# GIS - Lecture notes All

Geographic Information Systems (George Washington University)

# Introduction to GIS 09/04/2014

- What is GIS?
  - Software, Hardware, Data, Methods, People
  - *GIS is a computer-based system to aid in the collection, maintenance, storage, analysis, output and distribution of spatial data and information (1)*
- Functions of GIS
  - Identify a Location, and tell you where you are
  - Identify Distributions, Relationships, and Trends
  - Integrate Data from Diverse sources
  - Combine and overlay data to solve spatial problems
  - Map and model future events
- What is GIS… Applications <Check BB>
  - *Essential tools in business, government, education, and non-profit organizations (3)*
  - *Instrumental in addressing some of our most pressing societal problems (3)*
  - *Help identify and address environmental problems by providing crucial information on where problems occur and who are affected by them (4)*
- GIS Reality
  - All maps are concerned with two elements of reality; Locations and Attributes
    - **Locations**: Positions in 2 dimensional space
    - **Attributes**: Qualities or Quantities, such as city names or population figures
- Two Types of GIS Data
  - Vector
    - Polygon, Line, Point

- *Define discrete objects (33)*
- *Use points, lines, and polygons to represent the geometry of real world entities (33)*

- Raster
  - *Identifies and represents grid cells for a given region of interest (34)*
  - *"wall-to-wall" (arrayed in a row and column pattern) coverage of a study region (34)*
  - *Cell values are used to represent the type  or quality (34)*
  - *Used commonly with variables that may change continuously across a region like elevation, slope, average rain fall, mean temperature, etc. (34)*
  - *Also commonly used to represent discrete features*

GIS Components

- A Map is made up of Layers
- Layers can be vector or Raster
- Vector layers contain features
- Features can take the form of points, lines, and polygons, and know collectively as Vector Data
- Layers contain features, and each feature is linked to a row of information in the Attribute table

# Cartographic Design 1 09/04/2014

- **Cartography**: the making and study of maps in all their aspects
- **Maps**: the visual representation of a spatial relationship
- **Spatial**: pertaining to the distribution, distance, direction, area, and other aspects of space on the earth surface

**Basic Map Types**

- General reference: include various geographic features with no specific emphasis on any feature(s) or attribute(s)
  - Not publication quality
  - Nothing really stands out
- Thematic: illustrates specific theme(s)
  - Generally one theme, but there can be multiple
  - Tells a story. How it changed over time
  - Many forms and ways to create a thematic map.
- Special purpose: Developed for very specific purpose
  - Generally one theme

**Principles of Map Design**

- To get your point across in a reasonable manner. Make it look as pleasing as possible. Communicating data is the more important than the design
  - Purpose
    - what should be included in the final output
    - nice, clean communication in the map
  - Geographic Space/Expanse
    - Country, area
    - Make optimal use of the layout of the area

- Available Data
    - Data doesn’t exist for everything. There may be limited data. Work with what you have.
    - It will affect quality
- Map Scale
    - Scale of data will affect the level of detail
    - Screen resolution, quality is affected by map scale
- Audience
    - Choose presentation method to suit your audience
    - A good map should be easy to read and decipher
- Conditions of Use
    - Where the map will be used
    - Know the size, color before you know making a map
        - You can’t just zoom into the map (ex. Magazine)
        - Always make black and white map black and white. Don’t just change it to gray scale
    - Depending on where the map is used, the level of resolution is different.
- Technical Limits

**Map Composition**

- Map Body: Main window that contains the map information
    - Geographical reference-base information
    - Context: floating or cropped
    - Positioning: Right, left, centered

    - Zoom: Give the study area room to breathe.  in closely to your study area
        - Zoom in closely to your study area
        - Make sure it fits well in the space. Don’t waster map space and make sure it is centered.
- Inset/overview Map
    - Ex.: US- only way to show Alaska, Hawaii, etc.
        - Hawaii, Alaska scale is different. Scale bars are important
    - When would you use an Inset map?
        1. To show a primary map area in relation to a larger more recognizable area
        2. To enlarge important or congested areas
            - Zoom into the congested area
        3. To show alternate thematic topics that are related to the maps theme, or different dates of the same theme
- Title
    - All maps need a *title*
    - Short, snappy- but not overly abbreviated
    - Do not use the word “Map” in title
    - Use subtitles for complicated themes
        - Adjust size and length of the title
    - Can use a bounding box if the background is busy
    - Should e <u>largest</u> text on the map
    - Try to limit to one line

- **Typographic Guidelines**

- Avoid decorative fonts
- Consistency in font type (vary size, not style)
- Legibility
- Pronounced labels for more important features
- Do not passively accept the default label settings
- Spell-check

- Legend
    - Thematic maps <u>NEED</u> legends
        - Need it to communicate. Need explicit indication, so the audience knows
    - Not ALL symbols need to feature in the legend, only those central to the main theme
    - Symbol sizing should be <u>consistent</u> (the Gold Ratio ~1:1.618 is often used for polygon symbols)
- Scale
    - Verbal: 1 inch represents 1 mile (not as common)
    - Representative Fractions: the Ratio of map distance to earth distance, and indicates the extent to which a geographic region has been reduced from its actual size
        - 1:63,360
        - 1:100,000
    - Scale Bar: resembles a ruler, that can be easily used to measure distances on a map. It's ability to indicate distance, as well as it's ability to withstand enlargement and reduction, make it preferable for use on a thematic map. *Defined as the ratio of the distance on the map to corresponding distance on the ground (132)*
- Orientation/direction indicator
    - North Arrow
        - Appropriate size

    - Grid or Graticule (132)
        - Make sure to label axis
        - *Graticule represents constant latitude and longitude, set of coordinate lines, and lines may appear curved.*
        - *Grid are lines of constant coordinates, and appear straight on most maps. Lines are drawn in both x and y directions*
        - *They are useful because they provide a reference against which location may be quickly estimated.*
    - Only 1 needed for reference
- Map metadata
    - Author: who or what agency created the map
    - Date: the date the map was created
    - Source: cite any source data used in the map ie. ESRI, NASA, FEMA, etc.
    - Coordinate System: state the coordinate system used for the main map. (ex. UTM Zone 18N, NAD83)

**Visual Hierarchy & Balance**

- **Visual Hierarchy**: refers to the order of the graphical representation of your map information
    - Transparency, hash marks, circle vs. circle with a black outline, etc.
    - Mixture of organization, restacking.
- **Balance**: refers to the organization of map elements and the empty space, resulting in visual harmony and equilibrium

**Contrast and the Use of Color**

- Border- Enough contrast with the study area, but not too much.
- Art, but conservative art
- Don’t use pink and blue for land mass.
- Light hue of blue for water

- **Thematic Mapping**

- A map with a theme or a story
- A thematic map shows the spatial distribution of one or more specific data themes for standard geographic areas
    - Counties, zip codes, countries, census tracks
- Thematic maps can portray physical, social, political, cultural, economic, sociological, agricultural, or any other aspects of a city, state, region, nation, or continent
- The maps are different, but they show data aggregated to some geographic unit of analysis and the information is represented as a color unit that conveys the magnitude or type of variable involved.

**Spatial Distribution of Data**

- Simple set of models for geographic phenomenon
- This is the arrangement (spread, pattern), of thematic phenomena in geographic space
- Geographic Phenomena can be arranged along the following lines:
    - **Discrete**
        - Presumed to occur at distinct locations and space with empty space in between
            - Address, location of a hospital, school building
    - **Continuous**
        - Occur throughout a geographic area of interest
            - temperature, sales tax, pollen level
    - Both can be described as abrupt or smooth
        - **Abrupt**: change suddenly
        - **Smooth**: change gradually
        - Pattern followed by natural phenomenon
- Concept vs. the "Data"

- The way you record your data can have an effect on how it is spatially represented on a map

**Levels of Data Measurement**

- Each of these levels build upon features of the previous level
- Qualitative Data
    1. **Nominal** (Categorization)
        - Categorized into groups, names groups, but no numerical groups are associated with the data.
        - These features differ from each other, but not in a quantifiable manner
            - Ex: Religions, restaurant names, different types of schools and hospitals
- Quantitative Data
    2. **Ordinal*** (Categorization and Ordering)
        - Ex: Ranking cities
    3. **Interval** (Ordering & Explicit Values, arbitrary zero)
        - Has order and equal intervals
        - Counts of intervals (income, years of education, number of votes
        - You can add and subtract, but can’t multiply or divide
        - Ex: Measure of temperature
            - Zero degrees doesn’t mean there is an absence of temperature. zero is arbitrary
    4. **Ratio** (Ordering & Explicit Values, non-arbitrary zero)
        - Interval data, which has a non-arbitrary zero
        - Add, subtract, multiply, and divide
        - Ex: Income ratio

            - zero means nothing

- **<u>Thematic Mapping Techniques</u>**
- **Choropleth**
  - Used to portray data collected for Enumeration Units (Counties, states, or countries)
  - Software groups that data classes
  - Applies color data to your data
  - *Depicts quantitative information for areas (134)*
  - Suited to:
    - Abrupt data
    - Calculation of typical values
      - Ex: Median income for a state or county. It forces data to be abrupt
  - Disadvantages:
    - Doesn't show variation WITHIN mapping unit
      - Ex: Highs and lows in income- if you map the entire city you're not going to see through the highs and lows of income
    - Based on arbitrary boundaries
      - Doesn't' coincide with the patterns in the data
      - Ex: Diseases are unlikely related to your zip code
  - Considerations:
    - Standardization of Data
      - Raw data totals are adjusted for the enumeration units we are working with
    - Unevenly sized Enumeration Units

- **Cartogram**
  - Sacrifice true area and other thematic variable
    - Ex: County is resized based on its voter population. Visual effect of deemphasizing states with a lower population over states that contribute to the overall election.
- **Color Progression**
  - Should be sued to depict the data properly
  - **Single hue progression**
    - Fade from dark shade of a chose color to a light or white shade of the same hue
      - Darker- greatest number of the data set
      - Lighter- least number of the data set
    - Most common method to map magnitude
    - Two variables may be shown to the use of two overprinted single color schemes
      - Hues are typically red to white for the first data set and blue to white for the second data set. They are overprinted to produce various hues
  - **Bi-polar color progression**
    - Two opposite hues to show a change in value  from negative to positive on either end of central tendency, such as the mean of the variable being mapped.
    - Ex: Temperature
      - Dark blue for cold and dark red for hot
  - **Blended hue color progression**
    - Related hues to blend together two endpoint hues
    - Ex: elevation changes
  - **Partial spectral color progression**
    - Used to match two distinct set of data

    - Blend two adjacent opponent hues and show the magnitudes of the missing data classes
- **Full-spectral color progression**
    - Contains hues from Blue through read
    - Common on weather maps, relief maps. NOT recommend for other maps
- **Value progression**
    - monochromatic
    - Any color may be used. The archetype is Black to white with intervening shades of grade that represents magnitude

**Proportional Symbol**

- Scaling symbols in proportion to the magnitude of the data around a central point. Can be used for actual (city) or conceptual points (center of an enumeration unit, such as counties or states)
- Suited to:
    - Raw Data Totals
- Disadvantages:
    - Can become crowded on maps with small enumeration units

**Isarithmic (Isopleth Maps)**

- Isarithmic map (contour map) is created by interpolating a set of isolines between sample points of known values (ex: contour map)
    - Height or temperature contours
- Isopleth Map is a special kind of isarithmic map in which the sample points are associated with enumeration units
    - “display lines of equal value and they are used to represent continuous surfaces”
    - “Lines typically do not cross”
    - “Isopleths are typically estimated surfaces”
- Suited to:

    - Smooth Data Totals
        - Gradual change from place to place
- Considerations
    - Standardization of Data
    - Finer level of enumeration units more suitable

**Dot Mapping**

- One dot is set to be equal to a certain amount of phenomena, and ideally the dots are placed where that phenomena are most likely to occur
- “show quantitative data”
- Suited to:
    - Raw Data Totals
- Disadvantages:
    - If you do not have access to ancillary information, such as satellite imagery, it is hard to have confidence in dot placement
- Good example of Dot Mapping- Race and Ethnicity

Data Classification

**1. Equal intervals**

- Divides the range of attribute values into equal-sized sub ranges
- Advantages
    - Easy calculation
    - Easy interpretation
    - No gaps
- Disadvantages

        - Does not consider data distribution doing the number line

2. **Quantiles**
    - Equal number of observations are placed in each class
    - Advantages
        - Easy (manual) Calculation
        - Allows use of complete color spectrum
    - Disadvantages
        - Identical data values MAY be placed in different classes
        - Fails to consider how dada is distribution along the number line
3. **Natural Breaks**
    - Classes are based on natural groupings inherent in the data through an examination of the histogram
    - Advantages
        - Minimized the difference between data values in the same class, and maximize the differences between classes
    - Disadvantages
        - Data ranges are usually uneven
4. **Mean-Standard Deviation**
    - The classification scheme shows you how much a feature's attribute value varies from the mean
    - Class breaks are created using the values and two color ramps helps emphasize the values shown
    - Advantages
        - Considers how data is distributed along the number line
        - If data are normally distributed, then the mean is a natural dividing point

- Disadvantages
    - ONLY works well for data that are normally distributed

5. Manual
6. Optimal
    - More mathematically perfected version of natural breaks to make sure that the similar values are placed in the same class
7. **Geometrical Interval**
    - Classification scheme where the class breaks are based on class intervals that have a geometrical series
    - An algorithm creates these geometrical intervals by minimizing the square sum of element per class
    - Ensures that each class range has approximately the same number of values with each class and that the change between intervals is fairly consistent
    - Designed to accommodate continuous data
    - Produces a result that is visually appealing and cartographically comprehensive
    - It minimizes variants within classes and work reasonably well on data that is not normally distributed
    - Also referred to as smart quantiles

Databases and Database Management 09/04/2014

- **Database**
  - Structured tables
  - A Set of structured or related data
    - Two Types of Data to be managed in GIS
      - **Spatial**: Where things are and how we capture that
      - **Tabular Data**: What things are
      - **Conventional Databases:** Only non-spatial entities
- **Database Management system**
  - A DBMS is a software application designed to organize the efficient and effective storage, retrieval, indexing and reporting of data.
  - Ex: Data model, data load capability, indexes (all databases have them), query language, security (controlled access to the data, different types of data entry), controlled updates, back-up recovery, administrative tools, applications, application programming interfaces
  - Manages interactions between clients and raw data
  - Manages different versions of data
  - Provides secure access
- Database Data Models
  - **The Relational Database Model** (properties)
    - Each COLUMN has a unique name
    - Column entries MUST be drawn from same domain
      - All have to string, integer, float, or date
      - Everything has to be organized
    - Columns can be in ANY order
    - Only ONE entry per cell

        - No merging
    - Each ROW must be distinctive
    - NULL values are allowed
        - Some data are missing. Zero and Null are different. Null means something bad happed: damaged, not working, but zero means nothing.
- Terminology
    - Table: database
    - Row: row or record
    - Column: field or attribute field
    - Unique ID: Primary Key or Index

 **Querying:** facilitated by a simple system so you can interrogate the database to find information. Finding information in a logical way without having to search manually.

    - Relational databases have simple query systems
    - Queries expressions are based on relational algebra
    - SQL (structured Query Language)
- **Queries**
    - Primary method of data retrieval
        - Highlights information on the table and map and give you the option to sub select the data rolls that satisfy that query.
    - Create new information, but doesn’t change the older/existing information
    - Types of Queries
        - **Aspatial**: Non-spatial. Traditional interrogation of the database.
            - Set algebra (<,>, =, <>, …)
            - **Boolean** Queries  (OR, AND, NOT)

        - Can also be used for true and false
    - **Spatial**: Unique. Ability to use geometry or where something is located in order to identity or grab data.
  - **Structured Query Language (SQL)**
    - Structure Query Language is a standard computer language for accessing and managing databases
    - SQL was initially developed by the IBM Corp, but is supported by a number of software vendors.
  - Selection by Attributes
    - The AND conditions typically decreases the number of records and add restrictive criteria
    - The OR condition typically increases the number of records and is considered as an inclusive condition
    - The NOT is the negation operation and can be interpreted as meaning "select records that do not meet the conditions following the NOT"
    - <Venn Diagram in BB Slideshow>
    - AND's, OR's and NOT's can have complex effects when used in combined conditions.
    - The order is important in the query
- **Spatial Queries**: A GIS lets you select spatial features based on their location relative to other spatial features.
  - In ArcMap, a spatial query can be done using **Select by Location** dialog box
  - You can use a variety of selection methods to select the point, line, or polygon features in one layer that are near or overlap the features in the same or another layer
  - Two concepts
    - **Adjacency**: operations used to identify
      - When features touch another feature
      - Shared line or node required
    - **Containment**:
      - Something that has to be completely inside

**Joins**: allow you to take extra information from the outside and bring that information to GIS and impend it. It. allows you to take the information and make it spatial.

- Linking Spatial Data
    - **Primary Key**
        - Unique identifier for EACH row of information a particular data file. (Like a SSN)
    - **Foreign Keys**
        - Non-unique identifier that carries information that may be linked to primary key.
- Relational databases
    - For table join you don’t have to constantly create new tables
    - Unique keys can….

**ArcCatalog**: an application that allows you to Explore, Access, Manage, and Build geographic data. Allows you to browse GIS data. Every layer is not one file, it takes about 7-9 files.

- Seamless view of geographic data, similar to window’s explorer
- Icons communicate the role of individual GIS elements
- No Microsoft clutter. Only see GIS data

- Functions of ArcCatalog
    - Create and format new data
    - Search for data
    - Determine geographic extent
    - Determine data quality
    - Launch GIS operations

**Geodatabases**

- Table, feature class, Raster dataset

- Contain:
    - Feature Datasets
    - Relationship Classes
        - Define how rows in one table can be associated with rows in another table
    - Domains
        - List or range of valid values for attribute columns
    - Spatial relationships and rules
        - Topologies, along with network datasets, address locators, terrains, cartographic representations
    - Map Layers
    - Base Maps
- Types
    - ArcSDE
        - Large scale IT level database
        - Flexible
        - Orcale/SQL/PostgreS,
        - Multi-user geodatabase
    - File
        - Collection of GIS datasets
        - Flexible
        - Each dataset stored separately
        - 1Tb

        - Cross-platform. Not exclusive to window
    - Personal
        - 1st iteration. Original geodatabase type
        - one user, two use type things. Used by small group of people.
        - Microsoft Access
        - 2 GB
        - Windows only
- Geodatabase Layout
    - <On BB Slideshow>
        - Feature dataset is a folder
- Raster Data Organization
    - **Raster datasets** manage very large, continuous image datasets
    - **Mosaic datasets** allow you to store, manage, view, and query collections of raster image data
    - **Raster catalogs** manage well defined grids of raster data
    - **Raster attribute** columns in tables store pictures or scanned documents as attributes in tables
- Creating Geodatabases
    - You have to dig through it before you can add to it and you have to work on ArcCatalog
- Extended capailties of geodatabases
    - Attribute domains: specify a list of valid values or a range of valid values for attribute columns. Use domains to help ensure the integrity of attribute values> domains are often used to enforce data classifications
    - Relationship classes: build relationships between 2 tables using a common key. Find the related rows in a second table based on rows selected in the original
    - **Subtypes**: manage a set of attribute subclasses in a single table. This is often used on feature class tables to manage different behaviors on subsets of the same feature types

    - **Versioning**: it allows multiple users to edit the same data in an ArcSDE geodatabase
- History of Geospatial Data Types
    - Color code databases
- What is Metadata?
    - Documentation of the content, quality, conditions of the data
        - Who made it? Who distributes it?
        - What is the subject, processing?
        - When and where was it collected?
        - Why and how was it collected?
        - How much does it cost?
        - Who much does it cost?
        - How is it referenced to the real world?
        - What's the quality of the data?
        - Who should I contact if I have questions?
        -

# Open Sources GIS

**Open Source software**: computer software for which the human-readable source code is made available under a copyright license (or arrangement such as the public domain) that meets the Open Source Definition

- This permits users to use, change, and improve the software, and to redistribute it in modified or unmodified form. It is very often developed in a public, collaborative manner.

**Open Source Conditions** (10 conditions of the open source definitions)

- **Free Redistribution**: the software can be freely given away or sold. (This was intended to expand sharing and use of the software on a legal basis.)
    - It does not have to be free, but for free distribution. Difference between free distribution and free
- **Source code**: the source code must either be included or freely obtainable. (Without source code, making changes or modifications can be impossible.)
- **Derived Works**: redistribution of modifications must be allowed. (To allow legal sharing and to permit new features or repairs.)
    - Similar to free redistribution
- **Integrity of the Author's Source Code**: licenses may require that modifications are redistributed only as patches.
    - Building block of what the source is. It cannot be changed.
- **No discrimination Against Persons or Groups**: no one can be locked out.
    - No discrimination. Everybody has access
- **No discrimination Against Fields of Endeavor**: commercial users cannot be excluded.
    - No discrimination. Businesses can use and they are not precluded to make more money. Its anybody's.
- **Distribution of License**: the rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.
    - Legal stuff
- **License Must Not Be Specific to a Product**: the program cannot be licensed only as part of a larger distribution
    - Must be specific

    - **License Must Not Restrict Other Software**: the license cannot insist that any other software it is distributed with must also be open source.
        - Not restricted to other software. Can't tie it or bind to other things that will constrict you.
    - **License Must Be Technology-Neutral**: no click-wrap licenses or other medium-specific ways of accepting the license must be required.
        - Must be technology neutral. Work on many platforms.
- Puts you in a stronger position
- **Open Source Geospatial Foundation**
    - Not-for-profit organization
    - Mission is to support and promote the collaborative development of open geospatial technologies and data
    - Provide financial support

**Participation, interaction, and immersion**

- **Public Participation in GIS (PPGIS)**: having the community involved.
    - **VGIS (Volunteered Geographic Information)**: The harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals
        - Goochild, 2007
    - **Crowdsourcing**: Outsourcing a task to the community at large, via an open call.
        - Crowd- general public
        - Sourcing information from the public
        - Data capture
        - Technology design

**Featured Platforms**

- **OSM** (Open Street Map): Free, open editable map of the world
    - Gathered by GPS units or hand drawn by people from all over the world
    - Superficial use of that data. Map behind a glass window. You can't touch or extract the roads for analysis.
    - "OSM is a project aimed squarely at creating an providing free geographic data such as street maps to anyone who wants them"
    - A lot of political implications of drawing maps

- **Ushahidi** (means testimony)
  - It was built for information collection, visualization and interactive mapping.
    - Have the information and collection sit on top of a city map
  - Website that initially developed to map reports of post-election violence in Kenya
- **Quantum GIS**
  - Official project of OSGeo
    - Current version: Chugiak 2.4
  - Runs on Linux, Unix, Mac OSX, and Windows
  - Support Vector and Raster
    - QGIS Desktop, QGIS Browser, QGIS Server
- **TileMill**
  - Web base
- CartoDB
  - For people who are afraid of
  - Much easier to work with and user friendly

**Where/How do we collect data?**

- **GPS**
  - Systematic ground surveys using a handheld GPS unit and a notebook, digital camera
- **Local Knowledge**:
  - Augmenting already present vector data
  - Using "Walking Papers" to augment/validate traced work
  - You can add knowledge
- **Imagery**:
  - Trace features (no importing of names due to copyright)
  - Sources- purchased or donated imagery, bing
  -

# Data Creation 1

## Two Main Methods of Data collection

- **Primary data collection**: direct measurement
  - Raster (raw data capture)
    - Digital satellite: remote sensing images
    - Digital Aerial: aerial photographs
  - Vector
    - GPS: measurements
      - Capturing
    - Survey: measurements
      - Most detailed way to capture vector data
    - Geocoding: addresses
      - Between primary and secondary
- **Secondary data collection**: derivation from other sources; data derived from a product; sources are those reused from earlier studies, or obtained from other systems
  - Raster
    - Scanned maps or photographs
    - Digital elevation models from topographic map contours
      - Can be primary model
  - Vector
    - Existing Vector data- Coverages, Autocad
    - Manual digitizing from scanned maps

**Raster Data Capture (Primary)**

- **Remote Sensing**: the measurement of physical, chemical, and biological properties of objects without direct contact
  - Satellite imagery
  - Aerial photography
- Getting Remotely Sensed data into GIS
  - Considerations- season differentials, sometimes just bad quality
    - Registration/Georeferencing
      - Control points in image and on ground or from GIS layer
      - Georeferencing- Take the lost/old image and line it up with a new/modern landscape
  - Generating GIS layers with RS
    - Classification results
      - Use the computer to dig up images that are similar.
    - "heads-up digitizing"
      - ex: Open Source Mapping
  - Updating GIS layers with RS
    - Currency
      - Provides coverage over large spatial areas quickly
    - Editing
      - Use old maps and update that with new remotely sensed data

**Raster Data Capture (secondary)**

- **Scanning**: Using a scanning device to convert hardcopy analog media into digital images
  - **Georeference**- take a broken image and fix it

- Geographic wallpaper
- Template for vectorization
- Why scan?
    - Reduce wear and tear, improve access, provide DB storage
    - Provide geographical context
    - Scan prior to vectorization
        - Creates a raster data set that can be vectorized (automated and manual)

**Vector Data Capture (Primary)**

- **Ground Surveying**: determined by measuring angles and distances from known points
    - GPS (Global Positioning System)
        - GPS System (USA)
        - GLONASS (Russia- 2009)
        - Galileo (Europe- currently testing, fully operational 2019)
        - COMPASS (China)
        - IRNSS (India)
- **Surveying**: the art and science of measuring the surface of the earth and its features.
    - **Geodetic surveys**: take into account the true shape of the earth
    - **Plane surveys**: treat the earth as a flat surface
    - **Horizontal surveys**: determine the position of features on the ground
    - **Vertical surveys**: determine the elevation of heights of features
- **GPS**: a constellation of Global Positioning System satellites orbiting the earth is used to determine the position(s) of ground receivers

- Post processing GPS Data
    - Projections
    - Differential Correction
- GPS in ArcMap
- DNR Garmin
    - Developed by the Minnesota Department of Natural Resources
- How does it work?
    - Trilateration
        - Method of determining relative position using triangles
        - Have to have at least 3
    - For height reading you have to have at least 4

- **Geocoding**: the process of finding associated geographic coordinates (often expressed as latitude and longitude) from other geographic data, such as street addresses, or zip codes (postal codes).

**Vector Data Capture (Secondary)**

- Manual digitizing
- Heads up digitizing
    - From a map on your screen
        - i.e. scanned in amp or aerial photograph
    - very often the way to go now

- Vector Data Capture Errors
    - Post-vectorization editing is required to clean up errors
        - Importance of clean input data

        - Build topology process: editing and cleaning
        - Fix dangling vertices, snapping to nodes, etc.
        - Set tolerances
    - Interactive vectorization- snap to features
    - Undershoots and overshoots, invalid polygons , silver polygon
- **Topology**: the mathematics and science of geometrical relationships used to validate the geometry of vector entities
    - Typical Topological relationships
        - Connectivity( lines)
        - Directionality (Lines)
        - Adjacency (polygon)
            - Share a wall, no space between them; notion of overlaps
        - Exhaustive (polygon)
        - Planar topology (no overlaps)
        - Non-planar  (overlaps allowed)
        - Dangles
    - Intra-layer relationships
        - Overlap
        - connectivity
- **Topological relationships**: the properties of geographic objects that do not change when forms are bent, stretched, or undergo similar transformations
- Classification/Reclassification (Raster)
    - Raster reclassification assigns output values that depend on the specific set of input values

    - Based on matching input cell values to a reclassification table
    - Usually a single output value applies to a range of input values
- **Vectorization**: Converting raster to vector data
    - Batch vectorization (automated process)
        - Simple (spaghetti) lines are built from pixels

**Importance of Data quality**

- Data quality can be defined as "fitness for use" of data for an intended purpose
- One advantage of GIS is to integrate diverse data sets. However, data may be used in ways not foreseen by their producers and by users without the knowledge or experiences to judge whether an application is appropriate
- Potential pitfall of GIS programs is their ability to hide underlying data quality issue
- Understanding data quality is important because without it we do not know what confidence to put in the results of our analysis

- **Error**: Data quality refers to the relative accuracy and precision of a particular set of data
    - Error encompasses both the imprecision of data and its inaccuracies
    - Results are only as good as the input data
        - "garbage in-garbage out"
- Describing error
    - **Accuracy**: is the degrees to which information on a map or in a digital data base matches true or accepted value
    - **Precision**: refers to the level of measurement and exactness of description in a GIS database. Consistency.
    - **Bias**: is consistent error throughout a dataset
- When is Error introduced?
    - Data collection phase
        - GPS instrument, human, satellite, aerial errors

    - Data input and editing phase
        - Errors in digitizing areas (How you trace it)
    - Methodological phase
- Broad Types of Error
    - Spatial errors
    - Attribute errors
    - Conceptual error
    - Procedural/Analytic errors (Logical Error)
- MAUP
    - Modifiable Areal Unit Problem
        - Problem associated with aggregate data sources
        - Use of arbitrary spatial units
            - Boundaries are arbitrary and don’t capture people or traits of people in any sort of comprehensive way
        - Problem of Scale
    - Ecological Fallacy: Errors due to performing analyses on aggregate data when trying to reach conclusions on the individual units
- Data Quality Conclusions
    - GIS products are models of reality. There is always a degree of error
    - Error in a coverage must always be addressed
    - Inherent error may e incorporated into how objects are actually represented
    - Data in a map is only as good as the least accurate data source
    - If all GIS products are flawed how do we deal with this?
        - Document it through Metadata

# Spatial Analysis

- **Spatial analysis**: the process by which we turn raw geographic data into useful information; it includes all the manipulations and methods that can be applied to geographic data, to add value, support decisions, and reveal patterns and anomalies not immediately oblivious from displaying that data.

- Effective spatial analysis requires an intelligent user, not just a powerful computer.
- Spatial analysis helps us in situations where our eyes might otherwise deceive us.

**Applications of GIS**

- Location
  - Finding things and where they are to other things
- Quantity
  - How many schools are in the city? Etc.
- Patterns
  - It is hard to find trends, relationships, etc. from exclusively tabular data
- Trends (temporal aspect)
- Conditions and relationships
  - Other spatial elements that surround the feature that affect them
  - What are stuff is happening that is contributing to the relationship
- Implications (planning for the future)
  - Another way of saying modeling (Simplified version/ representation of reality)
  - Kind of assuming the future

 **Principles of Spatial Analysis**

**Basic Principles**

 **1. Measurements**

- **Measurements**: simple numerical values that describe aspects of, or simple properties of geographic data, such as length, area, shape, OR, relationships between pairs of objects, like distance or direction.

- **Distance**: different calculation methods based on data format
    - Vector
        - Distance
            - Euclidean distance (this cell to this cell)
    - Raster
        - Distance
            - Euclidean distance
            - Manhattan distance (city blocks)
            - Proximity

 **2. Scope**: the geographical extent or area of the input data used to determine the values at output locations

- Characterization of Scope
    - Local
    - Neighborhood
    - Global (everything in our data, not the globe)

 **3. Queries**: primary method of data retrieval. They create new information, but doesn’t change the older/exist information

- Type of queries: spatial, spatial

 **4. Proximity functions**

- **Buffers**: the creation of zone(s) of interest (inclusion or exclusion) around an entity.
    - Useful for selection processes and visualization
    - Capture stuff in the buffer zone or remove stuff in the buffer
- Spatial join

- Step sister of “select by location”
- Highlighting stuff that is close to other stuff and impending to the table of the related layer.

- **Thiessen Polygons**: This method assumes the values of un-sampled area are equal to the value of the closest sampled point

 **5. Geoprocessing**: involves using data from one or more spatial layers and integrating it to form a new, single layer

- Spatial Data Integration is also known as geoprocessing

- **Polygon layer (?)**: This process creates a new layer by using a polygon layer (or selected polygons form a particular layer) as a cookie cutter on a point, line, or polygon shapefile. The output layer contains information from Layer A only. Layer B is only used to define the new boundary (Layer B is just a mask).
- **Intersect**: preserves only those features falling within the spatial extent common to both layers. The features of the input layer are intersected or sliced by the intersect layer. The attribute data from both layers are included in the new layer’s attribute table.
- **Dissolve**: Dissolving features in a layer coalesces features that have the same attribute value. This tool is extremely important if you are trying to create a new shapefile, a file with a coarser layer of geography than your starting files.
- **Union**: Union creates a new layer by combining 2 polygon layers. The new layer has data and shapes from both layers, including their intersection.
- **Merge**: It is similar to union; a new layer is created from multiple layers but their features are not intersected. Merge allows you to combine the features from 2 or more layers of the same geometric type.

**Advanced Principles (select)**

 **1. Neighborhood Functions: Moving Window**

- 1. The window is a configuration of cells used to specify the input values for the operation
- 2. The function visits each cell in the faster and calculates the specified statistic within the identified neighborhood
- 3. The result is usually associated with a cell at the center of the windows position

    - 4. The window then moves to the adjacent cell, and re-performs the process
    - Shapes: rectangle, circle, annulus, wedge

**2. Spatial Estimation**

- **Spatial Interpolation**: The procedure of estimating the values of properties at un-sampled sites within an area covered by existing observations
    - Spatial data usually are:
        - Stratified
        - Random- More unbiased
        - Patchy- Collect information from specific clusters of areas
        - Adaptive- common in physical sampling in the field.
- **Triangulated Irregular Network (TIN):** This method is used to construct Digital Elevation models. Adjacent data points are connected by lines to form a network of irregular triangles.

# Coordinate Systems 1

- **Coordinate system**: is a reference system used to represent the locations of geographic features, imagery, and observations such as GPS locations within a common geographic framework
- Each coordinate system is defined by:
  - Its measurements framework, either geographic (3-D) or projected (planimetric) (2-D)
  - Unit of measurement
  - The definition of the map projection (for projected coordinate systems)
  - Other measurement system properties such as a spheroid (ellipsoid) of reference; a datum; and projection parameters like one or more standard parallels, a central meridian
- **Three types of coordinate systems**
- 1. **Geographic coordinate systems**: global or spherical coordinate system such as latitude-longitude, typically expressed as (Degree Minute Seconds 9DD:MM:SS) or Decimal Degrees (DD)
  - 3-D systems
- 2. **Projected coordinate systems**: coordinate system that provides various methods to project the earths' spherical surface onto a 2-D artesian coordinate plane. Typically using feet or meters
- 3. **Grid Systems**: a grid placed over a map projection using a plane or Cartesian (x,y) coordinate system to locate features
- **Geographic Coordinate System or Graticule**
- **Latitude**: an angular distance, North or South of the equator measured from the center of the earth
- **Longitude**: is an angular distance, East or West of a point on the Earth's surface, measured from the center of the earth (prime meridian)
- Earth's dimensions
  - **Modeling**: simplified representation of reality
- **Three Approximations of Earth**
- Sphere
  - Small scale maps, countries, continents
- Ellipsoid- flattening
  - Large scale maps of smaller areas

- Geoid-topographic; map of gravitation anomalies
  - Reference surface for ground surveying of horizontal and vertical positions

- **Datum**
  - where is the middle of my sphere? The center.
- Every geographic coordinate system includes an angular unit of measure, a prime meridian, and a datum
- A datum defines the position of the spheroid relative to the center of the earth

**Map Projection Techniques**

- **Step 1**
  - Reduce the Earth's size to that of an imaginary glove
  - **Reference Globe**: a model of the earth at a reduced scale, that is used to project the landmasses and graticule onto a flat map

**Step 2**

  - Project the graticule from the reference globe onto the developable surface
    - Cylinder, cone, plane method
  - **Developable Surface**: a mathematically definable surface onto which the land masses and graticule are projected from the reference globe

- **Map Projection Characteristics**

**Class**: refers to the overall appearance of the graticule, once the projection process is complete

- Cylindrical Class
  - Characteristics
    - Lines of longitude are straight, equally spaced
    - Lines of latitude are straight, parallel and intersect lines of long at right angles
  - Distinguishing features

        - The spacing of the parallels distinguishes one type of cylindrical projection from another
- Conic Class
    - Characteristics
        - Lines of longitude are straight lines of equal length, radiating from a central point (poles)
        - Lines of latitude are concentric circular arcs centered around one of the poles
    - Distinguishing features
        - "Pie-Wedge" shape
        - The angular extent of the wedge, and the spacing of the parallels distinguish one conic projection from another
- Planar Class
    - Characteristics
        - Lines of longitude are straight, equally spaced that radiate from the center
        - Lines of latitude appear as concentric circles, centered about a point
    - Distinguishing features
        - Again, the spacing of the parallels distinguishes one type of planer projection from another

**Case**: The case of a projection relates to how the developable surface is positioned with respect to the reference globe

- It can be described as Tangent of Secant
    - Secant projection is better because you are capturing stuff more honestly and there is less overall distortion.
    - Tangent is less difficult to calculate

**Aspect**: the aspect of a projection deals with the placement of then projections center with respect to the earth's surface

- A projection can have one of three aspects:
    - 1. Equatorial

- 2. Polar
- 3. Oblique

 **Grid Systems**: a grid is placed over a map projection using a plane or Cartesian (x,y) coordinate system to locate features
  - Created for larger scale mapping
  - Divided into zones with only positive numbers (meters or feet)
  - Easier to calculate area, direction and distance
- 2 Grid System:
  - 1. Universal Transvers Mercator (UTM)
  - 2. State Plane Coordinate System (SPCS)

 USA UTM Zones
  - Projections: different geographic languages. Its not completely right or wrong

 State Plane Coordinate System
  - Specific for the US

 SPCS Zones Close Up

 **Distortion**: altering the size or shape of the earth’s landmasses and graticule for projection to a flat or planar surface
- How do we identify it?
  - Visual comparison
- How do we analyze it?
  - **Scale Factor**: numerical assessment of how the map scale at specific map location compares to the map scale at the standard port, or along a standard line
    - Scale Factor= Local (Map) Scale/Principle Scale
    - Local Scale: the scale computed at a specific location

    - Principle scale: the scale computed along the Standard Line or Point (true scale)
- **Scale**: the ratio of map units to earth units, with the map units standardized to 1
    - 1: 100,000
    - Scale is unit-less
- Tissot's Indicatrix
    - Visual aid to distortion
    - World map will always have the most distortion

**Projection Properties**

- There are Four Spatial Relationships that can be preserved or distorted by a particular map projection:

| Relationship | Projection |
|---|---|
| • Area | (Equivalent, or Equal Area) |
| • Angle (shape) | (Conformal) |
| • Distance | (Equidistant) |
| • Direction | (Azimuthal Projections) |

- **Equivalent Projections** (or, Equal Area)
    - Preserve Area
    - Area and Shape are usually preserved the most
- **Conformal Projections**
    - Preservation of Angular Relationships
        - Often used for the map of the US
- **Equidistant Projections**

- Preservation of Distance relationships.
- Accurate distance from the center of the projections or along the given lines. Tends to look more obscure.

- **Azimuthal Projections**
  - Preservation of direction

 **Choosing the Correct Map Projections**

1. Which spatial properties do you want to preserve?
   - Depends on what you are communicating

 2. Where is the area you're mapping? Is your data in a polar region? An equatorial region?

 3. What shape is the area you're mapping? Is it square? Is it wider in the east-west direction?

 4. How big is the area you're mapping?
   - On large-scale maps, such as street maps, distortion may be negligible because your map covers only a small part of the earth's surface
   - On small-scale maps, where a small distance on the map represents a considerable distance on the earth, distortion may have a bigger impact, especially if you use your map to compare or measure shape, area, or distance.

- Person's Guidelines (Based on Latitude/Class)
  - Equatorial regions – cylindrical
  - Mid Latitude-conic
  - Polar regions- planar
- Robinson's Guidelines (Based on Function/Properties)
  - Conformal- Analyzing, measuring, recording angular relationships. Use for navigation, piloting, surveying
  - Equivalent-Geographic comparisons across space. Use for thematic maps that represent proportions, either through color, or dot density
  - Planar- tracking the direction of movement
  - Equidistant- determination of distances

- **Model**: idealized and simplified representation of reality. Modeling is an abstracted version of reality

- A model could be a theory, a law, a hypothesis, and equation, or even a structured idea
- It is a the process of making a model
- Ex: globe, map

**Data Model**: a set of constructs for representing objects and processes in the digital environment

- A GIS data model is the foundation for managing data objects (points, lines, and polygons), inter-relationships between those objects, and associated attributes within a geodatabase
- A framework ("schema") built upon accepted standards that:
  - Models the behavior of real-world objects in a geodatabase
  - Reduces the barriers to data sharing and interoperability
  - Dramatically improves the integrity of geospatial data
  - Ultimately promotes effective GIS exploitation

**Spatial Data Model**: set of constructs for representing Geographical objects, data, processes, and relationships in the digital environment, for the purposes of analysis and complex problem solving

- Finding relationships among geographic features to understand and address any particular problem

**GIS Data Models**

- Vector
  - TIN
  - Network Model
  - Object Data Model
- Raster
- Tools for Model Building
  - Model Builder

        - ArcGIS Diagrammer

**Vector Data Model**

- **Vector**: a method of storing, representing or displaying spatial data in digital form. It consists of using coordinate pairs (x,y) to represent locations on the earth. Features can take the form of single points, lines, arcs, or closed lines
    - Point: Single Coordinate Pair
    - Lines: simple (set of coordinate pairs-nodes); Detailed (Multiple Pairs-nodes and Vertices)
    - Polygon: Set of connected line segments, with the same start/end point
- Vector Models in detail: spaghetti vector model, topological vector model
- **Topology**: the mathematics and science of geometrical relationship used to validate the geometry of vector entities
- **Topological relationships**: the properties of geographic objects that do not change when the forms are bent, stretched, or undergo similar transformations
- Vector Data Model
    - Typical topological relationships
        - Connectivity (lines)
        - Directionality (lines)
        - Adjacency (polygon)
        - Exhaustive (polygon)
        - Planar Topology (no overlaps)
        - Non-planar )overlaps allowed)
        - Dangles
            - 
    - Intra-layer Relationships
        - Overlap

        - Connectivity

**Raster Data Model**

- **Raster**: a method of storing, representing or displaying spatial data in digital form
    - It consists of using cell data (not necessarily square) arranged in a regular grid pattern in which each unit (pixel or cell) within the grid is assigned an identifying value based on its characteristics
- Cell Dimension
    - Length and width in surface units
- Level of Detail
    - Trade off between spatial detail and file size
- Spatial Precision
    - Positional accuracy is assumed to be no better than on half of the cell size
- Data Assignment
    - Point Physical Value
    - Statistical Value
    - Classification Data
    - Point, line, polygon reassignment
- Within Cell Variation

**Compare Vector to Raster**

- Vector
    - Advantages
        - Precision
        - Quality of Cartographic output

        - Storage capabilities
    - Disadvantages
        - Certain types of spatial analysis does not work on vector data
- Raster
    - Advantages
        - Ability to store and represent large amounts of information (detail)
    - Disadvantage
        - File size (and draw time) due to the large amount of data being stored
        - Lack advanced data structure characteristics (topology, network analysis)
        - Raster are just cells, so you can’t to topology

- **Advanced Models**
    - Digital Elevation Model
        - TIN (Triangulated Irregular Network)
            - Using height as a network of points
    - Network Model
    - Object Data Model

**Network Models**

- Geometric, Network datasets (transportation)

- **Object Data Modeling**
    - The key behind object data modeling is to look at a collection of geographic objects, and the relationship between those objects

- **The Data Modeling Cycle**

**Geodatabase Design is modeling**

- Conceptual
  - 1. Identify the information products that you will create and manage with your GIS
  - 2. Identify the key data themes based on your information requirements
  - 3. Specify scale ranges and spatial representations for each data theme at each scale
- Logical
  - 4. Decompose each representation into one or more geographic datasets
  - 5. Define the tabular database structure and behavior for descriptive attributes
  - 6.Define the spatial behavior, spatial relationships, and integrity rules for your datasets (Topology)
- Physical
  - 7. Propose a geodatabase design (data model or formal schema to realize the above)
  - 8. Design editing workflows and map display properties
  - 9. Assign responsibilities for building and maintaining each data layer
  - 10. Build a working prototype. Review and refine your design.
  - 11. Document your geodatabase deign

**Simplified Modeling Process for Problem Solving**

- Identify the problem
- Breakdown (simplify) the problem
- Organize the data required to solve the problem
- Develop a clear and logical flowchart using well defined operations
- Run the model and modify it if necessary

**Model Builder in ArcGIS**

**Logical Models**

- ESRI Industry Models
  - GIS for the Nation Schema
- UNSDI-T (UN Spatial Data Infrastructure- transportation)

09/04/2014

- Open Source Software
  - What it is
  - 4-5 things about it. Conditions
    - not free, but freely accessible
- Data collection
- Reater, vector
- Primary, secondary
- How does GPS work
  - Trilateration!!! How stellites figure out where yo uare
- Georeferenicng (Rubbersheeting)
- Geocoding
- Topology
- Describing error
  - Accurarcy, precision, bias
- MAUP
- Princpiples of Spaial analysis
  - Basic principles
  - Advanced principles
- Scope
- Coordinate system types
  - Three diffenrt types
    - Geogprahic, projected, grid
- Class****
- Aspect
- Case
- Grid systmes
  - UTM **
- Distortion
- Projection properties