

GIS CAPACITY BUILDING EXPERT FOR UNION OF MUNICIPALITIES IN AKKAR AND NORTH LEBANON

BEGINNER COURSE MANUAL

JANUARY 2021

GIS Center, Faculty of Engineering

University of Balamand

>>This document has been developed by the University of Balamand as part of the MASAR project implemented by ACCD and funded by the European Union.<<

TABLE OF CONTENT

TABLE OF CONTENT	3
CHAPTER 1: WHAT IS GIS?	6
1.1 What is GIS?	6
1.2 What can a GIS do?	7
1.3 New trends and directions in GIS	8
1.4 GIS Concepts	8
1.4.1 Representing real-world objects on maps	8
1.4.2 Map scale	10
1.4.2.1 What is map scale?	10
1.5 Data quality	11
1.6 Citing GIS data sources	14
CHAPTER 2: GEO-ETHICS AND DATA PRIVACY	15
2.1 Ethical issues	15
2.2 Code of Ethics for GIS Professionals by URISA	16
2.2.1 Obligations to Society	16
2.2.2 Obligations to Employers and Founders	17
2.2.3 Obligations to Colleagues and the Profession	18
2.2.4 Obligations to Individuals in Society	19
CHAPTER 3: GLOSSARY OF COMMON GIS TERMS	20
CHAPTER 4: GIS DATA TYPES	29
4.1 Spatial Data	29
4.1.1 Vector data	29
4.1.2 Raster data	29
4.1.3 Imagery	29
4.1.4 TIN	29
4.1.5 Terrain data	30
4.2 Attribute Data	30
4.3 Metadata	31
CHAPTER 5: DATA MODELS FOR GIS	32
5.1 Raster Data Models	32
5.1.1 Advantages/Disadvantages of the Raster Model	36
5.2 Vector Data Models	38
5.2.1 Vector Data Models Structures	39
5.2.2 Advantages/Disadvantages of the Vector Model	44

5.3	Satellite Imagery and Aerial Photography	45
5.3.1	Satellite Imagery	46
5.3.2	Aerial Photography	47
CHAPTER 6: ArcGIS SOFTWARE COMPONENTS		50
6.1	Introduction to ArcGIS Desktop	50
6.2	Lab: Introduction to ArcGIS	51
6.2.1	Procedure	51
6.2.1.1	Downloading your data and creating your working directory	51
6.2.1.2	Accessing class resources	52
6.2.1.3	Starting with ArcCatalog and connecting to your data:	52
6.2.1.4	Exploring the ArcGIS Help System	53
6.2.1.5	Exploring ArcMap tools	53
6.2.1.6	Displaying data in ArcMap	54
6.3	Lab: Exploring Data Structures	56
6.3.1	Procedure:	58
6.3.1.1	The ArcCatalog graphical interface	58
6.3.1.2	Coverage structure	59
6.3.1.3	Shapefile Structure	62
6.3.1.4	Creating a point shapefile from tabular data	62
CHAPTER 7: INTRO TO COORDINATE SYSTEM AND PROJECTION		66
7.1	Referencing Data to Real-World Locations Using ArcGIS	66
7.1.1	Introduction	66
7.1.2	Geographic Coordinate Systems	67
7.1.3	The Components of a Geographic Coordinate System	67
7.1.3.1	Angular units of measure	68
7.1.3.2	Prime meridian	68
7.1.3.3	Spheroid	68
7.1.3.4	Datum	69
7.1.4	Working with Data in Different Geographic Coordinate Systems	70
7.2	Projected Coordinate Systems	70
7.2.1	Map projections	71
7.2.2	Projecting Coordinates onto Developable Surfaces	72
7.2.2.1	Cylindrical projections	72
7.2.2.2	Conic projections	73

7.2.2.3	Planar projections	73
7.2.3	Projections and Distortion	74
7.2.3.1	Conformal	74
7.2.3.2	Equal Area	74
7.2.3.3	Equidistant	75
7.2.3.4	Azimuthal	75
7.2.3.5	Gnomonic	76
7.2.3.6	Compromise	77
7.2.4	Reasons for Using a Projected Coordinate System	77
7.3	Integrating Data with Different Coordinate Systems	78
7.3.1	Working with data with different coordinate systems	78
7.4	Lab: Map Projection and Coordinate Systems	79
7.4.1	Overview	79
7.4.2	Procedure	80
7.4.2.1	Exploring a Cartesian coordinate system	80
7.4.2.2	Displaying data layers of different projections in ArcMap	80
7.4.2.3	Re-projecting GIS data with ArcToolbox	81
7.4.2.4	Creating layouts with multiple data frames:	82
7.4.2.5	Examining the effect of the different projections	84
CHAPTER 8: CAD TO GIS		85
8.1	DWG to Geodatabase	85
8.2	Editing for Cleaning and QA/QC	93
8.2.1	ESRI Topology Rules in ArcGIS 10	94
8.2.1.1	Add your data to a geodatabase	94
8.2.1.2	Create topology data set	95
8.2.1.3	Add topology rules	95
8.2.1.4	Inspect your data	96
8.2.1.5	Correct your errors	96
CHAPTER 9: GEODATABASE		99
9.1	Types of geodatabases	100

CHAPTER 1: WHAT IS GIS?

1.1 What is GIS?

GIS stands for geographic information system. In practical terms, a GIS is a set of computer tools that allows people to work with data that are tied to a particular location on the earth. Although many people think of a GIS as a computer mapping system, its functions are broader and more sophisticated than that. A GIS is a database that is designed to work with map data.

A GIS is built from a collection of hardware and software components.

- **A computer hardware platform.** Due to the intensive nature of spatial data storage and processing, a GIS was once limited to large mainframe computers or expensive workstations. Today, it can run on a typical desktop personal computer.
- **GIS software.** The software varies widely in cost, ease of use, and level of functionality but should offer at least some minimal set of functions, as described in the next few paragraphs. In this book, we study one particular package that is powerful and widely used, but others are available and may be just as suitable for certain applications.
- **Data storage.** Some projects use only the hard drive of the GIS computer. Other projects may require more sophisticated solutions if large volumes of data are being stored or multiple users need access to the same data sets. Today, many data sets are stored in digital warehouses and are accessed by many users over the Internet. Compact disc writers and/or USB portable drives are highly useful for backing up and sharing data.
- **Data input hardware.** Many GIS projects require sophisticated data entry tools. Digitizer tablets enable the shapes on a paper map to be entered as features in a GIS data file. Scanners create digital images of paper maps. An Internet connection provides easy access to large volumes of GIS data. High-speed connections are preferred because GIS data sets may constitute tens or hundreds of megabytes or more.
- **Information output hardware.** A quality color printer capable of letter-size prints provides the minimum desirable output capability for a GIS system. Printers that can handle map-size output (36 in. × 48 in.) will be required for many projects.

- **GIS data.** Data come from a variety of sources and in a plethora of formats. Gathering data, assessing their accuracy, and maintaining them usually constitutes the longest and most expensive part of a GIS project.
- **GIS personnel.** A system of computers and hardware is useless without trained and knowledgeable people to run it. The contribution of professional training to successful implementation of a GIS is often overlooked. GIS software varies widely in functionality, but any system claiming to be a GIS should provide the following functions at a minimum:
 - Data entry from a variety of sources, including digitizing, scanning, text files, and the most common spatial data formats; ways to export information to other programs should also be provided.
 - Data management tools, including tools for building data sets, editing spatial features and their attributes, and managing coordinate systems and projections.
 - Thematic mapping (displaying data in map form), including symbolizing map features in different ways and combining map layers for display.
 - Data analysis functions for exploring spatial relationships in and between map layers.
 - Map layout functions for creating soft and hard copy maps with titles, scale bars, north arrows, and other map elements.

Geographic information systems are put to many uses, but providing the means to collect, manage, and analyze data to produce information for better decision making is the common goal and the strength of each GIS. This book is a practical guide to understanding and using a particular geographic information system called ArcGIS, to learn what a GIS is, what it does, and how to apply its capabilities to solve real-world problems.

1.2 What can a GIS do?

A GIS works with many different applications: land use planning, environmental management, sociological analysis, business marketing, and more. Any endeavor that uses spatial data can benefit from a GIS.

There are many applications of GIS in almost every human endeavor, including business, defense and intelligence, engineering and construction, government, health and human

services, conservation, natural resources, public safety, education, transportation, utilities, and communication.

In August 2014, the Industries section of the ESRI website at www.esri.com listed 62 different application areas in these categories, each one with examples, maps, and case studies. Instead of reading through a list here, go to the site and see the latest applications.

1.3 New trends and directions in GIS

The GIS industry has grown exponentially since its inception. Starting with the mainframe and then the desktop computer, GIS began as a relatively private endeavor focused on small clusters of specialized workers who spent years developing expertise with the software and data. Since then, the development of the Internet and the rapidly advancing field of computer hardware have been driving some significant changes in the industry.

1.4 GIS Concepts

A geographic information system (GIS) provides data structures and capabilities for storing, analyzing, managing, and publishing map data using a computer. Many different software packages have been developed to accomplish this task. Although this book focuses on a specific software product, many of the concepts are fundamental to all GIS programs, even though the details of implementation may differ. In this book, the first section of each chapter, Mastering the Concepts, focuses on these cross-cutting fundamental ideas. The second section, About ArcGIS, discusses specific details of ArcGIS Pro, a specific GIS software product published by the company Esri™ (originally the Environmental Systems Research Institute, Inc.).

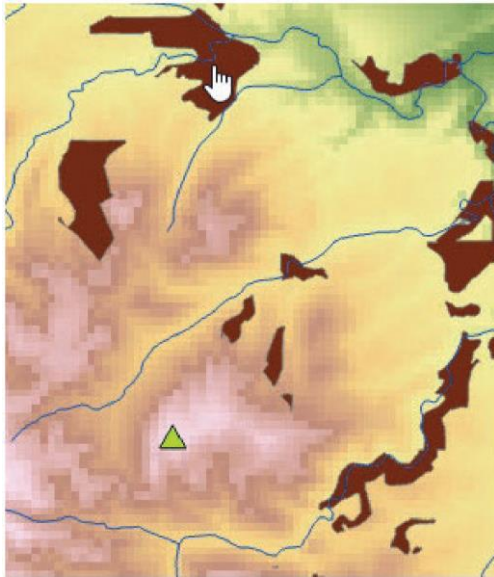
1.4.1 Representing real-world objects on maps

To work with maps on a computer requires storing different types of map data and the information associated with them. Geographic information falls primarily into two categories: discrete and continuous. Discrete data are objects in the real world with specific locations or boundaries, such as houses, cities, roads, or counties. Continuous data represent quantities that may be measured anywhere on the earth, such as temperature or elevation. Many different data

formats have been invented to encode data for use with GIS programs; however, most follow one of two approaches: the vector model, which is designed to store discrete data, or the raster model, which is designed to store continuous data. Both models are georeferenced, meaning that the information is tied to a specific location on the earth's surface using spatial (x-y) coordinates defined in a standard way. The familiar latitude and longitude values are one example. Data and software used to work with georeferenced information are encompassed by the general term geospatial.

The vector model uses three basic shapes to represent discrete objects, known as features. A point feature can represent a single location such as a well or a weather station. Line features represent linear objects like roads and rivers. A polygon feature represents a closed area such as a county or a state. Vector data sets include two parts: one or more x-y locations that place the feature on the earth's surface (its shape), and a table that stores attributes, or information about each feature, such as a state's name, abbreviation, population, and so on. A set of similar features stored together, such as the 50 states comprising the United States, is called a feature class. A feature class can store only one type of feature: points or lines or polygons, but never a combination, and the features should be the same type of thing. Counties and states, even though both are represented by polygons, should be stored in different feature classes.

The raster model breaks a geographic area into small squares known as cells or pixels, each one containing a number representing a value being measured or characterized. Some rasters store physical quantities like elevation, temperature, or precipitation. Others may store color and brightness values that make up aerial photographs, satellite images, or a scanned paper map. Even discrete features, such as land use or soil type polygons, may be represented by rasters if needed, even though the model is not optimized for that task.



This figure illustrates examples of geospatial data.

The green triangle is a point feature representing the location of a fire lookout tower. The blue line features represent streams. The brown shapes are polygons representing the preferred habitat areas for a species of snail.

The background shows a raster composed of $30\text{ m} \times 30\text{ m}$ square cells recording elevation values, with the green values representing the lowest elevations and the white cells the highest elevations.

Maps are often constructed from multiple data sets. Each data set is typically called a layer once it is placed into a map, and the layers are drawn in a hierarchical order defined by the cartographer, usually with raster or polygon layers on the bottom and point or line layers on top (so that the top layers do not cover up the bottom layers). The figure contains four different layers from bottom to top: a raster layer of elevation, a polygon layer of snail habitat (brown), a line layer portraying streams (blue), and a point layer representing the fire tower (green triangle).

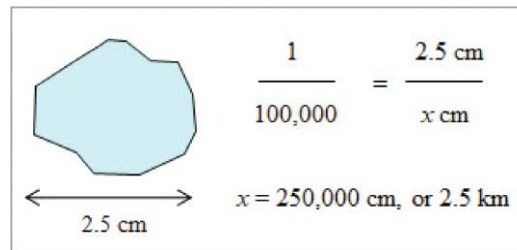
1.4.2 Map scale

The act of taking a set of GIS features with x-y coordinate values and drawing them on a screen or printing them on paper establishes a map scale. On a paper map, the scale is fixed at the time of printing. In a digital map configured for interactive display, the scale changes every time the viewer zooms in or out.

1.4.2.1 What is map scale?

Map scale is a measure of the size at which features in a map are represented. The scale is expressed as a fraction, or ratio, of the size of features represented by the map to the size of the actual objects on the ground. Because it is expressed as a ratio, it is valid for any unit of measure. For the common US Geological Survey topographic map, which has a scale of

1:24,000, one inch on the map represents 24,000 inches on the ground. The map scale can be used to determine the true size of any feature on the map, such as the width of a lake.



Measure the lake with a ruler and then set up a proportion such that the map scale equals the measured width over the actual width (x). Then solve for x. Keep in mind that the actual width and the measured width will have the same units, which can be converted if necessary.

Often people or publications refer to large-scale maps and small-scale maps. A large-scale map is one in which the ratio is large (i.e., the denominator is small). Thus, a 1:24,000 scale map has larger scale than a 1:100,000 scale map. A large-scale map shows a relatively small area, such as a quadrangle, whereas small-scale maps show bigger areas, such as states or countries.

The type of object used to represent features depends on the scale of the map. A river would be represented as a line on a map of the United States because at that scale it is too small for its width to encompass any significant area on the map. If one is viewing a USGS topographic map, however, the river encompasses an area and is represented as a polygon.

1.5 Data quality

Representing real-world objects as points, lines, polygons, or rasters always involves some degree of generalization, or simplifying the data for digital storage, such as turning a house into a rectangular polygon or even a point. No data file can capture all the spatial or attribute qualities of any object. The degree of generalization increases as the map scale decreases. On a standard topographic map, a river has a width and can be modeled as a polygon with two separate banks.

A city would be shown as a polygon area. For a national map, however, the river would simply be shown as a line, and a city would be shown as a point. Even a detailed representation of an object is not always a true characterization. Rivers and lakes can enlarge in size during spring flooding or shrink during a drought. The boundary of a city changes as the city grows. Users

of GIS data must never forget that the data they collect and use will contain flaws and that the user has an ethical and legal responsibility to ensure that the data used for a particular purpose are appropriate for the task. When evaluating the quality of a data set, geospatial professionals consider the following aspects.

Geometric accuracy describes how close the x-y values of a data set correspond to the actual locations on the earth's surface. Geometric accuracy is a function of the source scale and of how the data were captured. Surveying is one of the most accurate ways to position features. GPS units have an accuracy that ranges from centimeters to tens of meters. Maps derived from aerial photography or satellite imagery can vary widely in geometric quality based on factors such as the scale of the image, the resolution of the image, imperfections and distortions in the imaging system, and the types of corrections applied to the image. In the following figure, notice that the vector road in white is offset in places from the road as it appears in the aerial photo. These differences can arise from digitizing errors, geometric distortions from the camera or satellite, or other factors. Moreover, not every feature can be as precisely located as a road. Imagine mapping the land-cover types: forest, shrubland, grassland, and bare rock.



Where does one draw the line between shrubland and grassland? At what point does the shrubland become forest? Six people given this photo would create six slightly different maps. Some boundaries would match closely; others would vary as each person made a subjective decision about where to place each boundary.

Thematic accuracy refers to the attributes stored in the table. Some types of data are relatively straightforward to record, such as the name of a city or the number of lanes in a road. Even in this situation, the value of a feature might be incorrectly recorded. Other types of information can never be known exactly. Population data, for example, are collected through a process of surveying and self-reporting that takes many months. It is impossible to include every person.

Moreover, people are born and die during the survey process, or they are moving in and out of towns. Population data can never be more than an estimate. It is important to understand the limitations and potential biases associated with thematic data.

Resolution refers to the sampling interval at which data are acquired. Resolution may be spatial, thematic, or temporal. Spatial resolution indicates at what distance interval measurements are taken or recorded. What is the size of a single pixel of satellite data? If one is collecting GPS points by driving along a road, at what interval is each point collected? Thematic resolution can be affected by using categories rather than measured quantities: if one is collecting information on the percent crown cover in a forest, is each measurement reported as a continuous value (32%) or as a classified range (10–20%, 20–30%)? Temporal resolution indicates how often measurements are taken. Census data are collected every 10 years. Temperature data taken at a climate station might be recorded every 15 minutes, but it might also be reported as a monthly or yearly average.

Precision refers to either the number of significant digits used to record a measurement or the statistical variation of a repeated single measurement. Many people confuse precision with accuracy, but it is important to understand the distinction. Imagine recording someone's body temperature with an oral digital thermometer that records to a thousandth of a degree and getting the value of 99.894 degrees Fahrenheit. This measurement would be considered precise.

However, imagine that the reading is taken immediately after the person drinks a cup of hot coffee, which causes the thermometer to measure a value higher than true body temperature. Thus, the measurement is precise, but it is not accurate.

Logical consistency evaluates whether a data model or data set accurately represents the real-world relationships between features. In the real world, for example, two adjacent states share a common boundary that is exactly the same.



In a database, however, the states might be stored as two separate features with slightly different boundaries. Lines representing streets should connect if the roads they represent meet.

A line or a polygon boundary should not cross over itself. A county should not extend past the boundary of its state. It takes special effort to create and manage feature relationships within and between data sets, so not all data sets are logically consistent.

Evaluating the quality of a data set can be difficult, especially if the data were created by someone else. Professionals who create data usually also provide metadata, which store information about the data set, such as where it came from, how it was developed, who assembled it, how precise it is, and whether it can be given to another person. The user can use the metadata to decide whether a particular data set is suitable for a given purpose. Metadata may consist of a brief description of the most critical information, or they may include many pages of detail.

1.6 Citing GIS data sources

Ethical and professional considerations require that any map, publication, or report should cite the data source(s) used and give proper credit to the originators of the data. The metadata or the site from which the data were acquired are good sources of information for citations. The best practice is to record the citation when the data are obtained so the information is available when needed. Generally one cites only data that are publically available (free or purchased). Data created internally within the workplace need not be cited, although often the company name or logo will appear on the map. A data set provided once in response to a personal request should be cited as a personal communication. On a map, only one citation need be included for multiple data sets from the same source, but a report should list all data sets from each source.

Keep in mind that the place where data are found may not be their original source. A GIS administrator may have placed often-used data sets, such as Esri™ Data and Maps, on a university file server for easy access, but the fact that they were obtained internally does not free students from the need to cite them, and the original source should be cited, not the local server.

CHAPTER 2: GEO-ETHICS AND DATA PRIVACY

“Everything we do has an impact on others. Some of those impacts are negative and some are positive. Ethics is the philosophical framework that is used to maximize the good we do and minimize the harm”. Information is a source of power and, increasingly, the key to prosperity among those with access to it. Consequently, developments in information systems also involve social and political relationships and so make important ethical considerations in how information is used.

GIS should be used to enhance the quality of life, promote equity in access to knowledge for all members of society, reduce socio-economic gaps between members of the social system (or at least not widen them), and other "good purposes." The limitations of such reasoning and goals become readily evident when one attempts to apply such principles in everyday practice. Numerous gray areas exist in the use of GIS and determining what constitutes a beneficial versus a detrimental consequence is a value-laden judgment. In many instances, "fair, just, and equitable" conduct and results are not obvious. They often depend on the perspectives of those affected by the use of the information system.

2.1 Ethical issues

Discussing ethical issues for spatial information, GIS systems, and the GIS profession, rules of good behavior for GI and GIS associations are needed to define the term 'ethics'. Ethics refers to principles of human conduct, or morals, and to the systematic study of such human values, or the study of theories of conduct and goodness, and of the meanings of moral terms, which is called moral philosophy, the study of theories of conduct and goodness and the meanings of moral terms. These set of rules help to guide the human actions of an individual human being so that they are consistent with his or her values. Ethics concerns itself with questions such as when is an action ‘right’ or ‘wrong’ and what standards separate ‘good’ from ‘bad’. Whatever, the term "ethics" could be used in several different ways:

Ethics means the study of morals. It is also the name for that branch of philosophy concerned with the nature of morals and moral evaluation - e.g. what is right and wrong, virtuous or vicious, and beneficial or harmful (to others).

Formally, ethics is the study of morals and values. Morals - are rules that govern behavior. When we ask what is right and what is wrong, these are questions of morality. Values - are states of affairs that are desired by and for people.

2.2 Code of Ethics for GIS Professionals by URISA

Urban and Regional Information Systems Association-URISA in April 2003 defined the Code of Ethics for GIS Professionals. This code is based on the ethical principle of always treating others with respect and never merely as means to an end.

2.2.1 *Obligations to Society*

The GIS professional recognizes the impact of his or her work on society as a whole, on subgroups of society including geographic or demographic minorities, on future generations, and inclusive of social, economic, environmental, or technical fields of endeavor. Obligations to society shall be paramount when there is a conflict with other obligations. Therefore, the GIS professional will:

- i. Do the Best Work Possible
 - Be objective, use due care, and make full use of education and skills.
 - Practice integrity and not be unduly swayed by the demands of others.
 - Provide full, clear, and accurate information.
 - Be aware of the consequences, good and bad.
 - Strive to do what is right, not just what is legal.
- ii. Contribute to the Community to the Extent Possible, Feasible and Advisable
 - Make data and findings widely available.
 - Strive for broad citizen involvement in problem definition, data identification, analysis, and decision-making.
 - Donate services to the community.
- iii. Speak Out About Issues
 - Call attention to emerging public issues and identify appropriate responses based on personal expertise.

- Call attention to the unprofessional work of others. First, take concerns to those persons; if satisfaction is not gained and the problems warrant, then additional people and organizations should be notified.
- Admit when a mistake has been made and make corrections where possible.

2.2.2 *Obligations to Employers and Founders*

The GIS professional recognizes that he or she has been hired to deliver needed products and services. The employer (or founder) expects quality work and professional conduct. Therefore, the GIS professional will:

i. Deliver Quality Work

- Be qualified for the tasks accepted.
- Keep current in the field through readings and professional development.
- Identify risks and the potential means to reduce them.
- Define alternative strategies to reach employer/founder goals, if possible, and the implications of each.
- Document work so that others can use it. This includes metadata and program documentation.

ii. Have a Professional Relationship

- Hold information confidential unless authorized to release it.
- Avoid all conflicts of interest with clients and employers if possible, but when they are unavoidable, disclose that conflict.
- Avoid soliciting, accepting, or offering any gratuity or inappropriate benefit connected to a potential or existing business or working relationship.
- Accept work reviews as a means to improve performance.
- Honor contracts and assigned responsibilities.
- Accept the decisions of employers and clients, unless they are illegal or unethical.
- Help develop security, backup, retention, recovery, and disposal rules.
- Acknowledge and accept rules about the personal use of employer resources. This includes computers, data, telecommunication equipment, and other resources.
- Strive to resolve differences.

iii. Be Honest in Representations

- State professional qualifications truthfully.
- Make honest proposals that allow the work to be completed for the resources requested.
- Deliver an hour's work for an hour's pay.
- Describe products and services fully.
- Be forthcoming about any limitations of data, software, assumptions, models, methods, and analysis.

2.2.3 *Obligations to Colleagues and the Profession*

The GIS professional recognizes the value of being part of a community of other professionals. Together, we support each other and add to the stature of the field. Therefore, the GIS professional will:

i. Respect the Work of Others.

- Cite the work of others whenever possible and appropriate.
- Honor the intellectual property rights of others. This includes their rights in software and data.
- Accept and provide fair critical comments on professional work.
- Recognize the limitations of one's own knowledge and skills and recognize and use the skills of other professionals as needed. This includes both those in other disciplines and GIS professionals with deeper skills in critical sub-areas of the field.
- Work respectfully and capably with others in GIS and other disciplines.
- Respect existing working relationships between others, including employer/employee and contractor/client relationships.
- Deal honestly and fairly with prospective employees, contractors, and vendors.

ii. Contribute to the Discipline to the Extent Possible

- Publish results so others can learn about them.
- Volunteer time to professional educational and organizational efforts: local, national, or global.

- Support individual colleagues in their professional development. Special attention should be given to underrepresented groups whose diverse backgrounds will add to the strength of the profession.

2.2.4 *Obligations to Individuals in Society*

The GIS professional recognizes the impact of his or her work on individual people and will strive to avoid harm to them. Therefore, the GIS professional will:

i. Respect Privacy

- Protect individual privacy, especially about sensitive information.
- Be especially careful with new information discovered about an individual through GIS-based manipulations (such as geocoding) or the combination of two or more databases.

ii. Respect Individuals

- Encourage individual autonomy. For example, allow individuals to withhold consent from being added to a database, correct information about themselves in a database, and remove themselves from a database.
- Avoid undue intrusions into the lives of individuals.
- Be truthful when disclosing information about an individual.
- Treat all individuals equally, without regard to race, gender, or other personal characteristics not related to the task at hand.

Codes of ethics are created in response to actual or anticipated ethical conflicts. A few of the guidelines that are unique to the GIS profession include the encouragement to make data and findings widely available, to document data and products, to be actively involved in data retention and security, to show respect for copyright and other intellectual property rights, and to display concern for the sensitive data about individuals discovered through geospatial or database.

CHAPTER 3: GLOSSARY OF COMMON GIS TERMS

- Analysis: The process of identifying a question or issue to be addressed, modeling the issue, investigating model results, interpreting the results, and possibly making a recommendation.
- Annotation: Descriptive text used to label coverage features. It is used for display, not for analysis. One of the feature classes in a coverage used to label other features. Information stored for annotation includes a text string, the location at which it is displayed, and a text symbol (color, font, size, etc.) for display.
- Attribute: A characteristic of a feature in a Geographic Information System (GIS). Each identifiable feature has attributes. One common attribute of all geographic features is its position. Other attributes depend on the type of feature. Example: a road may have a name or designation number, pavement type, width, number of lanes, etc. Each attribute has a range of possible values called its domain. A column in a database table.
- Attribute table: A tabular file containing rows and columns. In a GIS, attribute tables are associated with a class of geographic features, such as wells or roads. Each row represents a geographic feature. Each column represents one attribute of a feature, with the same column representing the same attribute in each row.
- Cartesian coordinate system: A two-dimensional, planar coordinate system in which x measures horizontal distance and y measures vertical distance. Each point on the plane is defined by an x,y coordinate. Relative measures of distance, area, and direction are constant throughout the Cartesian coordinate plane.
- Column: The vertical dimension of a table. A column has a name and a data type applied to all values in the column.
- Connectivity: The topological identification of connected arcs by recording the from- and to-node for each arc. Arcs that share a common node are connected.
- Contiguity: The topological identification of adjacent polygons by recording the left and right polygons of each arc. continuous data A surface for which each location has a specified or derivable value. Typically represented by a grid, tin, or lattice (e.g., surface elevation).
- Contour: A line connecting points of equal surface value. contour interval The difference in surface values between contours.
- Coordinate: A set of numeric quantities that describe the location of a point in a geographic reference system. A coordinate pair describes the location of a point or node in two dimensions

(usually x-y), and a coordinate triplet describes a point in three dimensions (x-y-z). A series of points (two or more) is used to describe lines and the edges of polygons or areas. Coordinates represent locations on the Earth's surface relative to other locations.

- Coordinate system: A reference system used to measure horizontal and vertical distances on a planimetric map. A coordinate system is usually defined by a map projection, a spheroid of reference, a datum, one or more standard parallels, a central meridian, and possible shifts in the x- and y-directions to locate x,y positions of point, line, and area features. A common coordinate system is used to spatially register geographic data for the same area.
- Data capture: The collection of GPS attributes and position information for geographic features. Database: A logical collection of interrelated information, managed and stored as a unit, usually on some form of mass- storage system such as magnetic tape or disk. A GIS database includes data about the spatial location and shape of geographic features recorded as points, lines, areas, pixels, grid cells, or tins, as well as their attributes.
- Database design: The formal process of analyzing facts about the real world into a structured database model. Database design is characterized by the following phases: requirement analysis, logical design and physical design.
- Data integrity: Maintenance of data values according to data model and data type. For example, to maintain integrity, numeric columns will not accept alphabetic data.
- Data model: The result of the conceptual design process. A generalized, user-defined view of the data related to applications. A formal method of describing the behavior of the real-world entities. A fully developed data model specifies entity classes, relationships between entities, integrity rules and operations on the entities.
- Data set: A named collection of logically related data items arranged in a prescribed manner. data type: The characteristic of columns and variables that defines what types of data values they can store. Examples include character, floating point and integer.
- Database Management System (DBMS): A set of computer programs for organizing the information in a database. A DBMS supports the structuring of the database in a standard format and provides tools for data input, verification, storage, retrieval, query, and manipulation. Examples of database management systems include Access, SQL Server, Oracle, Informix, and DBASE.
- Datum: A set of parameters and control points used to accurately define the three-dimensional shape of the Earth (e.g., as a spheroid). The datum is the basis for a planar coordinate system. For example, the North American Datum for 1983 (NAD83) is a common

datum for map projections and coordinates within the United States and throughout North America.

- Digital Elevation Model (DEM): A digital representation of a continuous variable over a two-dimensional surface by a regular array of z-values referenced to a common datum. Digital elevation models are typically used to represent terrain relief. Also referred to as 'digital terrain model' (DTM). An elevation database for elevation data by map sheet from the National Mapping Division of the U.S. Geological Survey (USGS). The format of the USGS digital elevation data sets. USGS DEMs are produced by the Survey Branch of the United States Department of the Interior, consisting of a regular array of elevations referenced in the Universal Transverse Mercator (UTM) coordinate system. These data correspond to the standard 1:24,000-scale 7.5 x 7.5- minute quadrangles or 1:250,00
- Digitize: To encode geographic features in digital form as x,y coordinates. The process of using a digitizer to encode the locations of geographic features by converting their map positions to a series of x,y coordinates stored in computer files. Pushing a digitizer button records an x,y coordinate. A digitized line is created by recording a series of x,y coordinates.
- Ellipsoid: In geometric terms, a closed surface of which all planar sections are ellipses. In GIS and mapping practices, an ellipsoid is a specific mathematical representation of the earth that more closely approximates the shape of the surface than a sphere does.
- Feature: In a GIS, a physical object or location of an event. Features can be points (a tree or a traffic accident), lines (a road or river), or areas (a forest or a parking lot).
- Feature attribute table: A table used to store attribute information for a specific coverage feature class. feature class: A classification describing the format of geographic features and supporting data in a coverage. Coverage feature classes for representing geographic features include point, arc, node, route-system, route, section, polygon and region. One or more coverage features are used to model geographic features; for example, arcs and nodes can be used to model linear features such as street centerlines. The tic, annotation, link, and boundary feature classes provide supporting data for coverage data management and viewing. The conceptual representation of a geographic feature. When referring to geographic features, feature classes include point, line, area, and surface.
- Field: In a database, another term for column.
- Geodatabase: An object-based GIS data model developed by ESRI for ArcGIS that stores each feature as rows in a table. Personal geodatabases store data in a Microsoft Access .mdb

file. Corporate geodatabases store data in a DBMS such as SQL Server or Oracle. This data structure supports rules-based topology and allows the user to assign behavior to data.

- Geoid: The particular equipotential surface which coincides with mean sea level (MSL), and which may be imagined to extend through the continents. This surface is everywhere perpendicular to the force of gravity.
- Global Positioning System (GPS): A system of satellites and receiving devices used to compute positions on the Earth.
- Grid: Two sets of parallel lines intersecting at right angles in a plane coordinate system. A geographic data model representing information as an array of equally sized square cells arranged in rows and columns. Each grid cell is referenced by its geographic x,y location. See also raster and grid cell.
- Grid cell: A discretely uniform unit that represents a portion of the Earth, such as a square meter or square mile. Each grid cell has a value that corresponds to the feature or characteristic at that site, such as a soil type, census tract, or vegetation class.
- Layer: A thematic set of spatial data described and stored in a digital database or map library. Layers organize a database or map library by subject matter (e.g., soils, roads, and wells). Conceptually, layers in a database or map library environment are exactly like coverages.
- Legend: The reference area on a map that lists and explains the colors, symbols, line patterns, shadings, and annotation used on the map. The legend often includes the scale, origin, orientation, and other map information. The symbol key used to interpret a map.
- Map: An abstract representation of the physical features of a portion of the Earth's surface graphically displayed on a planar surface. Maps display signs, symbols, and spatial relationships among the features. They typically emphasize, generalize, and omit certain features from the display to meet design objectives (e.g., railroad features might be included in a transportation map but omitted from a highway map).
- Map extent: The rectangular limits (x_{min}, y_{min} and x_{max}, y_{max}) of the area of the Earth's surface displayed using a GIS. Map extent is specified in the coordinate system of the coverage or other geographic data set used. Typically, the extent of the geographic database (or a portion of it defined by a zoomed-in view) defines the map extent for display. The geographic extent of a geographic data set specified by the minimum bounding rectangle (i.e., x_{min}, y_{min} and x_{max}, y_{max})
- Map projection: A mathematical model that transforms the locations of features on the Earth's surface to locations on a two- dimensional surface. Because the Earth is three-

dimensional, some method must be used to depict a map in two dimensions. Some projections preserve shape; others preserve accuracy of area, distance, or direction. See also coordinate system. Map projections project the Earth's surface onto a flat plane. However, any such representation distorts some parameter of the Earth's surface be it distance, area, shape, or direction.

- Map query: The process of selecting information from a GIS by asking spatial or logical questions of the geographic data. Spatial query is the process of selecting features based on location or spatial relationship (e.g., select all features within 300 feet of another; point at a set of features to select them). Logical query is the process of selecting features whose attributes meet specific logical criteria (e.g., select all polygons whose value for area is greater than 10,000 or select all streets whose name is 'Main St.'). Once selected, additional operations can be performed, such as drawing them, listing their attributes or summarizing attribute values.
- Map scale: The reduction needed to display a representation of the Earth's surface on a map. A statement of a measure on the map and the equivalent measure on the Earth's surface, often expressed as a representative fraction of distance, such as 1:24,000 (one unit of distance on the map represents 24,000 of the same units of distance on the Earth). 'Large scale' refers to a large fractional value such as 1/24,000. A large-scale map shows a small geographic area in greater detail. 'Small scale' refers to a smaller fractional value such as 1/1,000,000. A small-scale map shows a large geographic area in less detail. map units The coordinate units in which a geographic data set (e.g., a coverage) is stored in a GIS. Map units can be inches, centimeters, feet, meters, or decimal degrees.
- Metadata: Data about the content, quality, condition, and other characteristics of data.
- Orthophoto: Photograph which results from processing aerial photographs to remove distortion and displacement. Like maps, they have one scale (even in varying terrain) and like photographs they show the terrain in actual detail (not by lines and symbols). True distances, angles and areas may be measured directly without making corrections for image displacement.
- Point: A single x,y coordinate that represents a geographic feature too small to be displayed as a line or area; for example, the location of a mountain peak or a building location on a small-scale map. A coverage feature class used to represent point features or to identify polygons. It is not recommended to have point and polygon features in the same coverage. When representing point features, the x,y location of the label point is also the location of the feature. When identifying polygons, the label point can be located anywhere within the polygon.

- Polygon: A coverage feature class used to represent areas. A polygon is defined by the arcs that make up its boundary and a point inside its boundary for identification.
- Projection file (.prj): A coverage file that stores the parameters for the map projection and coordinate system of a geographic data set (e.g., a coverage). A text file containing input and output projection parameters that can be used to convert a geographic data file from one coordinate system to another.
- Raster: A cellular data structure composed of rows and columns for storing images. Each unit in the grid is assigned a value associating it with the corresponding attribute data. Selection of grid size forces a tradeoff between data resolution (and detail) and system storage requirements. Data can be converted to vector data through the process of vectorization.
- Raster models: Rasters come in a wide variety of formats from simple grids to complex, multi-band imagery. Some commonly used models include GRIDS, DOQQ, geoTIF, DEM, MrSID, JPEG2000, and various forms of satellite imagery.
- Record: In an attribute table, a single 'row' of thematic descriptors. In SQL terms, a record is analogous to a tuple. A logical unit of data in a file. For example, there is one record in the ARC file for each arc in a coverage.
- Relate: An operation that establishes a temporary connection between corresponding records in two tables using an item common to both (i.e., relate key). Each record in one table is connected to those records in the other table that share the same value for the common item.
- Remote sensing: Acquiring information about an object without contacting it physically. Methods include aerial photography, radar, and satellite imaging.
- Resolution: Resolution is the accuracy at which a given map scale can depict the location and shape of geographic features. The larger the map scale, the higher the possible resolution. As map scale decreases, resolution diminishes and feature boundaries must be smoothed, simplified, or not shown at all. For example, small areas may have to be represented as points. Distance between sample points in a lattice. Size of the smallest feature that can be represented in a surface. The number of points in x and y in a grid or lattice (e.g., the resolution of a U.S. Geological Survey one-degree DEM is 1201 x 1201 mesh points).
- RMS error: Root mean square error. A measure calculated when registering a map to a digitizer, indicating the discrepancy between known point locations and their digitized locations. The lower the RMS error, the more accurate the digitizing or transformation.
- Row: A record in an attribute table. The horizontal dimension of a table composed of a set of columns containing one data item each. A horizontal group of cells in a grid, or pixels in an

image. Rubber sheeting: A procedure to adjust coverage features in a non-uniform manner. Links representing from- and to- locations are used to define the adjustment.

- Satellite image: A picture of the Earth taken from an Earth-orbital satellite. Satellite images may be produced photographically or by on-board scanners (e.g., MSS).
- Scale bar: A map element that shows the map scale graphically.
- Shapefile: Vector file format developed by ESRI and originally used by ArcView. This data model features distinct polygon structure and does not support topography. A shapefile is actually composed of a group of files. The minimum number required to define a shapefile is 3, with file extensions .shp, .shx, and .dbf.
- Sliver polygon: A small areal feature commonly occurring along the borders of polygons following the topological overlay of two or more coverages.
- Slope: A measure of change in surface value over distance, expressed in degrees or as a percentage. For example, a rise of 2 meters over a distance of 100 meters describes a 2% slope with an angle of 1.15. Mathematically, slope is referred to as the first derivative of the surface.
- Snapping: The process of moving a feature to coincide exactly with coordinates of another feature within a specified snapping distance, or tolerance.
- Spatial analysis: The process of modeling, examining, and interpreting model results. Spatial analysis is useful for evaluating suitability and capability, for estimating and predicting, and for interpreting and understanding. There are four traditional types of spatial analysis: topological overlay and contiguity analysis, surface analysis, linear analysis, and raster analysis.
- Spatial data: Information about the location and shape of, and relationships among, geographic features, usually stored as coordinates and topology.
- Symbol: A graphic pattern used to represent a feature.
- Table: A set of data elements that has a horizontal dimension (rows) and a vertical dimension (columns) in a relational database system. A table has a specified number of columns but can have any number of rows. A table is often called a relation. Rows stored in a table are structurally equivalent to records from flat files in that they must not contain repeating fields.
- Theme: The overall topic of a map layer in which the spatial variation of a single phenomenon is illustrated (e.g., classified vegetation, shaded relief, land use, land ownership). A user-defined perspective on a coverage, grid, tin, or image geographic data set specified, if

applicable, by a coverage name and feature class or data set name, attributes of interest, a data classification scheme, and theme-specific symbology for drawing.

- **TIFF**: Tagged interchange (image) file format. An industry-standard raster data format. TIFF supports black-and-white, gray-scale, pseudo-color, and true-color images, all of which can be stored in a compressed or uncompressed format. TIFF is commonly used in desktop publishing and serves as an interface to numerous scanners and graphic arts packages.
- **TIN**: Triangulated irregular network. A surface representation is derived from irregularly spaced sample points and breakline features. The tin data set includes topological relationships between points and their neighboring triangles. Each sample point has an x,y coordinate, and a surface or z-value. These points are connected by edges to form a set of non-overlapping triangles used to represent the surface.
- **Topographic map**: A map containing contours indicating lines of equal surface elevation (relief), often referred to as topo maps. Often used to refer to a map sheet published by the U.S. Geological Survey in the 7.5-minute quadrangle series or the 15-minute quadrangle series.
- **Topology**: The spatial relationships between connecting or adjacent coverage features (e.g., arcs, nodes, polygons, and points). For example, the topology of an arc includes its from- and to-nodes, and its left and right polygons. Topological relationships are built from simple elements into complex elements: points (simplest elements), arcs (sets of connected points), areas (sets of connected arcs), and routes (sets of sections, which are arcs or portions of arcs). Redundant data (coordinates) are eliminated because an arc may represent a linear feature, part of the boundary of an area feature, or both. Topology is useful in GIS because many spatial modeling operations don't require coordinates, only topological information. For example, to find an optimal path between two points requires a list of the arcs that connect to each other and the cost to traverse each arc in each direction. Coordinates are only needed for drawing the path after it is calculated.
- **UTM**: The Universal Transverse Mercator Map projection. A special case of the Transverse Mercator projection. Abbreviated as the UTM grid it consists of 60 North-South zones, each 6 degrees wide in longitude.
- **Vector**: Data type comprised of x-y coordinate representations of locations on the earth that take the form of single points, strings of points (lines or arcs), or closed lines (polygons) known as features. Each feature has an associated attribute code for identification. Data can be converted to raster data through a process known as rasterization.

- WGS-84: A set of parameters, established by the U.S. Defense Mapping Agency, for determining geometric and physical geodetic relationships on a global scale. The system includes a geocentric reference ellipsoid; a coordinate system; and a gravity field model.
- Z-value: The value of a surface at a particular x,y location (e.g., elevation). Often referred to as spot values or spot elevations.
- Zoom: To enlarge and display greater detail of a portion of a geographic data set.

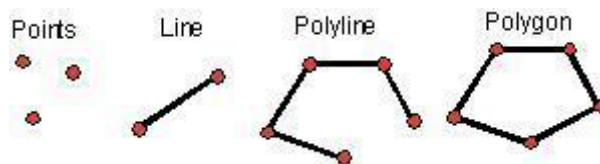
CHAPTER 4: GIS DATA TYPES

4.1 Spatial Data

4.1.1 Vector data

- i. Point Data — layers containing by points (or “events”) described by x,y (lat, long; easting, northing)
- ii. Line/Polyline Data — layers that are described by x,y points (nodes, events) and lines (arcs) between points (line segments and polylines)
- iii. Polygon Data — layers of closed line segments enclosing areas that are described by attributes

Polygon data can be “multipart” like the islands of the state of Hawaii.



4.1.2 Raster data

Raster or grid data (matrices of numbers describing e.g., elevation, population, herbicide use, etc.

4.1.3 Imagery

Images or pictures such as remote sensing data or scans of maps or other photos. This is a special “grid” where the number in each cell describes what color to paint or the spectral character of the image in that cell. (to be used, the “picture” must be placed on a coordinate system, or “rectified” or “georeferenced”).

4.1.4 TIN

Triangular Irregular Networks – used to discretize continuous data.

4.1.5 Terrain data

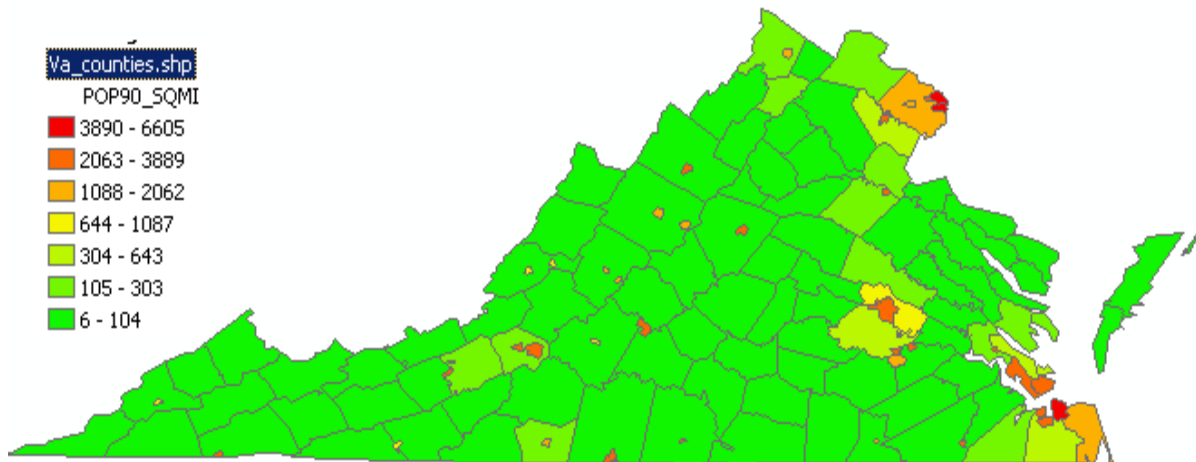
Terrain datasets built from LIDAR and other point clouds.

4.2 Attribute Data

Attribute data are non-spatial characteristics that are connected by tables to points, lines, “events” on lines, and polygons (and in some cases GRID cells)

- i. A point, vector, or raster geologic map might describe a “rock unit” on a map with a single number, letter, or name, but the associated attribute table might have
 - age
 - lithology
 - percent quartz
 - etc, for each rock type on the map.
- ii. Most GIS programs can either plot the polygon by the identifier or by one of the attributes.





The above examples from the following project show two ways to portray census data in Virginia. In the top image, each county/city gets a name and unique color, and in the bottom, the population density per square mile is read from the layer's attribute table and plotted using a different color for each class of density.

4.3 Metadata

- i. Metadata are the most forgotten data type
- ii. Absolutely necessary if you're going to use data, or if someone is going to use your data later (or your derivative information)
- iii. Contains information about
 - scale
 - accuracy
 - projection/datum
 - data source
 - manipulations
 - how to acquire data

CHAPTER 5: DATA MODELS FOR GIS

To visualize natural phenomena, one must first determine how to best represent geographic space. Data models are a set of rules and/or constructs used to describe and represent aspects of the real world in a computer. Two primary data models are available to complete this task: raster data models and vector data models.

5.1 Raster Data Models

The raster data model is widely used in applications ranging far beyond geographic information systems (GIS). Most likely, you are already very familiar with this data model if you have any experience with digital photographs. The ubiquitous JPEG, BMP, and TIFF file formats (among others) are based on the raster data model. Take a moment to view your favorite digital image. If you zoom deeply into the image, you will notice that it is composed of an array of tiny square pixels (or picture elements). Each of these uniquely colored pixels, when viewed as a whole, combines to form a coherent image (Figure 5.1).

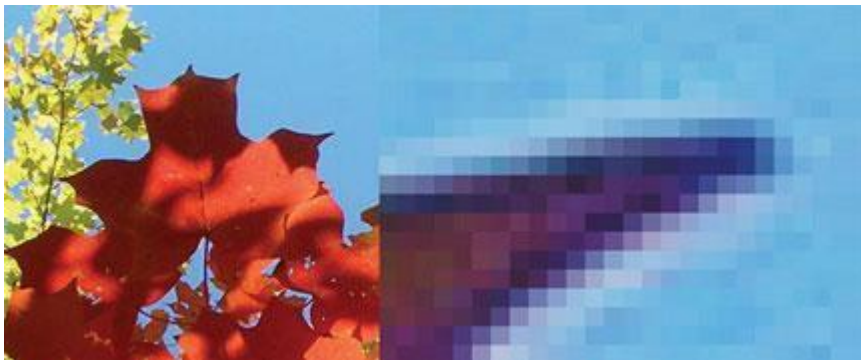


Figure 5.1: Digital Picture with Zoomed Inset Showing Pixilation of Raster Image

Furthermore, all liquid crystal display (LCD) computer monitors are based on raster technology as they are composed of a set number of rows and columns of pixels. Notably, the foundation of this technology predates computers and digital cameras by nearly a century. The neo-impressionist artist, Georges Seurat, developed a painting technique referred to as “pointillism” in the 1880s, which similarly relies on the amassing of small, monochromatic “dots” of ink that combine to form a larger image (Figure 5.2).

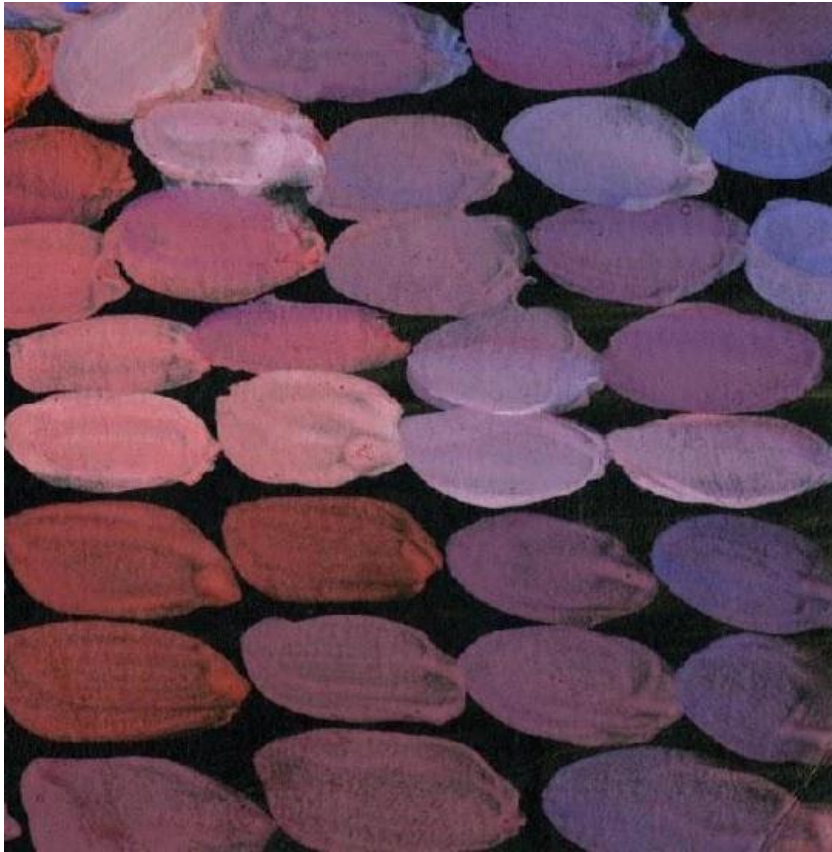


Figure 5.2: Pointillist Artwork

The raster data model consists of rows and columns of equally sized pixels interconnected to form a planar surface. These pixels are used as building blocks for creating points, lines, areas, networks, and surfaces. Although pixels may be triangles, hexagons, or even octagons, square pixels represent the simplest geometric form with which to work. Accordingly, the vast majority of available raster GIS data are built on the square pixel (Figure 5.3). These squares are typically reformed into rectangles of various dimensions if the data model is transformed from one projection to another (e.g., from State Plane coordinates to UTM [Universal Transverse Mercator] coordinates).



Figure 5.3: Common Raster Graphics Used in GIS Applications: Aerial Photograph (left) and USGS DEM (right) (Source: Data available from U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD.)

Because of the reliance on a uniform series of square pixels, the raster data model is referred to as a grid-based system. Typically, a single data value will be assigned to each grid locale. Each cell in a raster carries a single value, which represents the characteristic of the spatial phenomenon at a location denoted by its row and column. The data type for that cell value can be either integer or floating-point. Alternatively, the raster graphic can reference a database management system wherein open-ended attribute tables can be used to associate multiple data values to each pixel. The advance of computer technology has made this second methodology increasingly feasible as large datasets are no longer constrained by computer storage issues as they were previously.

The raster model will average all values within a given pixel to yield a single value. Therefore, the more area covered per pixel, the less accurate the associated data values. The area covered by each pixel determines the spatial resolution of the raster model from which it is derived. Specifically, the resolution is determined by measuring one side of the square pixel. A raster model with pixels representing 10 m by 10 m (or 100 square meters) in the real world would be said to have a spatial resolution of 10 m; a raster model with pixels measuring 1 km by 1 km (1 square kilometer) in the real world would be said to have a spatial resolution of 1 km; and so forth.

Care must be taken when determining the resolution of a raster because using an overly coarse pixel resolution will cause a loss of information, whereas using overly fine pixel resolution will result in significant increases in file size and computer processing requirements during display and/or analysis. An effective pixel resolution will take both the map scale and the minimum mapping unit of the other GIS data into consideration. In the case of raster graphics with coarse spatial resolution, the data values associated with specific locations are not necessarily explicit in the raster data model. For example, if the location of telephone poles were mapped on a coarse raster graphic, it would be clear that the entire cell would not be filled by the pole. Rather, the pole would be assumed to be located somewhere within that cell (typically at the center).

Imagery employing the raster data model must exhibit several properties. First, each pixel must hold at least one value, even if that data value is zero. Furthermore, if no data are present for a given pixel, a data value placeholder must be assigned to this grid cell. Often, an arbitrary, readily identifiable value (e.g., -9999) will be assigned to pixels for which there is no data value. Second, a cell can hold any alphanumeric index that represents an attribute. In the case of quantitative datasets, attribute assignment is fairly straightforward. For example, if a raster image denotes elevation, the data values for each pixel would be some indication of elevation, usually in feet or meters. In the case of qualitative datasets, data values are indices that necessarily refer to some predetermined translational rule. In the case of a land-use/land-cover raster graphic, the following rule may be applied: 1 = grassland, 2 = agricultural, 3 = disturbed, and so forth (Figure 5.4). The third property of the raster data model is that points and lines “move” to the center of the cell. As one might expect, if a 1 km resolution raster image contains a river or stream, the location of the actual waterway within the “river” pixel will be unclear. Therefore, there is a general assumption that all zero-dimensional (point) and one-dimensional (line) features will be located toward the center of the cell. As a corollary, the minimum width for any line feature must necessarily be one cell regardless of the actual width of the feature. If it is not, the feature will not be represented in the image and will therefore be assumed to be absent.

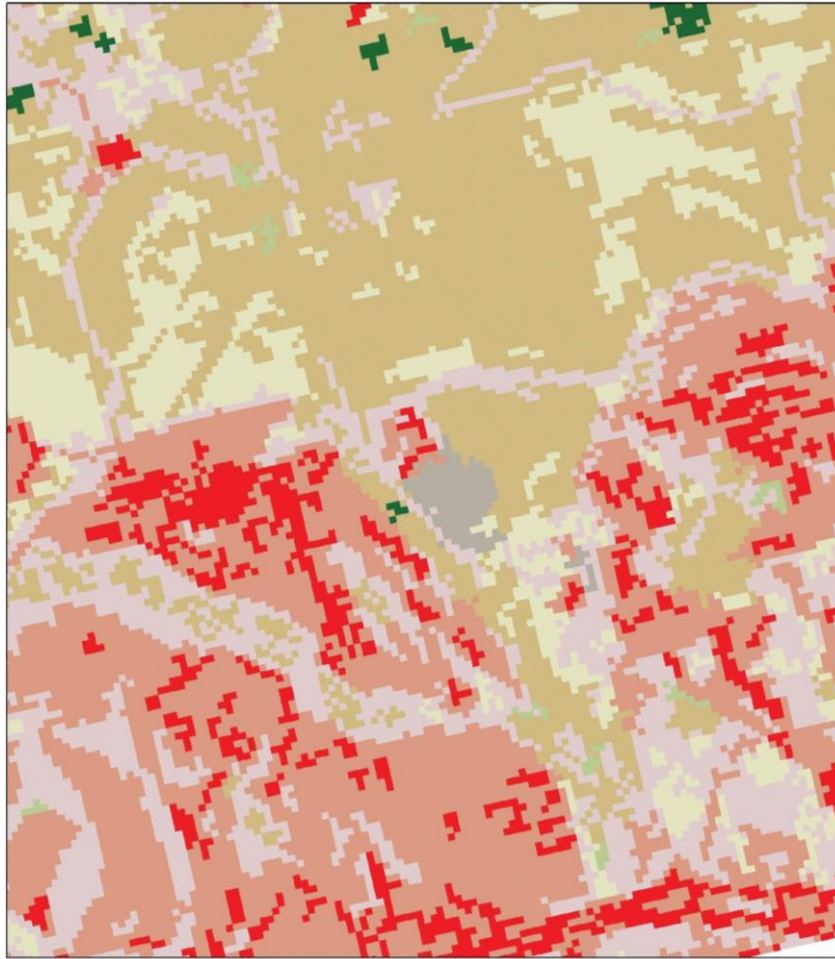


Figure 5.4: Land-Use/Land-Cover Raster Image (Source: Data available from U.S. Geological Survey, Earth Resources Observation, and Science (EROS) Center, Sioux Falls, SD.)

5.1.1 Advantages/Disadvantages of the Raster Model

The use of a raster data model confers many advantages. First, the technology required to create raster graphics is inexpensive and ubiquitous. Nearly everyone currently owns some sort of raster image generator, namely a digital camera, and few cellular phones are sold today that don't include such functionality. Similarly, a plethora of satellites is constantly beaming up-to-the-minute raster graphics to scientific facilities across the globe. These graphics are often posted online for private and/or public use, occasionally at no cost to the user.

Additional advantages of raster graphics are the relative simplicity of the underlying data structure. Each grid location represented in the raster image correlates to a single value (or series of values if attributes tables are included). This simple data structure may also help explain why it is relatively easy to perform overlay analyses on raster data. This simplicity also lends itself to easy interpretation and maintenance of the graphics, relative to its vector counterpart.

Despite the advantages, there are also several disadvantages to using the raster data model. The first disadvantage is that raster files are typically very large. Particularly in the case of raster images built from the cell-by-cell encoding methodology, the sheer number of values stored for a given dataset result in potentially enormous files. Any raster file that covers a large area and has somewhat finely resolved pixels will quickly reach hundreds of megabytes in size or more. These large files are only getting larger as the quantity and quality of raster datasets continue to keep pace with the quantity and quality of computer resources and raster data collectors (e.g., digital cameras, satellites).

A second disadvantage of the raster model is that the output images are less “pretty” than their vector counterparts. This is particularly noticeable when the raster images are enlarged or zoomed (refer to Figure 5.1). Depending on how far one zooms into a raster image, the details and coherence of that image will quickly be lost amid a pixilated sea of seemingly randomly colored grid cells.

The geometric transformations that arise during map reprojection efforts can cause problems for raster graphics and represent the third disadvantage to using the raster data model. Changing map projections will alter the size and shape of the original input layer and frequently result in the loss or addition of pixels (White 2006). White, D. 2006. “Display of Pixel Loss and Replication in Reprojecting Raster Data from the Sinusoidal Projection.” *Geocarto International* 21 (2): 19–22. These alterations will result in the perfect square pixels of the input layer taking on some alternate rhomboidal dimensions. However, the problem is larger than a simple reformation of the square pixel. Indeed, the reprojection of a raster image dataset from one projection to another brings change to pixel values that may, in turn, significantly alter the output information (Seong 2003). Seong, J. C. 2003. “Modeling the Accuracy of Image Data Reprojection.” *International Journal of Remote Sensing* 24 (11): 2309–21.

The final disadvantage of using the raster data model is that it is not suitable for some types of spatial analyses. For example, difficulties arise when attempting to overlay and analyze multiple raster graphics produced at differing scales and pixel resolutions. Combining information from a raster image with 10 m spatial resolution with a raster image with 1 km spatial resolution will most likely produce nonsensical output information as the scales of analysis are far too disparate to result in meaningful and/or interpretable conclusions. Also, some network and spatial analyses (i.e., determining directionality or geocoding) can be problematic to perform on raster data. of the major waterways in your county? Why?

5.2 Vector Data Models

In contrast to the raster data model is the vector data model. In this model, space is not quantized into discrete grid cells like the raster model. Vector data models use points and their associated X, Y coordinate pairs to represent the vertices of spatial features, much as if they were being drawn on a map by hand (Aronoff 1989). Aronoff, S. 1989. *Geographic Information Systems: A Management Perspective*. Ottawa, Canada: WDL Publications. The data attributes of these features are then stored in a separate database management system. The spatial information and the attribute information for these models are linked via a simple identification number that is given to each feature in a map.

Three fundamental vector types exist in geographic information systems (GIS): points, lines, and polygons (Figure 5.5). Points are zero-dimensional objects that contain only a single coordinate pair. Points are typically used to model singular, discrete features such as buildings, wells, power poles, sample locations, and so forth. Points have only the property of location. Other types of point features include the node and the vertex. Specifically, a point is a stand-alone feature, while a node is a topological junction representing a common X, Y coordinate pair between intersecting lines and/or polygons. Vertices are defined as each bend along with a line or polygon feature that is not the intersection of lines or polygons.

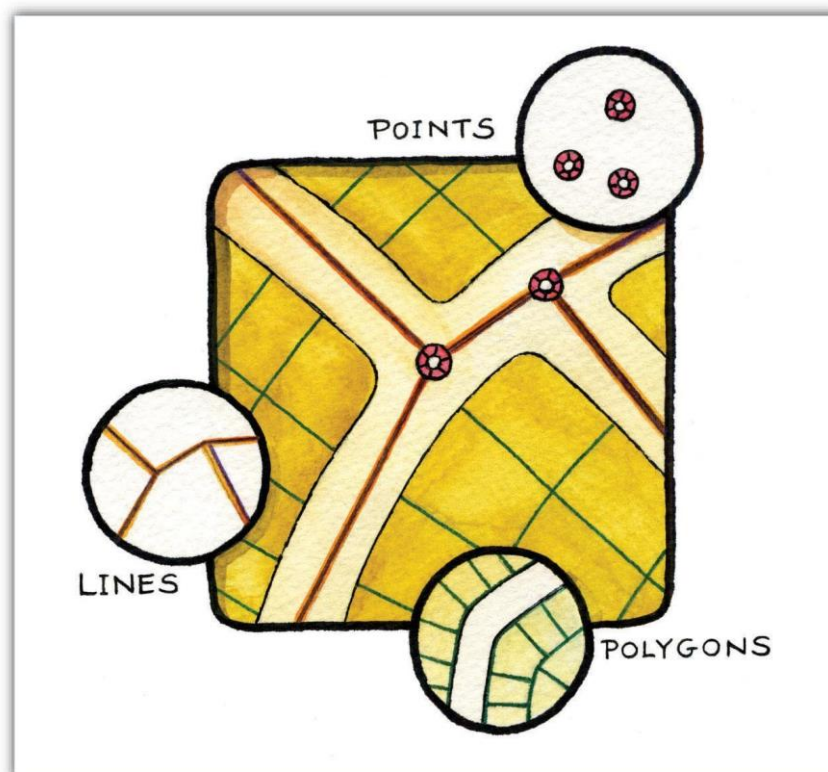


Figure 5.5: Points, Lines, and Polygons

Points can be spatially linked to form more complex features. Lines are one-dimensional features composed of multiple, explicitly connected points. Lines are used to represent linear features such as roads, streams, faults, boundaries, and so forth. Lines have the property of length. Lines that directly connect two nodes are sometimes referred to as chains, edges, segments, or arcs.

Polygons are two-dimensional features created by multiple lines that loop back to create a “closed” feature. In the case of polygons, the first coordinate pair (point) on the first line segment is the same as the last coordinate pair on the last line segment. Polygons are used to represent features such as city boundaries, geologic formations, lakes, soil associations, vegetation communities, and so forth. Polygons have the properties of area and perimeter. Polygons are also called areas.

5.2.1 *Vector Data Models Structures*

Vector data models can be structured in many different ways. We will examine two of the more common data structures here. The simplest vector data structure is called the spaghetti data model (Dangermond 1982). Dangermond, J. 1982. “A Classification of Software Components Commonly Used in Geographic Information Systems.” In *Proceedings of the U.S.-Australia Workshop on the Design and Implementation of Computer-Based Geographic Information Systems*, 70–91. Honolulu, HI. In the spaghetti model, each point, line, and/or polygon feature is represented as a string of X, Y coordinate pairs (or as a single X, Y coordinate pair in the case of a vector image with a single point) with no inherent structure (Figure 5.6). One could envision each line in this model to be a single strand of spaghetti that is formed into complex shapes by the addition of more and more strands of spaghetti. It is notable that in this model, any polygons that lie adjacent to each other must be made up of their own lines, or stands of spaghetti. In other words, each polygon must be uniquely defined by its own set of X, Y coordinate pairs, even if the adjacent polygons share the exact same boundary information. This creates some redundancies within the data model and therefore reduces efficiency.

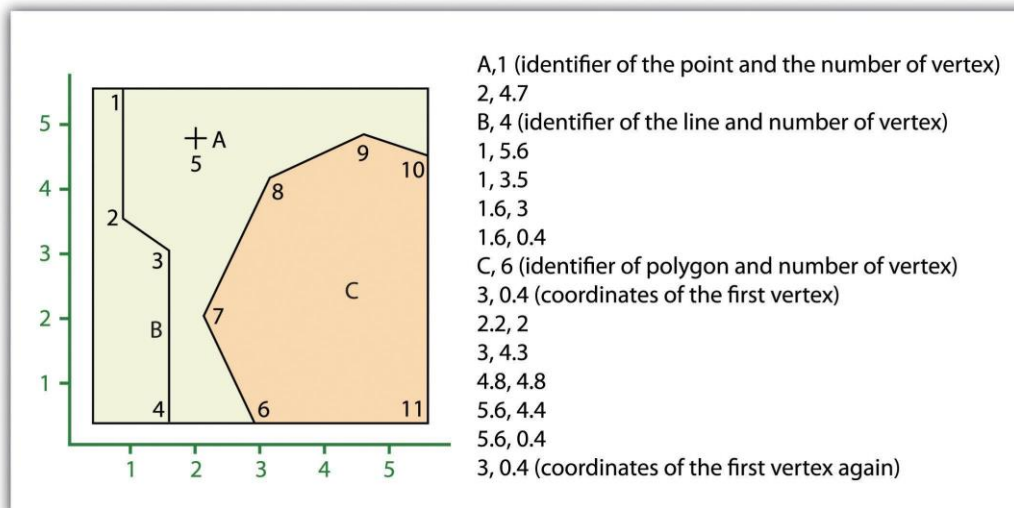


Figure 5.6: Spaghetti Data Model

Despite the location designations associated with each line, or strand of spaghetti, spatial relationships are not explicitly encoded within the spaghetti model; rather, they are implied by their location. This results in a lack of topological information, which is problematic if the user attempts to make measurements or analysis. The computational requirements, therefore, are very steep if any advanced analytical techniques are employed on vector files structured thusly. Nevertheless, the simple structure of the spaghetti data model allows for efficient reproduction of maps and graphics as this topological information is unnecessary for plotting and printing.

In contrast to the spaghetti data model, the topological data model is characterized by the inclusion of topological information within the dataset, as the name implies. Topology is a set of rules that model the relationships between neighboring points, lines, and polygons and determines how they share geometry. For example, consider two adjacent polygons. In the spaghetti model, the shared boundary of two neighboring polygons is defined as two separate, identical lines. The inclusion of topology into the data model allows for a single line to represent this shared boundary with an explicit reference to denote which side of the line belongs with which polygon. Topology is also concerned with preserving spatial properties when the forms are bent, stretched, or placed under similar geometric transformations, which allows for more efficient projection and reprojection of map files.

Three basic topological precepts that are necessary to understand the topological data model are outlined here. First, connectivity describes the arc-node topology for the feature dataset. As discussed previously, nodes are more than simple points. In the topological data model, nodes are the intersection points where two or more arcs meet. In the case of arc-node topology, arcs

have both a from-node (i.e., starting node) indicating where the arc begins and a to-node (i.e., ending node) indicating where the arc ends (Figure 5.7). Also, between each node pair is a line segment, sometimes called a link, which has its own identification number and references both its from-node and to-node. In Figure 5.7, arcs 1, 2, and 3 all intersect because they share node 11. Therefore, the computer can determine that it is possible to move along arc 1 and turn onto arc 3, while it is not possible to move from arc 1 to arc 5, as they do not share a common node.

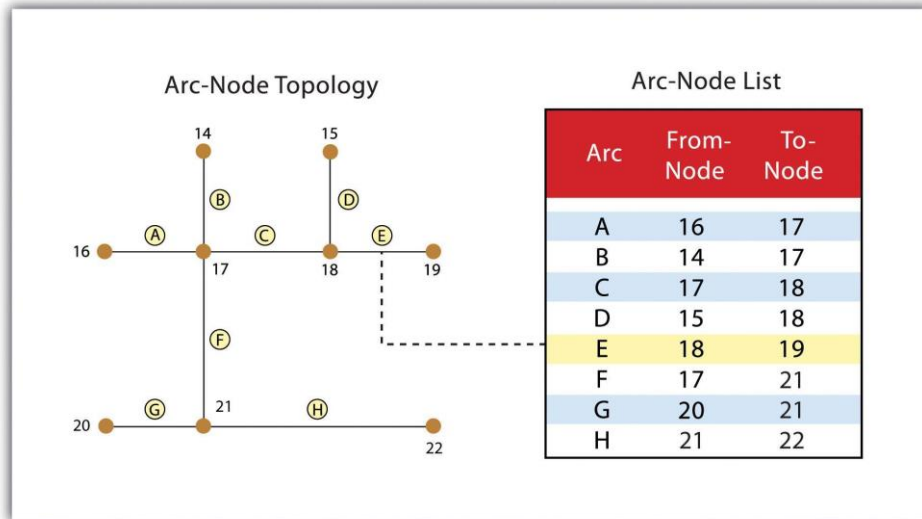


Figure 5.7: Arc-Node Topology

The second basic topological precept is the area definition. Area definition states that an arc that connects to surround an area defines a polygon, also called polygon-arc topology. In the case of polygon-arc topology, arcs are used to construct polygons, and each arc is stored only once (Figure 5.8). This results in a reduction in the amount of data stored and ensures that adjacent polygon boundaries do not overlap. In Figure 5.8, the polygon-arc topology makes it clear that polygon F is made up of arcs 8, 9, and 10.

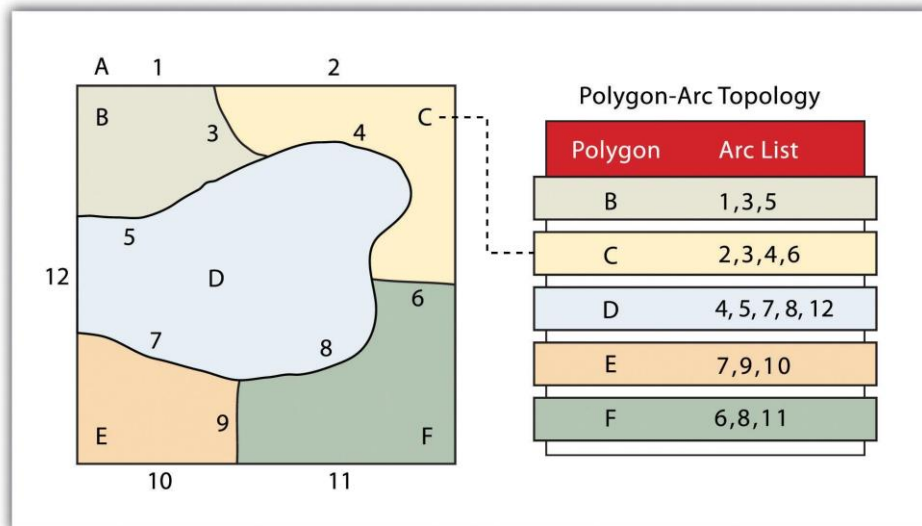


Figure 5.8: Polygon-Arc Topology

Contiguity, the third topological precept, is based on the concept that polygons that share a boundary are deemed adjacent. Specifically, polygon topology requires that all arcs in a polygon have a direction (a from-node and a to-node), which allows adjacency information to be determined (Figure 5.9). Polygons that share an arc are deemed adjacent, or contiguous, and therefore the “left” and “right” side of each arc can be defined. This left and right polygon information is stored explicitly within the attribute information of the topological data model. The “universe polygon” is an essential component of polygon topology that represents the external area located outside of the study area. Figure 5.9 shows that arc 6 is bound on the left by polygon B and to the right by polygon C. Polygon A, the universe polygon, is to the left of arcs 1, 2, and 3.

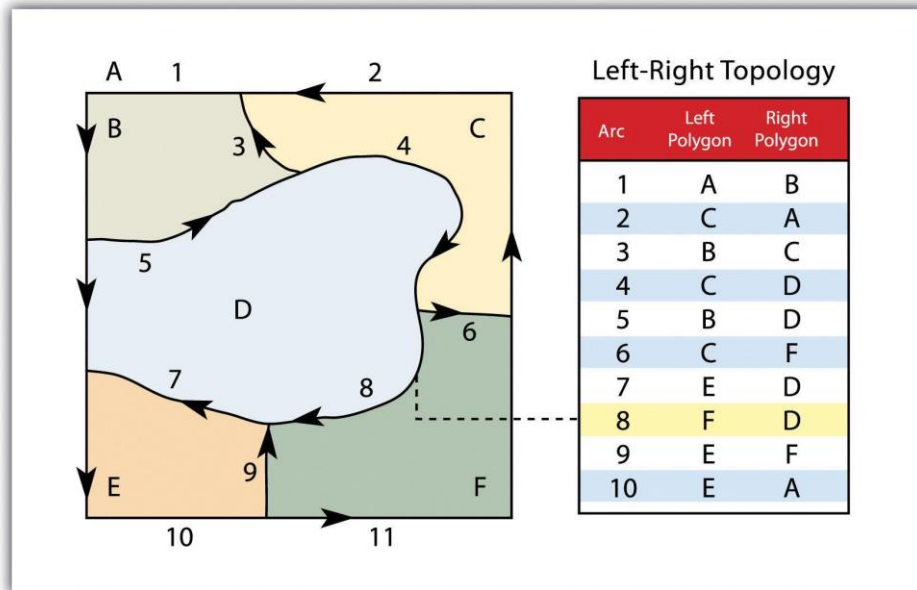


Figure 5.9: Polygon Topology

Topology allows the computer to rapidly determine and analyze the spatial relationships of all its included features. Also, topological information is important because it allows for efficient error detection within a vector dataset. In the case of polygon features, open or unclosed polygons, which occur when an arc does not completely loop back upon itself, and unlabeled polygons, which occur when an area does not contain any attribute information, violate polygon-arc topology rules. Another topological error found with polygon features is the sliver. Slivers occur when the shared boundary of two polygons do not meet exactly (Figure 5.10)

In the case of line features, topological errors occur when two lines do not meet perfectly at a node. This error is called an “undershoot” when the lines do not extend far enough to meet each other and an “overshoot” when the line extends beyond the feature it should connect to (Figure 5.10). The result of overshoots and undershoots is a “dangling node” at the end of the line. Dangling nodes aren’t always an error, however, as they occur in the case of dead-end streets on a road map.

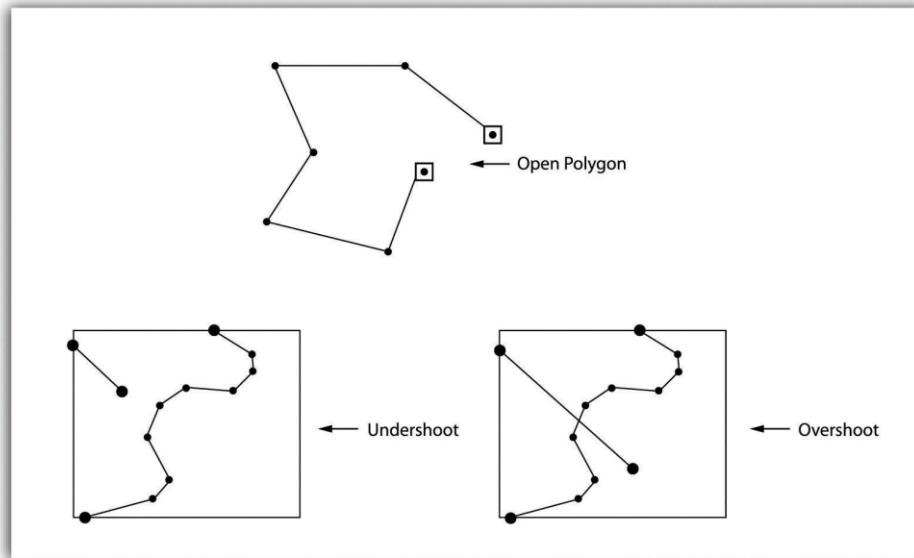


Figure 5.10: Common Topological Errors

Many types of spatial analysis require the degree of organization offered by topologically explicit data models. In particular, network analysis (e.g., finding the best route from one location to another) and measurement (e.g., finding the length of a river segment) relies heavily on the concept of to- and from-nodes and uses this information, along with attribute information, to calculate distances, shortest routes, quickest routes, and so forth. Topology also allows for sophisticated neighborhood analysis such as determining adjacency, clustering, nearest neighbors, and so forth.

Now that the basics of the concepts of topology have been outlined, we can begin to better understand the topological data model. In this model, the node acts as more than just a simple point along a line or polygon. The node represents the point of intersection for two or more arcs. Arcs may or may not be looped into polygons. Regardless, all nodes, arcs, and polygons are individually numbered. This numbering allows for quick and easy reference within the data model.

5.2.2 Advantages/Disadvantages of the Vector Model

In comparison with the raster data model, vector data models tend to be better representations of reality due to the accuracy and precision of points, lines, and polygons over the regularly spaced grid cells of the raster model. This results in vector data tending to be more aesthetically pleasing than raster data.

Vector data also provides an increased ability to alter the scale of observation and analysis. As each coordinate pair associated with a point, line, and polygon represents an infinitesimally exact location (albeit limited by the number of significant digits and/or data acquisition methodologies), zooming deep into a vector image does not change the view of a vector graphic in the way that it does a raster graphic (see Figure 5.1).

Vector data tend to be more compact in the data structure, so file sizes are typically much smaller than their raster counterparts. Although the ability of modern computers has minimized the importance of maintaining small file sizes, vector data often require a fraction of the computer storage space when compared to raster data.

The final advantage of vector data is that topology is inherent in the vector model. This topological information results in simplified spatial analysis (e.g., error detection, network analysis, proximity analysis, and spatial transformation) when using a vector model.

Alternatively, there are two primary disadvantages to the vector data model. First, the data structure tends to be much more complex than the simple raster data model. As the location of each vertex must be stored explicitly in the model, there are no shortcuts for storing data like there are for raster models (e.g., the run-length and quad-tree encoding methodologies).

Second, the implementation of spatial analysis can also be relatively complicated due to minor differences in accuracy and precision between the input datasets. Similarly, the algorithms for manipulating and analyzing vector data are complex and can lead to intensive processing requirements, particularly when dealing with large datasets.

5.3 Satellite Imagery and Aerial Photography

A wide variety of satellite imagery and aerial photography is available for use in geographic information systems (GIS). Although these products are basically raster graphics, they are substantively different in their usage within a GIS. Satellite imagery and aerial photography provide important contextual information for a GIS and are often used to conduct heads-up digitizing whereby features from the image are converted into vector datasets.

5.3.1 Satellite Imagery

Remotely sensed satellite imagery is becoming increasingly common as satellites equipped with technologically advanced sensors are continually being sent into space by public agencies and private companies around the globe. Satellites are used for applications such as military and civilian Earth observation, communication, navigation, weather, research, and more. Currently, more than 3,000 satellites have been sent to space, with over 2,500 of them originating from Russia and the United States. These satellites maintain different altitudes, inclinations, eccentricities, synchronies, and orbital centers, allowing them to image a wide variety of surface features and processes (Figure 5.11).



Figure 5.11: Satellites Orbiting the Earth

Satellites can be active or passive. Active satellites make use of remote sensors that detect reflected responses from objects that are irradiated from artificially generated energy sources. For example, active sensors such as radars emit radio waves, laser sensors emit light waves, and sonar sensors emit sound waves. In all cases, the sensor emits the signal and then calculates the time it takes for the returned signal to “bounce” back from some remote feature. Knowing the speed of the emitted signal, the time delay from the original emission to the return can be used to calculate the distance to the feature.

Passive satellites, alternatively, make use of sensors that detect the reflected or emitted electromagnetic radiation from natural sources. This natural source is typically the energy from

the sun, but other sources can be imaged as well, such as magnetism and geothermal activity. Using an example we've all experienced, taking a picture with a flash-enabled camera would be active remote sensing, while using a camera without a flash (i.e., relying on ambient light to illuminate the scene) would be passive remote sensing.

The quality and quantity of satellite imagery are largely determined by their resolution. There are four types of resolution that characterize any particular remote sensor (Campbell 2002). Campbell, J. B. 2002. *Introduction to Remote Sensing*. New York: Guilford Press. The spatial resolution of a satellite image, as described previously in the raster data model section, is a direct representation of the ground coverage for each pixel shown in the image. If a satellite produces imagery with a 10 m resolution, the corresponding ground coverage for each of those pixels is 10 m by 10 m or 100 square meters on the ground. Spatial resolution is determined by the sensors' instantaneous field of view (IFOV). The IFOV is essentially the ground area through which the sensor is receiving the electromagnetic radiation signal and is determined by the height and angle of the imaging platform.

5.3.2 Aerial Photography

Aerial photography, like satellite imagery, represents a vast source of information for use in any GIS. Platforms for the hardware used to take aerial photographs include airplanes, helicopters, balloons, rockets, and so forth. While aerial photography connotes images taken of the visible spectrum, sensors to measure bands within the nonvisible spectrum (e.g., ultraviolet, infrared, near-infrared) can also be fixed to aerial sources. Similarly, aerial photography can be active or passive and can be taken from vertical or oblique angles. Care must be taken with aerial photographs as the sensors used to take the images are similar to cameras in their use of lenses. These lenses add curvature to the images, which becomes more pronounced as one moves away from the center of the photo (Figure 5.12).

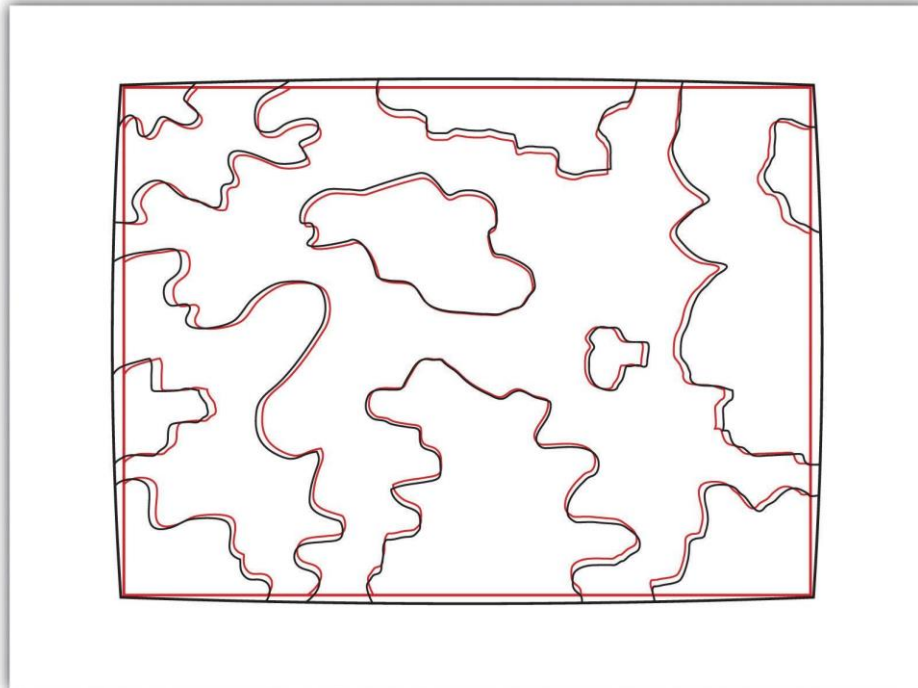


Figure 5.12: Curvature Error Due to Lenticular Properties of Camera

Another source of potential error in an aerial photograph is relief displacement. This error arises from the three-dimensional aspect of terrain features and is seen as apparent leaning away of vertical objects from the center point of an aerial photograph. To imagine this type of error, consider that a smokestack would look like a doughnut if the viewing camera was directly above the feature. However, if this same smokestack was observed near the edge of the camera's view, one could observe the sides of the smokestack. This error is frequently seen with trees and multistory buildings and worsens with increasingly taller features.

Orthophotos are vertical photographs that have been geometrically “corrected” to remove the curvature and terrain-induced error from images (Figure 5.13). The most common orthophoto product is the digital ortho quarter quadrangle (DOQQ). DOQQs are available through the US Geological Survey (USGS), which began producing these images from their library of 1:40,000-scale National Aerial Photography Program photos. These images can be obtained in either grayscale or color with 1-meter spatial resolution and 8-bit radiometric resolution. As the name suggests, these images cover a quarter of a USGS 7.5 minute quadrangle, which equals an approximately 25 square mile area. Included with these photos is an additional 50 to 300-meter edge around the photo that allows users to mosaic many DOQQs into a single, continuous image. These DOQQs are ideal for use in a GIS as background display information, for data editing, and heads-up digitizing.



Figure 5.13: Orthophoto (Source: Data available from U.S. Geological Survey, Earth Resources Observation, and Science (EROS) Center, Sioux Falls, SD.)

CHAPTER 6: ArcGIS SOFTWARE COMPONENTS

6.1 Introduction to ArcGIS Desktop

The geographic information system software package **ArcGIS** was developed by the Environmental Systems Research Institute (ESRI). ArcGIS is designed to create, develop, and interact with new and existing geographic data. It's designed to be a complete and integrated system for geographic data processing.

The desktop form of ArcGIS is available with three levels of functionality. The most basic level is called **ArcView**, which allows for many map making, visualization and map analysis capabilities. For creating and editing spatial data that go into these analyses, **ArcEditor** adds capabilities on top of ArcView. Finally, more advanced visualization and analysis tools are available at the **ArcInfo** level. At all levels, users interact with Desktop ArcGIS through three interface components: ArcMap, ArcCatalog, and ArcToolbox.

ArcMap is used to perform map and data-based tasks. These include geographic and attribute data creation, editing, data viewing and querying, and map production.

ArcCatalog is designed to browse, organize, and document geographic data in a Windows Explorer-like fashion. Operations such as previewing, copying, moving, renaming, or deleting can be performed within this module.

ArcToolbox is the data management and geo-processing module embedded within ArcMap and ArcCatalog. Task wizards have been created for the most commonly performed geo-processing operations. Some of the functionality includes: importing and exporting, overlays, buffering, and statistical calculations.

Although ArcGIS is the dominant software in the GIS industry, other GIS software packages are available:

- i. Common commercial software:
 - Idrisi (<http://www.clarklabs.org/>)
 - Mapinfo (<http://www.pbinsight.com/welcome/mapinfo/>)
 - ERDAS Imagine (<http://www.erdas.com/>)
- ii. Common freeware / open source software:
 - Google Earth (primarily for visualization purposes)

- R (<http://www.r-project.org/>)
- GRASS (<http://grass.fbk.eu/>)
- ILWIS (http://www.itc.nl/Pub/Home/Research/Research_output/ILWIS_-_Remote_Sensing_and_GIS_software.html)
- SAGA (<http://www.saga-gis.org/en/index.html>)
- QUANTUM GIS (<http://www.qgis.org/>)
- SPRING (<http://www.dpi.inpe.br/spring/english/index.html>)

6.2 Lab: Introduction to ArcGIS

In this exercise you will investigate some of the general capabilities of Desktop ArcGIS, its built-in help system, and the lab structure and setup. Specifically, you will learn some of the more basic capabilities of the software, which include downloading and reading in data and defining symbols to use for data display. To place these capabilities in the context of how they might be used in a GIS project, you will also use the web to find and download two existing GIS data sets for importing into ArcGIS.

Learning Objectives:

- To practice accessing your working directory and CTools to setup the lab data
- To gain a basic understanding of the ArcGIS modules
- To learn to navigate and utilize the ArcGIS help system
- To begin exploring the functionality of various toolbars in ArcGIS
- To become familiar with downloading GIS data from the Internet

6.2.1 Procedure

6.2.1.1 Downloading your data and creating your working directory

- The labs must be downloaded from Google Classroom one day before the class. After downloading the data, you must unzip it to C:\Temp on the machine you are working on.

- Create a folder dedicated to the entire course lab work. To create the folder, click on the C:\ drive, in the right window, right-click, select new, and then select folder, name it “ACCD_GIS”.

Important - your folder names cannot contain spaces if they are going to store data for ArcGIS.

Please remember – NO SPACES!

6.2.1.2 Accessing class resources

- Go to google classroom
- Download the zipped file called Lab1_data.zip and lab1.pdf
- **Unzip the lab data.** Navigate to your folder in Windows Explorer. Double-click on the Lab1_data.zip file that you saved to this directory. In the window that opens, the “Lab1_data” folder should be highlighted.
- Click the “Extract all files” button. Check that the “Files will be extracted to this folder” input box is set to the ACCD_GIS folder you created earlier.
- Click “Extract” to decompress the Lab1_data folder into your ACCD_GIS folder.
- Navigate to your Lab1_data folder to verify the files there.

This process is required at the start of each lab. Please refer back to this lab throughout the training if you cannot recall the procedure.

6.2.1.3 Starting with ArcCatalog and connecting to your data:

ArcCatalog was designed to browse, organize, and document geographic data in a format similar to Windows Explorer. Operations such as previewing, copying, moving, renaming, or deleting can be performed within this module. ArcCatalog is usually the best place to start working with ArcGIS.

In order to get started using ArcCatalog, you need to direct the system to the location of your data by **connecting to your working folder**. To accomplish this, do the following:

- Start ArcCatalog.exe by selecting Start > All Programs > GIS Applications > ArcGIS Desktop > ArcCatalog.exe
- Connect to the folder containing your working data (C:\ACCD_GIS\lab1_data) by selecting the icon as seen below or by choosing File > Connect Folder option



- Navigate to the folder where the data are located by **single-clicking the ►** next to the path name, start with (C:\), then work down to your working directory (lab1_data)
- Select your working directory (lab1_data) so that it is highlighted, click OK
- You will now be able to access this location directly from within ArcCatalog without navigating through all of the folders in your M: drive (see the new connection item at the top of your directory tree). You should also be able to see your data sets at this point.

This process is required at the start of each lab, or session in which you use ArcGIS. Please refer back to this lab throughout the semester if you cannot recall the procedure.

6.2.1.4 Exploring the ArcGIS Help System

In this step you will be exploring the Desktop help system in ArcGIS, which contains a large amount of information on how to perform specific operations and on related concepts. Throughout the semester you will be required to use the help system to solve problems or answer your own questions. The goal of this section is simply for you to learn how to access the help system.

- In the ArcCatalog window, find the ArcGIS Desktop Help under the Help menu. The help files are the same for ArcCatalog, ArcMap, and ArcToolbox
- The new display gives three tab options: Contents, Favorites, and Search. Familiarize yourself with some of the operations that each tab allows.

6.2.1.5 Exploring ArcMap tools

ArcMap is used to perform map and data-based tasks. These include geographic and attribute data editing, data viewing and querying, and map production using many data layers.

- Launch ArcMap from the main toolbar in ArcCatalog with the icon shown below.
- Select *OK*, to start a new blank map.



- Familiarize yourself with the navigation tools within ArcMap as well as the different drop down menus. You can hover your cursor over each icon on the toolbar for a short description.
- Alternatively, you can try out each of the buttons in the ArcMap toolbar. To do this you will first need to add data to your ArcMap session.
- In ArcMap, under the **File** dropdown (upper left corner) select *Add Data / Add Data*, navigate to your working directory, you should find several files, including two files named **Mohafaza.shp** and **Caza.shp**, add both to your project. You can select more than one file at a time by simultaneously holding *shift* or *ctrl* and selecting the file – then click *Add*.



- There is also a shortcut for adding data on the ArcMap toolbar. Look for this icon:
- With both files added to your ArcMap session, determine the function of each of the icons seen in the graphic below.

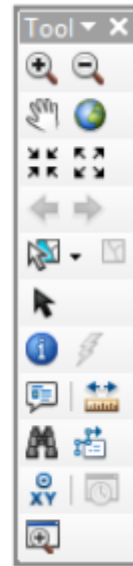


6.2.1.6 Displaying data in ArcMap

ArcMap is used to perform map and data-based tasks. These include geographic and attribute data creation and editing, data viewing and querying, and map production. Here you will use ArcMap to work with the data you downloaded.

- Exit out of your previous ArcMap session and then launch a new ArcMap session from the main tool bar in ArcCatalog with the icon shown below (Start with a “new empty map”)
- Add **all the existing data** that you downloaded from ACCD_GIS folder by clicking on the plus icon, selecting the data (remember how to use *Ctrl* to add multiple files) and clicking *Add*
- Use the **navigation tools** within ArcMap to explore your data. If you’re not sure which tool does what, *pause the cursor over it* for a second and a help tag will display the title of the tool.

Navigation Tools



- First select zoom to Full-extent (world icon) to see all of your data, then right-click on the *Akkar_Mohafaza.shp* dataset within the Table of Contents (left frame within ArcMap) and select “Zoom to layer” (this will change the zoom-level to the Akkar Area). Use “zoom” and “pan” in the navigation toolbar to adjust the view to your liking.
- Experiment with the order of data layers by dragging (holding a left-click while moving the layer) different datasets above and below one another in the table of contents.
- You can use the table of contents to **turn on and off** the various data sets (you can accomplish this by adding and removing checks from the box to the left of the file name).
- Altering symbology can be accomplished by *right-* or *left-clicking* on the actual symbol in the table of contents.
- While you’re adjusting colors and symbols (we will cover this fully in later labs) you want your audience to be able to read your data easily – so, for example, make sure you can differentiate between Assisted Living and Skilled Nursing facilities.
- Two additional tools that may be of assistance when viewing data include the **overview** and **magnifier** windows. These tools can be found under the *Windows pull-down menu*. Explore these tools with your datasets.
- In the table of contents, right-click on the layer *Cadastral.shp* and select *Open Attribute Table* option. The table that opens shows the “information” (in rows and columns; also known as attributes) that has been stored for the big-island dataset.

6.3 Lab: Exploring Data Structures

This lab will explore the data structures used to implement the two most common GIS spatial data models: *vector* and *raster*. These data models define how the data are created, stored, manipulated, and displayed. In this lab you will explore how ArcGIS deals with data in each of these models. In addition, you will become familiar with common formats used for storage, and how to exchange and import geographical data.

The *vector data model* (points, lines, and polygons) can be stored in several different file formats. In ArcGIS, the common file formats are **coverages**, **shapefiles**, and **geodatabases**.

Coverages are native to ESRI's original *Arc/Info* software, and are still sometimes used in contemporary GIS applications. Coverage information is stored in a topologically structured manner (you will work with topology in a future lab). *Topology is explicitly encoded information about the spatial connections between features*. By encoding spatial relationships in this way, the data structure provides a means to query and analyze the data for information about adjacency and contiguity without calculating the connections each time an analysis is performed. The geographic and attribute portions of coverages are stored separately and linked together via database relations that are managed with relational tables stored in an *INFO directory*.

ESRI's help system provides a great deal of illustrated introductory information about coverages and other file formats. For coverages, look in the ArcGIS Desktop Help system under the *Contents* tab, check *Geodata | Data Types | Coverages | What is a coverage?* For an illustrated description of topology, see *Geodata | Data Types | Topologies | Topology in ArcGIS*.

Shapefiles are the native data format of *ArcView 3.x* GIS, which was created as a user-friendly but less analytically powerful, supplement to ArcInfo. They are similar to coverages in that the geographic and attribute portions are stored separately (geography in the SHP file and attributes in the DBF file) and linked together by information stored in the SHX file. They are different from coverages because the spatial or geographic elements are stored as individual objects with no explicit representation of adjacency and contiguity built into the data structure (i.e. there is no topology information). Therefore, if topology-based analyses are required, the software needs to calculate spatial relationships on-the-fly (depending on the size of the dataset, this can be very cumbersome).

You can find more information regarding shapefiles in the Help system. Check *Geodata | Data Types | Shapefiles | What is a shapefile? & Shapefile file extensions*. Also, check *Geodata | Data Types | Tables | A quick tour of tables and attribute information & Tabular data sources* for diagrams and further explanation.

Geodatabases are the contemporary data structures for ESRI products, and are native to ArcGIS. Therefore, they will be the primary focus of vector data development throughout the remainder of the semester. They are similar to both coverages and shapefiles, in that topological structuring is possible (like coverages) and geographic elements are stored as individual objects (like shapefiles). They are unique from coverages and shapefiles in several ways:

- Unlike coverages, in which topology is automatically defined by the system, topology in geodatabases is user defined by selecting and applying (i.e., validating) topology from a suite of rules that specify the desired spatial relationships.
- Geodatabases can have ‘behaviors’ detailed within the data structure to better model real-world relationships.

There are two types of geodatabases freely available to all Arc users, Personal geodatabases and File geodatabases. File geodatabases are more current and exhibit several advantages over personal geodatabases, so we will be using them for the remainder of our labs. Some of these advantages are as follows:

- In Personal geodatabases, all data (both geographic and attribute) can be stored in a single MDB file (derived from the Microsoft Access database engine). As such, they are tied to the Windows operating system. File geodatabases are native to ArcGIS and stored as a file system folder. They are compatible across platforms.
- The maximum size of a Personal geodatabase file is 2GB; however, the effective size is between 250 to 500 MB as performance slows after that. File geodatabases can hold up to 1 TB of data *per dataset* (each File can store many datasets), and have significantly faster performance than any data formats we have discussed thus far.

It is worth noting that Geodatabases created in ArcGIS 10.x, cannot be open or edited in earlier ArcGIS versions. Geodatabases created in ArcGIS 9.x can be opened and edited in ArcGIS 10.x.

Look in the Help system under *Geodata | Geodatabases | Managing Geodatabases | An overview of the geodatabase | A quick tour of the geodatabase & Essential readings about the geodatabase* for further explanation.

GRID is the native ESRI format to store data using the *raster data model*. The grid format is a typical example of a raster data structure, in that it stores a matrix of cells that are organized into rows and columns. Each cell is assigned a single value to represent categories or numerical attributes at each location. The *Spatial Analyst* extension in ArcGIS offers several tools for manipulation and computation of data that are stored as raster data models. ArcGIS has the ability to display many other types of non-native raster data, including Imagine files (.img), American Standard Code for Information Interchange or ASCII (.asc), Tiff (.tif) and Joint Photographic Experts Group or JPEG (.jpg). To see the complete list, check *Geodata | Data Types | Rasters and images | Supported raster data*. In addition, check the section *Fundamentals of raster data*.

Learning Objectives:

- To explore and understand data structures used by ArcGIS
- To gain an introduction to the various topology options
- To comprehend the various raster display options and how they affect your data
- To learn to join tables
- To identify the parameters that define a raster data set
- To become familiar with some tools for importing data.

6.3.1 Procedure:

6.3.1.1 The ArcCatalog graphical interface

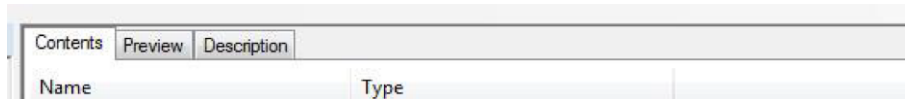
The latest version of ArcGIS can access all the data structures listed on the previous page, plus a number of other structures that are not native to ESRI software. These include a number of databases, tables, and image formats. ArcCatalog is an object-oriented interface that displays each data type using icons similar to those seen below.



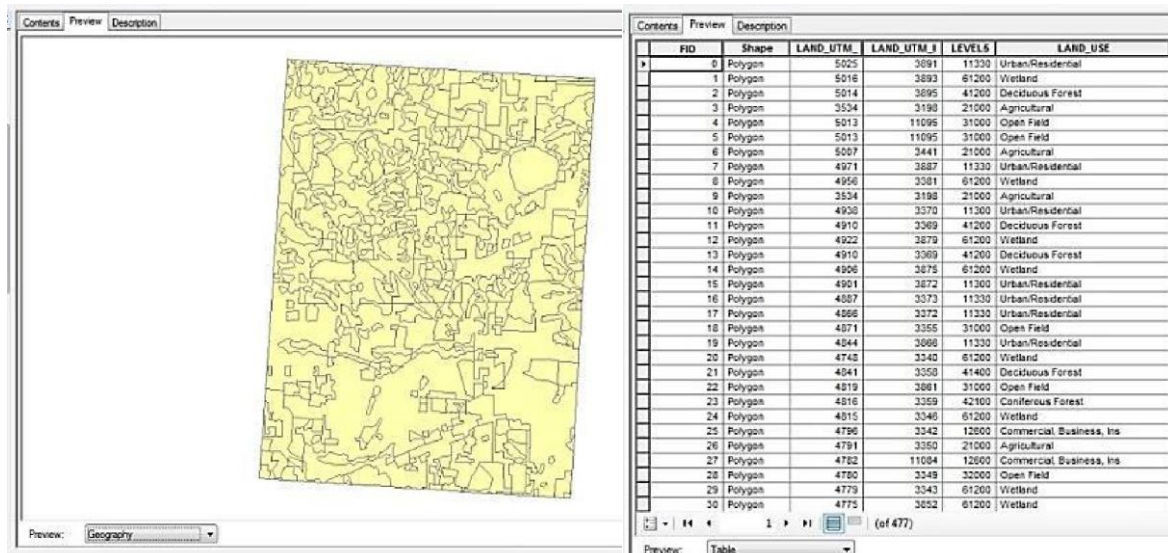
6.3.1.2 Coverage structure

The coverage structure consists of information stored in a system defined in a topologically structured manner (i.e., the data structure has explicitly encoded information about the spatial relationships between features). The geographic and attribute portions of coverages are stored separately and linked together via database relations that are managed with relational tables stored in the INFO directory. Although typically constructed to represent a single geographic feature type (points, lines, or polygons), the coverage can store multiple feature classes within itself. The various features within the coverage are part of a hierarchy constructed to represent the single feature type (i.e., points define the endpoints of arcs, arcs define the boundaries of polygons, and polygons are the feature type of interest). In this section, you will explore the various levels of the coverage within ArcCatalog and Windows Explorer.

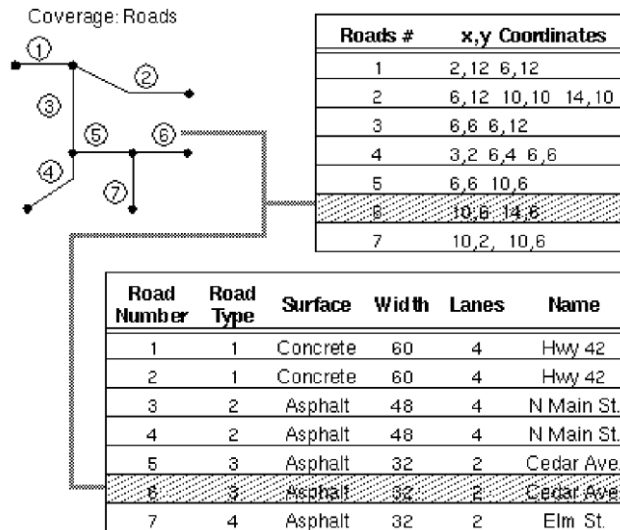
- Within ArcCatalog, connect to your “Lab02_data” directory and observe the files with it. Note what coverage is available for your use (the coverage is composed of various feature types).
- Explore the coverage by clicking on the pluses (+) and minuses (-) next to the file name or by double clicking on the file name.
- Use the *Preview* tab to explore both geography and table (right-click on feature class and open *Item Description* if using ArcCatalog via ArcMap).



- To change between geography and table view, use the pull-down option at the bottom of ArcCatalog within the *Preview Tab*:



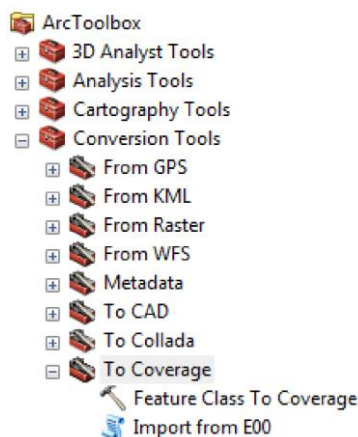
- In the Geography view, use the *geography tools* to explore the various feature classes within the coverages. For example, the arc feature class consists of lines (e.g., roads, rivers, or polygon boundaries); the point or label feature class consists of many points for locations (e.g., cities or study sites) or polygon attribute labels; the polygon feature class contains the geographic extent of all polygons (e.g., lakes or states); and the tic feature class contains the definition of geographic positioning or extent of the coverage.
- Now open *Windows Explorer*, navigate to your lab directory (i.e. ACCD_GIS/Lab02_data) and look for the coverage. Windows Explorer sees a Coverage as a folder because it stores files within itself. You will notice the individual files (arc data files) stored within the various coverages. These files are representative of the various feature classes (arc, aat – arcs, pat/nat – points or nodes, pal – polygon attributes, tic – geographic tics, etc.).
- The ‘INFO directory’ (or folder called “*info*”), is created automatically and is required in any workspace that contains coverages. The *info* directory contains the information needed to track and link the relationships between the various arc data files. An example of this linking process is seen in Figure 1. It shows how the arc attribute table, which contains the descriptive information about the roads, is related to the node attribute table, which contains the coordinates of the nodes (road intersections).



You will not actually see this image; this just gives you an idea of what is going on behind the scenes.

****Note:** The ‘info directory’ is the ‘director’ of the various components of your coverage. Without it, your coverage would be invalid and therefore, non-useable. A single *info* folder can store the information of several different coverages. For this reason, it is very important to use only ArcCatalog or another Arc/Info utility when copying or deleting coverages. ArcCatalog has the ability to maintain the relationships in the ‘info directory’ when copying files.

- One way to transport coverage structures is to compress them. The format is a single Interchange File that has the extension *.e00. The interchange file format is not exclusive to coverages - it can also be used to compress grids and tin files. The ArcToolbox offers options to decompress e00 files (*Conversion tools | To Coverage | Import from E00*).



6.3.1.3 Shapefile Structure

The shapefile structure is much simpler than either the geodatabase or coverage. Each shapefile consists of one feature class represented by a set of related files and does not use topological structuring. Data storage is limited to simple geographic and attribute information. While a shapefile may contain up to seven related files, the three of most importance are the .dbf, .shp, and .shx. Respectively, these files store the data's attribute information, geographic information, and internal relational tables linking the geographic and attribute information. In this section, you will explore the shapefile structure within ArcCatalog and Windows Explorer. For the full list of shapefile file extensions, please refer to *Professional Library* | *Geodata* | *Data Types* | *Shapefiles* | *Shapefile file extensions*.

- Within ArcCatalog, navigate to your working directory and the location of the three Tripoli shapefiles (Tripoli_roads.shp, Tripoli_places.shp and Tripoli_landuse.shp, remember what the icons mean from above) in the Lab02_data.
- Explore the various shapefiles – you should notice that selecting the shapefile itself does not reveal additional feature classes, as it does in a coverage. As stated above, one shapefile = one feature class, even though it takes several different files to create a single shapefile.
- Be sure to use the *preview* tab for both the *geography and table* options to explore your data – notice again that a set of navigation tools has become active for you to investigate your different data sets.
- Now open *Windows Explorer*, navigate to your lab directory and look for the various files that compose the shapefile. The individual files that comprise an individual shapefile will share the same name with different extensions (e.g., Tripoli_roads.shp, Tripoli_roads.shx, and Tripoli_roads.dbf, etc).




6.3.1.4 Creating a point shapefile from tabular data

Tabular data containing X/Y point locations can be easily transformed into point shapefiles (or other formats, for that matter). Check the file gas_stations.txt; the file contains a geographic location (X and Y columns) as well as a label or attribute for each set of coordinates:

"Type",	"X",	"Y"	→ Header
"Gas and diesel",	251372.4,	4690292.0	→ Values
"Gas",	250095.5,	4691953.7	→ Values

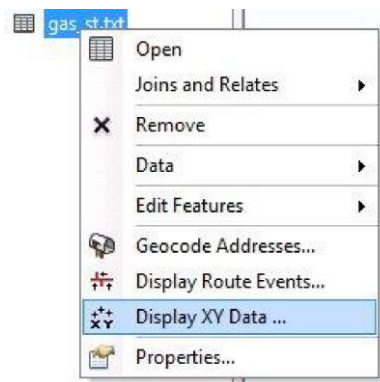
***Note that values must be separated by commas, and quotation marks are required to define text values.*

To create a shapefile from a text file with X/Y locations:

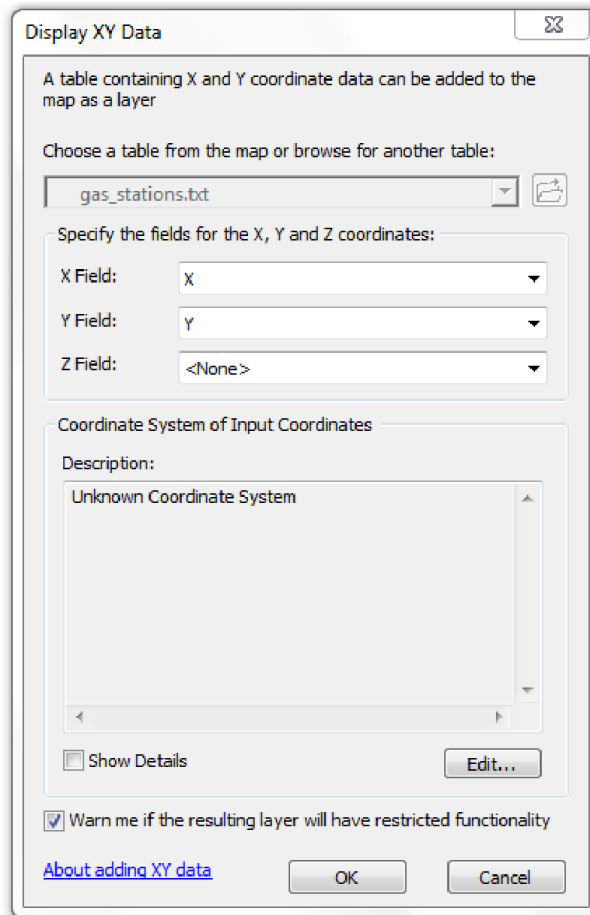
- Open ArcMap and add the table gas_stations.txt (remember the Add Data tool ). Notice that the Table of Contents mode changes from List By Drawing Order  to List By Source .


In the Table of Contents, listed by source, right-click on the table and from the new menu select the option Open to display the table. Review the table so you understand what you're adding.

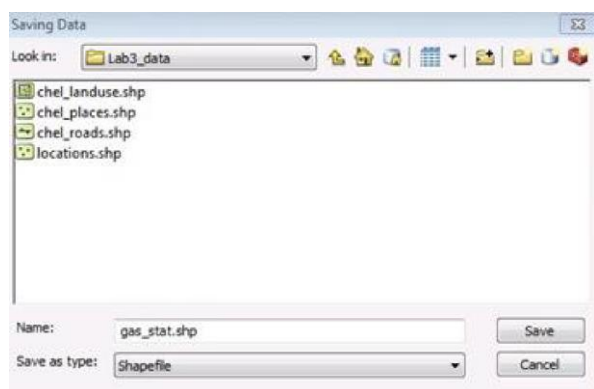
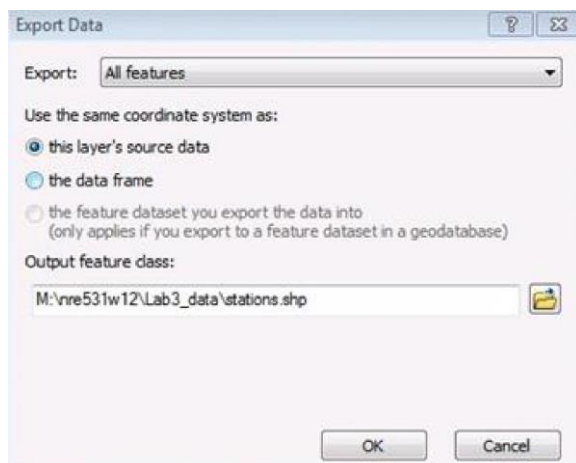
Right-click again on the table and select "Display XY Data;" this option allows you to display X/Y coordinates as a map layer.



A new window will open. Automatically the fields X/Y will be recognized (*Note, in case the software is not able to identify the X/Y fields automatically, the user can manually select the correct X/Y columns). Click OK. If you encounter a warning, click OK.



- A temporary map layer will be displayed as an event (e.g. “gas_stat.txt.Event”).
- To save your file permanently as a point shapefile, right-click the point map file you just created in the Table of Contents; select the option Data | Export Data.
- In the new window, select the option “this layer’s source data” [even though the coordinate system is unknown for this file, it is still good practice]. Click on the  to select your working directory. A new window will open.




- Provide a proper name; use something that reflects the data content. Avoid long names with more than 13 characters; do not leave spaces between words (e.g. NO gas stat.shp / YES to gas_stat.shp). In the option “Save as type” select Shapefile. Select “Save.” You will return to the previous window. To finish the process click OK.
- A window will appear asking if you would like to add the exported data to the map as a layer. Say Yes to display the file.

Microsoft Excel tables can also be used in ArcGIS. Excel tables can be added directly into ArcMap in the same way you did with the text file. If there is more than one worksheet in the file, the user needs to browse the Excel file and select the worksheet they would like to open. In ArcGIS, the first row of a worksheet will be used as the table header. Some of the naming recommendations for the first-row fields are:


- Field names must start with a letter.
- Field names must contain only letters, numbers, and underscores.
- Field names must not exceed 64 characters.

ArcMap can only read 255 characters per cell. Fields have to be consistent, all numeric; all text or all date:

	A	B	C	D
1	Latitude	Longitude	ID	
2	83.24	12.332	None	
3	82.56	11.011	2	
4	83.123	12.891	3	
5				
6				

A red circle with a diagonal slash through it, indicating that the data in this table is incorrect or invalid.

	A	B	C	D
1	Latitude	Longitude	ID	
2	83.24	12.332	1	
3	82.56	11.011	2	
4	83.123	12.891	3	
5				
6				

A green circle with a smiley face, indicating that the data in this table is correct and valid.

CHAPTER 7: INTRO TO COORDINATE SYSTEM AND PROJECTION

7.1 Referencing Data to Real-World Locations Using ArcGIS

7.1.1 Introduction

While a GIS represents reality, it is not reality. To be useful, a GIS map must accurately represent feature locations. To determine the location of features in the real world or on a map, you need a reference system, a standard framework for defining location.

In a GIS, the reference systems used to determine feature locations are called coordinate systems. Two of the most common types of coordinate systems used in a geographic information system (GIS) are:

- Geographic coordinate systems, such as latitude and longitude, which are global or spherical.
- Projected coordinate systems, which provide various mechanisms to project maps of the earth's spherical surface onto a two-dimensional Cartesian coordinate plane.

In this course, you will learn about these two types of coordinate systems, including how they work, how to identify them, and how to define or change the coordinate system for a dataset. By gaining a basic understanding of coordinate systems, you will be better able to manage your data in a way that increases the accuracy of your GIS maps and the work you do with them.

Learning Objectives:

After completing this course, you will be able to:

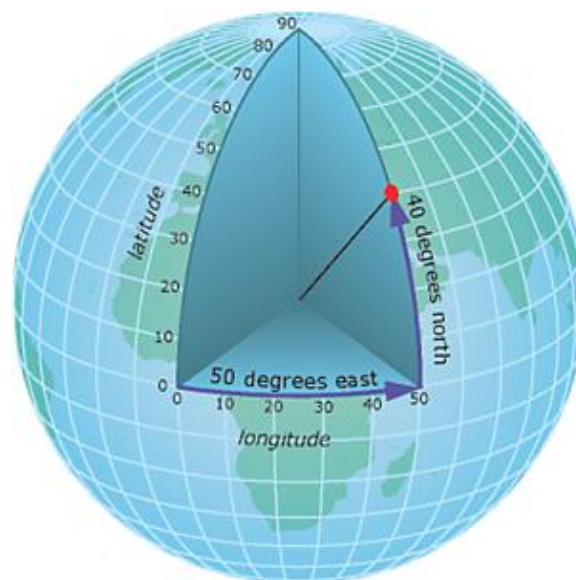
- Differentiate between geographic and projected coordinate systems.
- Choose an appropriate geographic transformation.
- Determine when to use a projected coordinate system.
- Evaluate the appropriateness of a projected coordinate system for a project and area of interest.
- Define the coordinate system for a dataset.

7.1.2 Geographic Coordinate Systems

The pairs of coordinate values that identify a feature on a map are relative to its geographic coordinate system (GCS); a coordinate system that uses a three-dimensional spherical model to identify points or areas on the surface of the earth.

Each coordinate system is commonly illustrated with a network of intersecting lines of latitude (parallels) and longitude (meridians) called the graticule. It is imaginary on the earth, but drawn on globes and maps for reference.

Each point is referenced by its longitude and latitude values. Longitude and latitude are angles measured from the earth's center to a point on the earth's surface. The angles often are measured in degrees (or in radians).



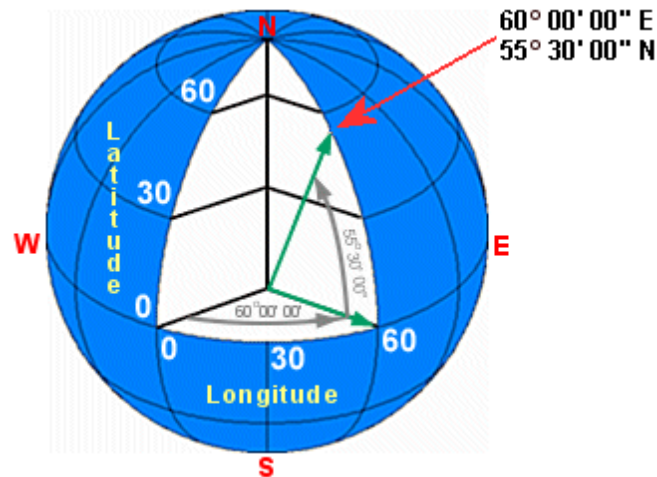
Latitude lines are parallel, run east and west around the earth's surface, and measure distances north and south of the equator. Longitude lines run north and south around the earth's surface, intersect at the poles, and measure distances east and west of the prime meridian.

7.1.3 The Components of a Geographic Coordinate System

Each geographic coordinate system is comprised of four primary components.

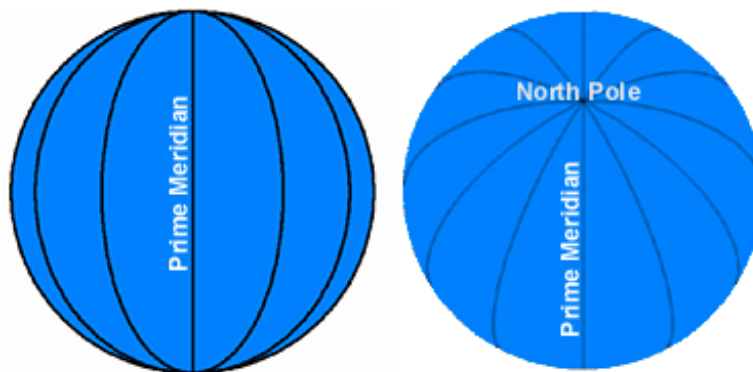
7.1.3.1 Angular units of measure

The graticule of latitude and longitude lines is an angular measurement system. All features on the earth's surface are located using measurements that are relative to the center of the earth. Latitude lines are parallel to each other, whereas longitude lines converge at the poles.



7.1.3.2 Prime meridian

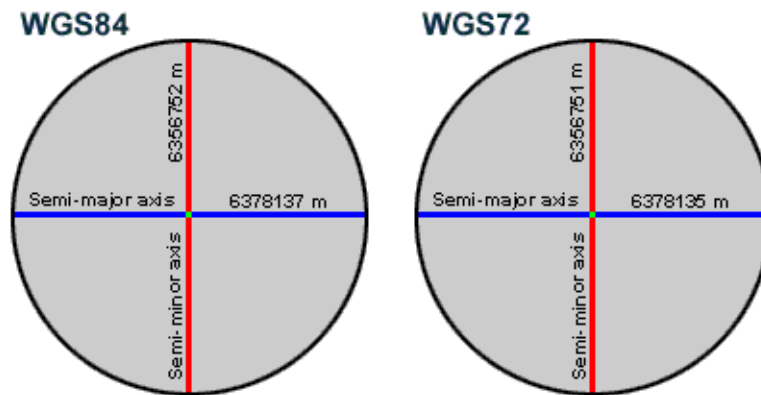
A prime meridian is the zero value for longitude. While the official prime meridian passes through Greenwich in southeast London, United Kingdom, the choice of prime meridian is somewhat arbitrary. It is not necessary to define a zero value for latitude because this is always the equator.



7.1.3.3 Spheroid

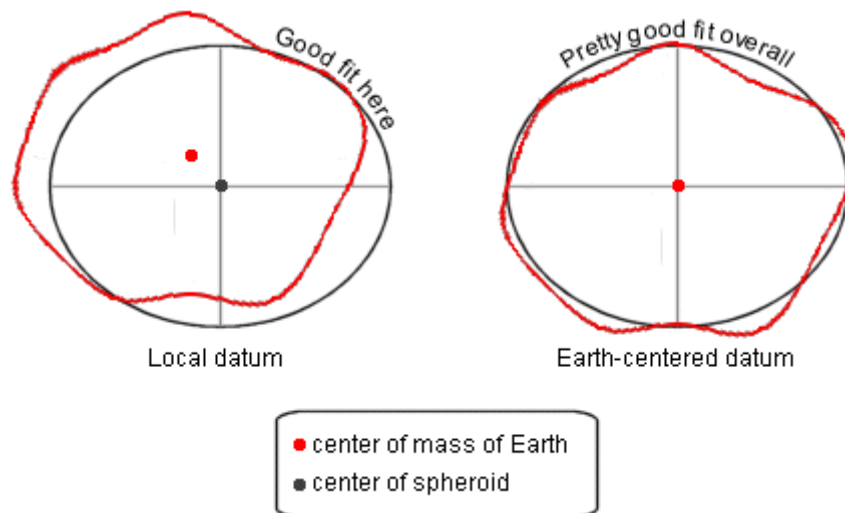
A spheroid is the mathematical model that estimates the size and shape of the earth. Because the earth's surface is not perfectly symmetrical, the semi-major and semi-minor axes that fit

one geographical region do not necessarily fit another one. This is one reason why there are multiple spheroids. Another reason is that, over time, technology has improved measurement techniques and increased overall accuracy.



7.1.3.4 Datum

A datum provides a frame of reference for measuring locations on the surface of the earth. It defines the origin and orientation of latitude and longitude lines. While a spheroid approximates the shape of the earth, a datum defines the position of the spheroid relative to the center of the earth. The underlying datum and spheroid to which coordinates for a dataset are referenced can change the coordinate values.



Each datum has a name and often a date associated with it, such as the European Datum of 1950, the Pulkovo Datum of 1942, or the Worldwide datum WGS 1984. Every datum is based on a specific spheroid because surveyors selected the spheroid that best represented their part

of the earth when they defined a new datum. This means that every datum has an associated spheroid.

The most commonly used datums in North America are:

- North American Datum 1927 (NAD 1927)
- North American Datum 1983 (NAD 1983)
- World Geodetic Survey 1984 (WGS 1984)

7.1.4 Working with Data in Different Geographic Coordinate Systems

A new data frame has no coordinate system. When you add your first layer to it, the data frame adopts that layer's coordinate system. For any subsequent layers that you add, one of two things happens:

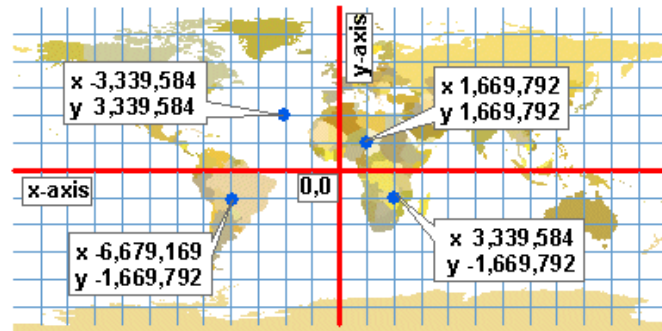
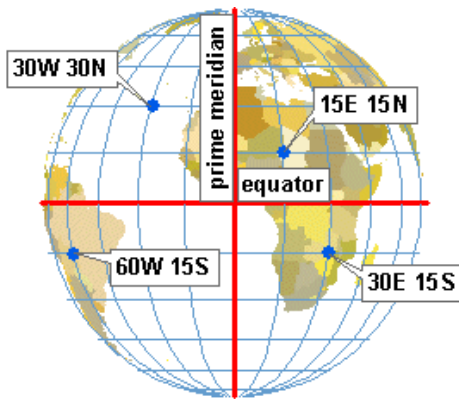
- If the new layer's geographic coordinate system already matches the data frame's geographic coordinate system, the layer is added without question.
- If the new layer's geographic coordinate system is different than the data frame's, ArcMap will warn you that the data may not line up until the coordinate system of the new layer is changed to match the geographic coordinate system of the data frame. This operation is called a geographic coordinate system (or datum) transformation.

7.2 Projected Coordinate Systems

To convert feature locations from the spherical earth to a flat map, the latitude and longitude coordinates from a geographic coordinate system must be converted or projected to planar coordinates. Every projected coordinate system (PCS) includes the following:

- A GCS serves as the starting point for transforming the coordinates to a flat surface.
- A linear unit of measure represents units on a planar surface and is usually meters or feet.
- A projection is the mathematical transformation used to convert from geographic coordinates to planar (projected) coordinates.

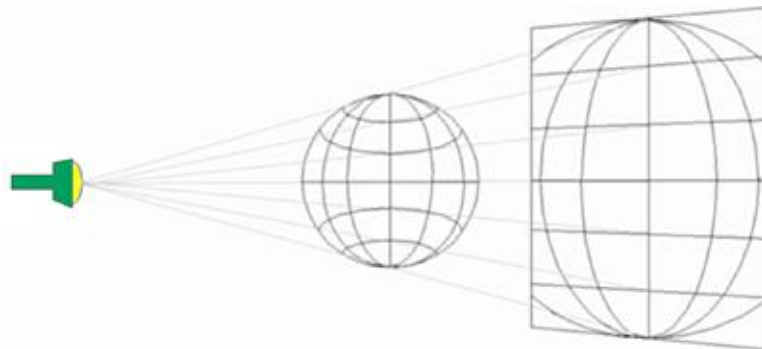
It may be helpful to equate longitude values with X and latitude values with Y. Data defined on a geographic coordinate system is displayed as if a degree is a linear unit of measure.



Round data is described with meridians, parallels, and latitude-longitude values. Flat data is described with x,y units. Projection parameters use both kinds of descriptions. The projection shown here is Plate Carrée.

7.2.1 Map projections

To create maps on paper or on your computer screen, you need to convert the earth's three-dimensional surface to a map's two-dimensional surface. This mathematical transformation is commonly referred to as a map projection. A map projection can represent the earth's entire surface or only a portion of it depending on your needs.



Map projections are so called because of the concept behind their construction. If you could project light from a source through the earth's surface onto a two-dimensional surface, you could then trace the shapes of the surface features onto the two-dimensional surface. This two-dimensional surface would be the basis for your map.

A spheroid cannot be flattened any more easily than an orange peel can be flattened—it will rip. Representing the earth's surface in two dimensions causes distortion in the shape, area,

distance, or direction of the data. Different projections cause different types of distortions. Some projections are designed to minimize the distortion of one or two of the data's characteristics. For example, a projection might maintain the area of a feature but alter its shape.

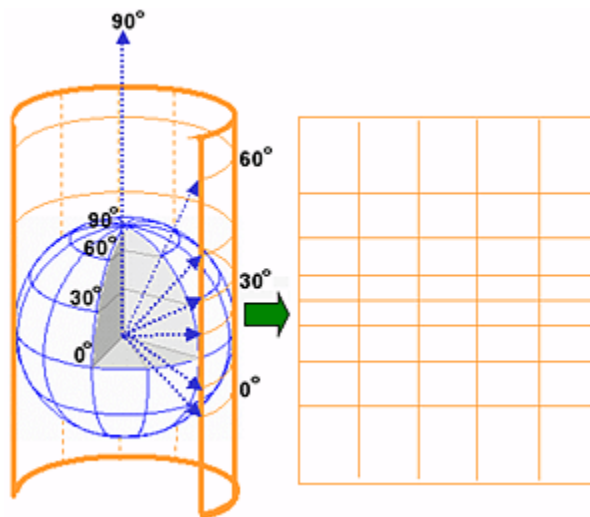
7.2.2 *Projecting Coordinates onto Developable Surfaces*

If you theoretically projected coordinates from a spherical coordinate system onto a surface that could be laid flat without distortion, you could use this surface as the basis of a map. This type of surface is called a developable surface, and cartographers use three of them: the cylinder, the cone, and the plane.

Converting a sphere to a flat surface results in distortion. This is the single most profound fact about projected coordinate systems—they distort the world.

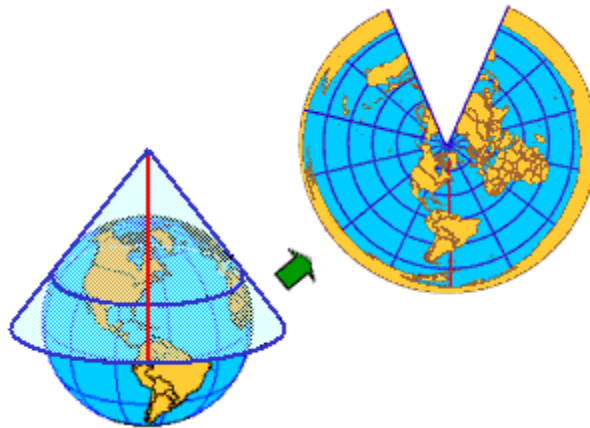
7.2.2.1 Cylindrical projections

Conceptually, cylindrical projections are created by wrapping a cylinder around a globe and projecting light through the globe onto the cylinder. Cylindrical projections represent meridians as straight, evenly-spaced, vertical lines; they represent parallels as straight, horizontal lines. Meridians and parallels intersect at right angles, as they do on the globe.



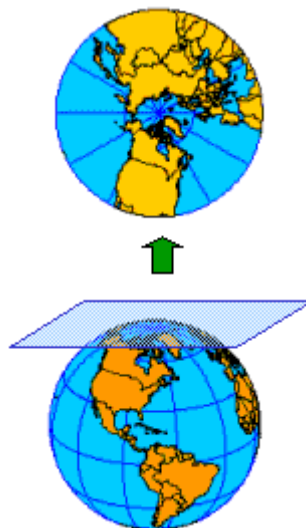
7.2.2.2 Conic projections

Conic projections are created by setting a cone over a globe and projecting light from the center of the globe onto the cone. The simplest conic projection contacts the globe along a single latitude line called the standard parallel. In general, distortion increases north and south of the standard parallel.



7.2.2.3 Planar projections

Planar projections, also called azimuthal projections, project map data onto a flat surface. When that point is either the north or south poles, longitude lines radiate outward from the pole at their true angle. Latitude lines appear as a series of concentric circles. Planar projections are used most often to map the polar regions.



7.2.3 *Projections and Distortion*

Because the earth is round and maps are flat, getting information from a curved surface to a flat surface involves a mathematical formula called a map projection, or simply a projection.

This process of flattening the earth will cause distortions in two or more of the following spatial properties:

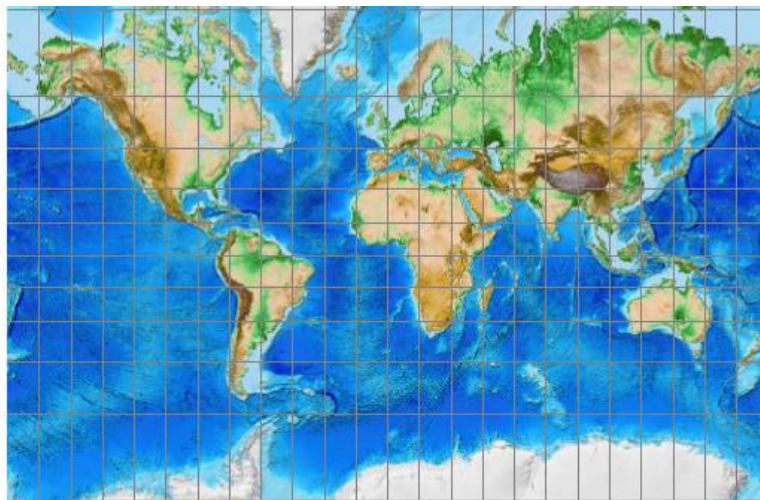
- Shape
- Area
- Distance
- Direction

Classifying projections according to the properties they preserve

Projections can be classified according to the developable surface on which they are based, but they can also be classified according to the spatial property or properties they preserve.

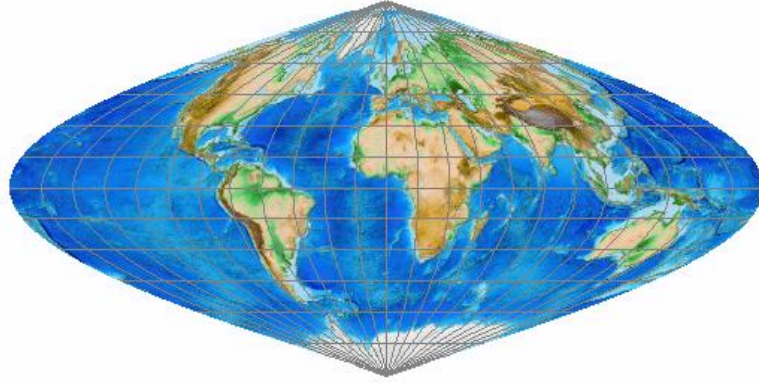
7.2.3.1 Conformal

Conformal projections preserve shape but not area. The Mercator projection is conformal in that angles and shapes within any small area are fairly accurately depicted.



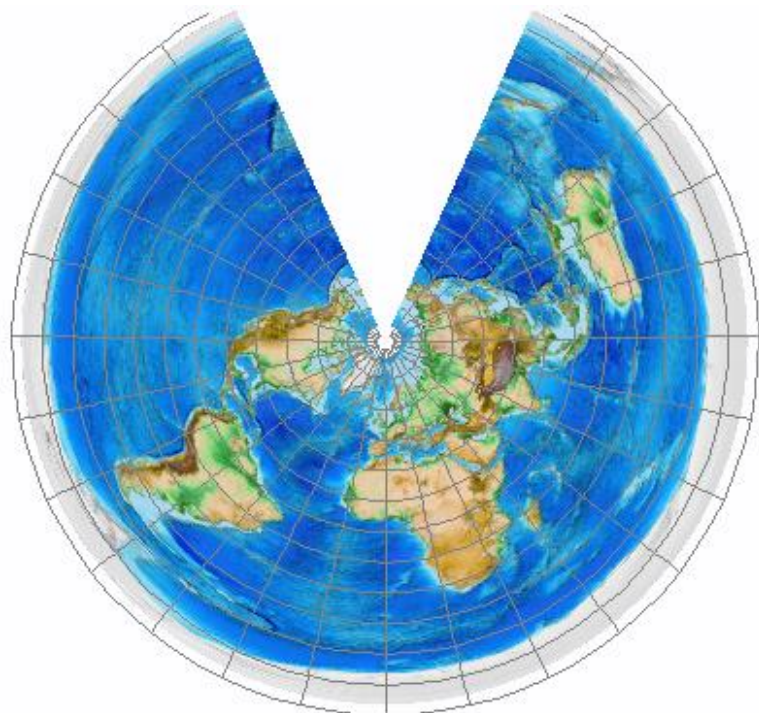
7.2.3.2 Equal Area

Equal Area projections preserve area but not shape. The Sinusoidal projection is an easily plotted equal-area projection for world maps.



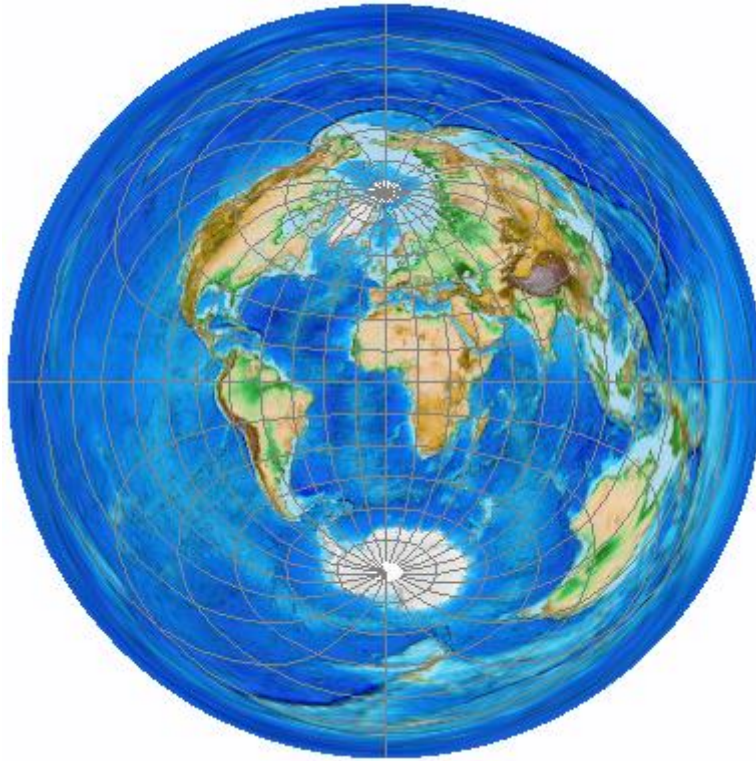
7.2.3.3 Equidistant

Equidistant projections preserve distance from one or two points to every other point. With the Equidistant Conic projection, distances are true only along all meridians and along one or two standard parallels.



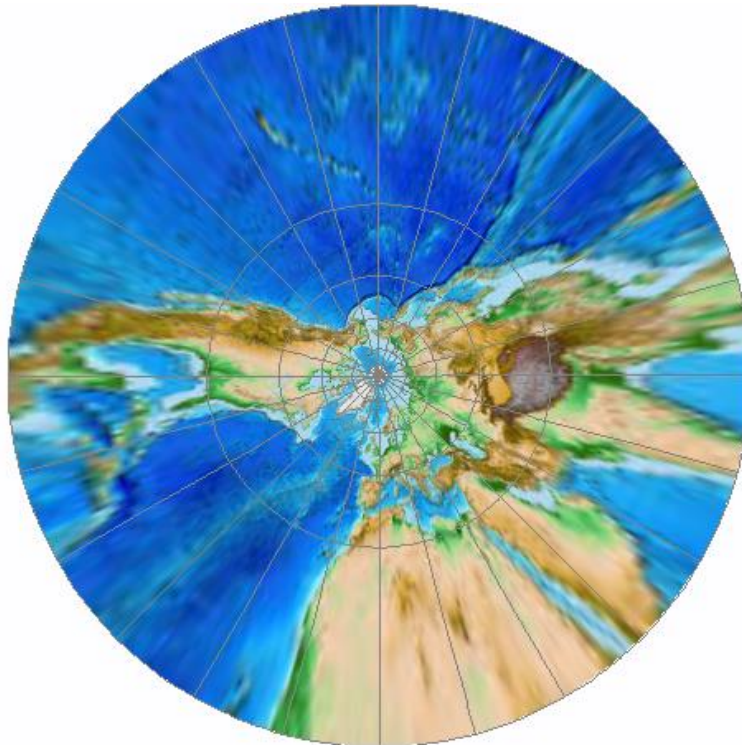
7.2.3.4 Azimuthal

Azimuthal projections preserve direction from one or two points to every other point. With the Azimuthal Equidistant projection, distances and directions to all places are true only from the center point of the projection.



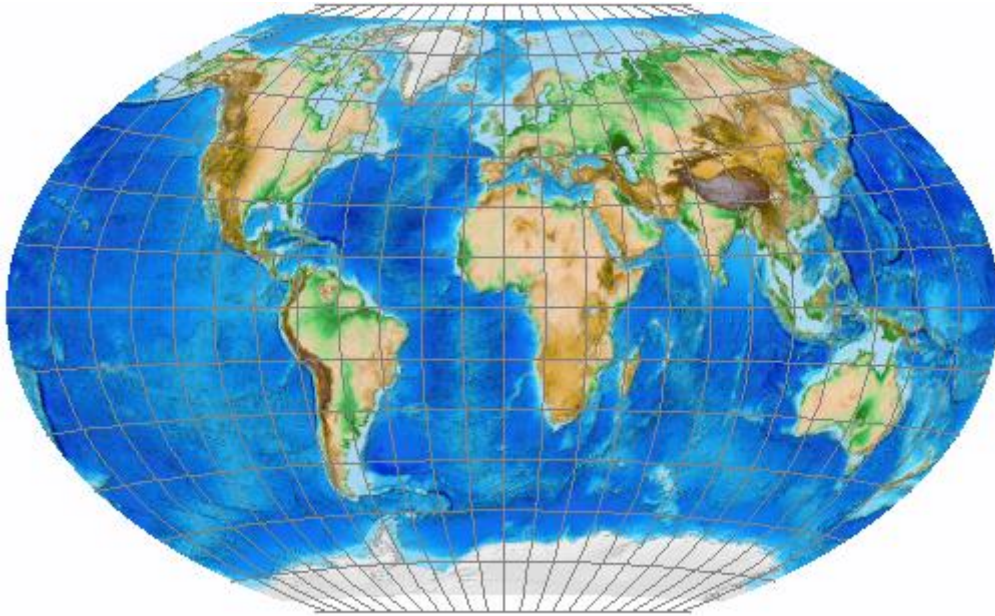
7.2.3.5 Gnomonic

Gnomonic projections preserve the shortest route (distance and direction), but cannot preserve area. With the North Pole Gnomonic projection, any straight line drawn on the map is on a great circle, but directions are true only from the center point of projection.



7.2.3.6 Compromise

Compromise projections try to balance shape and area distortion. No flat map can be both equal-area and conformal; you need a globe for that. The Winkel Tripel projection minimizes overall distortion but doesn't preserve any of the four spatial properties.



7.2.4 *Reasons for Using a Projected Coordinate System*

There are a number of reasons to use a projected coordinate system.

- i. To adhere to standards or mandates

Many organizations mandate using a particular projected coordinate system for all maps created by their employees. Some may have a limited number of projections that they use depending on the type of map that is being created and how that map will be used.

- ii. To make accurate measurements

Degrees are not consistent units of measure for area, distance, or direction. To make accurate measurements in your map and to ensure that the spatial analysis options you use in ArcMap calculate distance correctly, you need to apply a projection.

- iii. To preserve spatial properties

When creating a map in which area, distance, or direction need to be preserved, you will need to apply a projection that preserves the desired property. Note that shape is preserved in a geographic coordinate system.

- iv. To create a small-scale map

When creating a small-scale map, such as national or world map, your choice of map projection determines the overall appearance of the map. For example, with some projections, lines of latitude and longitude will appear curved; with other projections, they will appear straight.

7.3 Integrating Data with Different Coordinate Systems

Data can come from a variety of sources. You may find that the data you need for a particular project is in various coordinate systems, whether geographic or projected. When performing analyses with your data, you should project it all into the same projected coordinate system—one that is appropriate for your objective and area of interest. When simply viewing data together in a single data frame, you can add data with any combination of geographic and projected coordinate systems.

7.3.1 Working with data with different coordinate systems

The process of matching the coordinate systems of layers to the coordinate system of the data frame has many permutations. Use the information in the following table to determine the operation needed for each scenario below.

Data frame's coordinate system	Layer's coordinate system	Operation needed
GCS1	GCS1	None
GCS1	GCS1 + PCS1	None; ArcMap "unprojects" the layer on the fly
GCS1	GCS2	Datum transformation from GCS2 to GCS1
GCS1	GCS2 + PCS1	Datum transformation from GCS2 to GCS1 and on-the-fly unprojection of the layer
GCS1 + PCS1	GCS1	On-the-fly projection of layer to PCS1
GCS1 + PCS1	GCS1 + PCS1	None
GCS1 + PCS1	GCS1 + PCS2	On-the-fly reprojection of layer to PCS1
GCS1 + PCS1	GCS2	Datum transformation from GCS2 to GCS1 and on-the-fly projection of layer to PCS1
GCS1 + PCS1	GCS2 + PCS1	Datum transformation from GCS2 to GCS1
GCS1 + PCS1	GCS2 + PCS2	Datum transformation from GCS2 to GCS1 and on-the-fly reprojection of layer to PCS1

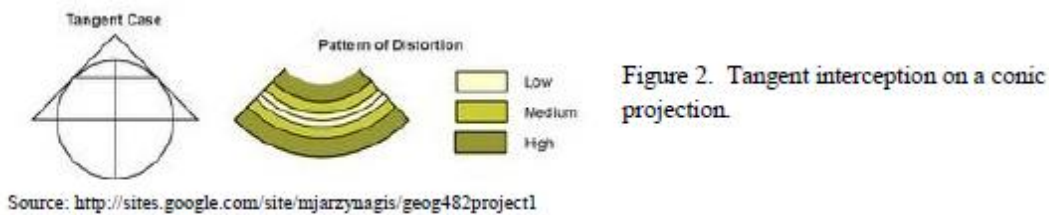
7.4 Lab: Map Projection and Coordinate Systems

7.4.1 Overview

Any planar coordinate system is based on a map projection, which is the orderly transfer of positions or places on the surface of the Earth to corresponding points on a two-dimensional surface, like a sheet of paper. Several elements go into defining a projection. To start with, Earth is nearly spherical. In fact, to better describe the shape of Earth, we use the geometrical figure called a spheroid, which is a flattened sphere and can also be called an ellipsoid. The elliptical nature of the spheroid is defined by two parameters: the semi-minor (polar) and the semi-major (equatorial) axes of the spheroid. These axes are measured from the center of the earth to either the North Pole (polar) or a point along the equator (equatorial). Over the years and with the advancement of measurement technologies, the distances measured for the semi-major and semi-minor axes have changed. The most commonly used spheroids for Lebanon is the Clarke spheroid of 1880 (Clarke 1880).

When projecting from a geographic coordinate system (latitude, longitude) to a planar coordinate system (UTM, state plane, etc.), there have been several standards developed that provide a frame of reference for measurement. These standards (datums) define the origin and orientation of the grid that makes up each particular coordinate system. Datums that are frequently used in Lebanon are defined based upon the Clarke 1880 spheroid or on the GRS80 spheroid and uses the center of Earth's mass as its origin. When specifying the parameters of a projection, you will typically specify either the spheroid or the datum but not both.

Finally, it must be understood that distortions of shape, area, distance and direction are inherent to all projections. Projections can be conformal, equal-area, equidistant, or true-direction. Projections may minimize distortion in various categories, but no projection can minimize distortion in all the categories. The attempt to minimize the distortions is approached by looking at various ways of projecting a spheroid onto a flat piece of paper. Projections that are cylindrical, conical, or planar in nature use various points of intersection or tangency with the spheroid upon which to minimize the distortion. In defining these points of tangency, or standard parallels, we are specifying the extent of our study area for which distortions will be minimized. The further away you move from these defined areas (standard lines), the more distortion you will encounter.



Learning Objectives:

- To explore the basic structure of a planar coordinate system;
- To learn to re-project geographic data into various coordinate systems;
- To understand the distortions of shape, area, distance, and direction that are inherent to the various types of projections;
- To work with multiple data frames in a map layout

7.4.2 Procedure

7.4.2.1 Exploring a Cartesian coordinate system

- Start ArcCatalog and connect to your working directory
- Launch ArcMap with the appropriate icon from the toolbar
- Add the Cadastral.shp and villages.shp from your lab03_data using the technique used in earlier labs. Click OK if you get the following error message:

** It is important to note that this error will be displayed any time a layer is added without a defined projection.

- Use the identify tool to determine the coordinates of any of the villages.shp.



7.4.2.2 Displaying data layers of different projections in ArcMap

- Start a new view within ArcMap by selecting the New Map File icon (do not save the previous view of the grid and triangle).
- Drag and Drop (or any other method to add data) villages.shp. Then add Mohafaza.shp from your working directory into the display area of ArcMap. Drop in the villages.shp

layer first. (If you add Mohafaza.shp first, the data will be slightly offset; this is due to an unusual discrepancy in converting between these two particular datums.)

- These layers are in two different projections (Note the “Geographic Coordinate Systems Warning”). ArcMap automatically re-displays all layers in the data view to the projection of the first layer added, also referred to as “projecting on-the-fly.” (In the future be sure to add the data with your projection of interest to the data frame first).
- To find out the projections of villages.shp and mohafaza.shp, right-click on the layer in the table of contents (TOC) and navigate to “Properties” then the “Source” tab.
- You can also find out the projection of a shapefile within ArcCatalog by right-clicking on the data layer, selecting “Properties”, and clicking “XY Coordinate System”. The projection is specified by the highlighted projected system in the list with full details in the “Current coordinate system” box below.

Make sure you understand the different functions of on-the-fly projection (what ArcMap just did to mohafaza.shp so that it could be displayed with villages.shp), projecting using ArcToolbox (what you are going to do next), and defining a projection. If the projection is not defined to begin with, ArcMap has no idea how to re-project the data or project them on-the-fly. (ArcMap will display data with undefined projections, but only as data on a meaningless coordinate system that cannot be mapped to any of the coordinate systems it knows).

When two layers have different coordinate systems, why would we bother to re-project them if ArcMap can automatically redisplay differently projected layers just fine using on-the-fly projection? (Answer: Because many analysis tools within ArcToolbox require datasets to have the same defined projection, and other GIS software, such as ArcView and ArcInfo, do not automatically redisplay differently projected layers).

7.4.2.3 Re-projecting GIS data with ArcToolbox

In this section, you will re-project a feature class, Caza, to observe how different projections affect the form of Lebanon.

- Launch ArcToolbox from the ArcCatalog toolbar or ArcMap toolbar. (The ArcMap Toolbar is shown below)
- Select Data Management Tools | Projections and Transformations | Feature | Project and double-click to open the tool.

- For the Input Dataset or Feature Class, navigate (using the folder button on the right) to the caza feature class within the Lebanon.gdb (located in your working directory in Lab04_data). This is the feature class you will be re-projecting throughout this section. Click Add.
- Type a name for your new geodatabase feature class under Output Dataset or Feature Class (something like caza_lamb for Lambert conformal conic).

Be sure you save this and all further projections to your working directory under the Lebanon.gdb geodatabase by clicking on the folder to the right and navigating to the appropriate directory. Click Save.

- Click the button next to Output Coordinate System.
- Expand the Projected Coordinate Systems folder (double-click or click the plus sign) and then select Lambert Conformal Conic

Note:

- If you are projecting in ArcCatalog, you may need to refresh ArcCatalog from the contents tab before you can see the new dataset; to do this, go to the directory of interest and press F5, then double check that the geodatabase feature class is created by inspecting it within ArcCatalog (use the preview tab to look at the newly created data – select the file then click preview).
- If you are projecting in ArcMap, open up ArcCatalog to inspect your geodatabase feature class (use the preview tab to look at the newly created data – select the file then click preview).

7.4.2.4 Creating layouts with multiple data frames:

Instead of opening up four separate ArcMap sessions, or clearing the data frame coordinate system (View | Data Frame Properties | Coordinate System tab | Clear button), this section will demonstrate how to add additional data frames to ArcMap. You will then use these four data frames to produce a single layout with four inset data frames, one for each projection of the states layer.

- Add a data frame to your map for each new re-projection by clicking Insert and selecting Data Frame.

- Follow the same re-projection procedure as above for each projection on the next page.

In general, each new data frame can be populated by using the Add function (in ArcMap) or drag and drop (from ArcCatalog).

- Note that when in data view, only one data frame can be viewed at once. To switch between data frames, you need to activate the frame you want to view by right-clicking on the frame's name (here, "New Data Frame") and selecting Activate from the menu.
- In layout view, you should be able to see the frames. (You may need to drag the new data frame off of the old one to see it).
- Arrange the data frames in your layout before adding other layout elements. It is a good idea to make the data frames equal sizes in this case by reducing the first to the size you think you want, then overlaying the new frame on top of it, stretching/shrinking it to match up exactly, then dragging it to where you want it. Make sure the two frames do not overlap anywhere data will be. Also make sure to leave room for other layout elements.
- Use the scale pull-down window to ensure that the scales are the same for the maps. If your scale is the same for each data frame, there is no need for multiple scale bars.

However, different projections will likely have different scales, requiring a scale bar for each data frame. Likewise, if your symbology is the same for the all data frames, there is no need for multiple legends. If there are differences, you need to be aware that map elements are linked to their data frame, so you need to know which data frame you are in when adding elements (e.g. don't make a scale bar for one frame and place it next to a different frame at a different scale).

- If you want to make all data frames with the same scale units, you can specify this in the Scale Line Properties by double clicking the scale. Under the Scale and Units tab, choose "Adjust width" from the When resizing pull down box and set your desired Division Value. Make sure your units are appropriate. Under the Numbers and Marks tab, you can specify the frequency of numbers and marks of the scale as appropriate for your chosen divisions and subdivisions (e.g. Frequency: "ends (and zero)").
- Re-project the caza feature class four more times using the following projection definitions and following the procedures from the steps above each time. Assign an appropriate name to each of the outputs from the projection procedures,

caza_<projection>. (Remember: never put spaces in the names of files, and keep names short!)

- 1) Stereographic Projection (Use the same procedure as above, but modify the projection).
- 2) Double Stereographic:
- 3) UTM Zone 36 N

7.4.2.5 Examining the effect of the different projections


Use ArcMap to explore the various outcomes of the projections processes

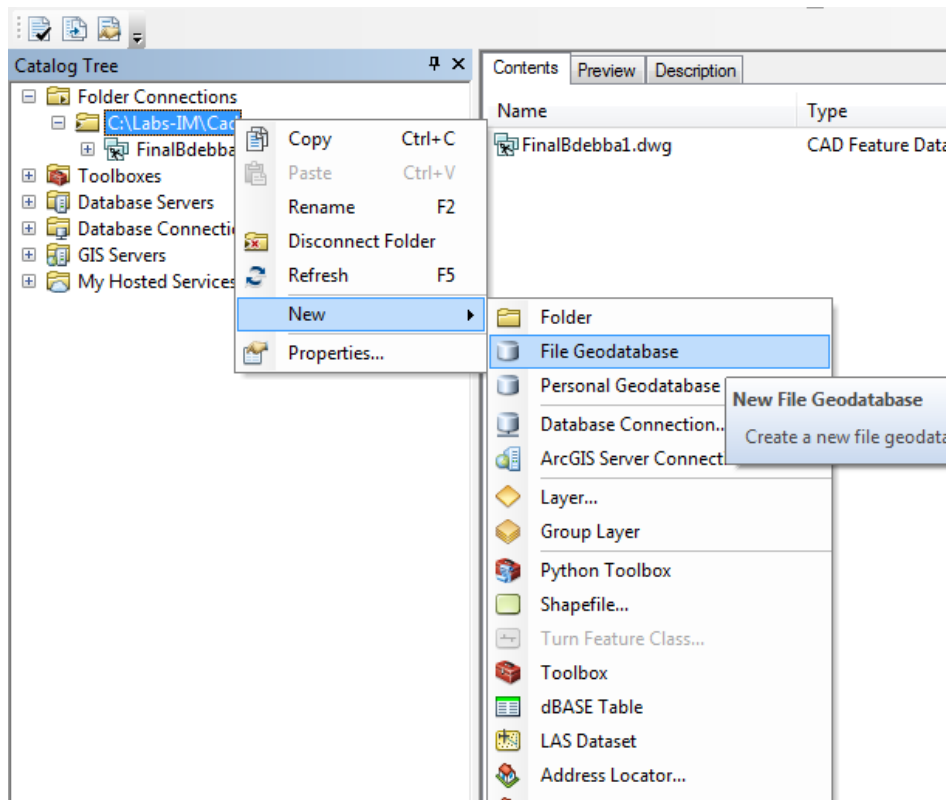
- Within each data frame of your ArcMap session, make sure you have added one of your projected views of Lebanon Caza, and make sure it is not projecting on-the-fly by checking the Data Frame Properties | Coordinate System tab.
- For each of the projections, you can Activate each of the data frames individually, toggle back and forth between the data frames to compare them (and look at the Layout view) and use the identify tool to fill in the table (please make sure you use the correct area attribute– the Shape_Area, and not Area).
- Note the differences between how the areas of Akkar caza changed with each changing projection. You should have an understanding of why these differences occur.

Map to Submit: You will show each of the four projections of Lebanon on one layout using multiple data frames.. Include map elements of title, north arrow, scale(s), legend(s), source, and “layout by”.

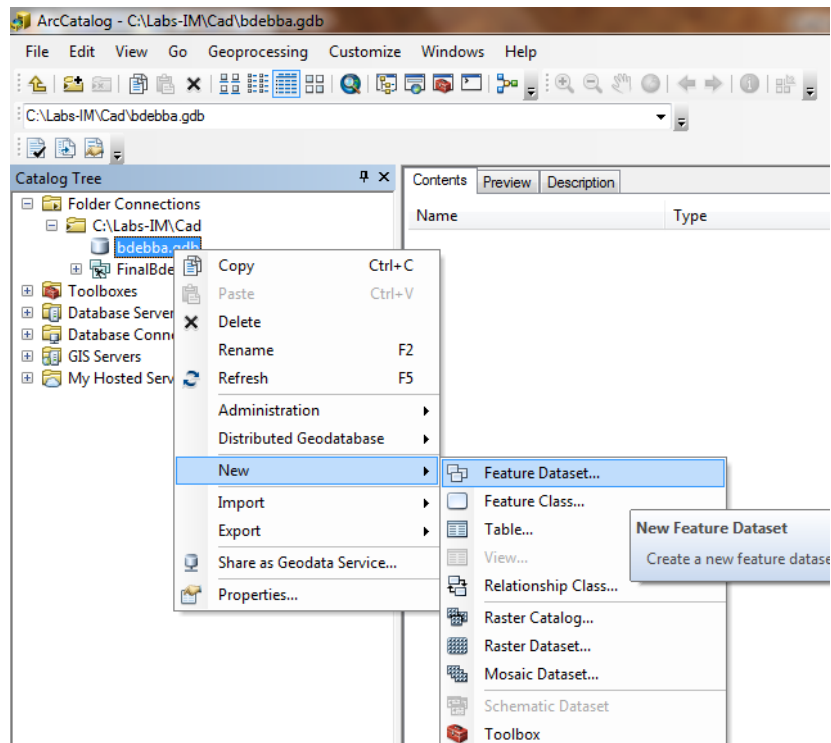
CHAPTER 8: CAD TO GIS

8.1 DWG to Geodatabase

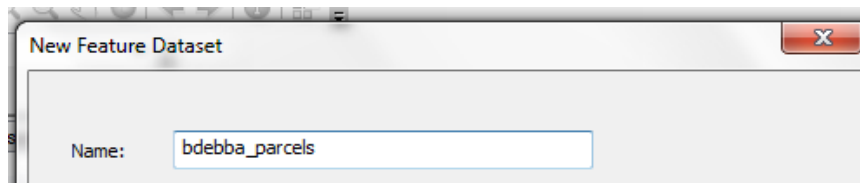
- Open “ArcCatalog 10.1” (Start > ArcGIS > ArcCatalog)
- Connect to  the folder containing your Autocad drawing.
(i.e. C:\Labs-IM\Cad\FinalBdebba1)
- Right click on the folder you just connected to; go to new – file geodatabase



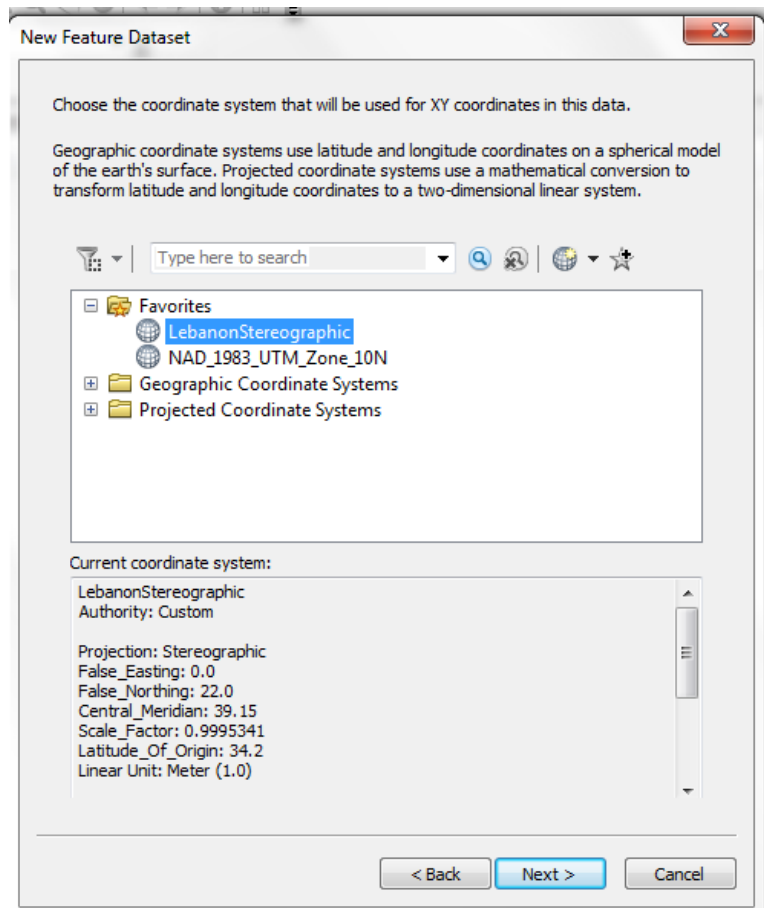
- Name the geodatabase you village name (i.e. Bdebba)
- Right click on you database you just created (bdebba.gdb) go to New – feature dataset.



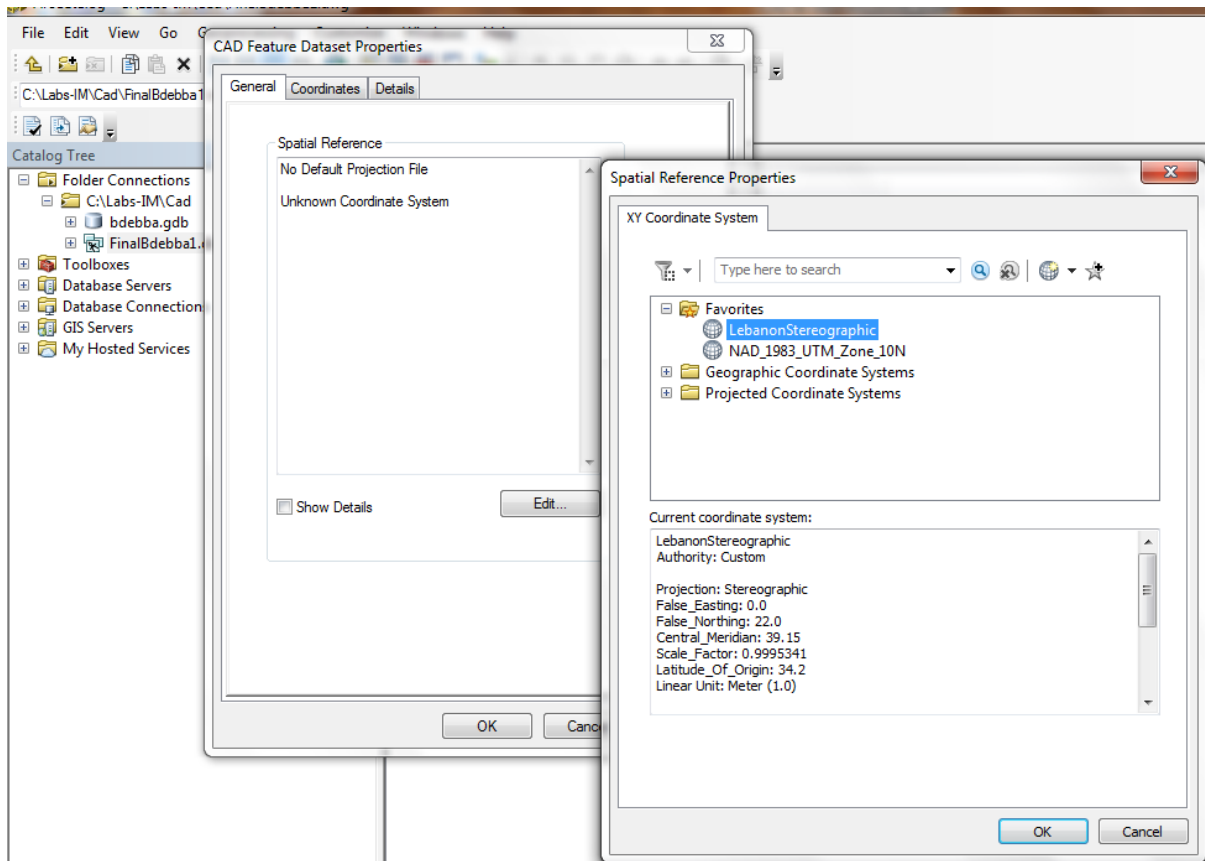
- Name: Village_parcel (Bdebba_parcel)




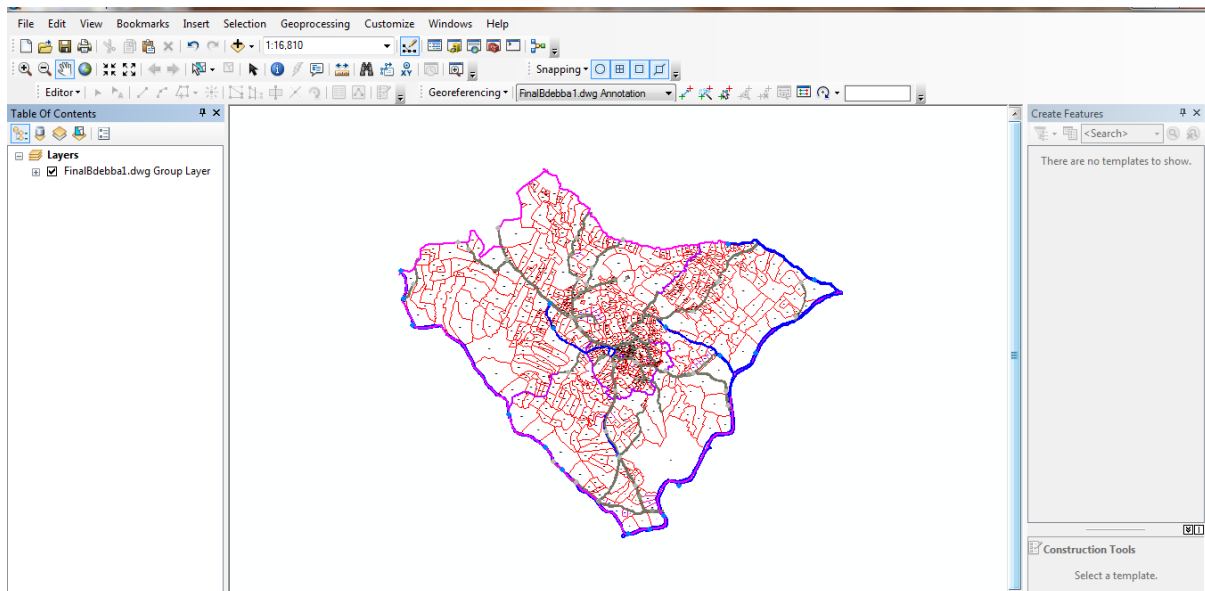
- Click on Next
- For the coordinate system: choose the appropriate projection depending on the location.
In our example we choose the “Lebanonstereographic” projection.
- Click on next. Disregard the Z-coordinate system and finish.




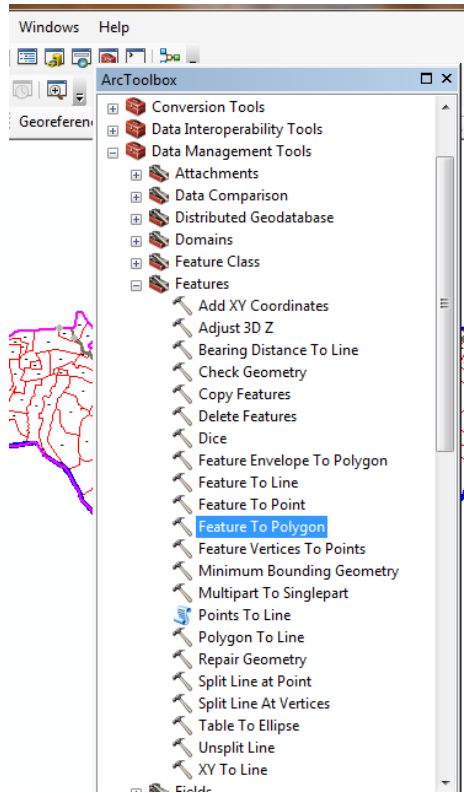
- Make sure your CAD map has a projection. Right click- properties. If the coordinate system is not defined: Click on Edit and choose “Lebanonsterographic” projection.




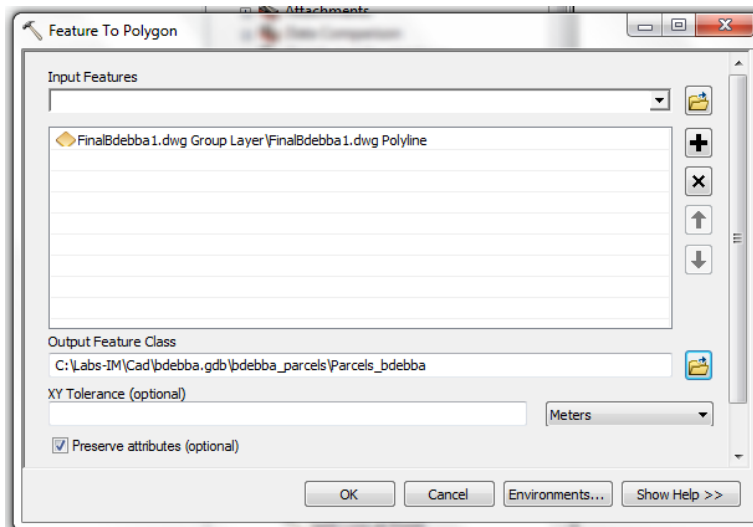
- Launch ArcMap 
- Add the AutoCAD map (finalbdebba1.dwg)



- To create the parcels polygon (عقارات) click on ArcToolbox 
- Go to Data Management Tools/ Features/ Features to Polygon

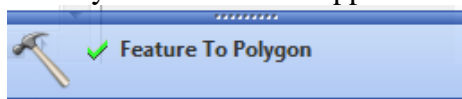


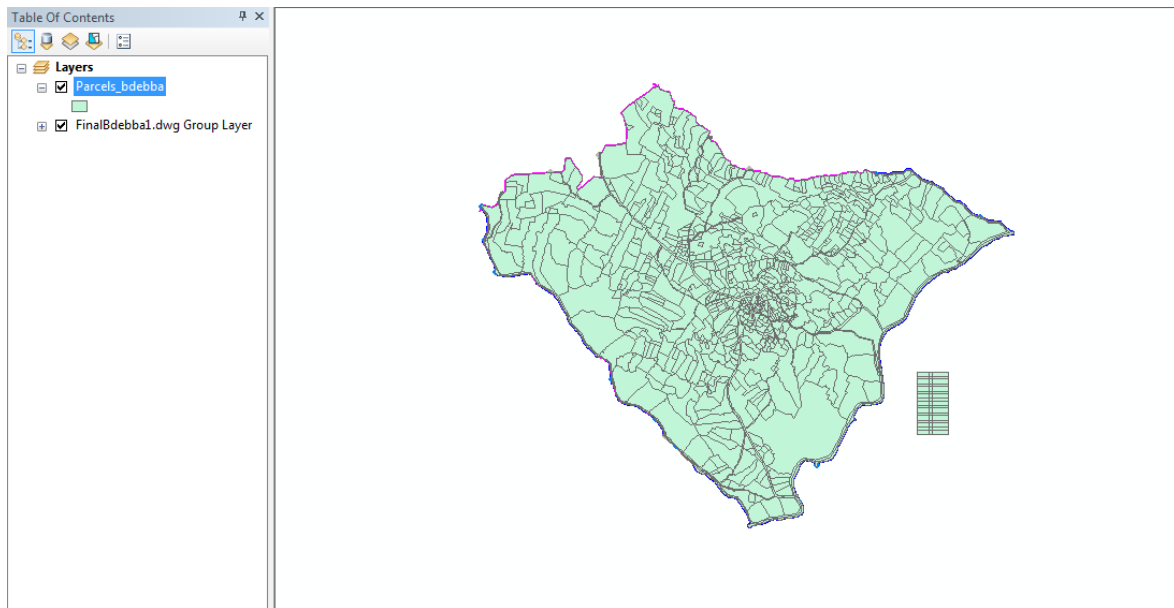
- “Input Feature”: autocaddrawing.dwg polyline
- For the output click on Browse  and look for the dataset you created in your geodatabase.



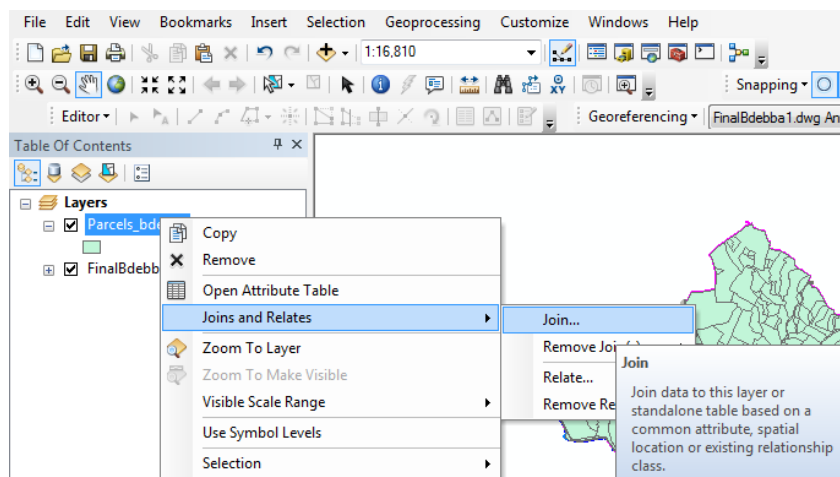
- Click on OK and wait for a few seconds.

If you followed the steps correctly this box should appear on the lower right of the screen

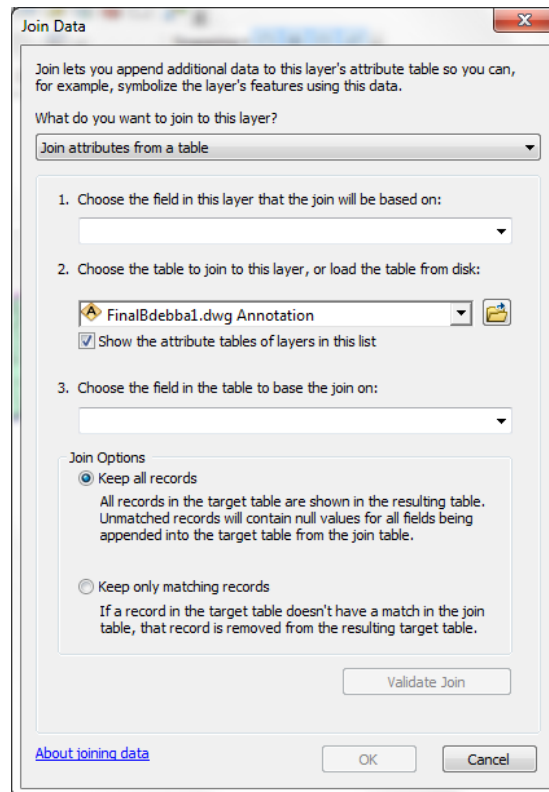




- Open the attribute table of the parcels created (Right click – Open attribute table) No lot number is present (لا رقم عقار) close the table
- Right click on the parcels shapefile, go to joins and relates- join

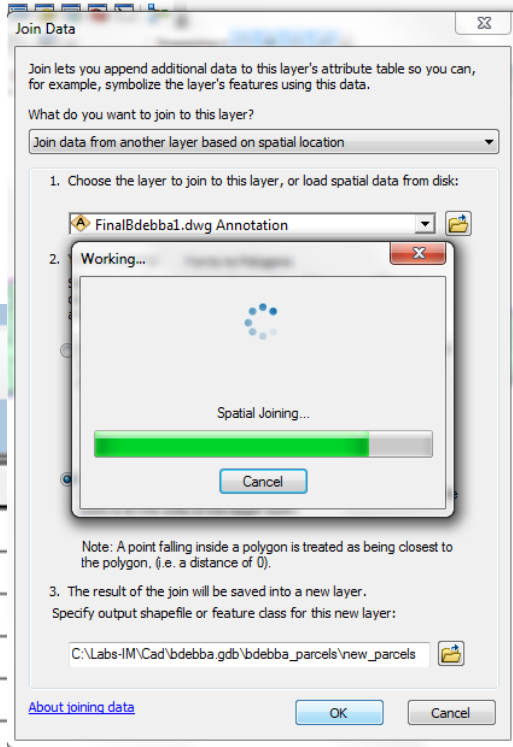


- This window will open

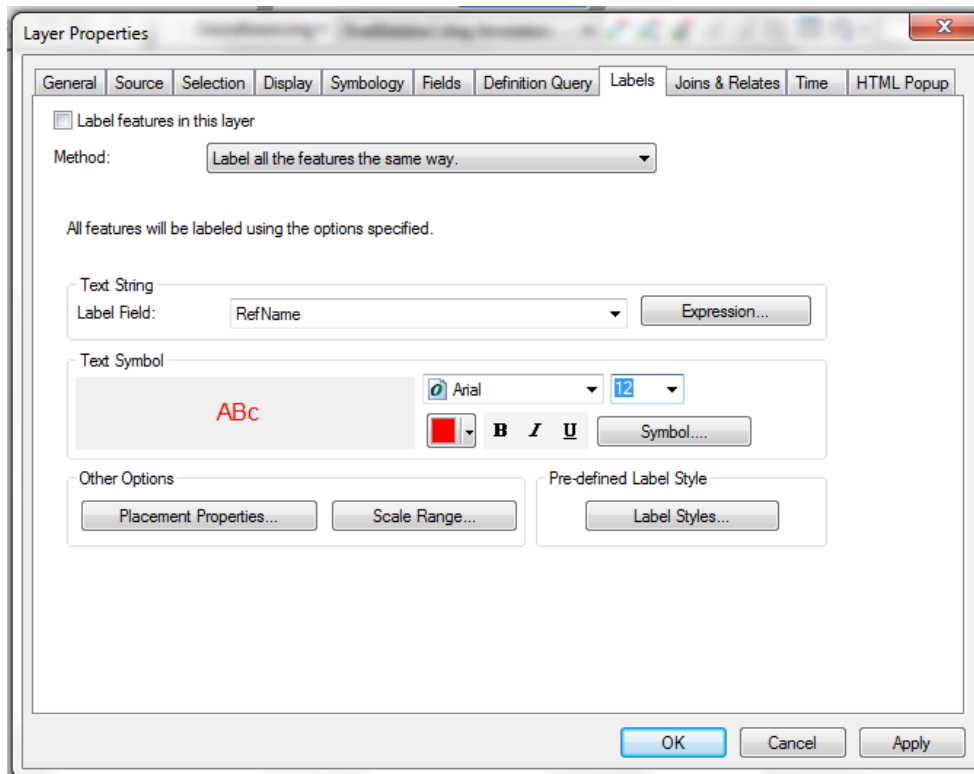


- What do you want to join to this layer? Join data from another layer based on spatial location.
- Autocadname.dwg Annotation
- 2nd option
- Save the joined output in your database.
- If you open the table now you will find the lot number under refName or text field.

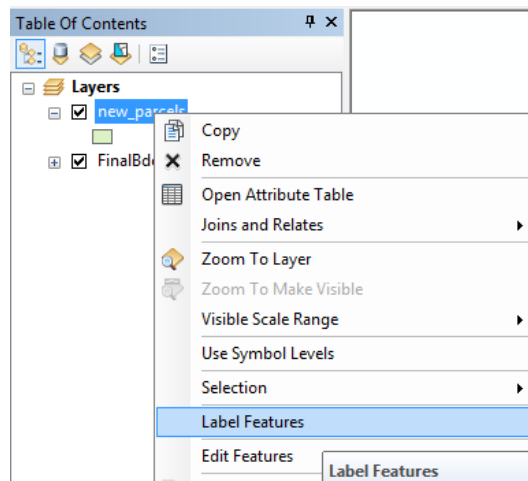
RefName	Style	Text
753	STANDAR	753
747	STANDAR	747
736	STANDAR	736
745	STANDAR	745
746	STANDAR	746
721	STANDAR	721



- Right click on the new parcels shapefile and go to Properties – Labels. Click on the dropdown option in the text string and choose either RefName or Text for the Label field. You are able to change the color, font and size of the labeling. Close the window.



- Right click on the shapefile and click “Label Features”

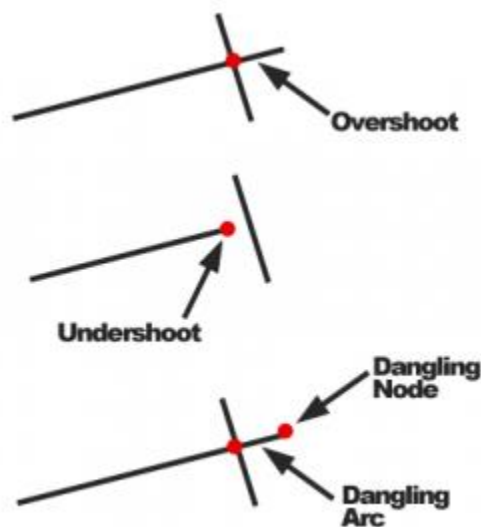


Now your conversion from Autocad to GIS is Complete!

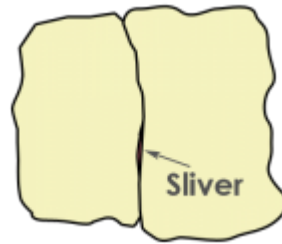
8.2 Editing for Cleaning and QA/QC

Unknowingly, errors can exist in GIS data. You may not even know. That’s why we can fix errors automatically with topology rules. For example, positional accuracy, digitizing error, data generalization, and age of data are often sources of data error.

Dangles, overshoots and **undershoots** are examples of errors.



SLIVERS are small, spurious gap often as a result of imprecise digitization of features in GIS.



OVERLAPS are polygons or lines that are on top of each other which could result in double-counting areas.



Also, when you have **two** conflicting data sets, **conflation (edgematching and rubbersheeting)** resolves conflicts by preserving the most accurate one.

8.2.1 ESRI Topology Rules in ArcGIS 10

Topology rules ensure connectivity between GIS datasets. It fills in gaps, removes slivers, and fixes overlapping polygons. When polygons share a boundary, topology rules enforce data integrity.

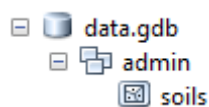
... But you already know that.

Let's get into the nuts and bolts of how to use topology in ArcGIS.

8.2.1.1 Add your data to a geodatabase

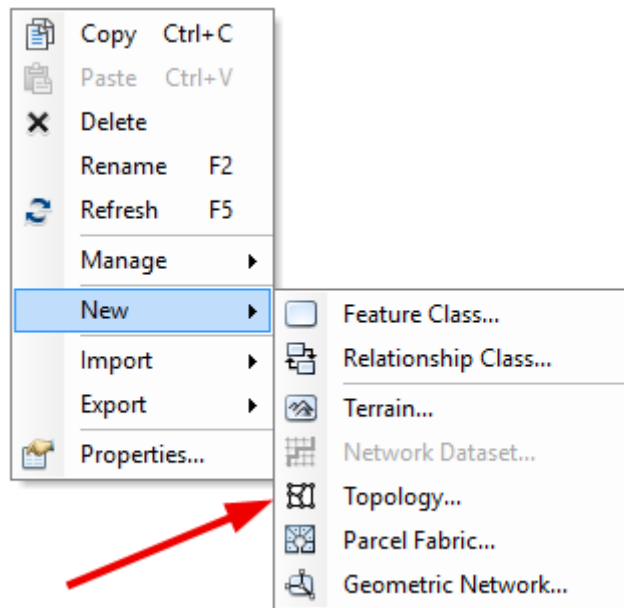
Creating topology rules in ArcGIS is just a couple of simple steps.

- First, add your data in a feature data set.



8.2.1.2 Create topology data set

- Once your data is in a geodatabase, you have to create a topology data set.

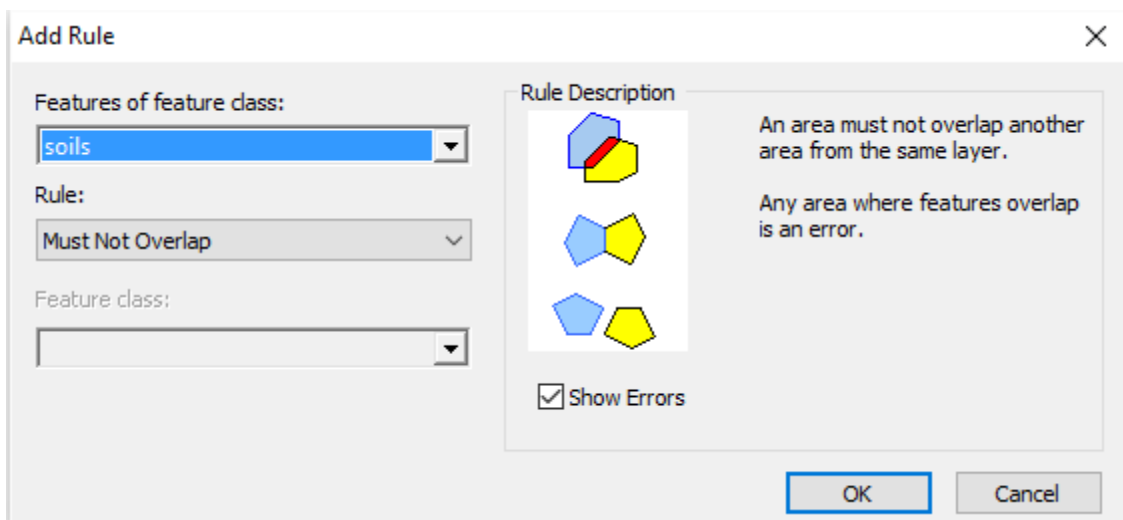


- Follow the instructions by selecting your feature class.

8.2.1.3 Add topology rules

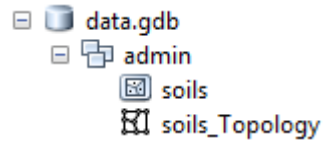
There are over 30 rules you can add to inspect the quality of your data.

- One-by-one, you add rules which govern exactly how to perform quality control to your spatial data. Below is a screenshot of adding an overlap rule.




Overlapping polygons can be resolved by the topology rule *“must not overlap”*.

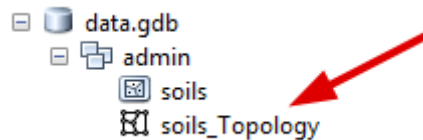
- Select yes to validate. This is how ArcGIS refreshes and ensures that it's using the most recent data errors. The geodatabase should have a topology dataset within it now:



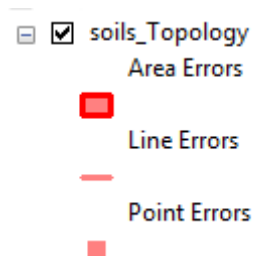
8.2.1.4 Inspect your data

Once your data is in a geodatabase and rules have been added, you can now inspect your data topology. ArcGIS topology fixing is interactive.

- One-by-one, you can go through errors and fix them with the error inspector.
- Click the Add Data button  Add Data.
- Add the topology data set.



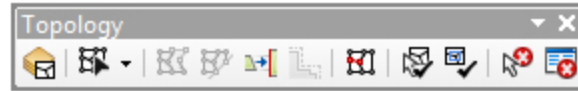
- The errors in your data will appear as area, line, and point errors.



8.2.1.5 Correct your errors

The ArcGIS topology tools support richer, complex error checks. Resolve topology issues with automatic or manual fixes.

- Add the Topology Toolbar by clicking Customize > Toolbars > Topology. If the Topology Toolbar is greyed out, you have to turn on Editing mode by clicking Editor > Start Editing.



- When you push the error inspector button, you can manually go through each error.



Error Inspector

- Going through the error inspector report, you can right-click individual errors for quick fixes.

Here are the types of topology fixes you can perform:

- **MERGING:** The overlapping portion will be subtracted from one polygon and added to the other polygon.
- **SUBTRACT:** The overlapping portion will be removed completely from both polygons.
- **CREATE FEATURE:** The overlapping portion will be removed completely from both polygons and added as an entirely new feature.

Preventing Errors by Using the Snapping Environment

Using snapping in GIS can prevent errors from happening altogether. The snapping environment determines the distance that newly digitized points, lines, and polygons will occupy the same location as existing features.

When new features are created in the editor, they will be automatically snapped to existing features. For example, you can snap to edges, points, ends, intersections, midpoints, and tangents.



What's Next?

When errors exist, topology tools clean geometry for GIS features.

Topology is especially useful for detecting and correcting errors of digitized data because it checks and validates the spatial relationship of neighboring and overlapping features.

Topology editing is a strong point in ArcGIS with a ton of options to fix editing errors.

CHAPTER 9: GEODATABASE

At its most basic level, an ArcGIS geodatabase is a collection of geographic datasets of various types held in a common file system folder, a Microsoft Access database, or a multiuser relational DBMS (such as Oracle, Microsoft SQL Server, PostgreSQL, Informix, or IBM DB2). Geodatabases come in many sizes, have varying numbers of users and can scale from small, single-user databases built on files up to larger workgroup, department, and enterprise geodatabases accessed by many users.

But a geodatabase is more than a collection of datasets; the term geodatabase has multiple meanings in ArcGIS:

- The geodatabase is the native data structure for ArcGIS and is the primary data format used for editing and data management. While ArcGIS works with geographic information in numerous geographic information system (GIS) file formats, it is designed to work with and leverage the capabilities of the geodatabase.
- It is the physical store of geographic information, primarily using a database management system (DBMS) or file system. You can access and work with this physical instance of your collection of datasets either through ArcGIS or through a database management system using SQL.
- Geodatabases have a comprehensive information model for representing and managing geographic information. This comprehensive information model is implemented as a series of tables holding feature classes, raster datasets, and attributes. In addition, advanced GIS data objects add GIS behavior; rules for managing spatial integrity; and tools for working with numerous spatial relationships of the core features, rasters, and attributes.
- Geodatabase software logic provides the common application logic used throughout ArcGIS for accessing and working with all geographic data in a variety of files and formats. This supports working with the geodatabase, and it includes working with shapefiles, computer-aided drafting (CAD) files, triangulated irregular networks (TINs), grids, CAD data, imagery, Geography Markup Language (GML) files, and numerous other GIS data sources.
- Geodatabases have a transaction model for managing GIS data workflows.

Each of these aspects of the geodatabase is described in more detail in the other topics in this section of the help.

9.1 Types of geodatabases

The geodatabase is a "container" used to hold a collection of datasets. There are three types:

- File geodatabases—Stored as folders in a file system. Each dataset is held as a file that can scale up to 1 TB in size. The file geodatabase is recommended over personal geodatabases.
- Personal geodatabases—All datasets are stored within a Microsoft Access data file, which is limited in size to 2 GB.
- Enterprise geodatabases—Also known as multiuser geodatabases, they can be unlimited in size and numbers of users. Stored in a relational database using Oracle, Microsoft SQL Server, IBM DB2, IBM Informix, or PostgreSQL.