Geography 360
Principles of Cartography

April 10, 2006

Outlines

- Map projections and distortion
- Sustainability indicator
  - What are common data types to represent the geographic phenomenon?
  - What are common ways to measure the geographic phenomenon as a basis for sustainability indicator?
  - How are they different: data, information, evidence, and knowledge?
- Data classification
  - Four cartographic abstractions
  - Four popular data classification schemes
  - Evaluating data classification schemes

Map projections and distortion

- Is distortion avoidable?
  - Is the scale factor invariable at all locations?
- Can all geometric properties (e.g. area, angle, distance) be preserved simultaneously?
  - Tissot's indicatrix for Mercator Lec_W3M\Mercator_Tissot.JPG
  - Compare between Albers equal-area conic and Lambert conformal conic Lec_W3M\USMap.jpg
- In which scale {large, small} map, is distortion more significant?
  - World map vs. State Plane for Washington State

Data type for geographic data

Digital (mathematical) representation of geographic phenomenon

- Space
  - Point: a coordinate pair
  - Line: a sequence of points
  - Polygon: a closed sequence of lines Lec_W3M\Vector.jpg
  - Surface: tessellation Lec_W3M\Vector_Raster.jpg
- Time
  - Date
  - Time interval
- Attribute
  - Category or type (nominal) → qualitative thematic map
  - Number → quantitative thematic map

If analogy of geometry can be used, what can be said about geometric dimensions for each of three?

Spatial data types

- For representing
  - GDP
  - Accessibility
  - The location of crop land
  - Mortality rate
  - Protected area
  - Hawaiian islands
  - Road intersections
  - Elevation
  - City
- Which spatial data types would you choose? Why?
- from the reading “Shapefiles white paper”

So how would you say about representing the geographic phenomenon compared to representing attributes?

Common ways to measure the geographic phenomenon

- Categories (e.g. rural/urban)
- Counts (e.g. population)
- Measures of central tendency
  - Space: centroid, mean center Lec_W3M\meanctr.pdf
  - Time: point, interval
  - Attribute: mean, median, mode
- Proportions (e.g. population density, mortality rate)

Do counts fit into the four levels of measurements {NOIR} well?

Data, information, evidence, knowledge

- Data: raw fact
- Information: data whose meaning is revealed by users (relevance)
- Evidence: information whose meaning is revealed by community (consistency)
- Knowledge: collection of information/evidence that enhance our understanding of the process of sustainable development (multifaceted)

Example: Climate change

Cartographic abstraction

- Selection:
  - Decide where (regions), when (time frame), and what (variables to be mapped) guided by purposes
- Simplification/generalization [chapter 6]
  - Operations of reducing details
  - Choose which/how to be removed
    - e.g. removing width dimensions of road segments
    - e.g. reducing the details of complex shorelines
  - Function of map scale and intended uses
    - See Figure 6.3
    - Political boundary in political map vs. highway map
  - Douglas-peuker algorithm (simplification for vector data)
    - See Figure 6.9

Cartographic abstraction

- Symbolization [chapter 4]
  - Represent data using visual cues
  - e.g. using Red to represent communist country
  - e.g. darker shade for higher value
  - e.g. higher number of dots for higher density
- Classification [chapter 5]
  - Group data attribute values into the predefined number of classes
  - Why classification?

Four data classification techniques

- Equal intervals
- Quantiles
- Mean-standard deviation
- Natural breaks (jenks optimal)

Equal intervals

- Range = max – min
- Class interval = Range / # classes
- Group attribute data into classes with the equal interval

Quantiles

- Rank attribute data in ascending order
- # observation = total # observation / # classes
- Place the equal number of observation in each class

Mean-standard deviation

- Compute mean and standard deviation
- Determine class boundary by adding standard deviation to or subtracting from the mean

Jenks algorithm

- Choose the number of classes
- Compute ADAM (the sum of absolute deviations about the mean for the entire data set) $\Sigma (x_i - m_a)^2$ where m_a is array mean
- For each iteration {
  - Determine class boundaries
  - Compute ADCM (the sum of absolute deviations about class mean) : $\Sigma \Sigma (x_i - m_c)^2$ where m_c is class mean
  - GADF (Goodness of absolute deviation fit) = 1 – ADCM/ADAM
- Repeat the iteration until GADF cannot be maximized further

GADF ranges from 0 to 1, with 0 representing the lowest accuracy (a one-class map) and 1 the highest accuracy (# class = # observation unit)

Attempt to minimize the variation within classes

Data classification

- Seattle housing price example: http://courses.washington.edu/geog258/lec4_rev.htm

- How does it work?
- What are advantages/disadvantages?

Which data classification methods would you choose if you …?

- Want to compare poverty level between two cities
- Want to know which area is the most deviated from typical value
- Want to identify natural subgroups
- Want to examine details in the group of highly concentrated values in histogram
- Want to know exact values in each enumeration unit

Which classification method

- Considers data distribution?
- Can be used for ordinal data as well?
- Can assist in selecting number of classes?

What are pitfalls in using …?

- Equal interval
- Quantiles
- Mean-standard deviation
- Natural break

Evaluating classification schemes

	Works best when…	But there is a pitfall in that…
Equal intervals	Rectangular distribution Want to compare?	Rectangular distribution is rare
Quantiles	Interested in flattened pattern of skewed distribution?	Hide the fact data is skewed
Standard deviations	Normal distribution Interested in how typical/untypical?	What if map users don't understand mean and standard deviation?
Natural breaks	Any distribution Flexible, intuitive	Breaks not always obvious Not good for comparison