center of the photo. Objects that are below the datum, are displaced inwards towards the center of the photo. This means, that relief displacement is not consistent throughout the photo, and can vary greatly for each feature.

The terrain distortions are radial in nature. That means that as you travel further away from the center point of a bird's eye view photo, the distortions grow greater. This relief distortion effects angles and distances on the aerial photo and is the major cause of non-constant scale on an aerial image. Therefore, the vertical aerial image taken over every terrain is not orthographic. Since there are very few perfectly flat locations on Earth, that means, that virtually all vertical aerial images are not orthographic, meaning they show relief distortion, and do not have constant scale. A photogrammetrist will work with the photo to reduce these distortions.

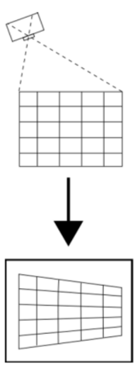
| **Key Facts** |
|---|

Terrain Distortions

- Terrain distortions are radial
- Relief distortions affect angles and distances on an aerial image
- Scale is not constant on aerial images
- A vertical aerial image taken over varied terrain is not orthographic

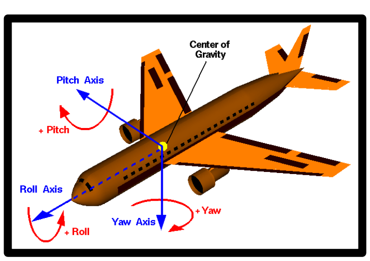
## Camera Tilt

The second source of error, camera tilt, is another major source of error in aerial imagery. Camera tilt is when the optical axis points in a non-vertical angle, causing a non-bird's eye view. This non-vertical angle result of a complex perspective convergence of features on the photo.



In this illustration, the photo is not looking straight down at the grid, thereby providing a perspective view. This perspective you created is perspective convergence, where the width and height of the grid cells shrink and converge at a point far off in the distance. To avoid this, it is important to always stabilize the camera and have the pilot fly as smooth and perpendicular to the ground avoiding any sudden moves.

In order to assist the camera in maintaining a perfect bird's eye view, intricate camera mounting systems are devised that keep the optical axis point directly below. The camera mounting systems are constantly measuring and adjusting for the three angles of rotation: pitch, roll, and yaw.




This image illustrates the relationship between pitch, roll, and yaw in relation to the center gravity of an airplane.

**Key Facts**

Camera Tilt Distortion

- Camera mounting systems are devised to keep optical axis pointing directly below.
- Tilt is characterized by three angles of rotation:
    - Pitch
    - Roll
    - Yaw

### Minor Errors

The other three errors, the minor errors, are radial lens displacement, systematic errors, and atmosphere distortion. The radial lens displacement represents the imperfections in the camera's lens surface. If the lens is not perfectly formed, it will change the energies angle so that it hits a different part of the sensor. A systematic error is an error that is constant across a system, or series of photos. Examples of systematic error are a deformed pattern that the file should lie flat upon. If the pattern is deformed, then the film cannot lay perfectly flat and the image is therefore distorted. Atmosphere distortion is when the light waves are bent as they travel to the atmosphere changing their properties. If the properties of the light are changed, then the resulting image will not display the true properties of light.

**Definition**

**Radial Lens Displacement:** imperfections in camera lens surface
**Systematic Error**: an error that is constant across a system
**Atmospheric Distortion**: bending of light

### Geometric Correction

To remove the geometric errors we spoke about previously, photogrammetry provides tools needed to remove this distortion from the photos. The correction process requires measurement of image coordinates, and their corresponding ground coordinates. Using that information, and photogrammetric methods, the end result is an orthographic image. An orthographic image is an image with terrain and perspective distortions removed, providing the highest possible accuracy with respect to the limitations we face when taking aerial photos.

This image is an example of an orthographic image. This is an image with terrain and perspective distortions removed. The image still shows the perspective view of Earth; however, the image has been modified through stretching and skewing to compensate for the errors inherent in a perspective photograph.

# Uses of Aerial Imagery

### Primary Uses of Aerial Imagery

There are three primary uses of aerial imagery: surveying topography, interpretation, and background use. Aerial images form the basis for surveying and topographic mapping. Aerial images properly corrected for geometric distortion, allow for accurate measurement of both horizontal, and vertical positions. Image interpretation allows experts to categorize what they see on the image and/or assign attributes to those features. Third, aerial imagery is often used as a backdrop for a map, to provide a nice visual context. For this section, we will focus on image/photo interpretation. In a future section, we will discuss how horizontal and vertical positions are derived from aerial image backspace. The third primary use, a backdrop for map, will not be discussed as it is easy to understand and does not need further discussion.

**Key Facts**

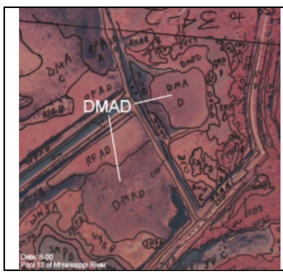
**Primary Uses of Aerial Imagery**

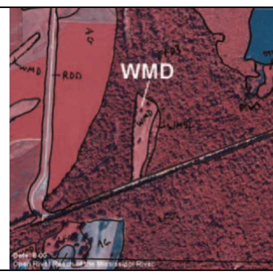
- Basis for surveying and topographic mapping
  - Measure X, Y, Z

- Image/Photo interpretation
    - Categorize or assign attributes to surface features
- Used as backdrop for map

## Photo Interpretation

Photo interpretation is the process of converting images to information often through categorization or assignment of attributes. Trained interpreters use the size, shape, color, brightness, texture, and relative and absolute location of features to interpret images.




These two infrared images have been categorized by trained interpreters. If you look closely, you can see that sections of each photo contain a unique texture, shape, color, and so on, that allow the interpreter to differentiate the features on the ground.

To properly use the characteristics of the terrain to interpret the photo, the photo interpreters should first develop some familiarity with the features of interest. This familiarity is often created through the interpreters visiting the site portrayed in the aerial image. This is known as ground-truthing.

The common result of photo interpretation is a thematic map. In order to make a thematic map, this requires that the interpreter establishes categories for the interpreted features to be placed within. Additionally, this requires the establishment of a minimum mapping unit. A minimum mapping unit is an agreed-upon area that is the smallest of features that can be considered its own separate, significant feature. For example, if we are mapping tree stands, and within a large tree stand there is a significantly different tree. If the minimum mapping unit is, say, 10 trees, the next significantly different tree will not be isolated as a separate feature. Photo interpretation is executed by either on-screen digitizing, or hardcopy interpretation which is later scanned into the computer, and in a digitized form.

**Key Facts**

**Photo Interpretation**

- Is the process of converting images to information

- Interpreters use the size, shape, color, brightness, texture, relative and absolute location of features to interpret images.
- To properly use the characteristics, the photo interpreters should develop some familiarity with the features of interest
- Interpreted photos are often compared with findings on the ground, Also Known As (AKA) Ground-Truthing

# Stereo Images

## What are Stereo Images?

It is important to introduce the concept of stereo images as they are important for measuring horizontal and vertical locations from aerial and satellite imagery.

Stereo images are two overlapping photos, called a stereo pair that may be used to determine elevation differences. Typically, when an aerial or satellite image collection campaign is underway, some endlap and overlap of the images are built-in. Endlap is overlap of the top and bottom of the photographs as the aerial platform travels in a line. Overlap, is where photos overlap on the sides, from where the aerial platform traveled in a parallel line. Endlap and overlap provide two important things. One, with sufficient endlap and overlap, if there is an error with one photograph, there should be enough coverage from surrounding photographs to fill in that space. Second, and the focus of this section, is the ability for us to exploit the phenomenon known as parallax, which provides a three-dimensional view from two overlapping flat photographs.

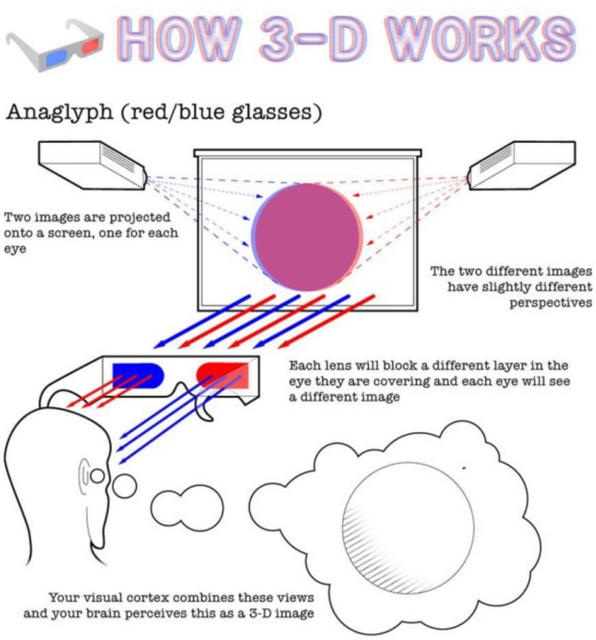
*Endlap* – Overlap in flight line
*Overlap* – Overlap in adjacent flight lines

## How Stereo Pairs Work

Images with sufficient endlap reveal elevation due to parallax. Parallax is a shift in relief displacement due to a shift in observer location. What that means is that the displacement of any given point is different on the left versus right ground views because relative geometry is different. So for instance, if you close your right eye, you'll see a more left perspective of an object are looking at, if you close your left eye and open your right eye, you will see more of the right perspective of the object. When you open both eyes, your visual cortex receives both images from the left and right perspective, and uses them together to create depth. We are simply exploiting the same phenomenon but using photographs.

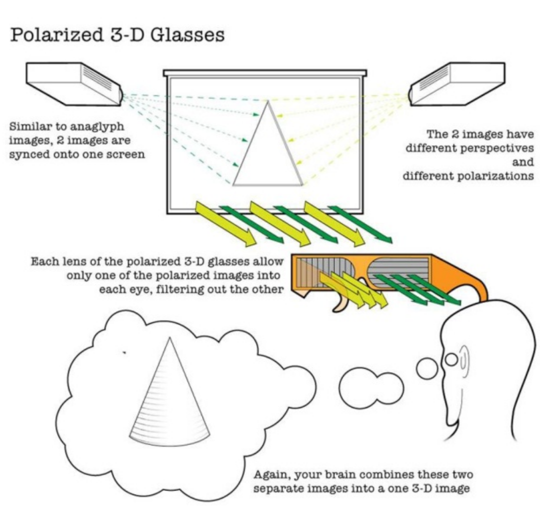
Take look at a few different ways that we can use technology to display two overlapping photos to trick the brain into thinking that we're looking at a three-dimensional object even though we're looking at two separate flat photos.

### How 3-D Works



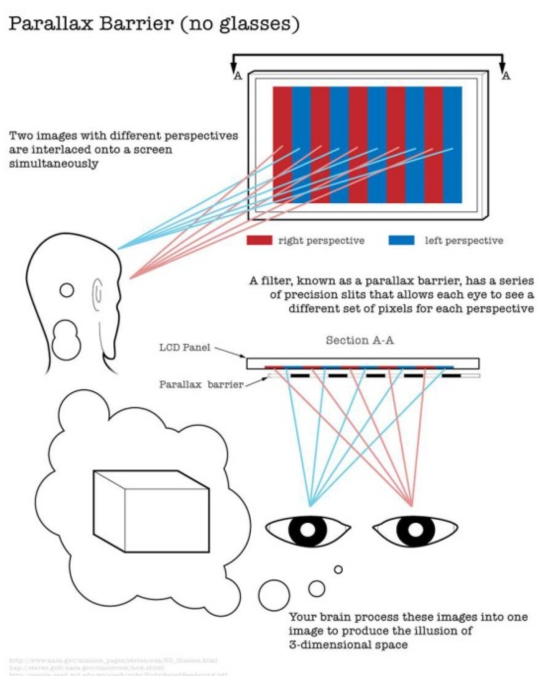

A classic way of showing flat images in 3-D is to use a pair of red/blue anaglyph glasses. The way it works, is that two images are projected onto a screen, one for each eye. The two images have a slightly different perspective. Each lens will block a different layer in the eye they were covering and each eye will see different image. This is because one image is projected in red, and the other image is projected in blue. Therefore, the glasses filter the red and blue light so that each eye only receives one image. Your visual cortex then combines these two different views in and your brain perceives this as a three-dimensional image.

### Polarized 3-D Glasses




Another way to begin to view 3-D from two flat images, is the through the use of polarized glasses. Similarly to anaglyph images, two separate images are projected onto one screen, and each image shows a slightly different perspective. One image is projected through a horizontally polarized piece of glass, and the other images projected through a vertically polarized glass. Each lens of the polarized 3-D glasses allows only one of the polarized images into each eye. Again, your brain combines these two separate images into one 3-D image. This is the method used at movie theaters today.

### Parallax Barrier




A third way that we can see a three dimensional image from two flat images, is through the use of a parallax barrier. For this to work, two images with different perspectives are simultaneously displayed on the screen. However, a filter, known as a parallax barrier, has a series of precision slits that allows each eye to see a different set of pixels for each perspective. Again, your brain processes these images into one image to produce the illusion of a three-dimensional space. In the consumer market, some new TVs are using this method, as well as the Nintendo 3DS.

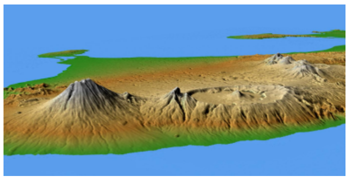
## Stereo Model

In GIS, when we have a three-dimensional perception of terrain that we see viewing a stereo pair, we call it a stereo model. In a stereo model, we observe parallax differences. Our brain converts these two perceptions of depth, and we can then use that perception of depth to make measurements, and calculations.

Using the information derived from stereo models, and active remote sensing systems, such as Lidar, we can create 3-D renderings in the computer that can be manipulated and rotated in real-time.

**Key Facts**

**Stereo Model**



- 3D perception of terrain that we see when viewing a stereo pair
- We observe parallax differences and our brain converts these to perceptions of depth.

# Satellite Images

### Satellite Images vs. Aerial Images

Satellite images have a similar idea as aerial images. They both have some of the same uses, such as serving as a basis for mapping, image interpretation, and being used as a backdrop for maps. However, there are some important differences between satellite images and aerial images.

The way in which satellite images differ in a positive way from aerial images is that satellite images typically have spectral ranges beyond aerial film. This means, that satellites can often record a much wider portion of the electromagnetic spectrum. Satellite images also have a very high perspective, much higher than aerial photography, and if you remember, the further away we are from an object, the less terrain distortion occurs. Therefore, satellite images have the potential to have much less terrain distortion than aerial images. Third, satellites can cover large areas in a single image, and have the potential to cover a much larger portion of the earth in a shorter amount of time. Negative aspects of satellite images are that they may be more expensive for a smaller area than aerial imagery. Additionally, satellite imagery may require specialized image processing software. Additionally, image acquisition must be scheduled days or weeks in advance as satellites follow a fixed flight schedule. Lastly, and this is changing rapidly, satellites typically have a lower effective resolution than aerial imagery, but are quickly catching up.

Satellite Images vs. Aerial Images

**Important Factors to Consider**

| Satellite | Aerial |
|---|---|
| Have spectral ranges beyond aerial film. | May require specialized image processing |
| Very high perspective<br>○ Reduces terrain distortion | Image acquisition must be scheduled days or weeks in advance<br>○ Follow fixed flight schedules |
| Covers large areas:<br>○ High cost for small area coverage. | Lower effective resolution |

## Basic Principles

Scanners on satellites, which are the mechanisms to record reflected energy, may operate a little differently than the cameras on aerial platforms. For instance, scanners and satellites can often be pointed in the size of the area viewed by each scanner which is known as the instantaneous field of view. Using the scanners, we can build 2-D images of a surface by recording reflected energy much like we do with aerial imagery. Additionally, satellites collect enough endlap that a three-dimensional image can later be created. To give you an idea of the capabilities of current satellites in use today, three remote sensing satellites will now be discussed.

### Ikonos

The Ikonos satellite is considered a high resolution remote sensing satellite. It was launched in September 1999, and provides 1 meter resolution panchromatic infrared images, and 4 meter visible/infrared images. The Ikonos scanner can cover a 13 km swath has askance back and forth across years. It takes between one to three days for the Ikonos satellite to revisit the same location on earth.

This is an Ikonos image covering a small portion of Athens, Greece. As you see, the resolution of the satellite is quite impressive considering that it is orbiting the Earth from space.

### Landsat

The Landsat satellite was launched in July of 1972 and initially was a system designed to gather data about the Earth's land surface. There are three imaging scanners on board Landsat satellites: video camera, multispectral scanner, and a thematic/ an enhanced thematic mapper. The Landsat satellites revisit the same location on earth between 16 to 18 days.

This is a Landsat image of Vancouver. The different types of objects that cover the land are easy to differentiate on this Landsat image.

**Key Facts**

**Landsat**

- Launched July 1972
- First system designed to gather data about Earth's land surface
- Three imaging scanners onboard
  - Video camera
  - Multispectral scanner
  - Thematic Mapper
  - Enhanced Thematic Mapper
- Revisit time is 16 - 18 days

### MODIS

The MODIS satellite collects data in a range of wavelengths from the visible through the thermal infrared portion of the electromagnetic spectrum. The MODIS satellite has a resolution between 250 m and 1 km, a very low resolution satellite. However, the MODIS satellite is very quick as it revisits the same location on earth between 1 to 2 days at a 1 km resolution.

This satellite image is an example of the MODIS satellite image. As the resolution of the image is so low, we can focus more on the physical aspects of the earth, and get a gorgeous view.

## Which to use Satellite or Aerial Images?

So in the end, when should you use satellite imagery and when should you use aerial imagery?

You need to take into account the following factors:

- What kind of spatial resolution is required? If you need a very high resolution photo, perhaps an aerial imagery is more appropriate. Conversely, if you want to focus more on large features of the earth, perhaps a satellite view would be more appropriate. Satellite images can capture a larger portion of the surface in a single frame than a fleet of airplanes.

- You must consider the price. For small areas of coverage, it is probably more economical to hire an aerial photography company. For very large area, perhaps a satellite is more economical. The next consideration is time constraints. Satellites do have sometimes lengthy revisit times that you may not want to wait on. This is where it is easier to put an airplane in the air to get coverage quickly.

- You must consider post processing. If you are covering a large area with aerial imagery, and you want to show a unified image, you may have to spend a large amount of processing time to mosaic, or paste the photos together into a single larger photo. For the spectral requirements, true color, and infrared is common for aerial imagery, but for satellite imagery they tend to also cover a larger portion of the electromagnetic spectrum which you may require.

- The last consideration is digital versus film format. Satellites can only offer data in a digital format; where as aerial images can be provided in either a film or digital format.