GEOG 204

LECTURE 10

Course Review

1

Final Exam: Mon, Dec 1st Cumulative, 45 mins

Tutorials:

- Today, I will be there, but no instruction
- No tutorials next week

Project Submission: By Fri Dec 5

Housekeeping

Presentation Title 2

2

Definition

- GIS: system used to store, manipulate, analyze, manage, and present geographic data
- Importance of GIS
  - Makes it possible to efficiently integrate, manage, and analyze geographical information from maps, remote sensors, data collectors and text.

3

3

Data Collection Classification

- Data collection can be classified by source
  - Primary Sources
    - captured by direct measurement specifically for use in GI systems
    - both raster and vector data can come from primary sources
  - Secondary Sources
    - reused from earlier studies or obtained from other systems
    - raster and vector data are created from maps, photographs, and other hardcopy documents

4

4

Primary Data Collection

- Raster data
  - Data are collected by remote sensing
    - Remote sensing is the measurement of physical, chemical, and biological properties of objects without direct contact
  - Information is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects.
    - Passive sensors rely on reflected solar radiation or emitted terrestrial radiation
    - active sensors (such as synthetic aperture radar) generate their own source of electromagnetic radiation
    - Sensors are mounted on earth-orbiting satellites or other airborne platforms

5

5

Primary Data Collection

- Vector data
  - Data are captured by ground surveying, GPS and LiDAR
  - Ground surveying is based on the principle that the location of any point can be determined by measuring angles and distances from other known points.
    - It is highly accurate but time consuming and expensive
  - The GPS consists of a system of 24 satellites each orbiting the Earth every12 hours and transmitting radio pulses at precisely timed intervals
    - A receiver on the ground must make exact calculations from the signals, the known positions of the satellites, and the velocity of light in order to determine its position
    - GPS was developed by the US. Russia has GLONASS; China has BEIDU; Europe has GALILEO

6

6

Secondary Data Collection

- Raster Data Capture
  - Scanners
    - A scanner is a device that converts hardcopy media into digital images
    - Documents, such as building plans, CAD drawings, property deeds, and equipment photographs are scanned to reduce wear and tear, to improve access, to provide integrated database storage, and to index them geographically (e.g., building plans can be attached to building objects in geographic space).
    - Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for vector data layers
    - Maps, aerial photographs, and images are scanned prior to vectorization and sometimes as a prelude to spatial analysis

7

7

Secondary Data Collection

- Vector Data Capture
  - The digitization of vector objects from maps and other geographic data sources by heads-up digitizing and vectorization, photogrammetry, and COGO data entry
  - Heads-up digitizing and vectorization
    - creates vectors selectively from raster data
    - digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor.
      - **heads-up digitizing** because the map is vertical and can be viewed without bending the head down.
    - Used to collect data for land parcels, buildings, and utility assets….

8

8

Secondary Data Collection

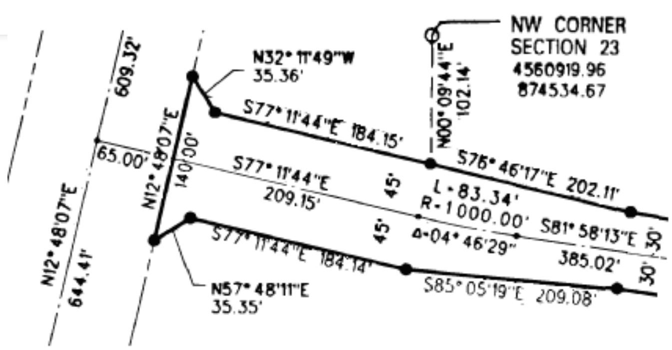
- Vector Data Capture
  - Photogrammetry
    - Measurements are taken from pictures, aerial photographs, and images
    - Measurements are captured from overlapping pairs of images using stereoplotters.
  - COGO and Other Data Entry
    - COGO is short for coordinate geometry and a method for data entry
    - Uses bearings and distances to define each part of an object
    - The COGO system is widely used in North America to represent land records and property parcels

9

9

Secondary Data Collection

- COGO descriptions for a road centerline and parcel boundaries adjoining the road



Source: ESRI

10

10

Sources of Error

- Precision:
  - If points +/- 25m on creation
    - Similarly +/- 25m error introduced on digitization
    - Conceivably 50m total error
- Accuracy:
  - Paper may have shrunk, stretched or torn
  - Symbols rearranged to prevent overlap
  - Map sheet boundaries
  - Human boredom, fatigue, humor or malice

11

Geographic phenomena

- Spatial Variation
  - Tobler's law
  - Homogeneity/Uniform distribution
  - Random distribution
  - Clustering
- Temporal(time) Variation

12

12

Representing geographic space

- The two basic forms
  - Objects and Fields
    - Vector and Raster
- The three geometry types
  - Points, lines, polygons
- The different data formats for Raster and Vector
  - e.g. shp, geotiff, GeoJSON, rst,

13

13

Location and Attributes (Fields)

Attribute data: allows us to ask the question … "what is it ?"

> Every layer has an associated table

> These are linked to spatial location by a code number

➢Attributes are stored in columns as *items*

➢Rows display the attributes for each feature = *records*

➢Entries may be text strings, integers, float (decimal) or dates

Item (Description)

Record (place) ->

POLYGON	ESA_1	SPC1	PCT1	SPC2	PCT2	AGE_CL	HT_CL_IN	SITE_IDX	CRNCL_CL	SitePrep	Dist	YearDist	Regen	STTEND
67		HW	40	S	40	2	1	16.6	8	B	R	1985	1999	F
133			0		0	0	0	0	0			0	0	
199		HM	40	HW	30	9	3	7.2	5		L	1980	0	
353		HW	90	BA	10	9	4	11.6	1	B	L	1980	1999	F
229		HW	70	HM	20	9	3	9.5	5	B	L	1980	1999	F
264		HM	50	HW	30	9	3	7.5	5	H	L	1980	1999	F
162			0		0	0	0	0	0			0	0	
393		HW	60	HM	20	9	3	8.5	5	H	L	1980	1999	R
165		HM	80	BL	20	9	3	7	4	H	L	1980	1999	R

14

Scale

- Scale: The ratio or relationship between a distance on a map and the corresponding distance the ground.
- Map details depend on the scale of the data
  - The larger the scale, the more the detail
    - Large scale is synonymous with fine scale (fine granularity)
  - The smaller the scale, the less the detail
    - Small scale is synonymous with course scale (course granularity)

15

15

Scale

- Spatial Scale **could be making reference to**
  - Resolution: the size of your pixel (raster)
  - Extent: the size of your study area
    - "small-scale" operation covers a small area
    - "large-scale" operation covers a large area
- Cartographic Scale
  - 1:1000000 (1 cm = 10 km)
- Scale Bar

16

16

Census Data

- Geographic Units
  - Dissemination Block
  - An area bounded on all sides by roads and/or boundaries of standard geographic areas. The dissemination block is the smallest geographic area for which **population and dwelling counts** are disseminated.
  - Dissemination Areas
    - composed of one or more adjacent dissemination blocks with an average population of 400 to 700 persons. It is the smallest standard geographic area for which **all census data** are disseminated.
    - Enumeration Areas: term used for the same geographic unit prior to 2001
  - Census Tracts
    - Larger areas that have a population between 2,500 and 8,000 persons.
    - They are located in centres of 50,000 or more in the previous census.
- What is a unique aspect about the 2011 Census?
- What are boundary files?
- What is the difference between a census block, dissemination area, and census tract?

17

17

	Nontopological	Topological	
		Featured-based	**Layer-based**
Logical operations	• Attribute database query		• Reclassification and aggregation
Arithmetic operations	• Change mapping • Summary statistics		
Overlay operations	• Address geocoding	• Overlay analysis	• Overlay analysis
Geometric property operations	• Calculation of areas, perimeters, and distances		• Network analysis
Geometric transformation operations	• Coordinate and geometric transformation • Surface interpolation		
Geometric derivation operations		• Buffering	

Source: C.P. Lo, A.K.W. Yeung (2007) Concepts and Techniques of Geographic Information Systems

18

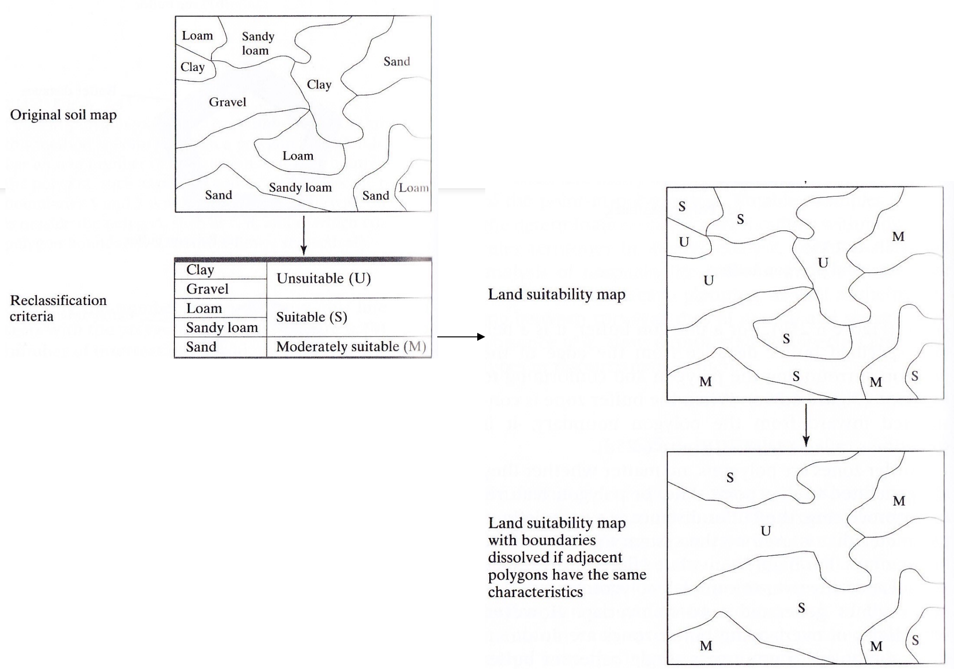
18

Topological Functions

- Reclassification
  - Database simplification process that reduces (modifies) the categories in attribute data
  - Two-step process
    - Nontopological: select attribute e.g. a range of values and assign them a new class
    - Topological: dissolve according to new classification scheme

19

19



Source: C.P. Lo, A.K.W. Yeung (2007) Concepts and Techniques of Geographic Information Systems

20

20

Topological Functions

- Topological Overlay Operators
  - Types:
    - **Point** layer in **polygon** layer
    - **Line** layer in **polygon** layer
    - **Polygon** layer in **Polygon** layer
  - Operators:
    - UNION, INTERSECT, IDENTITY, CLIP, ERASE, SPLIT, BUFFER

21

21

Raster Operations

- Raster operations: grouped according to the way raster cells are used in the analysis
  - Local Operations:
    - value of the cell in the output layer is a function of the cell at the same location in the input layer
  - Neighborhood Operations:
    - value of the cell in the output layer is a function of the cells neighboring the cell at the same location in the input layer
  - Extended Neighborhood Operations:
    - value of the cell in the output layer is a function of the cells neighboring and beyond the immediate neighborhood of the cell at the same location in the input layer
  - Regional Operations:
    - the output layer us generated by identifying cells that intersect with or fall within each region on the input layer

22

22

Local Operations

- Reclassification
  - Create a new raster layer by applying changes to the attribute values of the cells in the input layer
  - Logical or arithmetic operations

Binary masking; Classification reduction; Classification Ranking; Changing Measurement Scales

- Overlay Analysis
  - Logical or arithmetic operations
    - AND, OR, XOR; addition, subtraction, multiplication, division, assignment
  - Two or more input layers

23

23

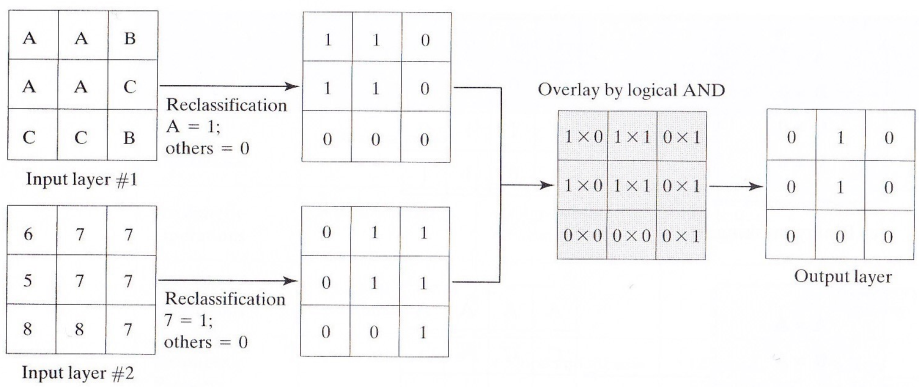
Raster Operations

	Local Operations	Neighborhood Operations	Extended Neighborhood Operations	Regional Operations
Logical Operations	• Reclassification			
Arithmetic Operations	• Reclassification	• Aggregation • Filtering	• Statistical analysis	
Overlay Operations	• Logical • Arithmetic			• Category-wide overlay
Geometric Property Operations		• Slope and aspects	• Distance, proximity, and connectivity	• Area • Perimeter • Shape
Geometric Transformation Operations			• Rotation • Translation • Scaling	
Geometric Derivation Operations			• Buffering • Viewshed analysis	• Identification and reclassification

24

24

Overlay Analysis

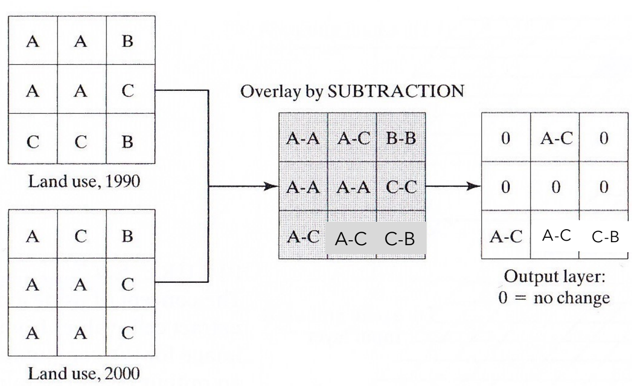


(a) Overlay by logical AND to find "A" and "7" in input raster layer

25

25

Overlay Analysis



(a) Overlay by arithmetic SUBTRACTION to detect land-use change

26

26

Levels or Scales of Measurement

- Nominal
  - Categorical data e.g. land use type, religious affiliation
- Ordinal
  - Ranked data , e.g. main, secondary, minor roads
- Interval:
  - Interval between any two units can be measured on scale. Zero value is assigned arbitrarily e.g. Celsius and Fahrenheit scales (80°F is not twice as hot as 40°F)
- Ratio:
  - interval data with an absolute zero value

27

27

Discrete and Continuous Variables

- A variable is a property or a characteristic of each a given phenomenon or object that can be measured
  - The resulting measurement or code is called a ***data value***
  - a variable which can theoretically fall between two values is called a **continuous variable.**
    - Has infinite number of possible values
    - E.g. height can be 178,178.78,179 centimeters
  - a variable which can be determined by counting is a called **discrete variable**
    - Number of children in a household can be 0, 1, 2, 3… (not 2.5 or 2.2)

28

28

Descriptive Statistics

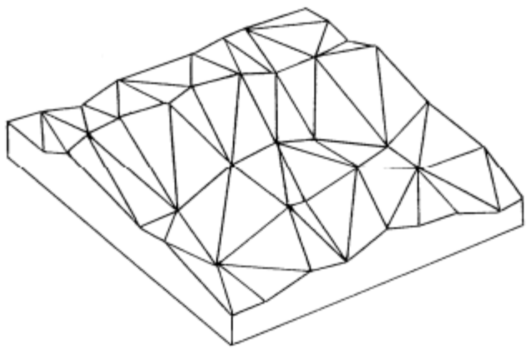
- Descriptive statistics provide concise, easily understood characteristics of a particular dataset
  - Measures of Central Tendency
  - Measures of Dispersion and Variability
  - Measures of Shape and Relative Position

29

Spatial Interpolation

- Triangulated Irregular Network (TIN)
  - Consist of z-value nodes that are connected by edges to form contiguous and non-overlapping triangles
  - The edges in TINs can be used to capture the position of linear features that play an important role in the definition of the surface (e.g. ridgelines or stream courses)

Triangulated Irregular Networks (TIN): vectors



TIN: a series of triangles capturing the topography .. x, y, z at nodes

Each triangle has a uniform slope and direction (aspect)

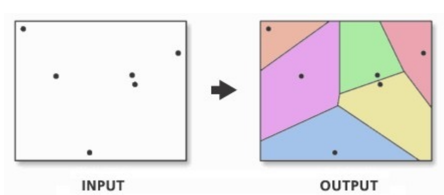
Advantage: **significant points or lines** can be encoded e.g. peaks, ridges, valleys

Disadvantage: more complex, needs more processing to generate, when a new point is added, the TIN needs to be rebuilt

31

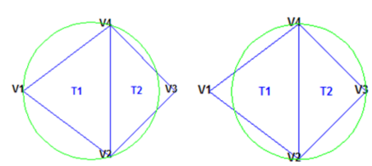
Spatial Interpolation

- Thiessen Polygons



Any location within a Thiessen polygon is closer to its associated point than to any other point input feature

All points are triangulated into a triangulated irregular network (TIN) that meets the Delaunay criterion: the circumcircle of every triangle is empty, that is, there is no other point in its interior.



The perpendicular bisectors for each triangle edge are generated, forming the edges of the Thiessen polygons. The location at which the bisectors intersect determine the locations of the Thiessen polygon vertices.

32

32

Spatial Interpolation

- Contouring
  - Contours are lines that connect locations of equal values for a given continuous phenomenon.
    - as elevation, temperature, precipitation, pollution, or atmospheric pressure.
    - Contour lines are often generally referred to as isolines but can also have specific terms depending on what is being measured (e.g. isobars for pressure, isotherms for temperature, and isohyets for precipitation)

33

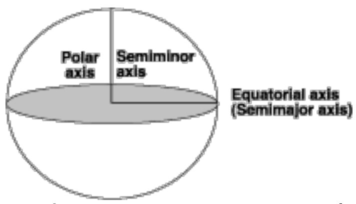
Spatial Interpolation

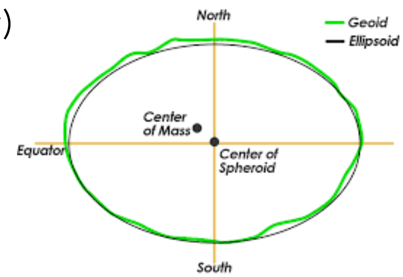
- Contouring
    - The distribution of the contour lines shows how values change across a surface.
      - Little change in a value, the lines are spaced farther apart.
      - Great change, the lines are closer together.

34

Modeling the Earth's shape

- the Ellipsoid/Spheroid
- the Geoid (approximates earth's gravity)





- the Mean Sea Level
  - Over oceans MSL = geoid; differ on landmass
- the Terrain

35

35

Types of Coordinate Systems

- Geographic coordinate systems
  - Global or spherical coordinate systems such as latitude-longitude.
    - Prime meridian
    - Latitude, Longitude
    - Parallels, meridians
    - Units of measure
- Projected coordinate systems
  - Examples?
- Datum
  - Examples?

36

36

Map Projections

- Basic Concepts
  - "Project" features on the Earth's curved surface onto a flat surface (e.g. sheet of paper)
  - Therefore "map projection" is about preserving the properties of real-world features when they are depicted on a 2D map.
    - The properties are:
      - Area
      - Shape
      - Distance
      - Direction

37

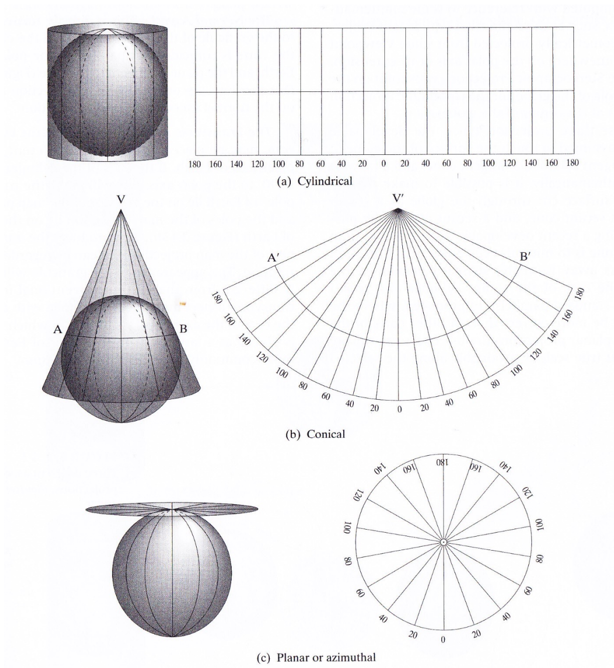
37

Map Projections Classes

- There are different ways to classify map projections
  - One is based on developable surface on which a network of meridians and parallels is projected
    - Cylindrical
    - Conical
    - Planar/Azimuthal

38

38



Three basic types of map projections

Source: C.P. Lo and A.K.W Yeung

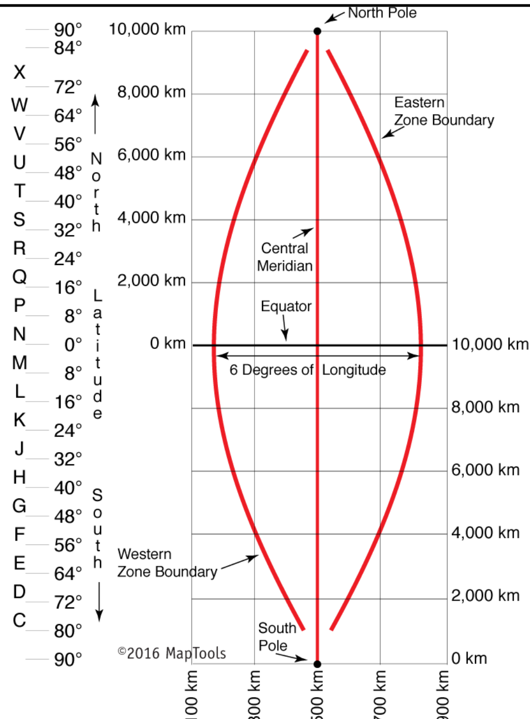
39

39

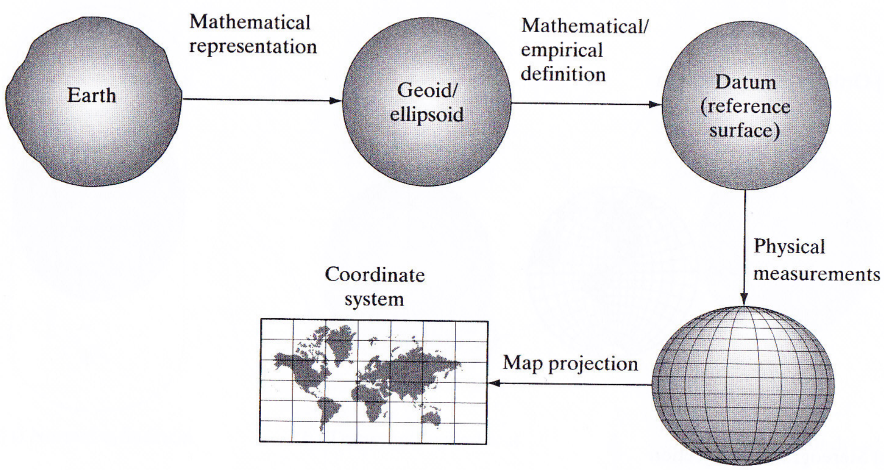
Major Projection Classes

- The other is based on the projection properties
  - Conformal
    - Local shapes/angles are correct, areas allowed to vary
  - Equal-area
    - Areas are correct, shapes allowed to vary
  - Equidistant
    - Distance to certain point correct
    - Areas and shapes allowed to vary
  - Azimuthal
    - Directions are accurate

40



41



Georeferencing

Source: C.P. Lo and A.K.W Yeung

42

42

Geocoding

- Geocoding
  - The process of assigning spatial locations to descriptive data for those locations
    - from location description to spatial locations
- The most common type of geocoding is *address matching*
- Address matching
  - requires two sets of data
    - The first data set contains individual street addresses in a table, one record per address
    - The second is a reference database that consists of a street map and attributes for each street segment such as the street name, address ranges, and postal code
  - Three phases: preprocessing, matching, plotting
    - To plot the location, the system performs linear interpolation

43

43

Data Quality: Sources of Errors in Data

- Sources of Errors:
  - **Human errors** include mistakes, such as reading an instrument incorrectly, and faulty judgments (e.g. ambiguous boundaries such as high water mark).
  - **Environmental characteristics**, such as variations in temperature can result in measurement errors
  - **Instrument errors** Measurements are as precise as the instrument's capabilities.
    - The smallest measurement that can be made is the instrument's resolution.

44

44

Elements of Data Quality

- Accuracy
  - Positional accuracy
    - closeness of locational information (usually coordinates) to the true position
  - Thematic/attribute accuracy
    - the closeness of attribute values to their true value
- Lineage
  - a record of the data sources and of the operations which created the database
    - how were they digitized, from what documents?
    - when were the data collected? By who?
  - is often a useful indicator of accuracy
- Logical consistency
  - refers to the consistency of the data model (particularly the topological consistency)
    - is the database consistent with its definitions?
    - is there exactly one label for each polygon?
    - are there nodes wherever arcs cross, or do arcs sometimes cross without forming nodes?

45

45

Elements of Data Quality

- Completeness
  - degree to which the data exhausts all the possible items
    - are all possible objects included within the database?
  - affected by rules of selection, generalization and scale
- Temporal quality
  - The quality of temporal attributes and temporal relationship of features.
- Data usability
  - Suitability to an application and its related functional requirement

46

46

Data Quality - Key Issues

- Key Concepts
  - Accuracy, Precision and Uncertainty
- Accuracy:
  - closeness of the measurements, computations to the true values (or values accepted to be true)
    - spatial data are a generalization of the real world, the "true value" is thus an estimate of the real world
  - ~ absence of errors

47

47

Data Quality - Key Issues

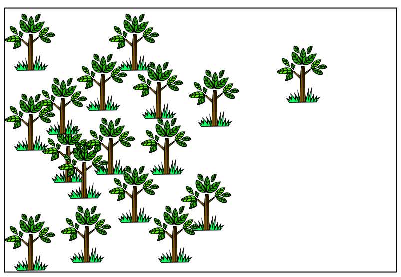
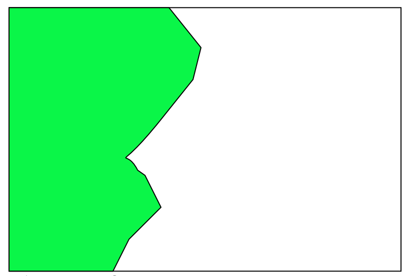
- Precision:
  - the number of decimal places or significant digits in a measurement
    - precision is not the same as accuracy - a large number of significant digits doesn't necessarily indicate that the measurement is accurate
    - Precision is related to how detailed a description one has
  - a GIS works at high precision, mostly much higher than the accuracy of the data itself

48

48

Data Quality - Key Issues

- Uncertainty: our imperfect and inexact knowledge of the world
  - Positional uncertainty
  - Attribute uncertainty
  - Definitional uncertainty
  - Measurement uncertainty

www.geog.ucsb.edu/~kclarke/G176B/Lecture07.ppt

49

49

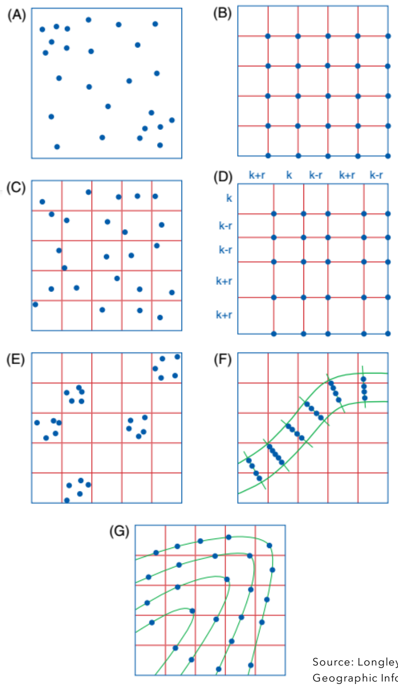
Sampling

- Provides knowledge about a whole population
  - i.e. make inference about a population from the sample data
- Larger sample sizes are more accurate representations of the whole
  - Large samples are costly: time, labour
    - Can be wasteful since we can statistically infer from appropriate samples
- A sampling strategy with the minimum bias is the most statistically valid

50

50

Sampling



Spatial sample designs: (A) simple random sampling, (B) systematic sampling, (C) stratified random sampling, (D) stratified sampling with random variation in grid spacing, (E) clustered sampling, (F) transect sampling, and (G) contour sampling.

Source: Longley, Paul A.; Goodchild, Michael F.; Maguire, David J.; Rhind, David W.. Geographic Information Science and Systems, 4th Edition. Wiley.

51

51

Random Sampling

- Random sampling: each member of the population has an equal chance of being selected
  - Advantages:
    - Can be used with large sample populations
    - Avoids bias
  - Disadvantages:
    - Can disproportionately represent some parts of the population at the expense of others

52

52

Systematic Sampling

- Systematic Sampling: Samples are chosen at regular intervals
  - Sample locations are evenly distributed for example every two metres along a transect line
    - systematic sampling implies a regularly spaced grid
      - Advantages:
        - It is more straight-forward than random sampling
        - Provides a good coverage of the study area
      - Disadvantages:
        - It is more biased: not all points have an equal chance of being selected
        - It may lead to over or under representation if there is periodicity in the data (e.g. sampling at the same interval as the location of erosion barriers along a beach. Or a city road grid)

53

53

Stratified sampling

- Stratified sampling: used when the parent population is made up of sub-groups that of interest.
  - Divide the sampling design into strata(classes), and then select a sample from each stratum
  - The strata are defined so that individuals inside each class are similar based on the characteristic believed to influence the phenomena

54

54

Stratified sampling

- Advantages:
  - If the proportions of the subgroups are known, the results are representative of the whole population
  - Correlations and comparisons can be made between subgroups
- Disadvantages:
  - The proportions of the subgroups must be known

Terms

- Choropleth Map
- Topology
- Topography
- Temporal
- GIS Functions/Operations vs Analysis Functions/Operations
- Geographic Coordinate Systems
- Projected Coordinate Systems