Lesson 1: Reviewing the Basics of Geospatial Data

## INTRODUCTION

## This lesson includes a review of geospatial data basics. You will learn about attributes related to geospatial data and view examples to further your understanding. Vector data and raster data models will be discussed and exhibited. You will gain an understanding of the various types of storage formats including shapefiles, rasters, and geodatabases and you will also have the opportunity to learn about basic cartographic and data presentation techniques. Coordinate systems and map projections are detailed within the lesson and include examples to help aid you in understanding these concepts.

## LESSON OBJECTIVES

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

1. Identify basic geospatial data elements.

2. Explain the various coordinate systems and their importance.

3. Differentiate vector and raster data formats.

## LEARNING SEQUENCE

|  |  |
| --- | --- |
|  | |
| Required Reading | Read the following:  Reviewing the Basics of Geospatial Data   * Attributes * Vector Data Models * Raster Data Model * Storage Formats * Coordinate Systems and Map Projections * Map Design and Map Elements |
| Assignments | Complete the following:   * Lab: Reviewing the Basics of Geospatial Data * Quiz: Reviewing the Basics of Geospatial Data |

## INSTRUCTION

**Review of the Basics of Geospatial Data**

## Types of Spatial Phenomena

If we are going to simplify reality and store the concept of reality in a geospatial data model, we need to understand the way in which spatial phenomena are structured. There are two types of spatial phenomena that we are going to discuss: discrete and continuous.

Discrete

A discrete spatial phenomenon is anything that exists that is individually distinguishable. It has well-defined boundaries and it is easy to see where it begins and ends. It does not exist between where we observe it to be. For example, a stream or lake is both easy to individually distinguish on the earth. The same thing goes for the roads. Roads have well-defined boundaries and are each easily distinguishable as individual roads.

**Streams and Lakes**  **Roads**

|  |  |
| --- | --- |
|  |  |

Continuous

A continuous spatial phenomenon is something that exists between our observations. It is data of a continuous nature that cannot be isolated as an individual. It is perhaps best explained through examples.

For instance, temperature is a continuous phenomenon that gently varies throughout space. A temperature reading at a single location does not represent a well-defined location where it is exactly that temperature. Instead, the temperature reading must be put in the context of the surrounding area as part of the larger "surface" of temperature readings; temperatures should be considered as part of a larger continuous spatial phenomenon.

Elevation is another example of a continuous spatial phenomenon. An elevation reading is only a single point in a larger surface.

**Temperature**   **Elevation**

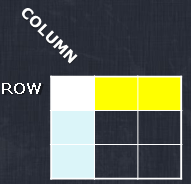
|  |  |
| --- | --- |
|  |  |

**Attributes**

## Attribute Data

Attributes are the non-spatial characteristics that describe spatial entities.

Attributes are commonly arranged in tables where a row is equivalent to one entity and a column is equivalent to one attribute or descriptor of that entity.

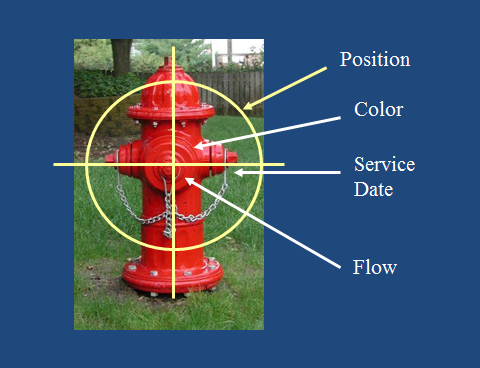


**Row** = 1 entity

**Column** = 1 attribute

Typically, each row relates to a single object in a spatial data model. It is also typical for each object to have multiple attributes that describe the object. All attributes are often displayed in a table format. Attributes can be stored on a computer using a flat file formator in a database management system.

For example, consider that we have a spatial data model that stores the location of fire hydrants. In order for each object to represent a fire hydrant, we would need to store their positions. In addition to positional information, we would also store attributes that describe those fire hydrants.

In this example, we are storing color, service, date, and flow as three attributes that describe this particular fire hydrant at this particular position on Earth. The position, color, service, date, and flow will be stored as one row in an attribute table that will contain four columns because there are four descriptors for this fire hydrant.

## Attribute Data Types

Computers fundamentally “think” differently than humans. While humans see numbers, letters, pictures, and sounds, a computer only sees zeros and ones, which is known as binary data. Therefore, we need a way to translate the numbers, sounds, and videos, as humans know it, to a form in which a computer can understand and store the information.

Computer scientists have created data structures that can be used to translate our information into a format which a computer can store in its memory. This data structure is known as a data type.

There are four typical data types that we use in GIS: integer, float/real, text/string, and date. In order to use the computer’s memory most efficiently, it is important that we specify which data type we are going to use to store information in the computer’s memory. It is important to let the computer know which operations are allowed for each data point stored in that memory location using a specific data type.

|  |
| --- |
| **Key Facts**  Attribute Data Types   * Attributes are stored in computer memory. * The *data type* of the attribute needs to be specified for efficient use of memory and determination of operation applicability. * There are four *typical* data types:   + Integer   + Float/Real   + Text/String   + Date |

**Vector Data Models**

## What are Vector Data Models?

A vector data model defines discrete objects. Examples of discrete objects are fire hydrants, roads, ponds, or a cadastral. A vector data model is broken down into three basic types: points, lines, and polygons. All three types of vector data are composed of coordinates and attributes.

Points

A point uses a single coordinate pair to define its location. Points are considered to have no dimension even though they may have a real world dimension. For the purposes of a GIS, no dimension is assumed. Each point has associated attribute information, and the information is attached to the center of the point. Examples of spatial phenomenon that would be modeled well as points are light poles, manhole covers, and crime locations.

Lines

A line vector type is defined by an ordered set of coordinates. Each line, and curve, is made up of multiple line segments, however, on occasion, curved lines are represented mathematically. There are two words that we need to define when discussing lines: a node, and a vertex.

A node is where a line begins or ends. A vertex is where a line changes direction. The smallest possible line will have two nodes, a start node, and an end node. Longer lines will have at least two nodes, and many vertices in between where the line changes direction. Attributes may be attached to the entire line, individual node, or individual vertices, therefore, each line may have multiple rows of attributes in the attribute table.

For example, if a line represents a road, each road segment between two intersections may have its own address information, such as the start address and the end address for that block. An intersection may have an attribute that describes where the intersection has a stop sign or stoplight. The other option is for the entire line to have one row of attributes no matter how complex the line. Examples of spatial phenomenon that are modeled well by lines are roads, pipelines, outlines of objects, and power lines.

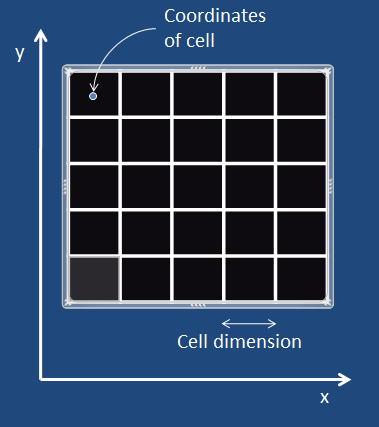
Polygons

The last vector data type is the polygon. A polygon is formed by a set of connected lines where the start and end point have the same coordinate. Because the start and end point have the same coordinate, the polygon will close and will have an interior region. Attribute information is attached to the center of the polygon no matter how complex the polygon. Examples of spatial phenomenon modeled well by polygons are lakes, cities, tree stands, and political boundaries.

**Raster Data Model**

What is a Raster Data Model?  
The raster data model best represents continuous objects such as temperature or elevation. A raster can take the form of a regular set of cells, like pixels in a photograph or it can appear as cells arranged in a grid pattern, which is also referred to as a matrix.

Each cell in the raster contains a single value, and the coordinate of each cell of the raster refers to the center of the cell. Therefore, the single value stored in each cell of the raster, applies to the entire cell in the raster matrix. Each cell can be defined by a cell dimension such as the cell width and height. Often, cells in a raster are square, so the cell width and cell height will be the same.



## Raster Resolution

In GIS, it is important that we know the resolution of the raster. The raster resolution is the cell size of each cell of the raster. Unlike how photographers represent resolution, as the number of megapixels their camera uses, in GIS, we are not as concerned about the number of cells, but of how much area on the ground each cell covers. There is a direct trade-off between resolution and file size, the cell coordinate is a center of the point cell, and the coordinate applies to the entire cell area.

It is important to reiterate that each raster cell represents a given area and the value assigned applies to the entire cell area. If there is more than one value, they can fall inside the raster cell, and then the raster cell may contain the average, central, most common, or only value covered by the cell.

|  |
| --- |
| **Key Points**  Raster Resolution   * The cell size is the resolution. * There is a trade-off between resolution and raster file size. * The cell coordinate is the center point of the cell. * The coordinate applies to the entire cell area. * Each raster cell represents a given area and the value assigned applies to the entire cell. * The raster cell value represents the average, central, most common or only value covered by the cell. |

**Storage Formats**

## Types of Storage Formats

Shapefiles

A shapefile is a common data format for storing vector data that spatially describes geometries such as points, lines, and polygons. In a shapefile, each object, or record, must also have attributes that describe them. It is important to note that a shapefile can only hold one geometric type. That means, that a shapefile can hold only points, or only lines, or only polygons.

Rasters

Rasters are stored in many different file formats. Common file formats are JPEG, TIF, bitmap, SID, and so on. Each format will have different advantages and disadvantages, therefore you should read the documentation of each one of these file formats when determining which one to store your raster file in.

Geodatabase

A geodatabase is a storage container for GIS data. Databases have many advantages over shapefiles such as speed, ability to hold vector and raster data, ability to restrict attributes to certain formats or range of values, ability to enforce rules on input, and many more.

The standard geodatabase contains two major structures that you should be aware of a feature dataset and a feature class.

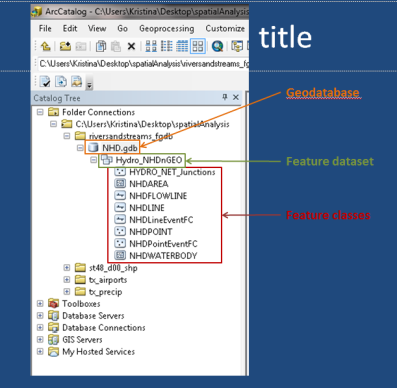
A feature data set is simply a collection of feature classes. You can think of a feature data set as a folder inside of the database. A feature data set will contain one or more feature classes and all those feature classes must have the same coordinate system. In other words, a feature data set is a logical grouping of feature classes inside of a geodatabase.

A feature class is a collection of common features and is analogous to a single shapefile. Each feature class must have the same geometry, just like shapefile.

**Example: National Hydro Data Set**

If we take a look at the National Hydro Data Set which is stored in a geodatabase, we can see how the feature data set and feature classes work together. The geodatabase is the parent level container for all of the information; the feature data set logically grouped together feature classes. Remember that all feature classes inside of the feature data set must have the same coordinate system.

**National Hydro Data Set**



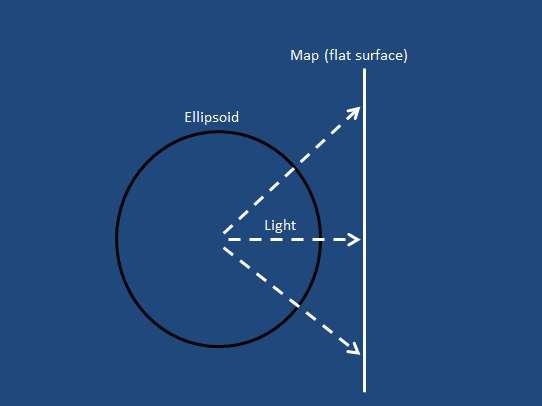
**Coordinate Systems and Map Projections**

## Why Do I Need to Know This?

It is extremely important that you gain an understanding of coordinate systems and map projections. GIS works with spatial data and with using data that needs to be placed at locations. You will need to understand the building blocks of how to define and visualize a location. It will be important to focus on map projections because although globes are great for visualization purposes, they are not practical for many uses. One reason globes are not practical is that they are not very portable. Additionally, a round Earth cannot fit without distortion onto a flat piece of paper, so you will need to understand how the earth is distorted in order to flatten it for the creation of a map.

## Map Projections

A map projection is defined as a systematic rendering of locations from the earth’s curved 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.

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.

## Projection Parameters

There are five map projection parameters:

* Standard Points and Lines
* Projection Aspect
* Central Meridian
* Latitude of Origin
* Light Source Location

A map projection property is defined as an alteration of area, shape, distance, and direction on a map projection. Map projection properties exist based on the conversion from a three dimensional object, for example the Earth, to a two-dimensional representation, such as a flat paper map. This conversion requires 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.

## 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.

**Map Design and Map Elements**

## Map Basics

A map is a way of representing our world-based on the knowledge and culture of the mapping society at a particular time in history. The purpose of a map is to transmit knowledge visually. Maps are often considered to be one of the three major modes of communication.

A cartographer is someone who designs and prepares a map for distribution. Cartographers are trained professionals in the field of cartography which is the art and science of making maps. Cartographers also study the philosophical and theoretical bases of the rules for making maps.

Cartography is a professional field that has existed for hundreds of years and it takes cartographers many years of apprenticeship to become skilled at their craft. However, with today’s computer technology allowing us to quickly manipulate, analyze, and visualize spatial information, mapmaking is now being shared by professions outside of cartography. Therefore, it is important that any user of a GIS be trained in cartography so that their maps will be effective tools for communication.

## Map Title

The title of a map should be dominant in size and is typically the largest text on a map. A good map title should focus the user’s attention on the purpose of the map. A good map title should also be brief but descriptive. Typically, a map title will include information such as where the map is focused, what information is being focused upon, and the timeframe for which the map is applicable. Map titles are typically placed at the top center of the production medium, however a map title can truly be placed anywhere so long as it is easily found by the map user. If the map is a figure in a larger document, you should not place the title on the map, but instead, place the title of the map in the caption.

|  |
| --- |
| **Key Facts**  **Map Title**   * Dominant in size * Focuses attention on map purpose * Brevity is desired * Typically includes where, what, and when * May omit if figure and caption exists |

## Map Body

The map body is the main focus of the map and contains the geographic features that are important to the message of the map. The map body is typically the largest map element on the map, and should dominate the user’s attention.

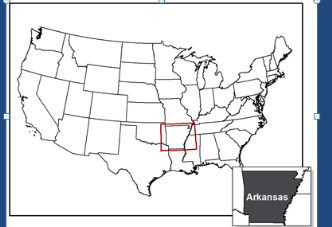
Typically, when designing a layout, the map body is placed first, and other elements are then placed around. However, do not be afraid to move or resize the map body to better accommodate other elements. Being that the map body is the element we want the user to focus on, it should be easy to find, dominant and of adequate size to effectively show the geospatial data.



## Inset Maps

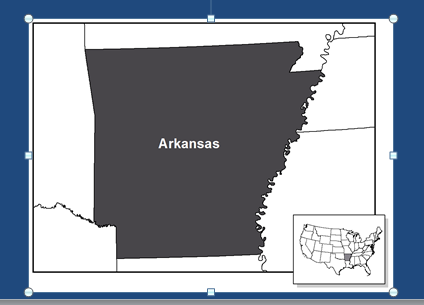
Inset maps are small ancillary maps that have a larger scale than the main map body. The role of an inset map is to show more detail in a map body of a smaller geographic area.

In the example provided, the United States of America is the main map body, and the smaller map of Arkansas in the lower right-hand corner, is the inset map which is showing a smaller area in more detail. To make it obvious to the map reader where the inset map is referring to on the main map, you should show an outline of the extents of the inset map of the main map body, or provide leader lines from the main map body to the inset map.



## Location Map

A location map is a small ancillary map that is at a smaller scale than the map body. The location map identifies a location of where the main map body is in a larger geographic context. A location map is to be used when the location of an area on the main map body is unfamiliar or not intuitive to map reader. In this illustration, the map of Arkansas is the main map body, and the smaller map of the lower right-hand corner is the location map. Similar to the inset map, there is a visual marker on the inset map that shows the map reader where the main map body is located.



## Map Scale

The map scale is used to measure linear relationships on the map. A map scale is typically included on a reference map, but is not required to be included on thematic maps. A map scale should only be included on a map when you want the user to measure the distance on the map, or the scale of the map is not intuitive to the map reader. There are three types of scales that we can place on a map: a graphic scale, a verbal scale, and a representative fraction scale.

### Graphic Scale

The graphic scale is perhaps the most common type of scale placed on maps. The graphic scale is a visual representation of the ratio at which the earth has been reduced. The graphic scale typically starts at zero, and measures out to a meaningful, typically round number. One major advantage of a graphic scale is that if the map is enlarged or reduced,say using a photocopier, then the graphic scale will scale with the enlargement or reduction, and will always be correct.



### Verbal Scale

The second type of scale is the verbal scale. The verbal scale is a statement that describes how a distance measured on the map relates to a distance measured on the ground. Again, it is important to use meaningful, typically round measurement units to make it easier for the map user to measure distances.

Example: Verbal Scale

*One inch on the map*

*equals twenty feet on the ground*

### Representative Fraction Scale

The third type of scale is the representative fraction scale sometimes referred to as the unit of scale. The representative fraction scale is a map scale that is used to represent units in centimeters, inches or feet in the form of a fraction or a ratio. This fraction or ratio, 1: x is used to indicate one unit on the map. The number to the left of the colon indicates that one unit on the map represents x units on the earth's surface indicated by the number to the right of the colon.

Example: Representative Fraction Scale

1:20

## Legend Map

The legend map element identifies unknown or unique map features succinctly. A legend may optionally have a title, or contain the title of the entire map. The legend needs to have representative symbols that are found on the map followed by a description of what each symbol represents.

The symbols on a legend should be the exact same color, shape, and size of the symbols shown the map. If the symbol on the map varies in size, the symbol in the legend should be the size of an average sized symbol on the map. Let’s look at two examples of legends.

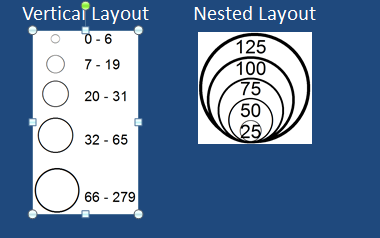
## General Reference Map

For a general reference map legend, it would display all symbols found on the map. The representative symbol should be to the left of the short description, and the legend can be organized vertically in one or more columns.

General Reference Map Legend

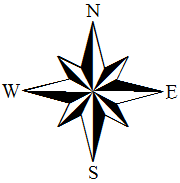
## Thematic Map

For a thematic map and in this case, a graduated symbol map, the graduated circle legend is to show how the size of the symbol changes with the value of the attribute that it is representing. Graduated symbol legends can be placed in a vertical layout, horizontal layout, or a nested layout.



Graduated Symbol Legends

## Directional Indicator

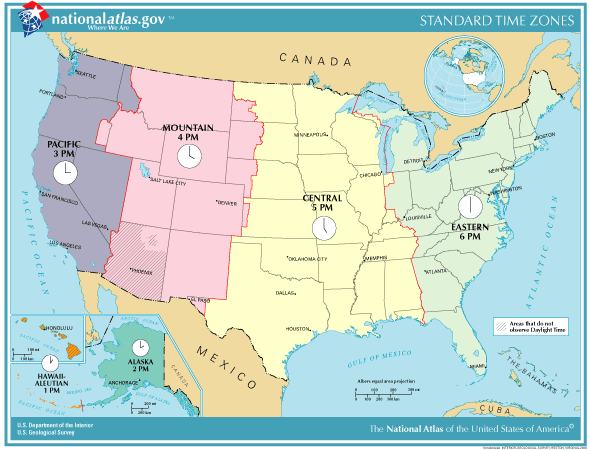
The directional indicator is often considered part of a legend, and may be placed inside the neat line around the legend, near the legend, or elsewhere on a map.

The directional indicator, commonly a north arrow, is necessary when north is not at the top of the map, or the map readers are unfamiliar with the area being displayed on the map.

Since map readers are typically familiar with orientation of large landmasses, or their own country or state, north arrows are often not necessary on small scale maps. The directional indicator should be reasonable in size, not dominant at all, but should still be easily found by the map reader.

## Labels

Labels communicate attribute or ancillary information directly on the map body, and related to map features on the map body. The purpose of label is to identify features on the map, and help users to orient themselves to the information being displayed on the map. Labels should be placed at locations that allow the map reader to easily associate each label with the feature it is labeling, and should be reasonable in size.



## Metadata

Metadata or credits are used to cite the sources of data sets used to create a map. It is also used to provide the map author’s information, the date the map was created, and other explanatory information about the creation of the map. Since you want the user’s focus to be on other aspects of the map, metadata should not be visually dominating.

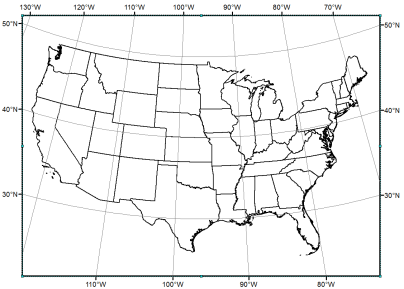
Metadata is generally placed along a bottom edge of the map and deemphasized. If the map reader wants to read the metadata, they will typically spend a little time searching for it.

|  |
| --- |
| **Key Facts**  Metadata (Credits) are used to:   * + Cite the sources of data sets used to create map.   + Provide the map author’s information.   + Include the date the map was created.   + Include any additional information needed. |

## Graticule Map

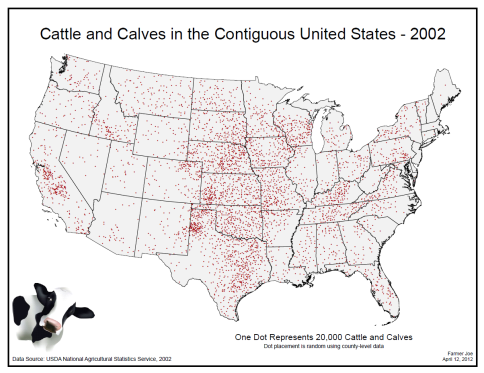
The graticule map element visually represents a coordinate system or location scheme. You should include the graticule on a map if the map reader will be referencing coordinate locations throughout the map.

You should use meaningful divisions on the graticule so that it is easy for the map reader to use it. Typically, graticules are omitted from thematic maps as the purpose of a thematic map is not to measure, but to look at spatial distributions and patterns of data.



## Neatline

The neatline is considered to be the frame of the map and should encapsulate the map and map elements if needed. The goal of the neatline is to provide a nice, clean frame for the map to live within, and to separate the map from surrounding items on the medium. The neatline is used to direct user’s eyes to the center of the map. Generally neatlines should not be visually dominant, but large enough so that the eye can use it as a frame.



## Ancillary Text and Objects

Ancillary text and/or objects are additional supporting information which provides a greater understanding of the topic of the map. A few examples of ancillary text or objects are text, pictures, sounds, movies, and graphs.

A couple of common reasons to include ancillary text on a map are to indicate data manipulation pertinent to the interpretation of the map and to indicate special cases or missing data.

|  |
| --- |
| **Key Facts**  Ancillary text and objects are used to:   * Provide a greater understanding of the topic of the map. * Help with the interpretation of the map. * Highlight special cases or missing data. |

## Map Elements: Best Practices

Best practices dictate that map elements are positioned and sized in accordance with their importance. The most significant items should be roughly at the center of the map or placed at the top of the page. This is typically why the title is at the top of the page, and the map body is in the center of the page.

Map elements should use as much space as possible within the neatline so that white space is reduced. Additionally, the map elements should be placed around the map so that the map has a visual balance that is pleasing to the eye.

|  |
| --- |
| **Key Facts**  Map elements should be:   * Positioned and sized based on their importance. * Used to reduce white space. * Balance the map layout. |

## SUMMARY

In this lesson you reviewed the basics of geospatial data. Types of spatial phenomena were identified and explained for your understanding. Vector data models were reviewed and examples were provided in this lesson. Storage format information was covered along with coordinate systems and map projections. Map design and elements were also recapped to reinforce your understanding of these concepts.

## ASSIGNMENTS

1. Lab: Reviewing the Basics of Geospatial Data

2. Quiz: Reviewing the Basics of Geospatial Data