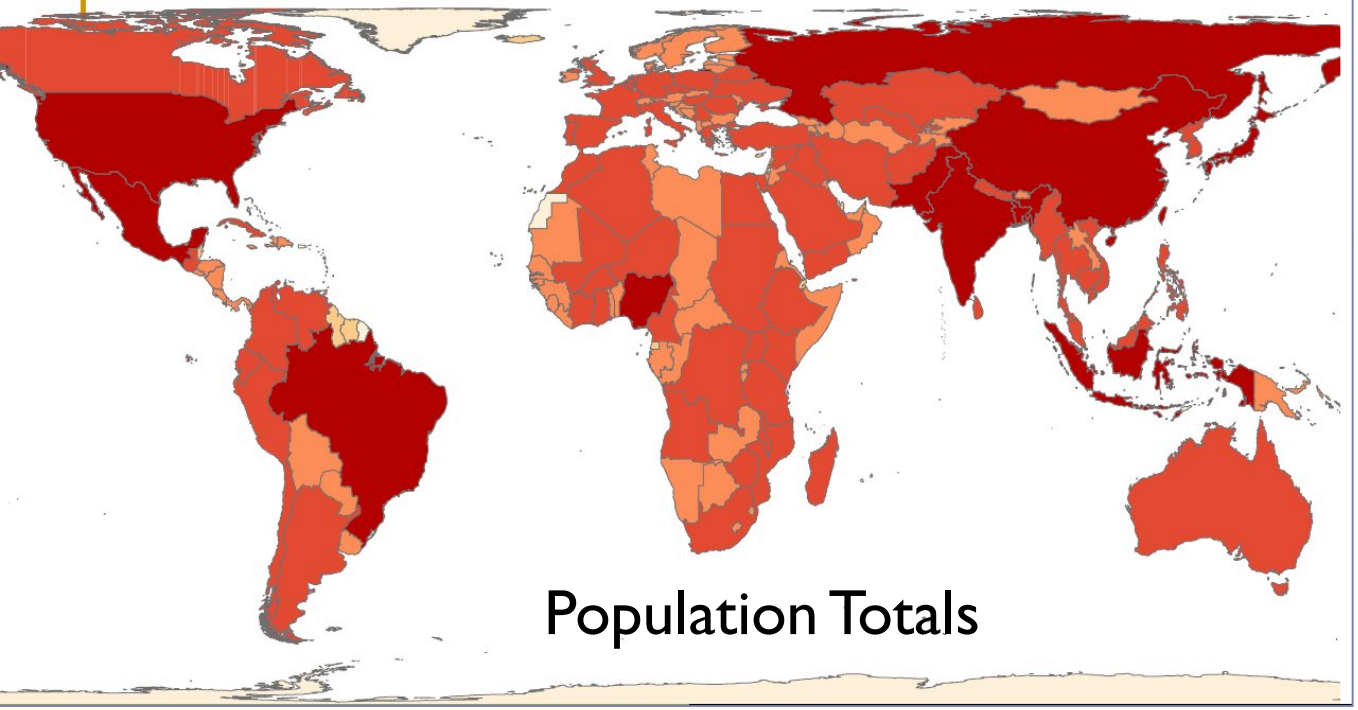
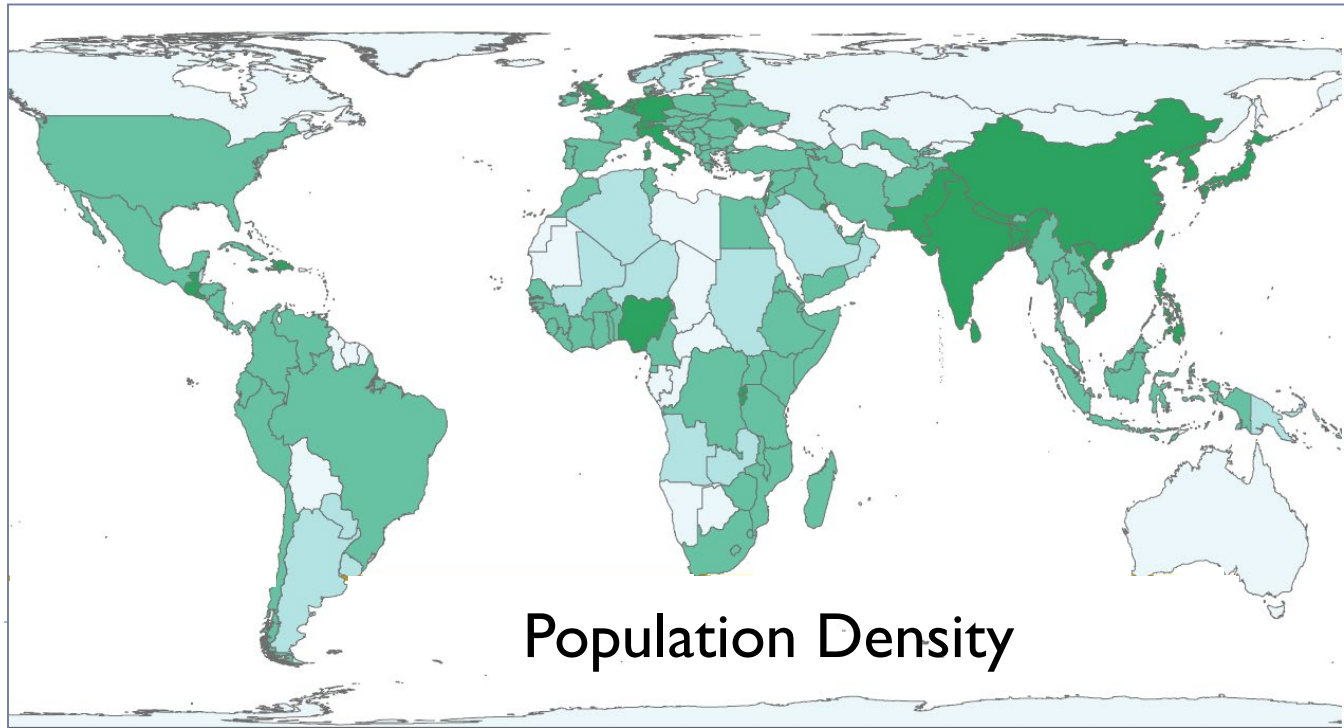
Classification I

GEOG380 FA2018

World population maps

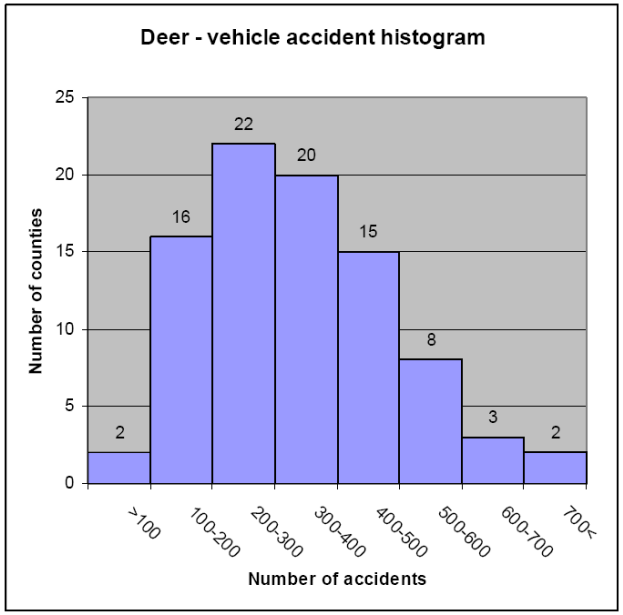


- Note:
  - Russia
  - South East Asia

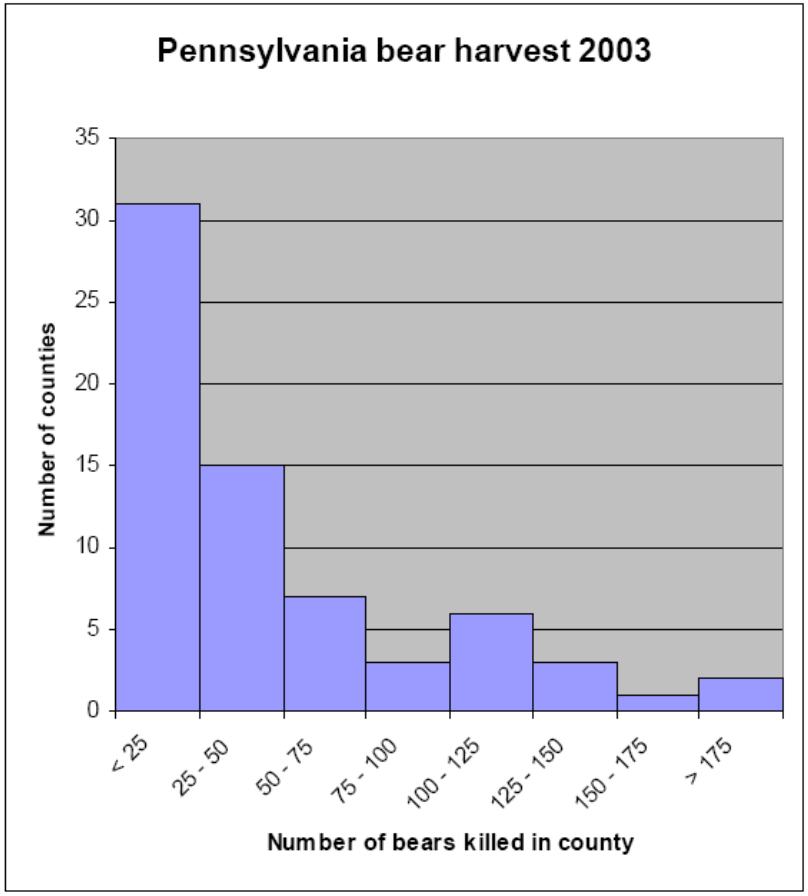


Revisit the histogram

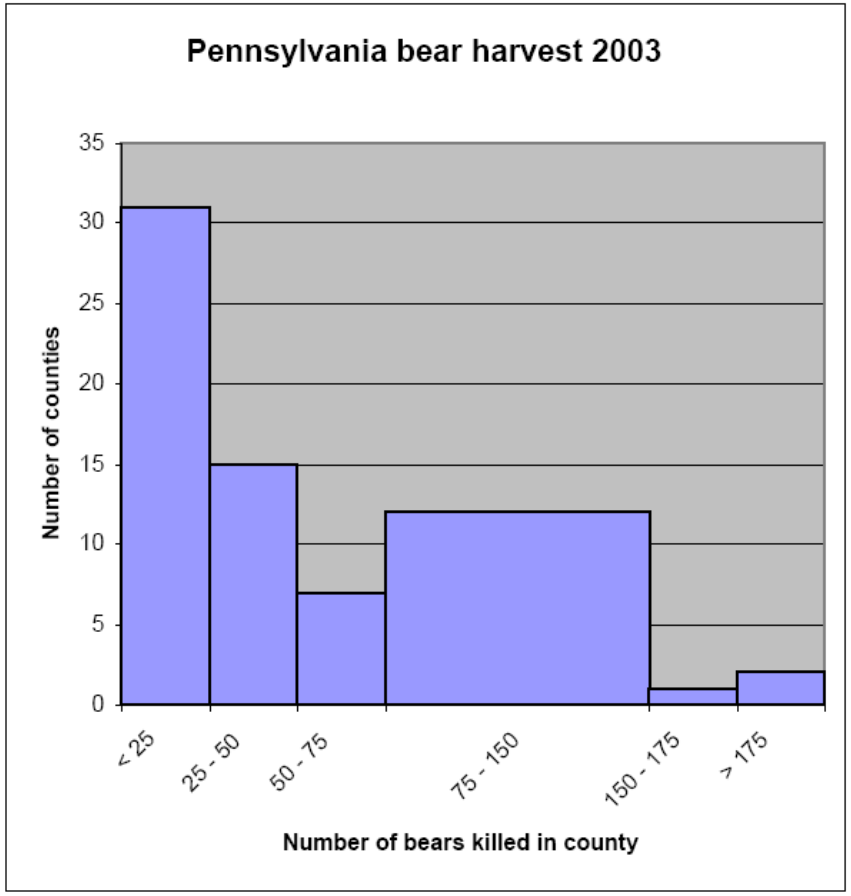
- Divide the range of the data into a series of equal intervals
- Count how many cases lie in each interval
- Plot the counts (or frequencies) as vertical bars



Importance of equal interval in histograms



Equal interval

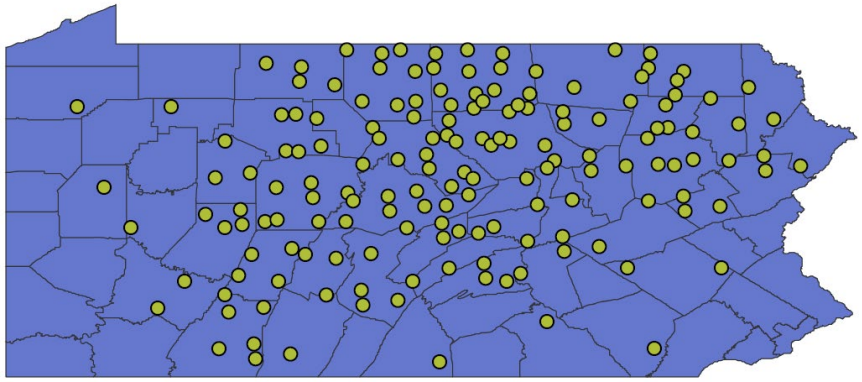


Irregular interval

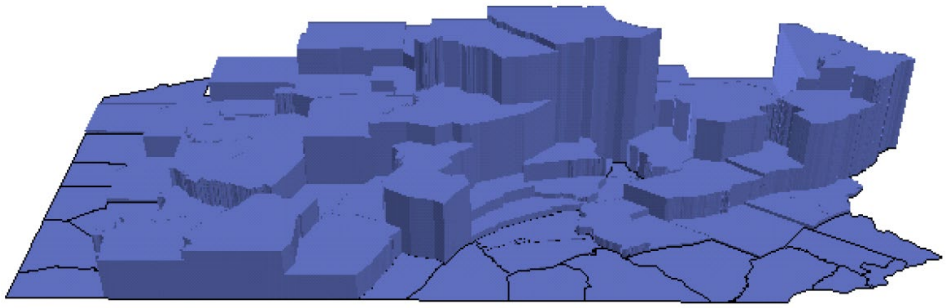
Q. Why is it problematic?

Think of the map as a spatial histogram

- Things we measure are distributed over space

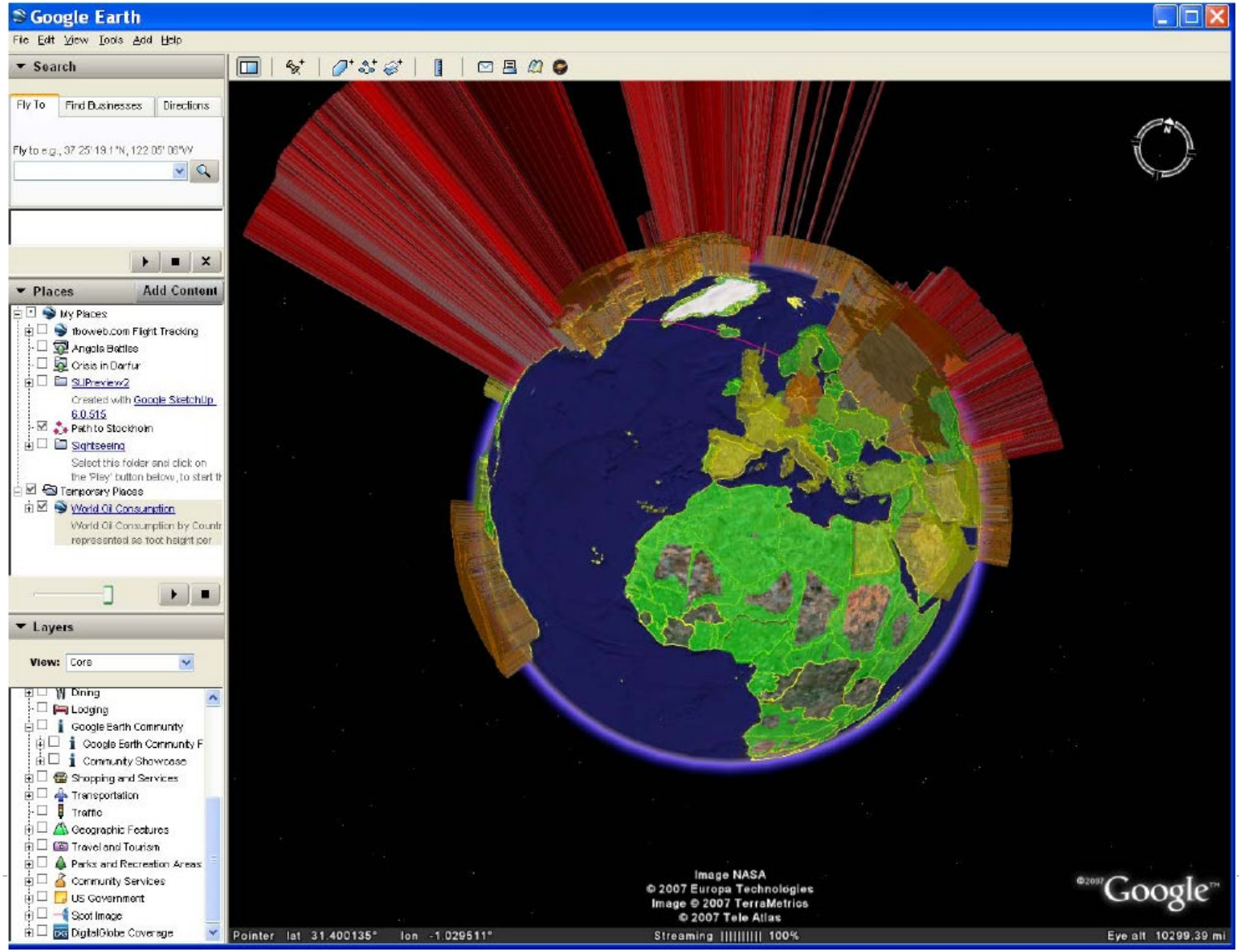


- A normalized map divides space into a pre-defined enumeration unit
  E.g., a block group



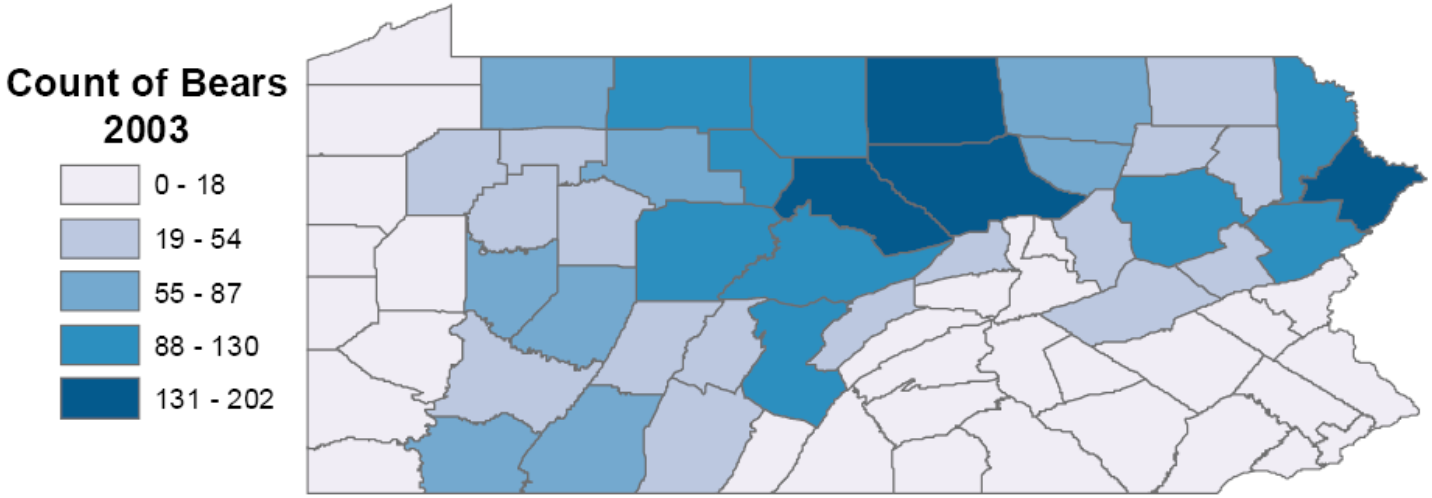
- Each enumeration unit gets a value
  - E.g. count of bears

FYI, 3D map symbolization



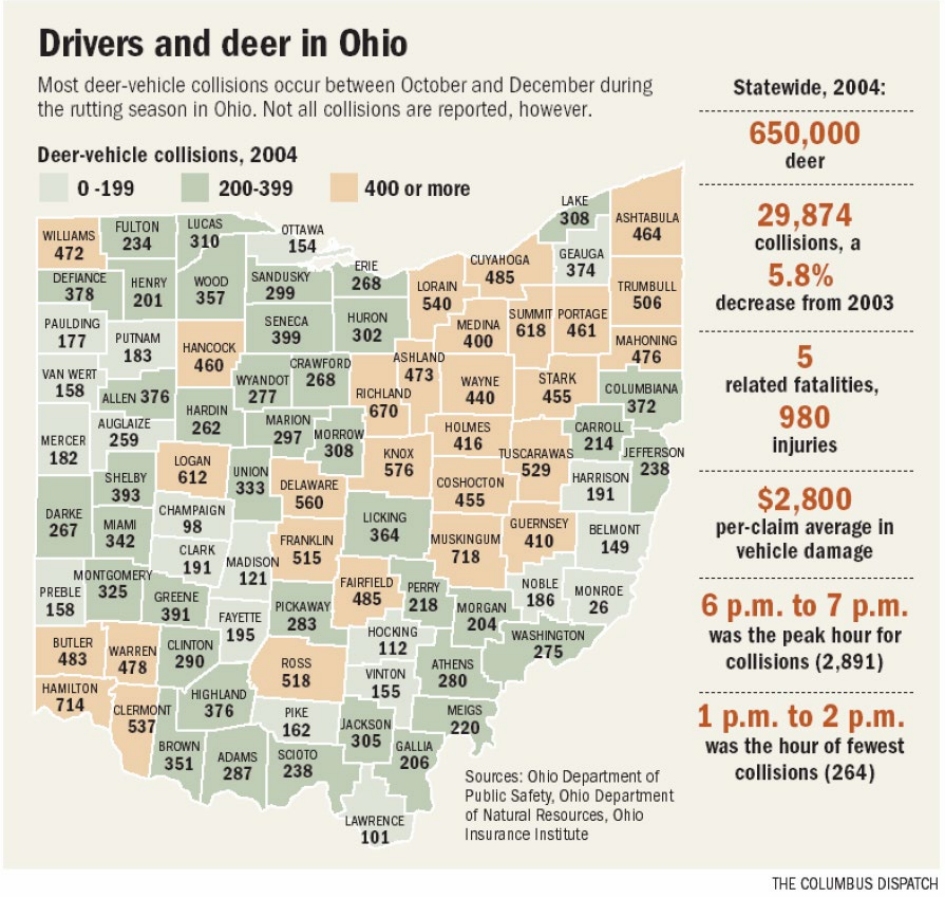
2D map symbolization

- Color or shades replace height of bars in histogram
- Each enumeration unit is a spatial interval
  - Remember: A histogram requires equal intervals to display correctly
  - Q. Do we have (geographically) equal intervals in maps?



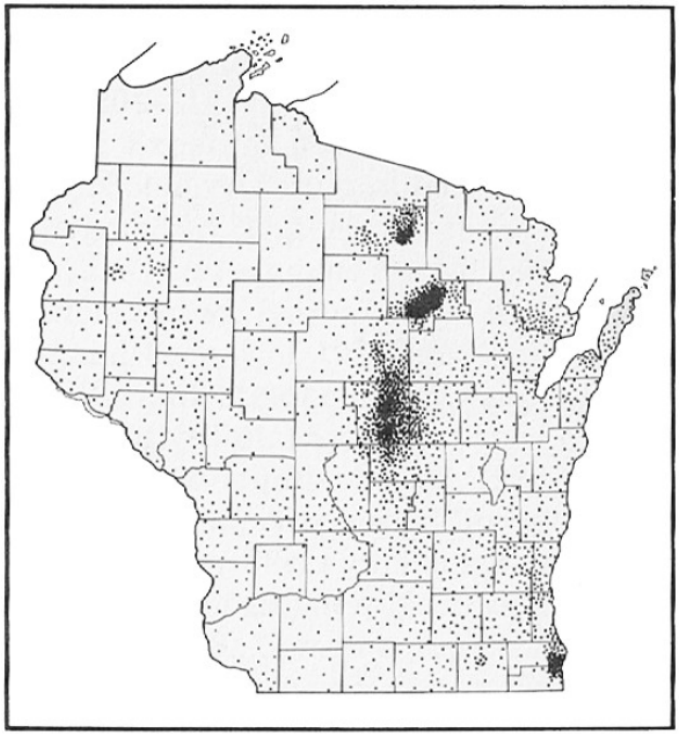
Utmost Rule #1

- Raw totals must be adjusted to account for enumeration size variation
- It is not acceptable to uncritically map total values in mapping
  - Recall the equal interval issue in histogram
- One exception is when enumeration unit size is very similar throughout the map
  - E.g. a county map of Ohio



Statistical Surface on a Map

- Any distribution that is mathematically continuous over an area and is measured on an ordinal, interval, or ratio scale
  - Examples: elevation, population density, temperature…
- Point, line, area, and volume symbols can be used for mapping



"Each dot represents 16.2 hectares (10,000 m²) of land" in potato production in Wisconsin, 1947

Q. Can you say how many dots are there? Then what should you do?

Solution: ?

- Normalization: express figures as *ratio,* not as raw values
  - e.g. population/area (or population density)

…or…

  - as Z-scores (this is also a rate; "how many standard deviations is a value above or below the mean" = deviation/std. dev.)

…or…

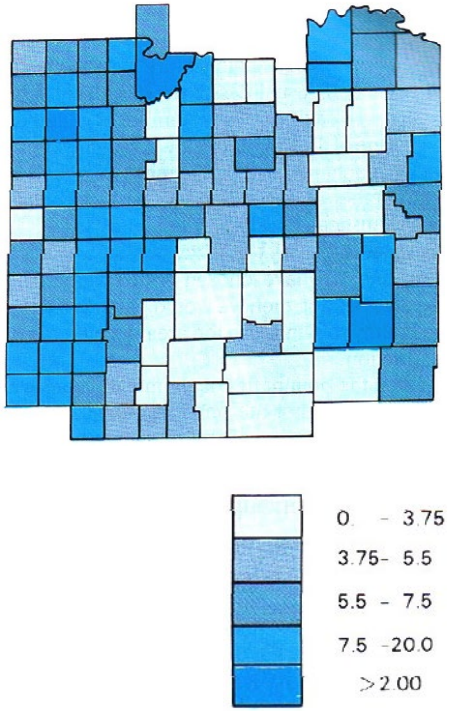
  - as a proportion (%) or rate of proportions (e.g. location quotient)

➔ *Choropleth mapping* (next slides)

Mapping data in pre-defined areal units

- Choropleth mapping
  - After Greek words *choros (place)*, *pleth (value)*

- *"A method of cartographic representation which employs distinctive color or shading applied to **areas** other than those bounded by isolines. These are usually statistical or administrative areas."*
  - ICA (International Cartographic Association)

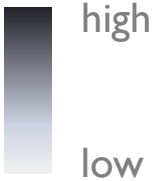


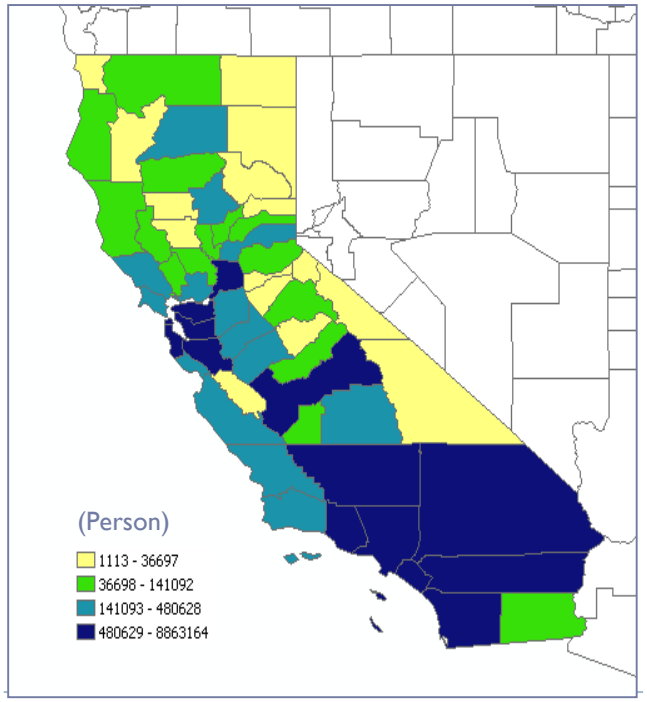
- Area symbols coincide with the data collection regions ← useful for classification
  - These regions are often called *enumeration units* or *districts* *(ex. a census block-group)*

http://kendallmaps.blogspot.com/2010/07/choropleth-maps.html

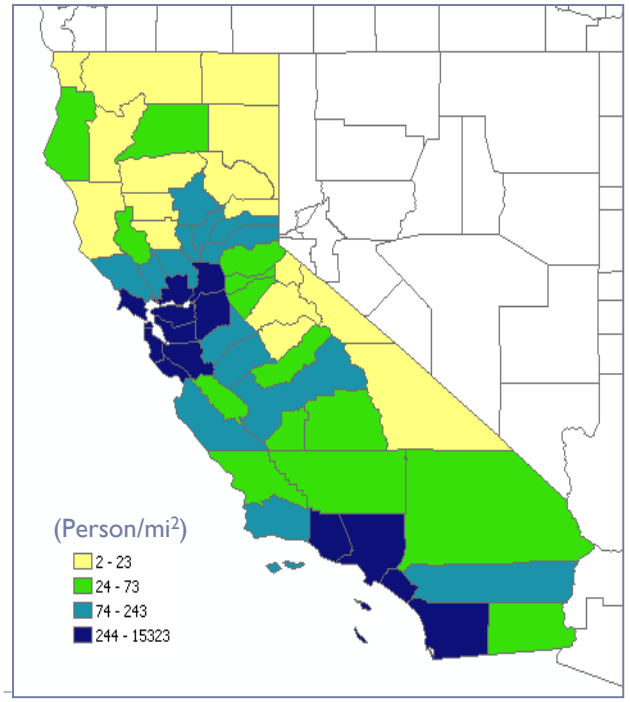
Two Kinds of Choropleth Maps

- Unclassified maps: graded symbology
  - http://fishermaniacsmaps.blogspot.com/2012/03/unclassed-choropleth-maps.html
- Classified maps: range-graded symbology
  - Population vs. population density of counties in CA





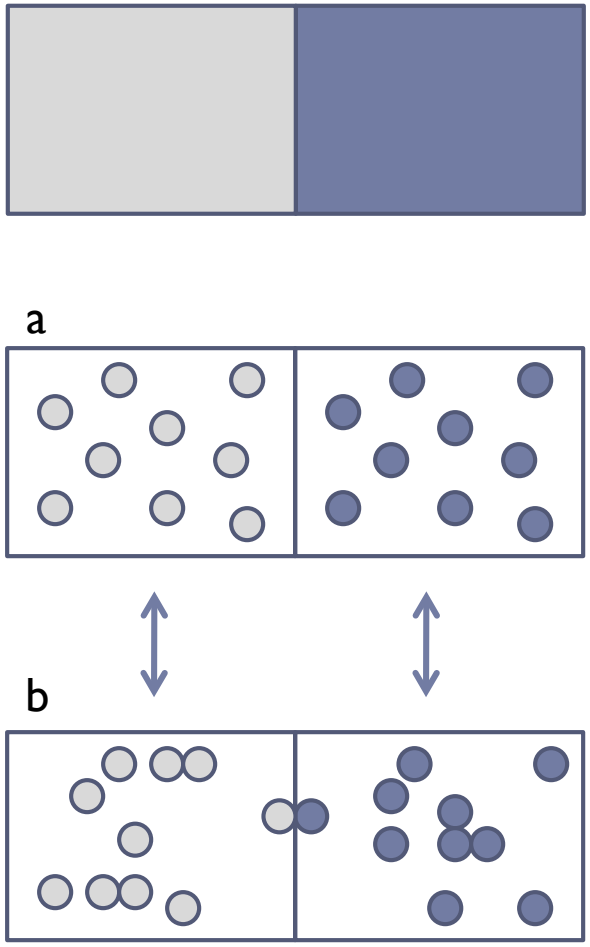
1990 Population, CA



1990 Population Density, CA

Use and Misuse

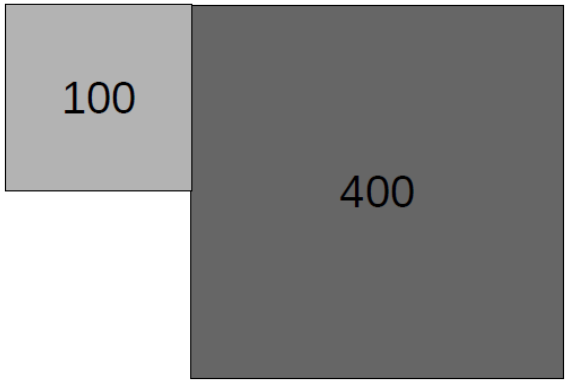
- A reader may
  - Obtain an actual value of an area
  - Observe the overall spatial pattern
  - Compare to other choropleth maps
- Ideally, phenomena should uniformly distribute within each enumeration unit and only change at the boundaries
  - However, there are not too many examples of the case!
- Likely to have a wrong impression that there is uniformity within the units and that breaks occur in the surface at the unit boundaries



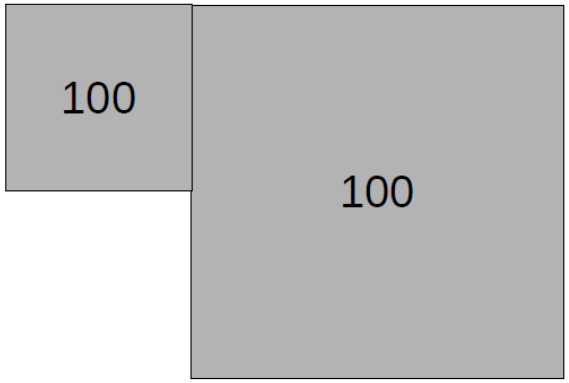
Q. What data between a and b made the choropleth map above?

Do NOT use absolute numbers alone!

- We should avoid the direct use of raw-total data
  - Size effect!
- Colors used may give the readers illusion of…
  - The density of the mapped area (even if showing raw values)
    - Size is also an effective visual variable
  - Q. Regional total – which one is better? For what reasons?



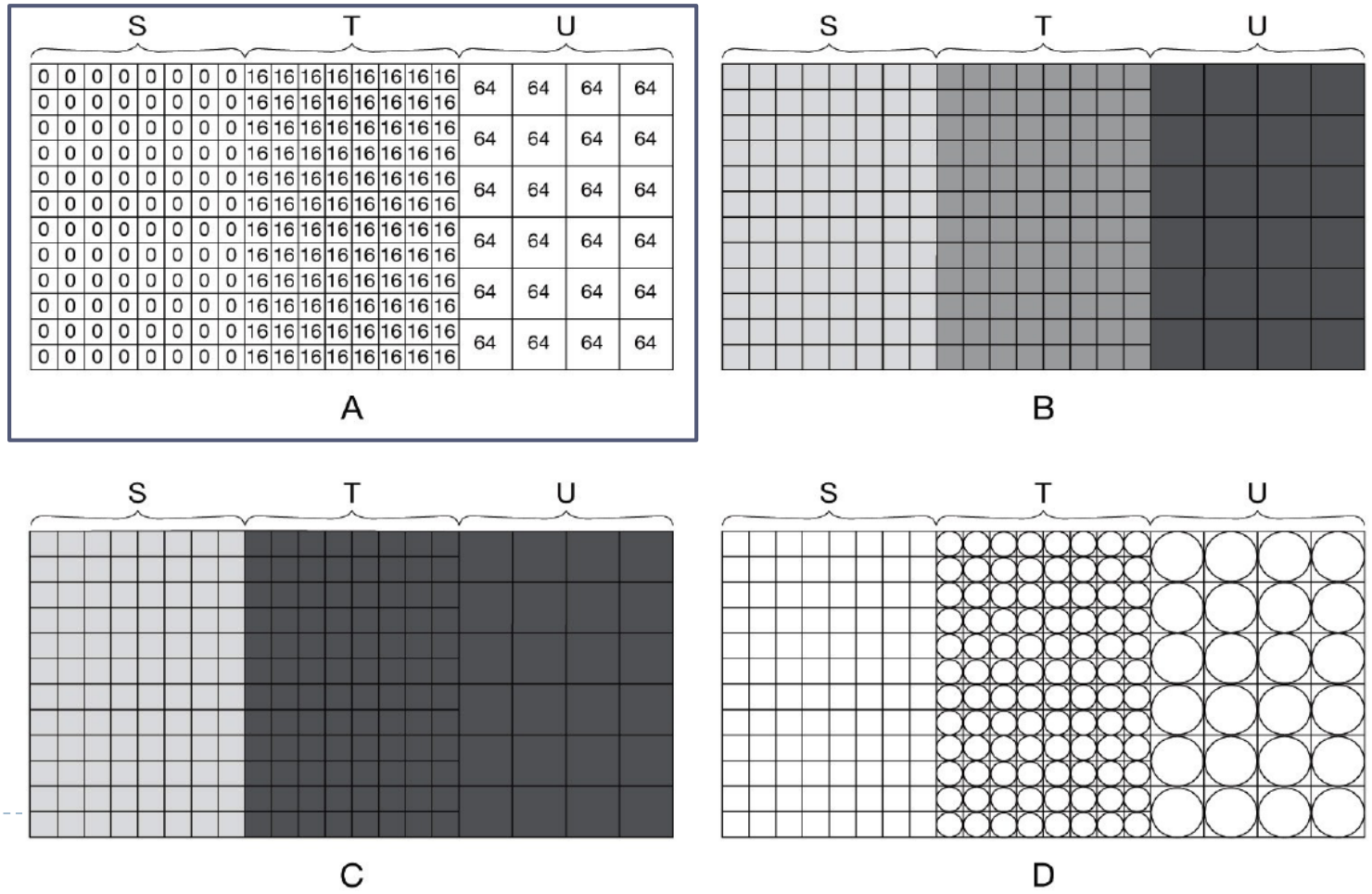
Total population



Population density

Group Activity:
Illusion of regional total and colors

- Q. What is the best design for symbology among B, C, and D to correctly represent data in A?

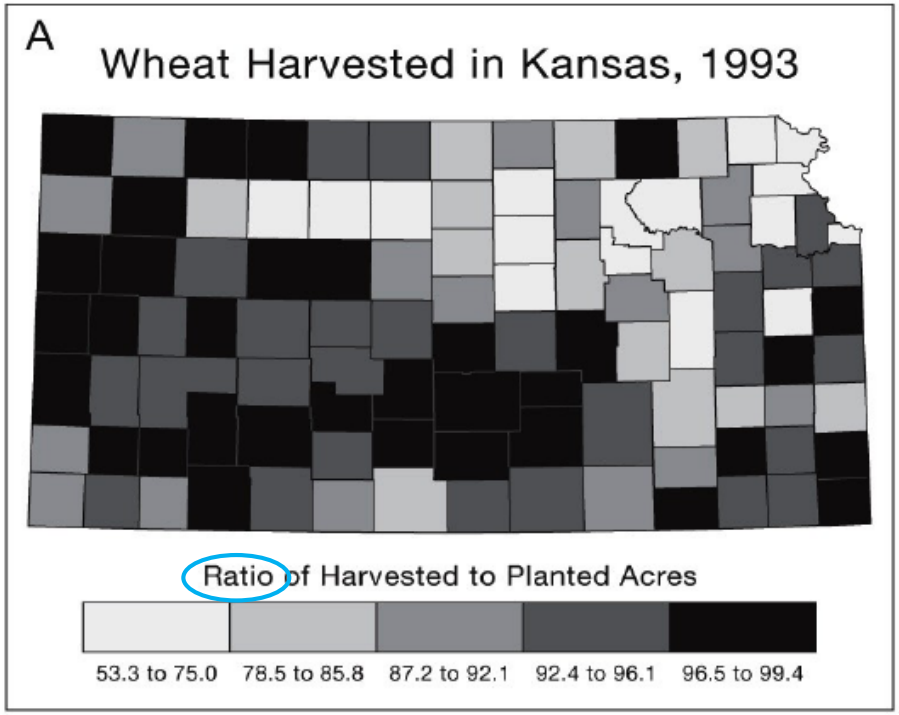


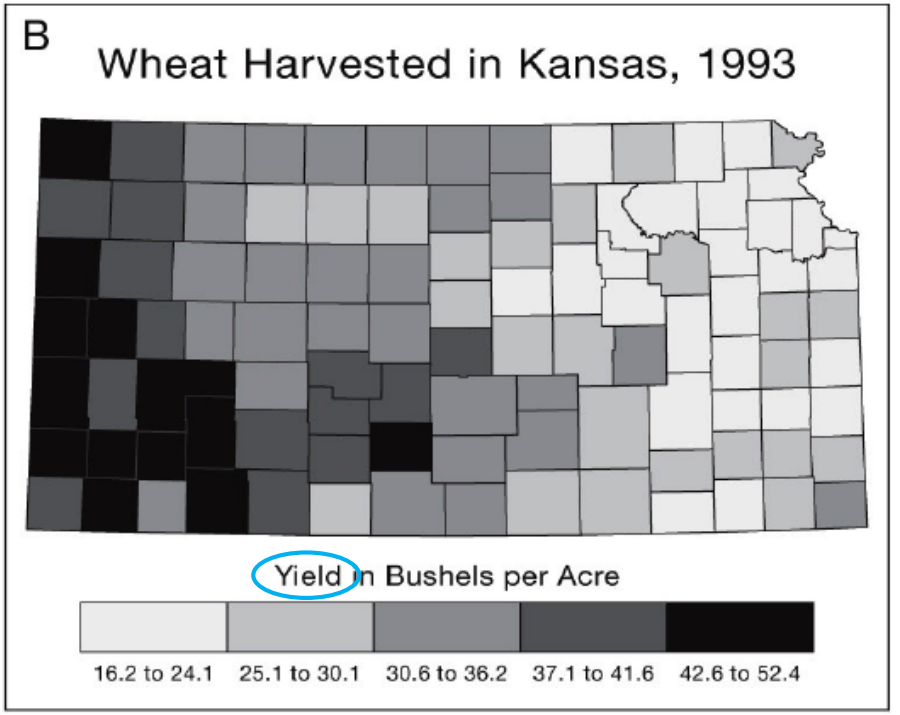
So, do NOT use absolute numbers alone!

- Map readers may perceive the color (value) and the areas of the mapped variable as the density of that variable because the area is completely filled with the color
  - When you use absolute numbers, they are not the density!
    - Ex. the two CA maps
- The unit of measurement is included to express the amount of the density value
  - Ex. 100mi^2, 1,000 person/km^2
- However, if the areas are more or less similar, a choropleth map of raw data or absolute numbers may be acceptable
  - Ex. the OH map

Group Activity:
Raw data vs. standardized data

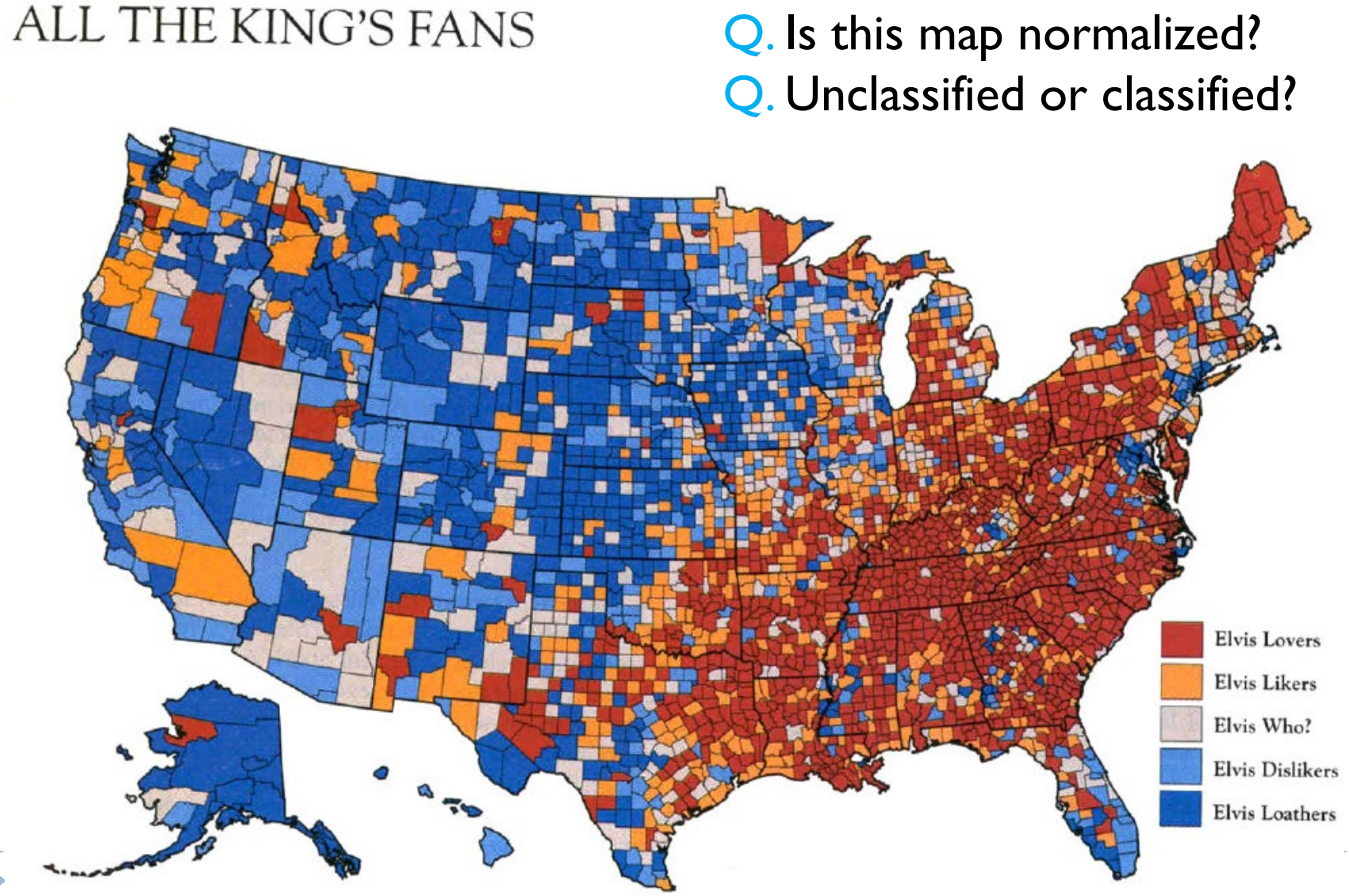
- Q. Which map is better to show spatial patterns of the phenomena?





(source: Slocum et al. 2009)

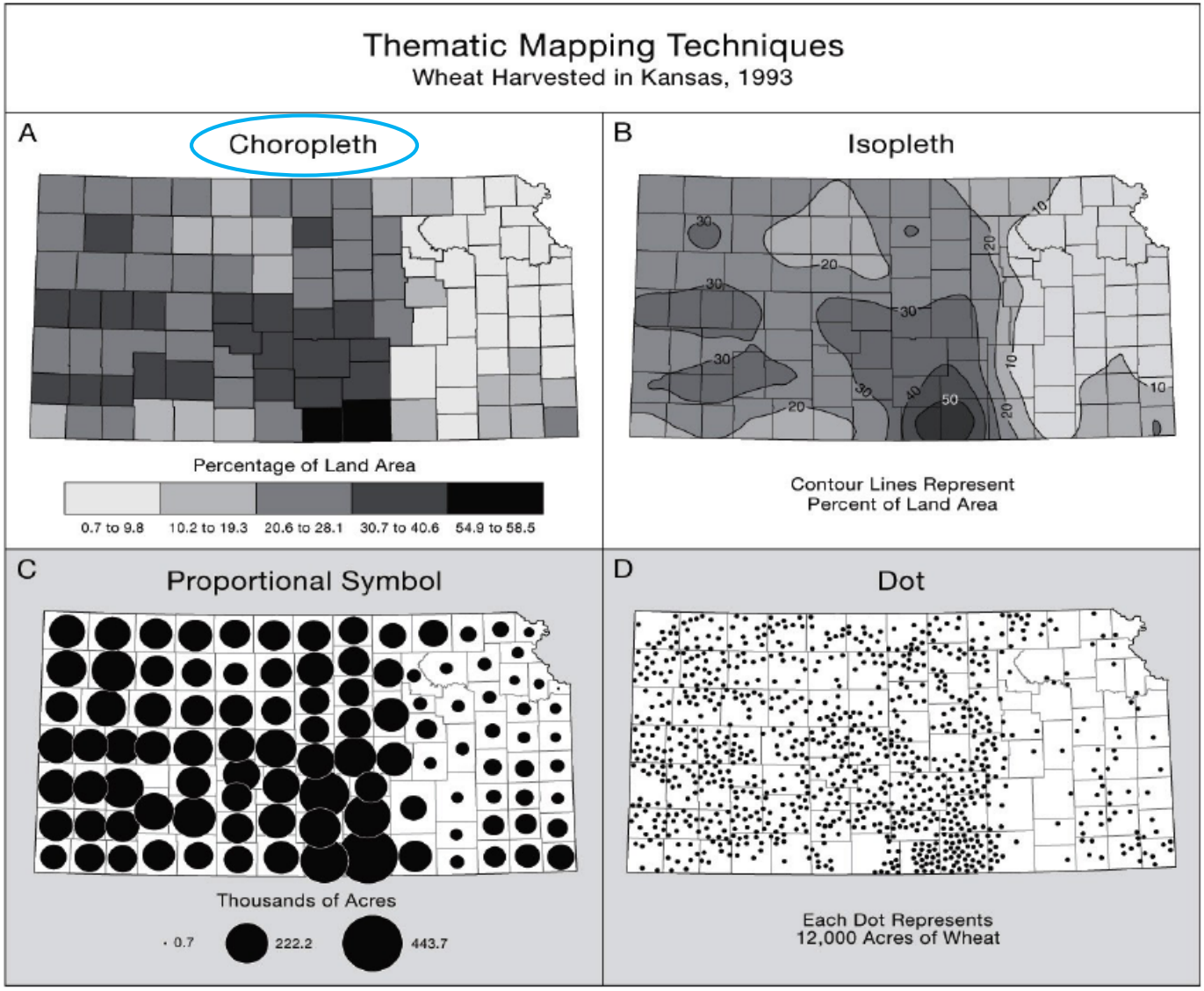
Q. Is this a choropleth map?



Q. Is this map normalized?

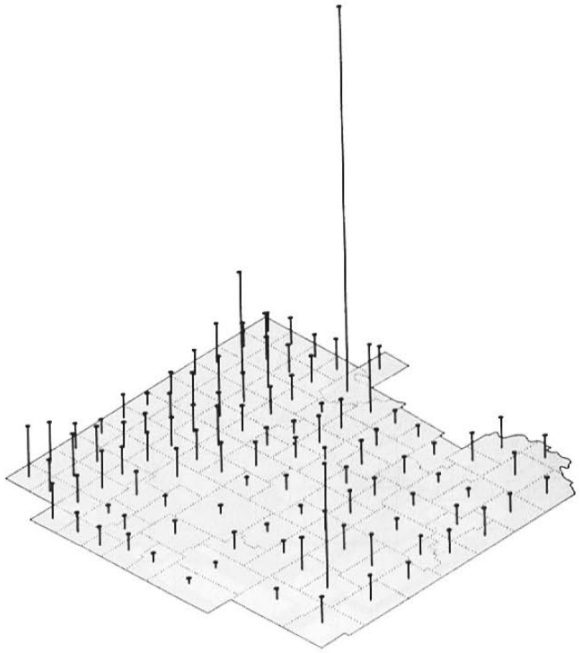
Q. Unclassified or classified?

Various techniques for thematic mapping

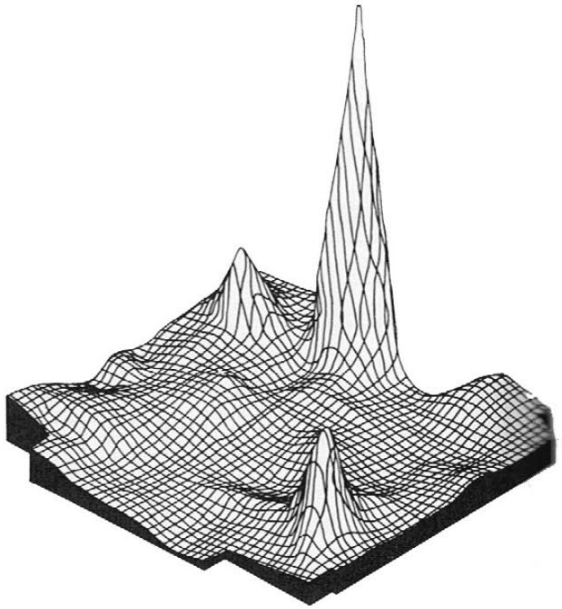


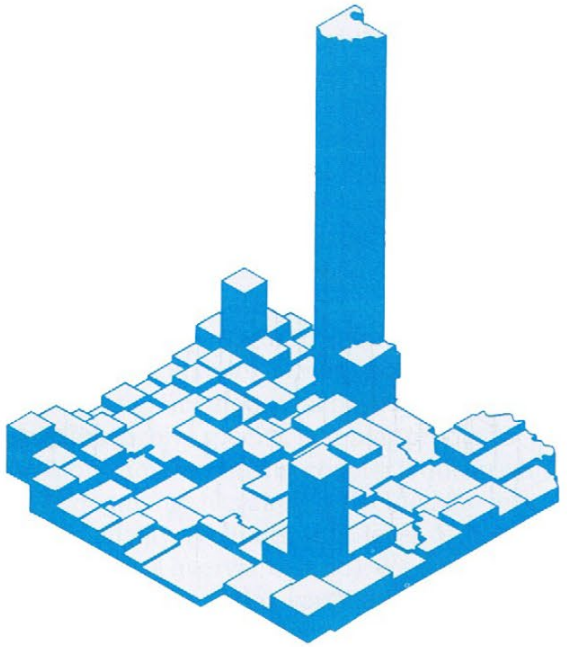
(source: Slocum et al. 2009)

Chorodot mapping: choropleth mapping in 3D

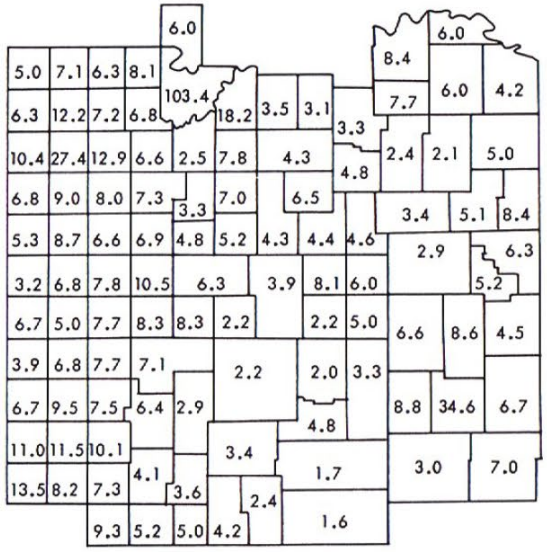


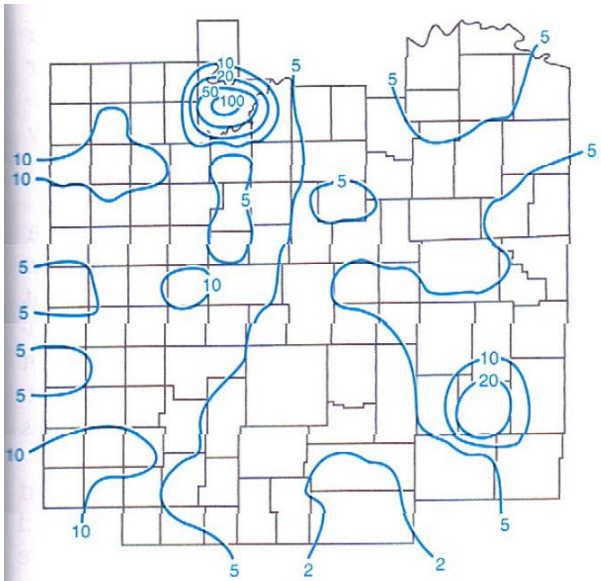
(absolute values)

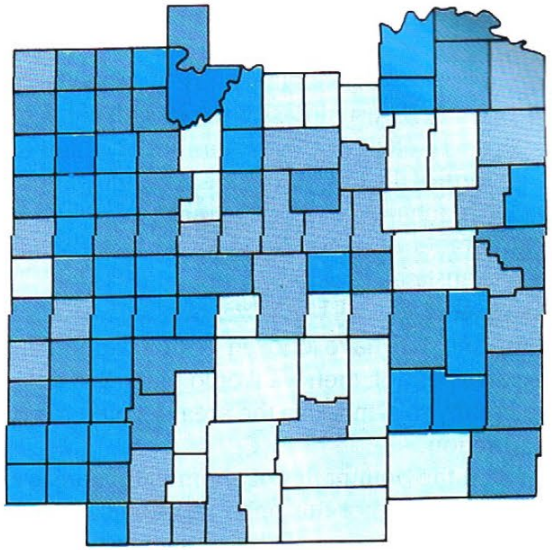




(classified values)







Demonstration

- Choropleth mapping with normalization using ArcGIS

Summary

- Importance of equal interval and normalization in classification
- Choropleth map and classification
  - Use and misuse
- Raw data vs. standardized data
- Other thematic mapping techniques than choropleth mapping

For next time…

- Reading
  - Ch. 4

- Lab2 (cont.)