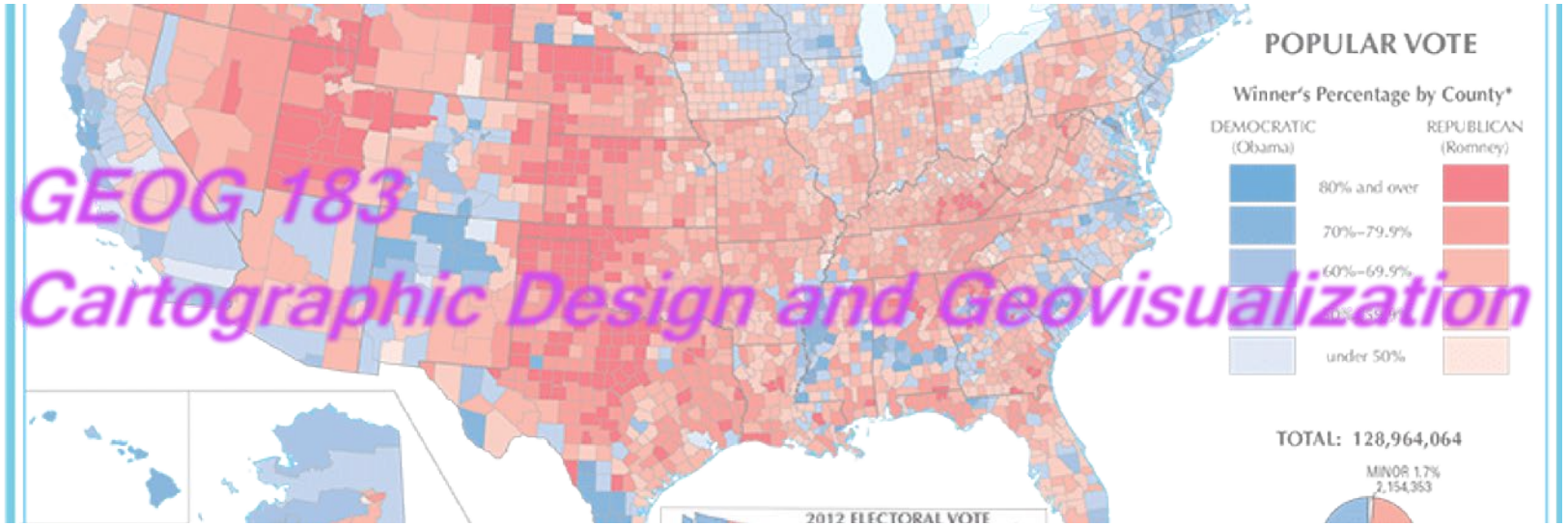
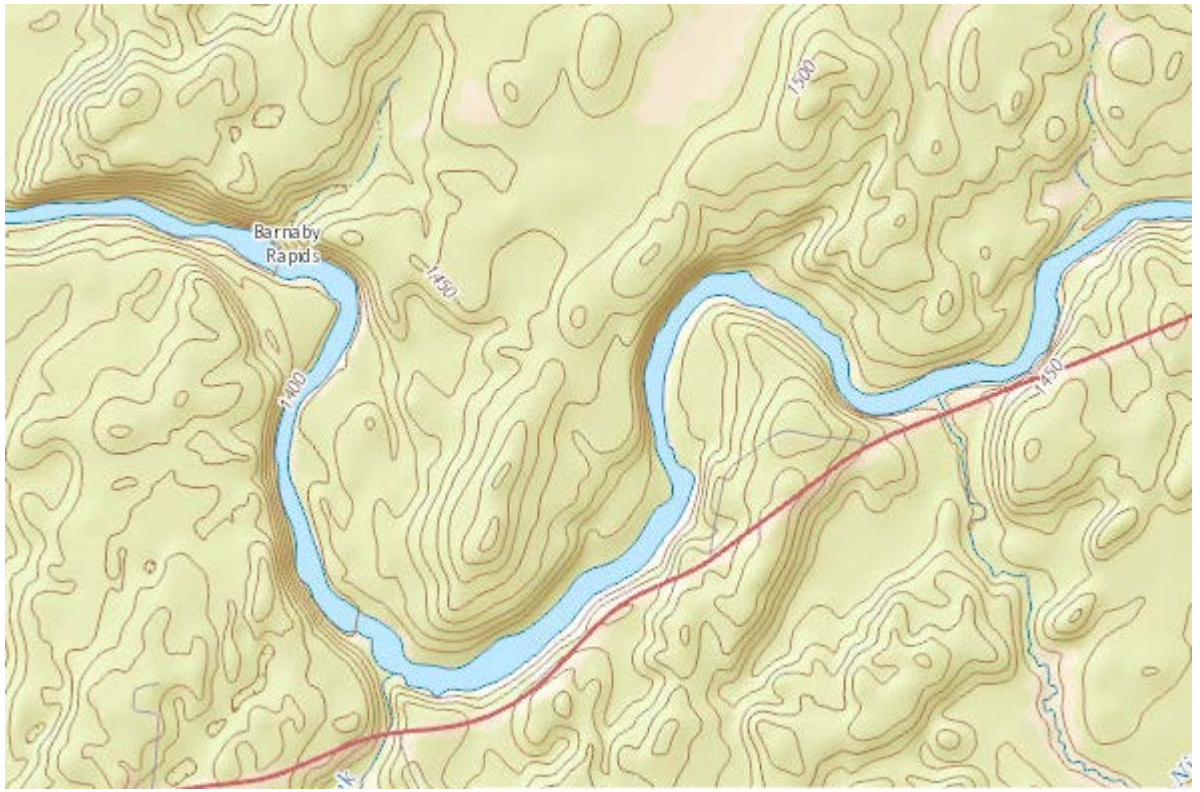


Geog183: Cartographic Design and Geovisualization Spring Quarter 2020

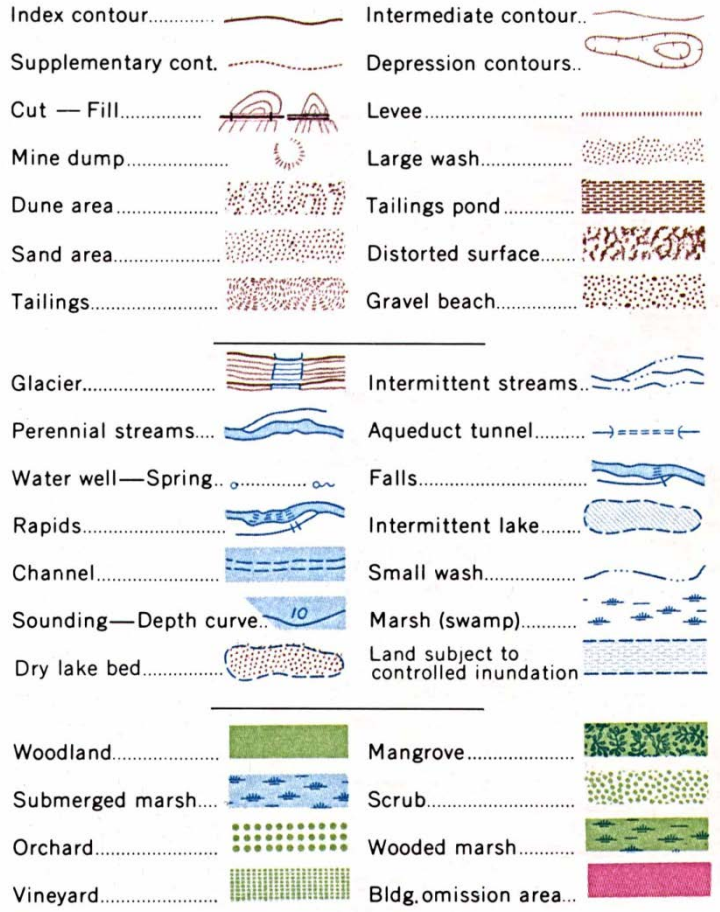
Lecture 5: Choropleth and bivariate maps and classification

Cartographic symbolization

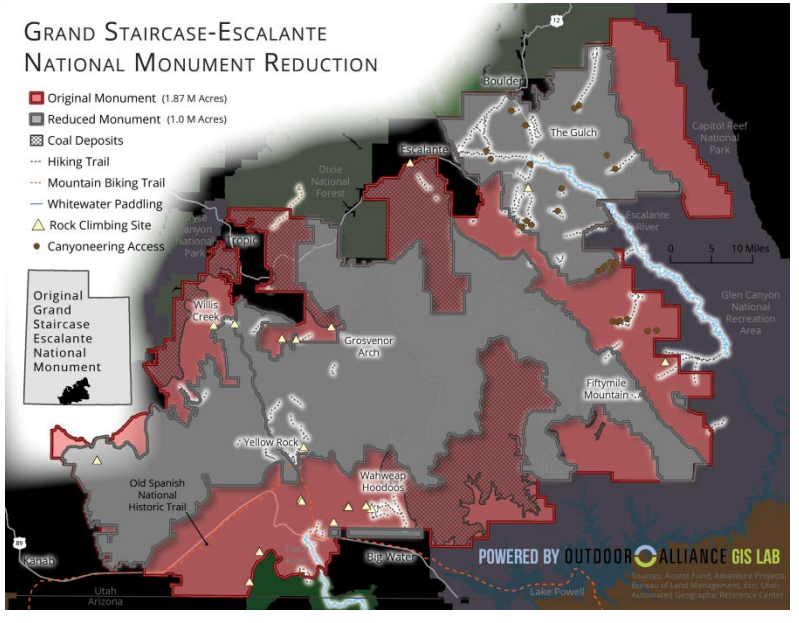


What is symbolization?

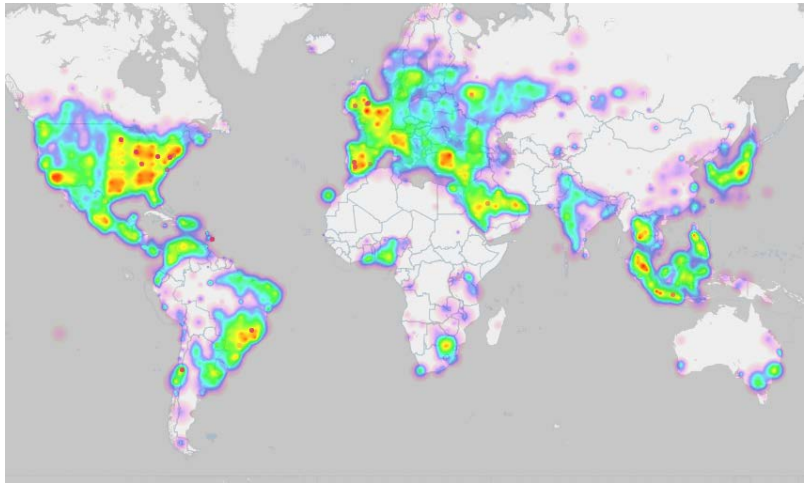
- Assumes map objects are thematic, and purpose of map is related to point, line, area, or volume objects
- Assumes choice of scale and granularity
- Assumes choice of continuity model
- Assumes choice of data level (adds cyclical, counts, fuzzy, bipolar)
- Goal is to use data to visualize phenomenon
- Make choices of cartographic method (e.g. choropleth, proportional symbol, isopleth, dot)
- Apply and adjust visual variables

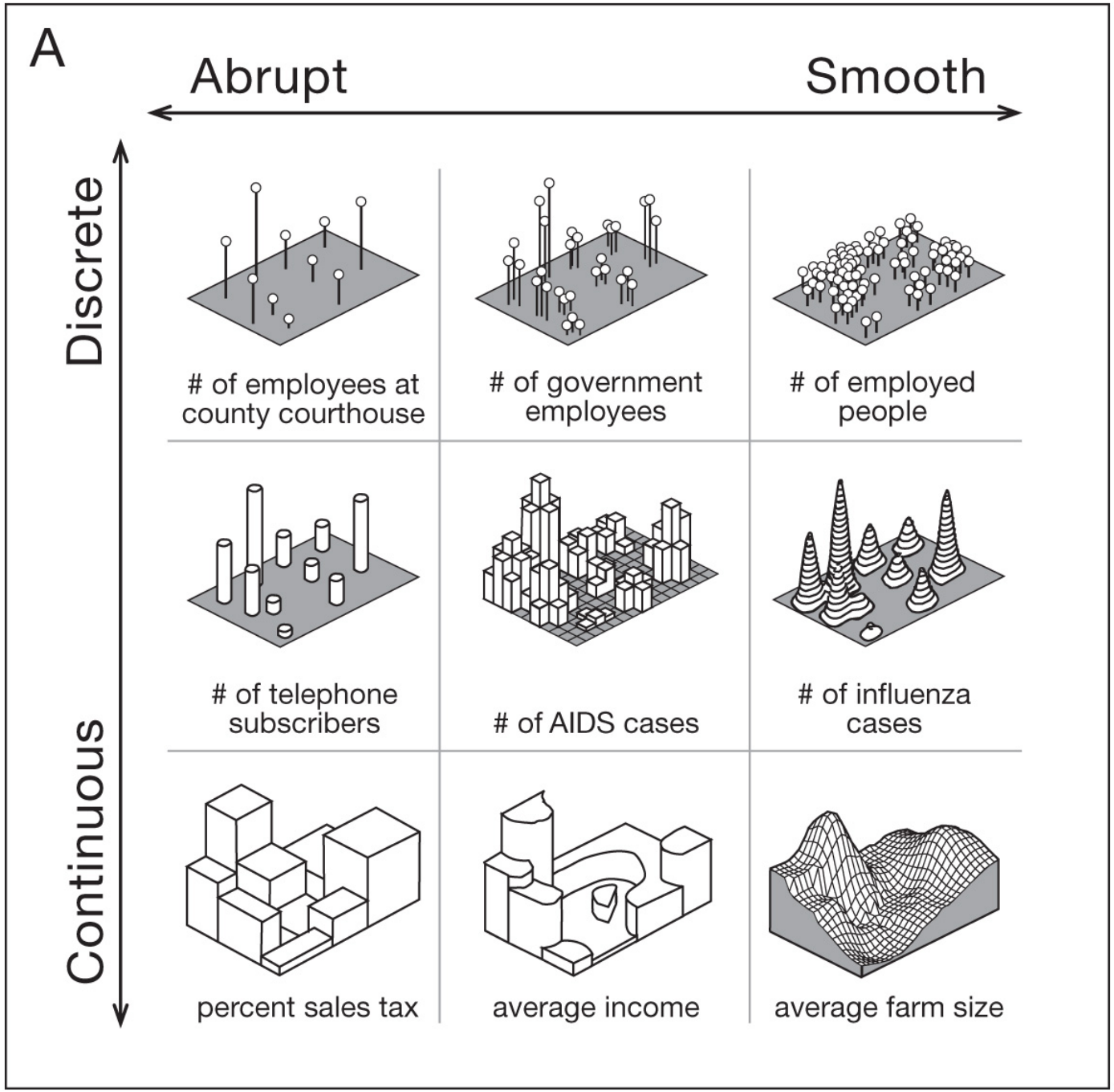


Fuzzy Features

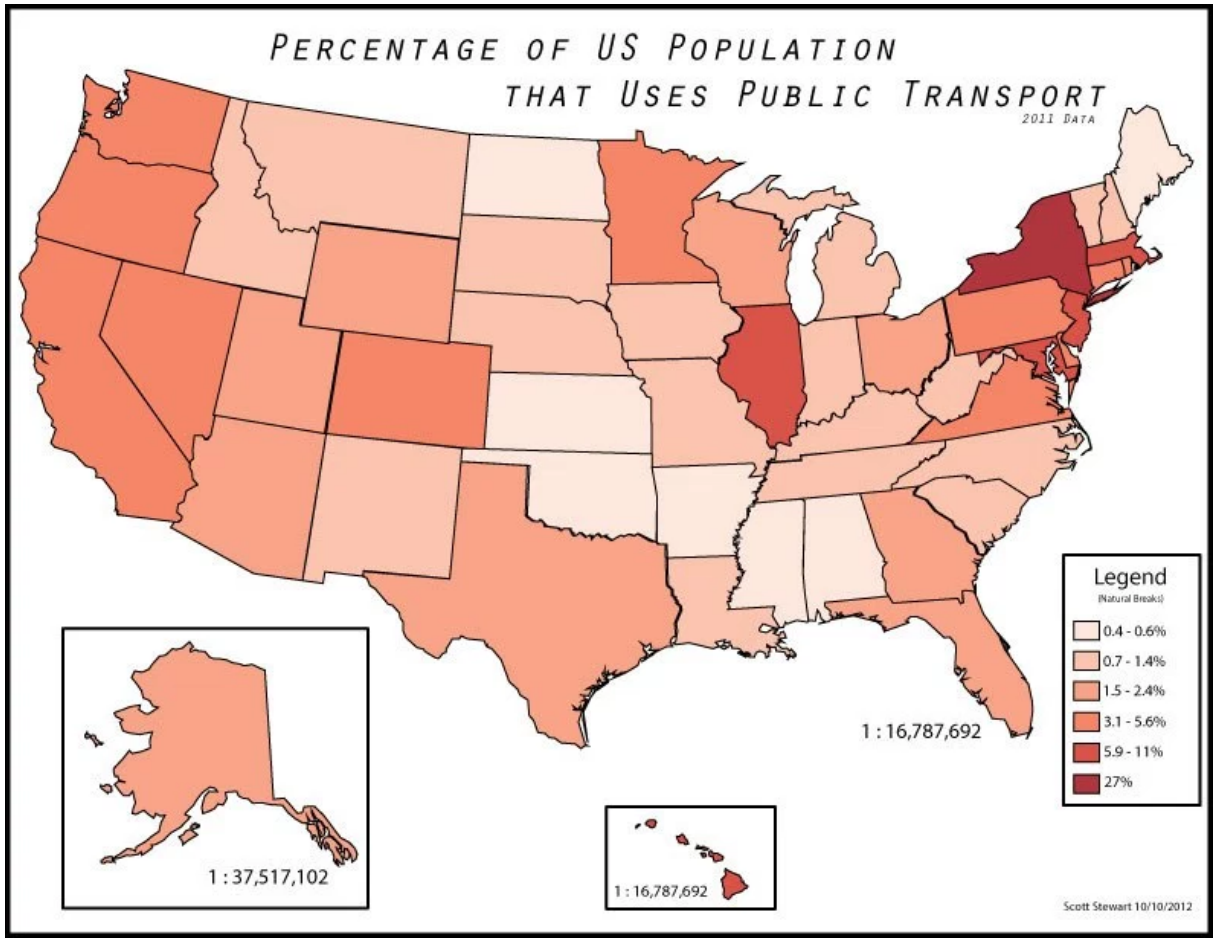


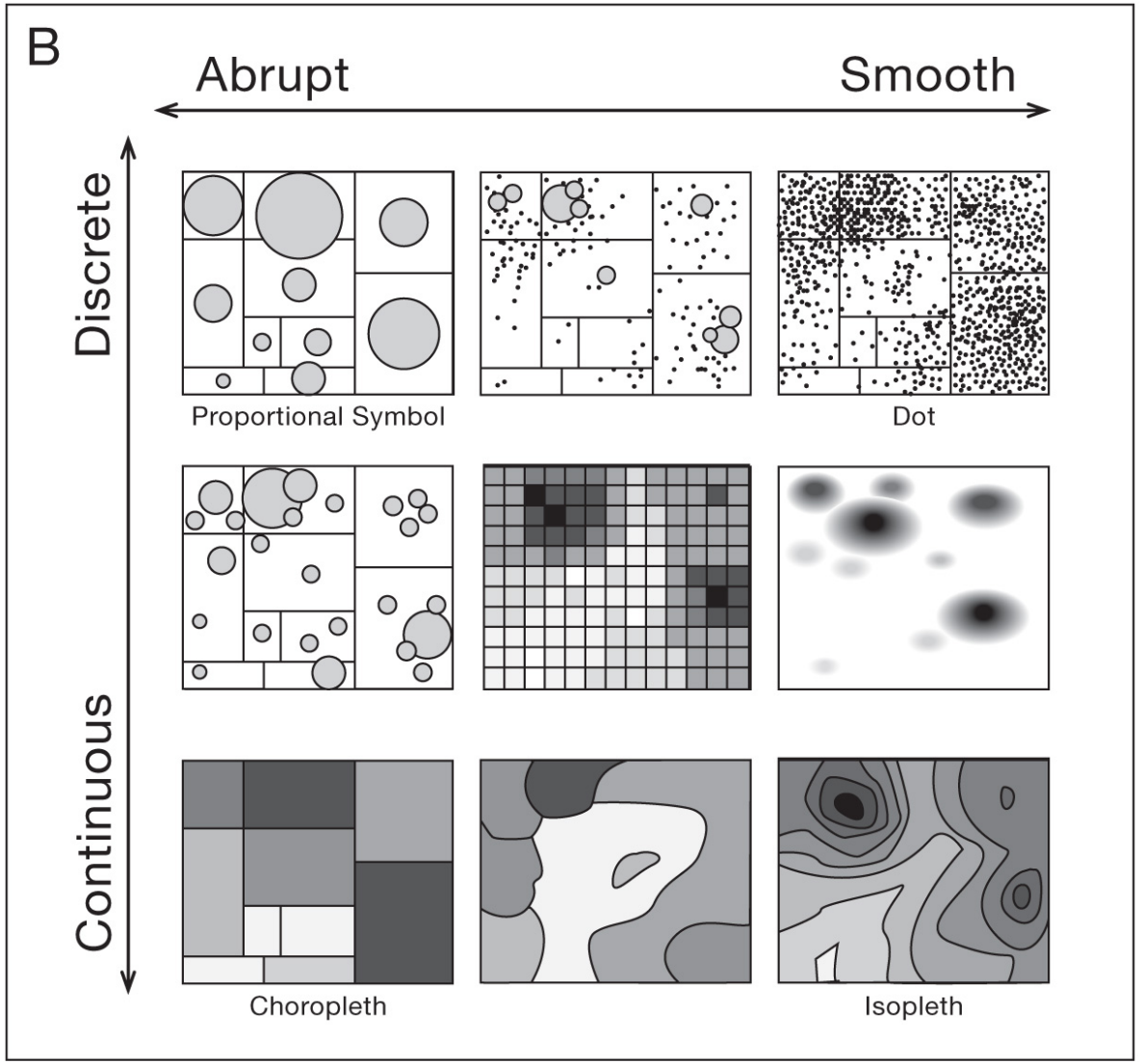
Continuity

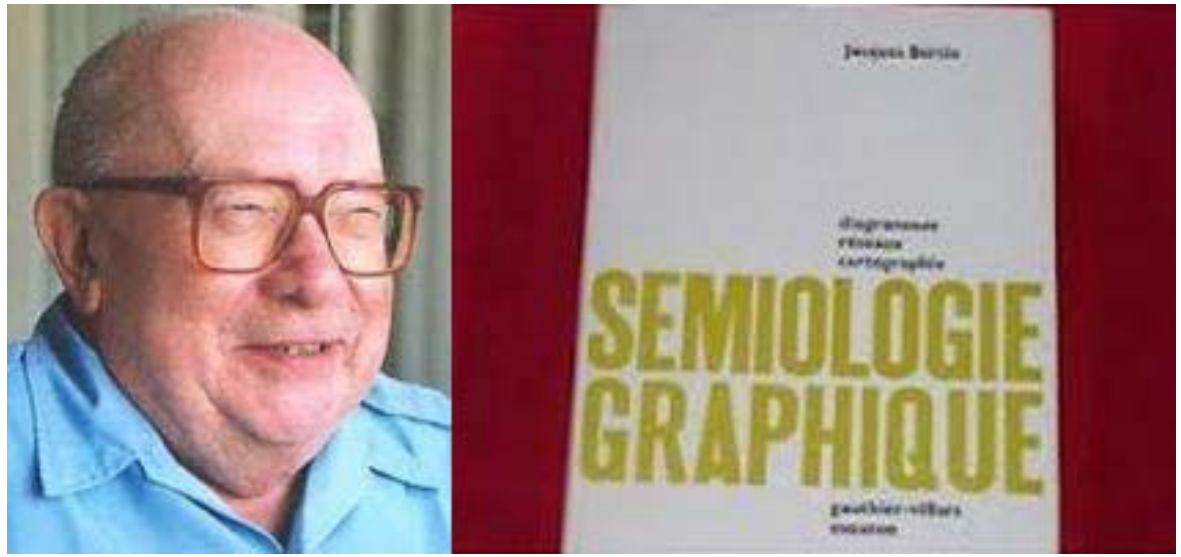




Dots, proportional symbols, isopleths, choropleths



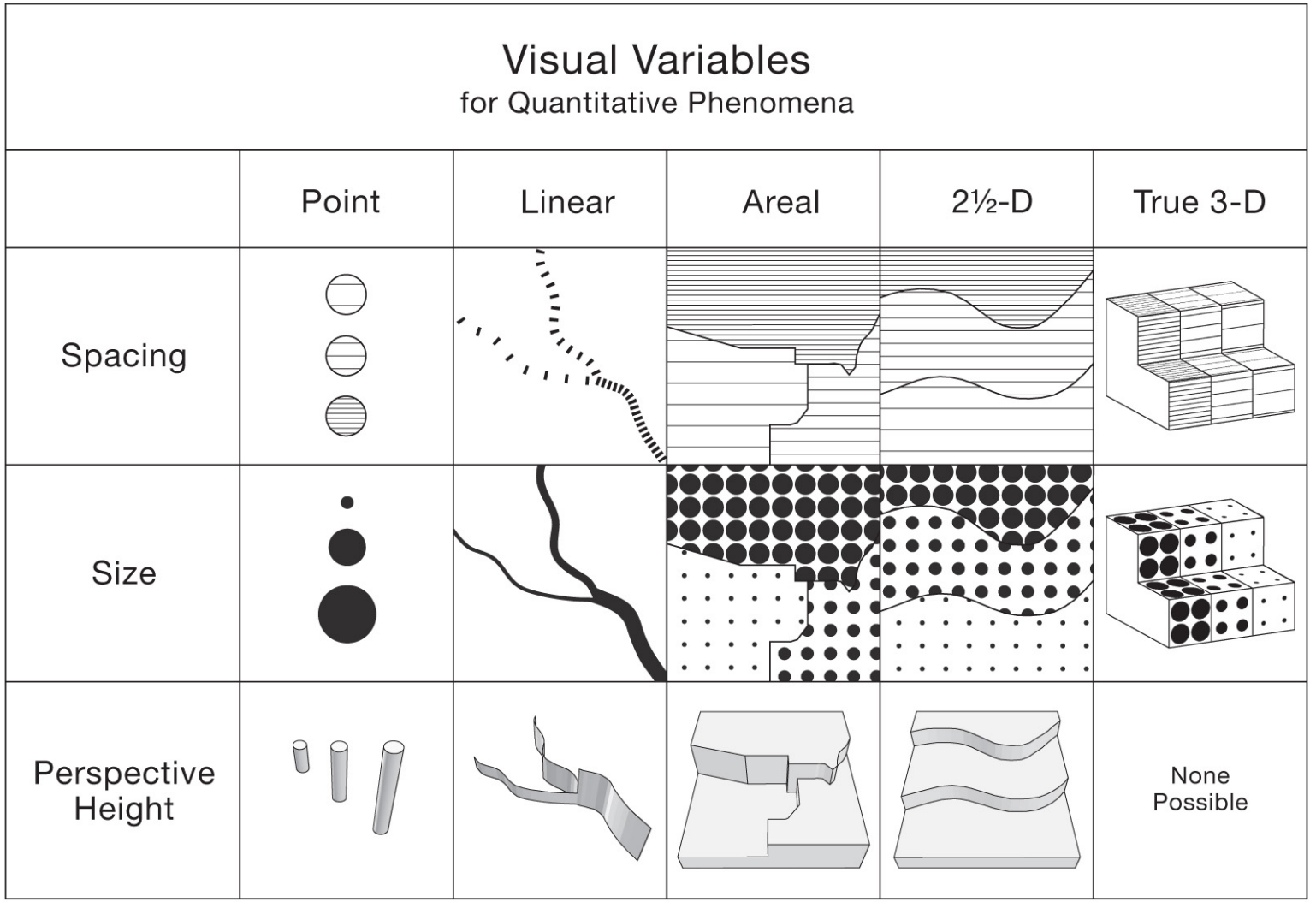




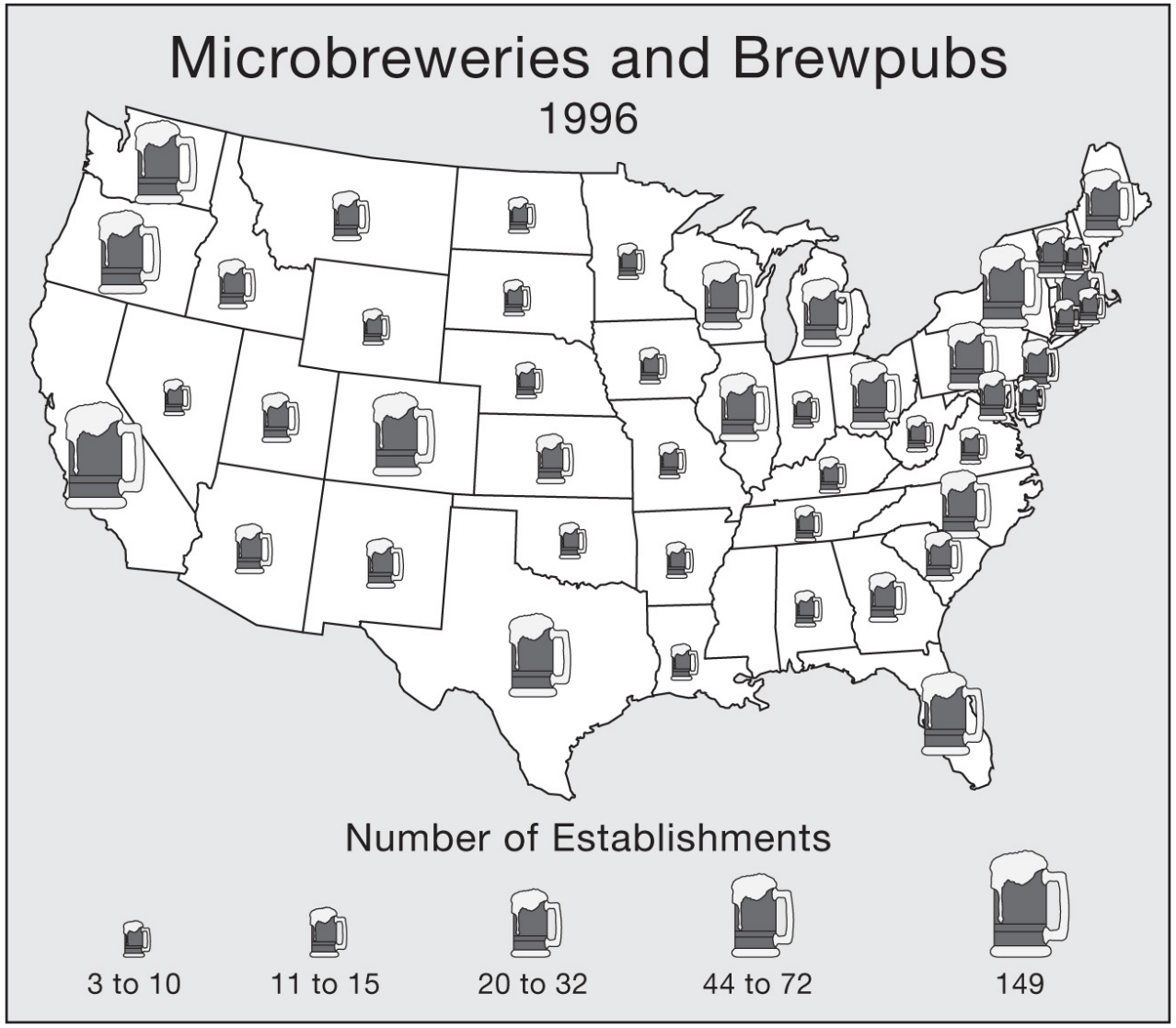
Visual Variables
for Quantitative Phenomena

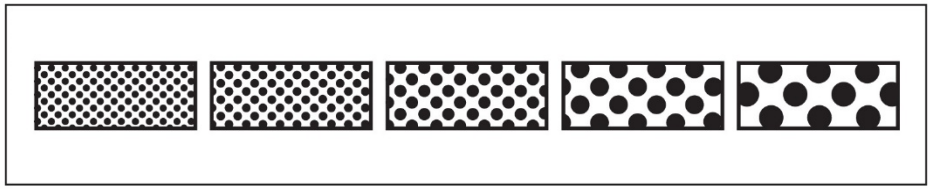
	Point	Linear	Areal	2½-D	True 3-D
Spacing					
Size					
Perspective Height					None Possible
Color (Hue)	Yellow, Orange, Red	Yellow, Orange, Red	Red, Yellow, Orange	Red, Orange, Yellow	Red, Orange, Yellow
Color (Lightness)	Light Green, Green, Dark Green	Light Green, Green, Dark Grn	Dark Green, Light Green, Green	Dark Green, Green, Light Green	Dark Green, Green, Light Green
Color (Saturation)	Muted Red, Red, Vivid Red	Muted Red, Red, Vivid Red	Vivid Red, Muted Red, Red	Vivid Red, Red, Muted Red	Vivid Red, Red, Muted Red

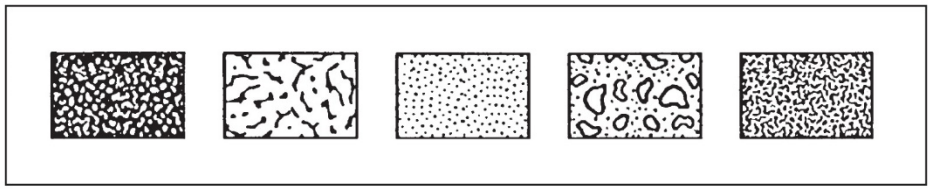
Bertin's Visual Variables



Pattern: Combining visual variables

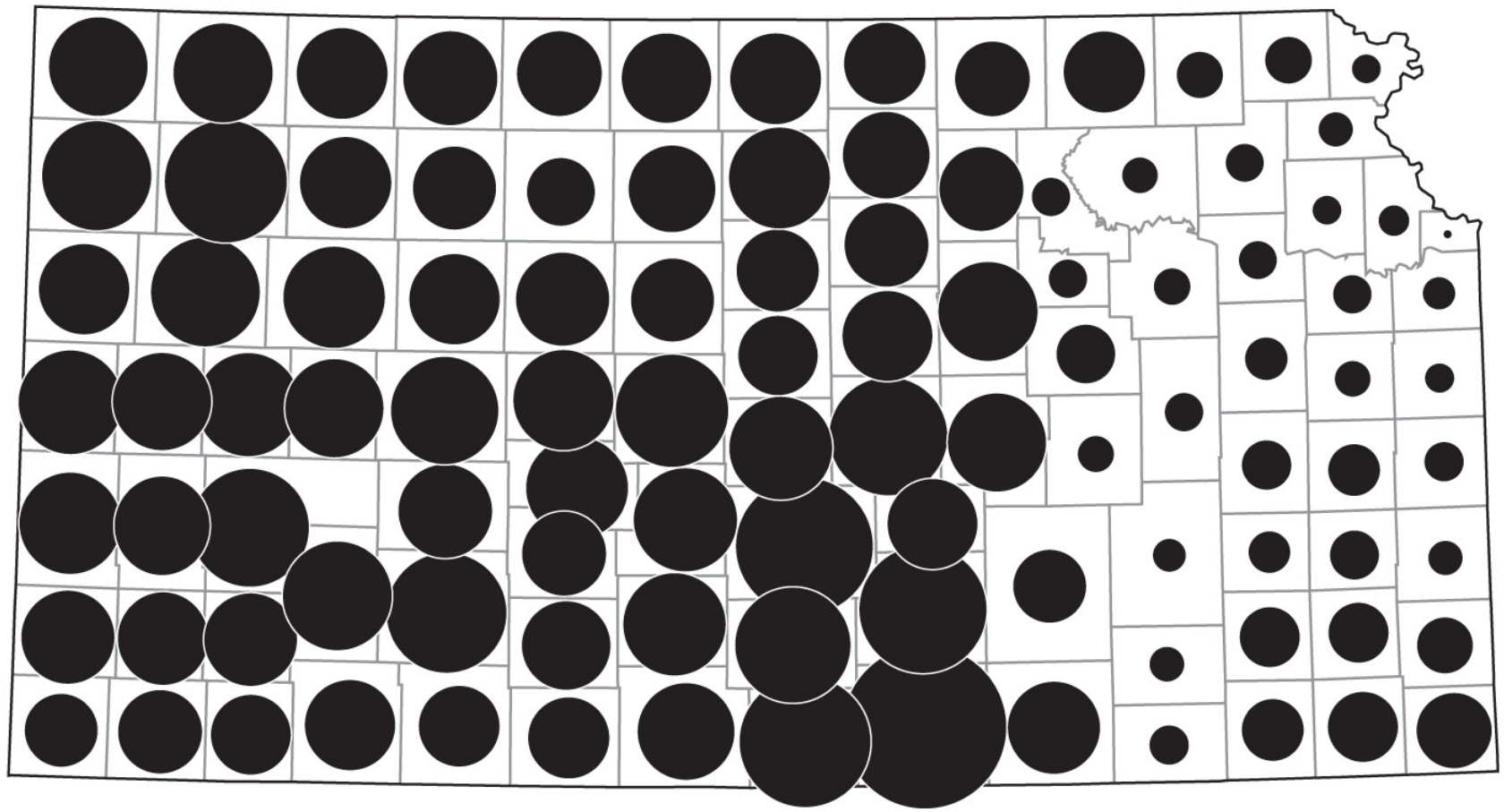






C

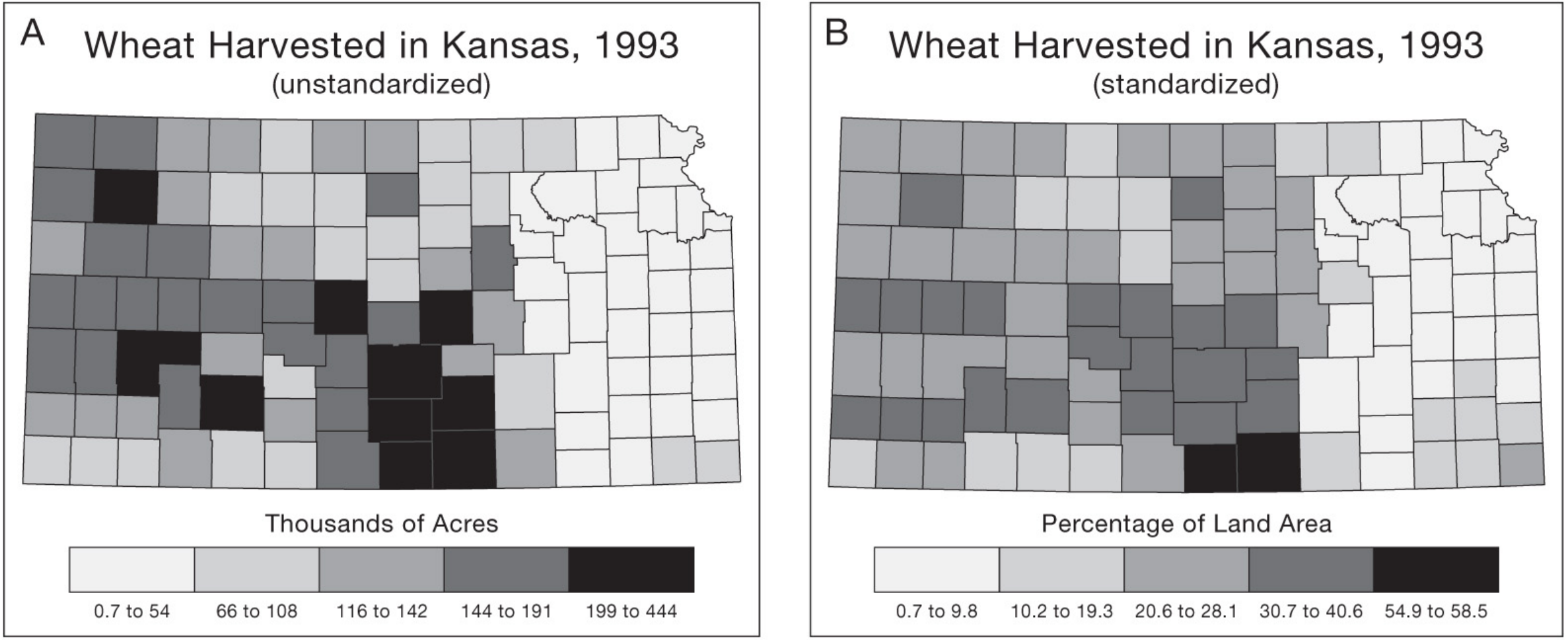
Proportional Symbol



Thousands of Acres

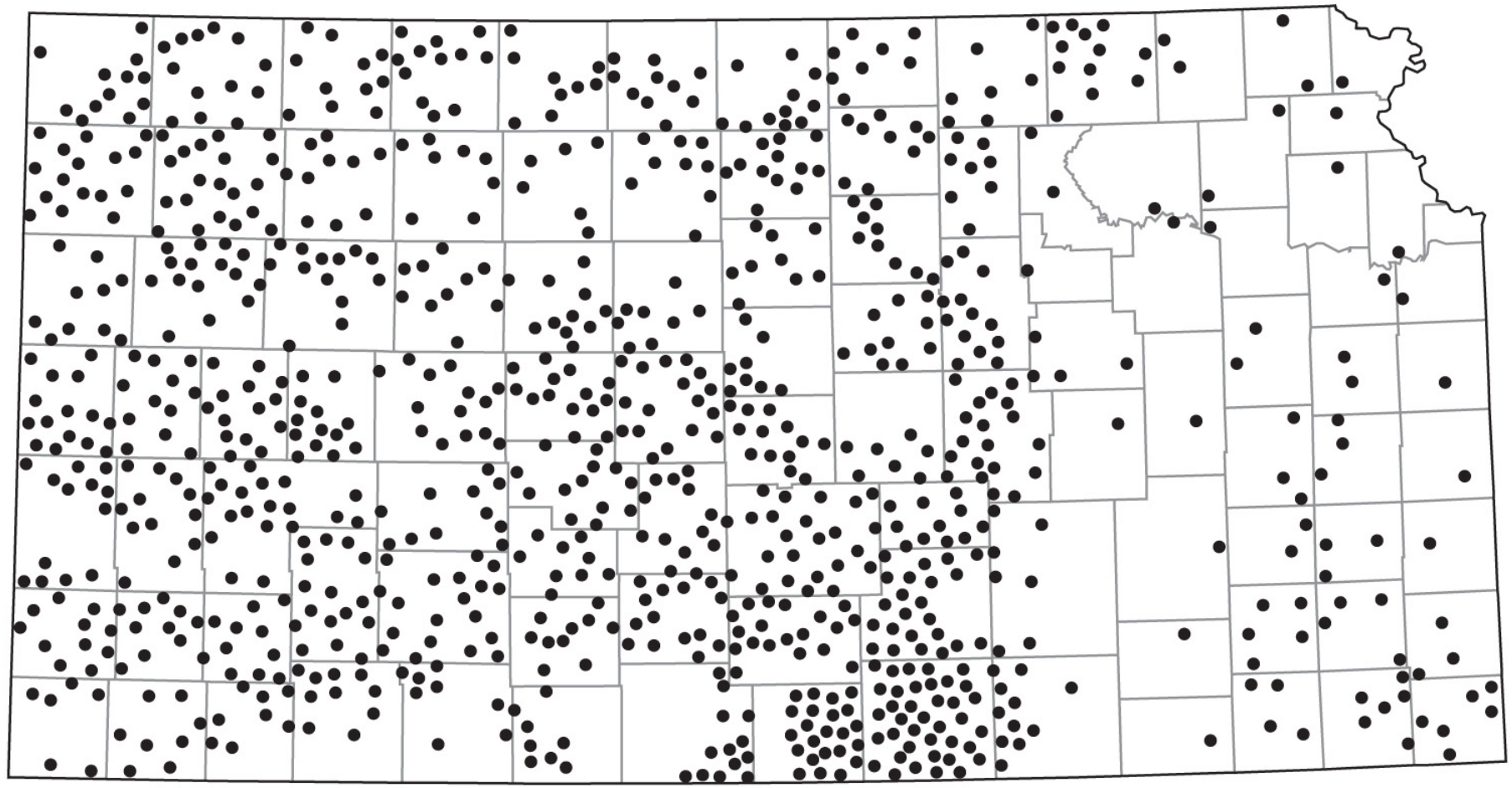
• 0.7 222.2 443.7

Choropleth



D

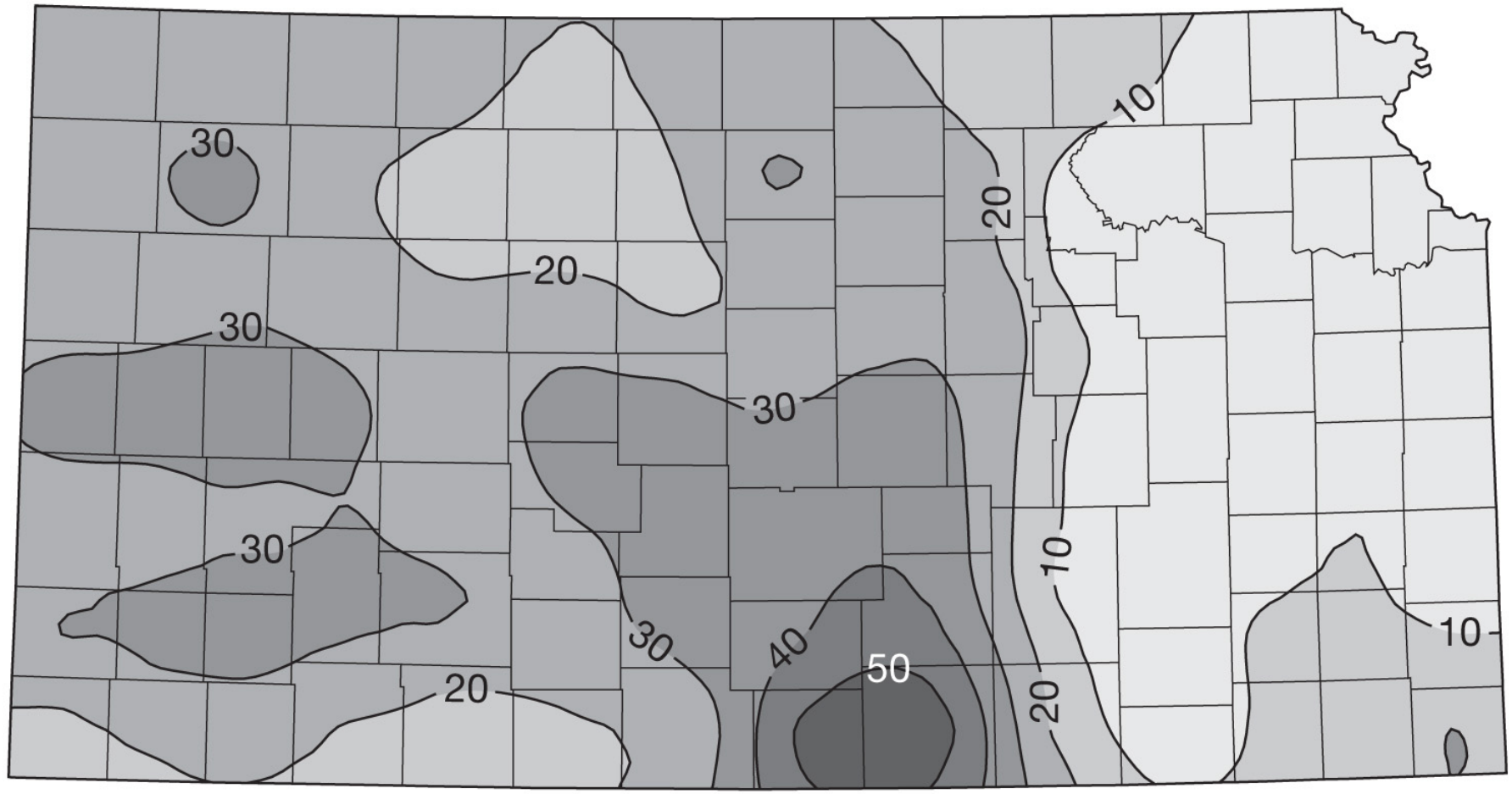
Dot



Each Dot Represents
12,000 Acres of Wheat

B

Isopleth



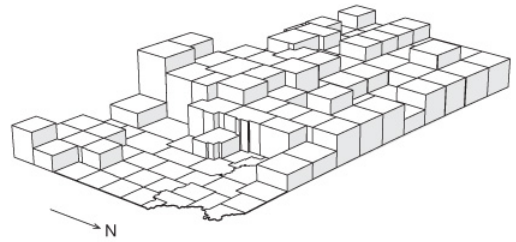
Contour Lines Represent
Percent of Land Area

Visual Variables

Wheat Harvested in Kansas, 1993

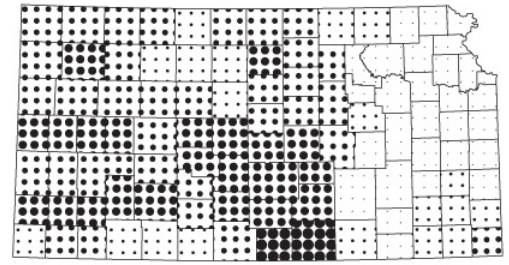
A

Perspective Height



B

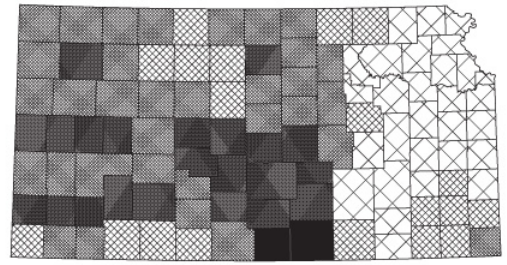
Size



Percentage of Land Area

C

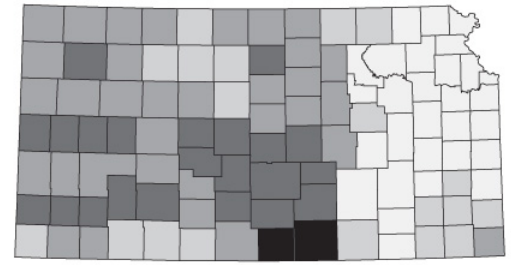
Spacing



Percentage of Land Area

D

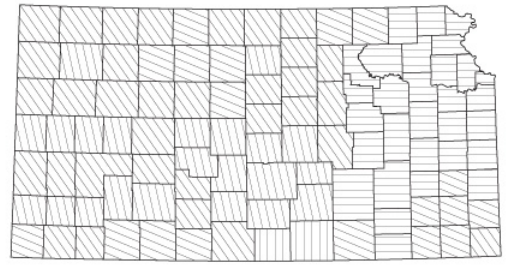
Lightness



Percentage of Land Area

E

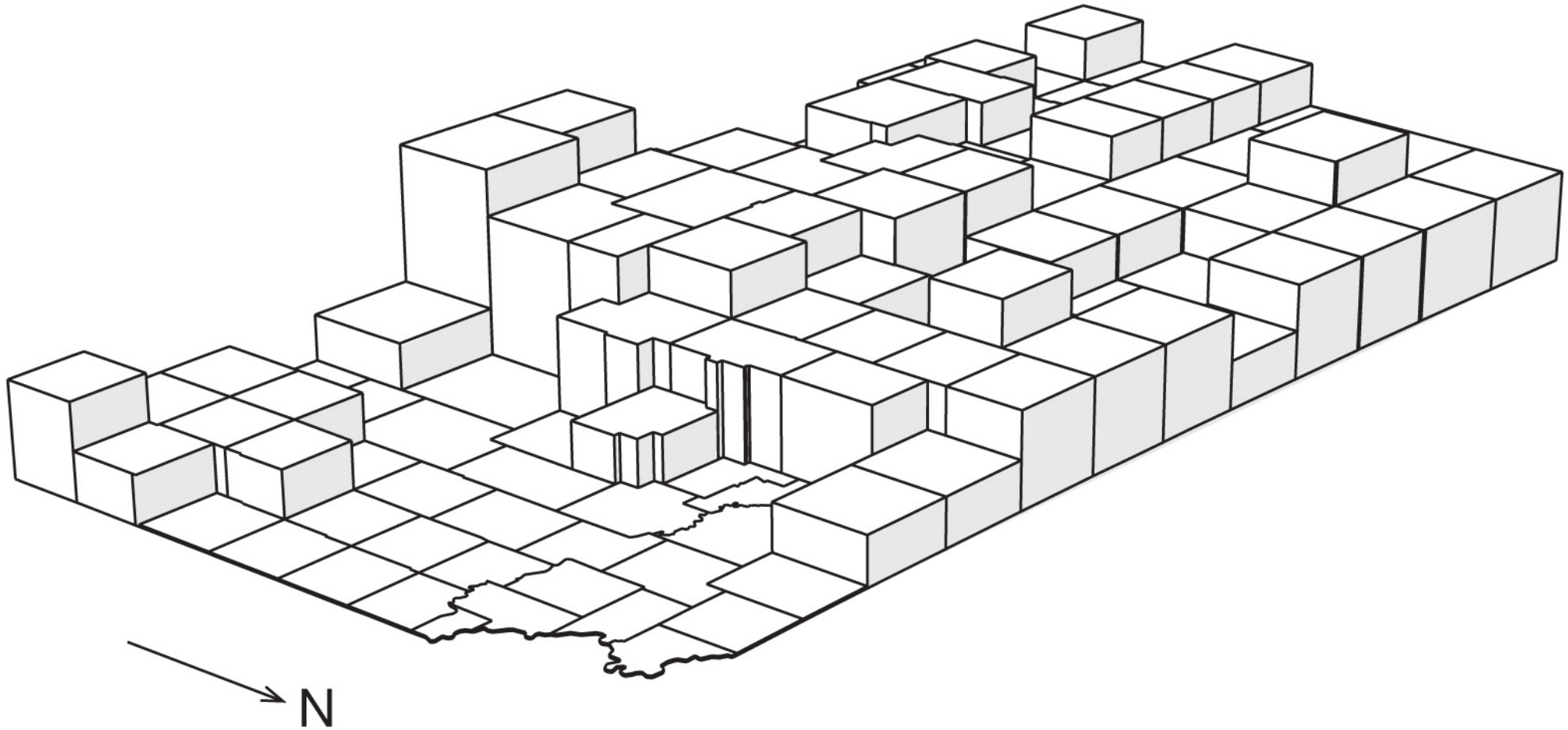
Orientation



Percentage of Land Area

A

Perspective Height



Slocum's Symbolization conclusion

	Nominal	Ordinal	Numerical
Spacing	P	M[c]	M[c]
Size	P	M	M
Perspective Height	P	M[a]	G[b]
Orientation	G	P	P
Shape	G	P	P
Arrangement	G	P	P
Lightness	P	G	M
Hue	G	G[d]	M[d]
Saturation	P	M	M

P = Poor M = Marginally Effective G = Good

[a] Since height differences are suggestive of numerical differences, use with caution for ordinal data.

[b] Hidden enumeration units and lack of a north orientation are problems.

[c] Not aesthetically pleasing.

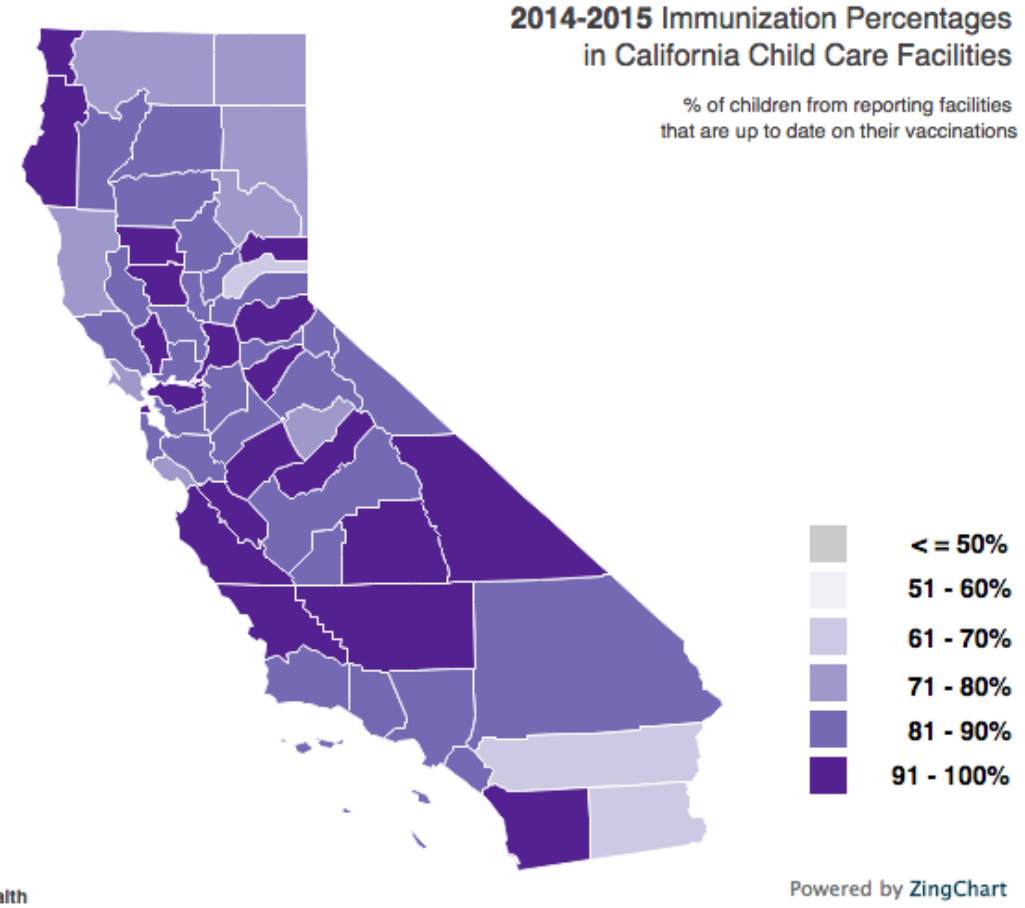
[d] The particular hues selected must be carefully ordered, such as yellow, orange, red.

Next, choropleth in detail

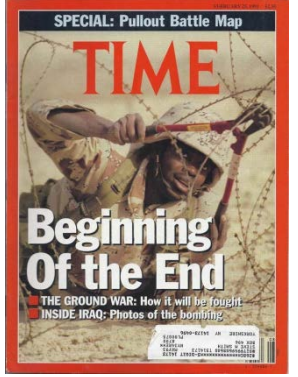
Greek χώρο ("area/region") + πλήθος ("multitude")

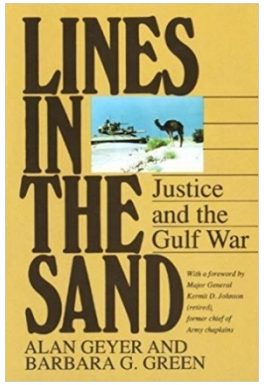
Key factors

- Resolution of base map
- Data
  - Source and processing
  - Classification
  - MAUP
  - Legend
- Symbolization
- Figure/ground and details

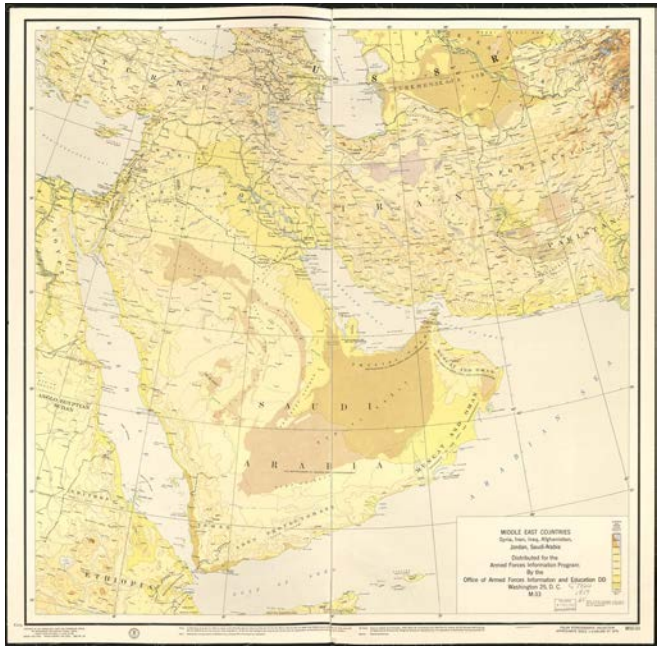


Units or “Geography”

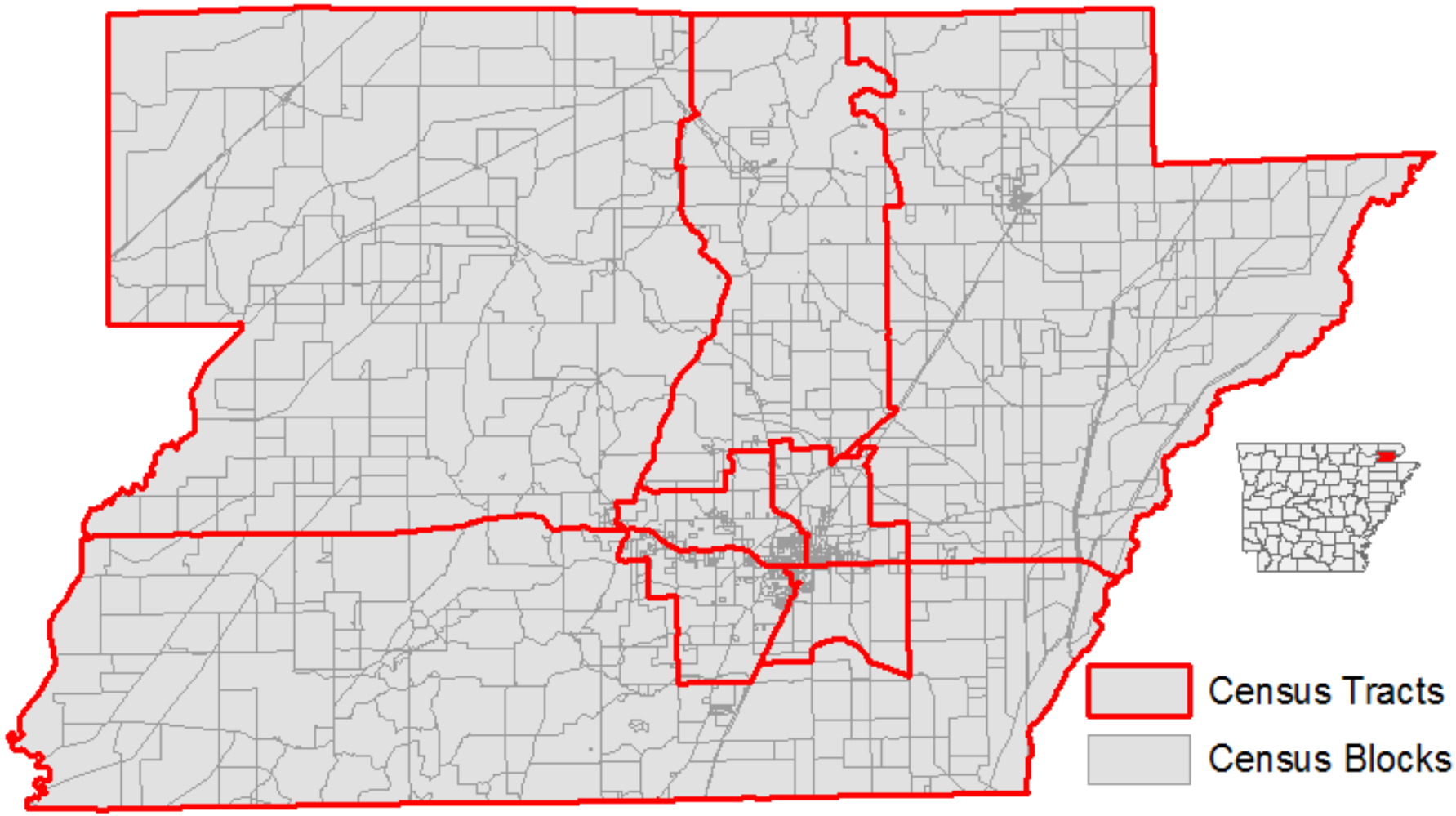




- Do the data collection units have meaning, or are they simply a sampling frame?
- Are the boundaries crisp, continuous or fuzzy?
  - Census divisions
  - Zip codes
  - Neighborhoods
  - Counties
  - States
  - Nations



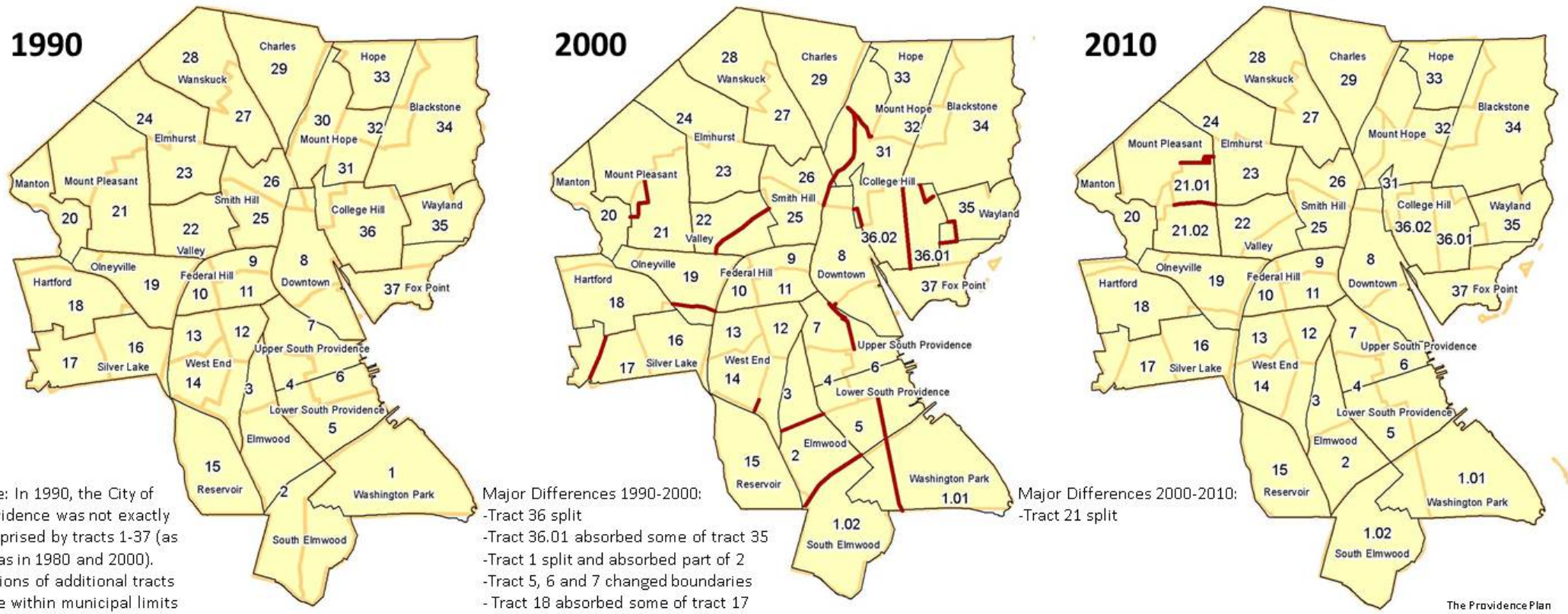
Census Blocks and Census Tracts
Greene County, Arkansas



Census units

Change in Census Tract Boundaries 1990-2010

Providence, RI Neighborhoods



Note: In 1990, the City of Providence was not exactly comprised by tracts 1-37 (as it was in 1980 and 2000). Portions of additional tracts were within municipal limits and portions of 1-37 were actually within other municipalities. 1990 tracts 1-37 are shown for simplicity.

Major Differences 1990-2000:
-Tract 36 split
-Tract 36.01 absorbed some of tract 35
-Tract 1 split and absorbed part of 2
-Tract 5, 6 and 7 changed boundaries
- Tract 18 absorbed some of tract 17
- Tract 20 absorbed some of tract 21
- Tract 25 absorbed some of tract 22
- Tract 31 absorbed some of tract 26
- Tract 30 no longer exists- absorbed by tracts 29, 31 and 33

Major Differences 2000-2010:
-Tract 21 split

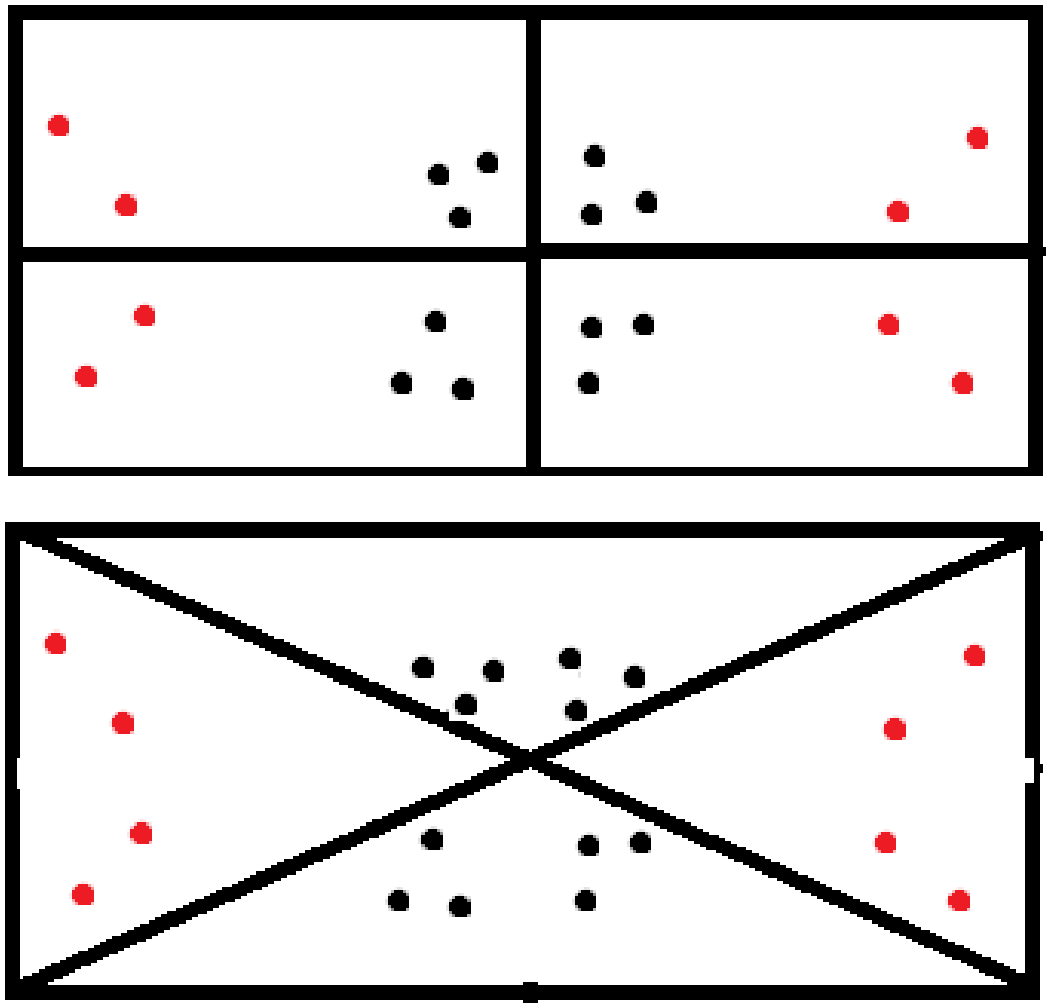
The Providence Plan
10 Davol Square, Suite 300
Providence, RI 02903
www.provplan.org

ProvPlan

MAUP

- Modifiable area unit problem
- Also called the ecological fallacy
- Source of statistical bias that can radically affect the results of statistical hypothesis tests
- Results when measures of spatial phenomena (e.g. population density) are aggregated into districts, usually for counting
- Openshaw (1983) "the areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating."

MAUP in a nutshell

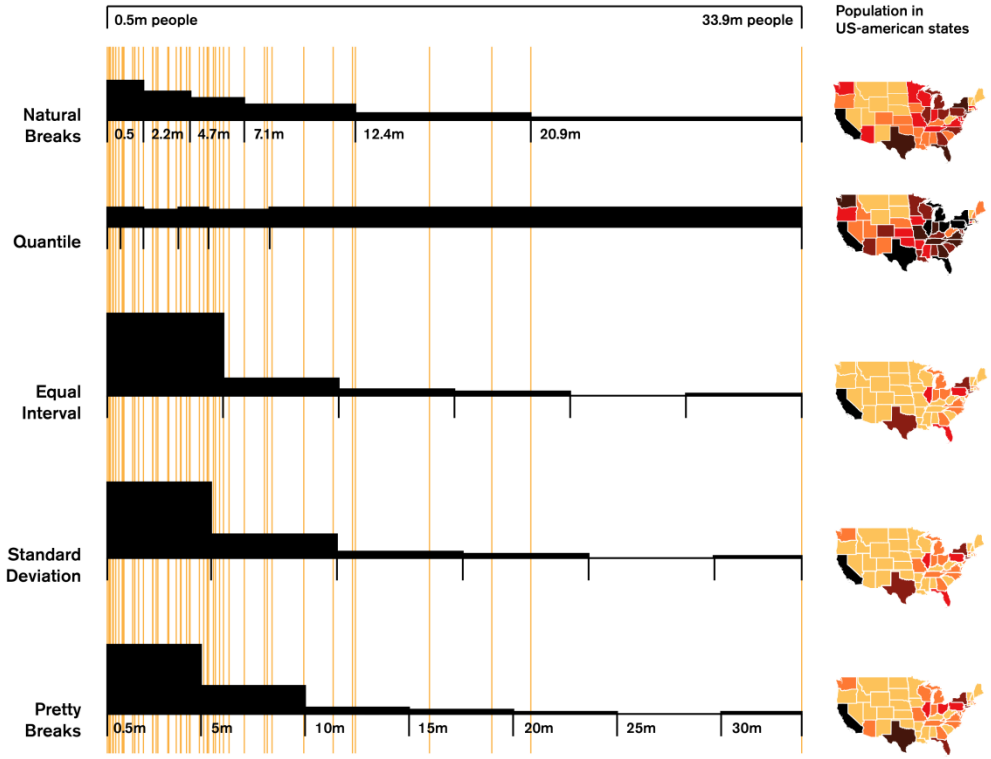


Uniform

Spread bias

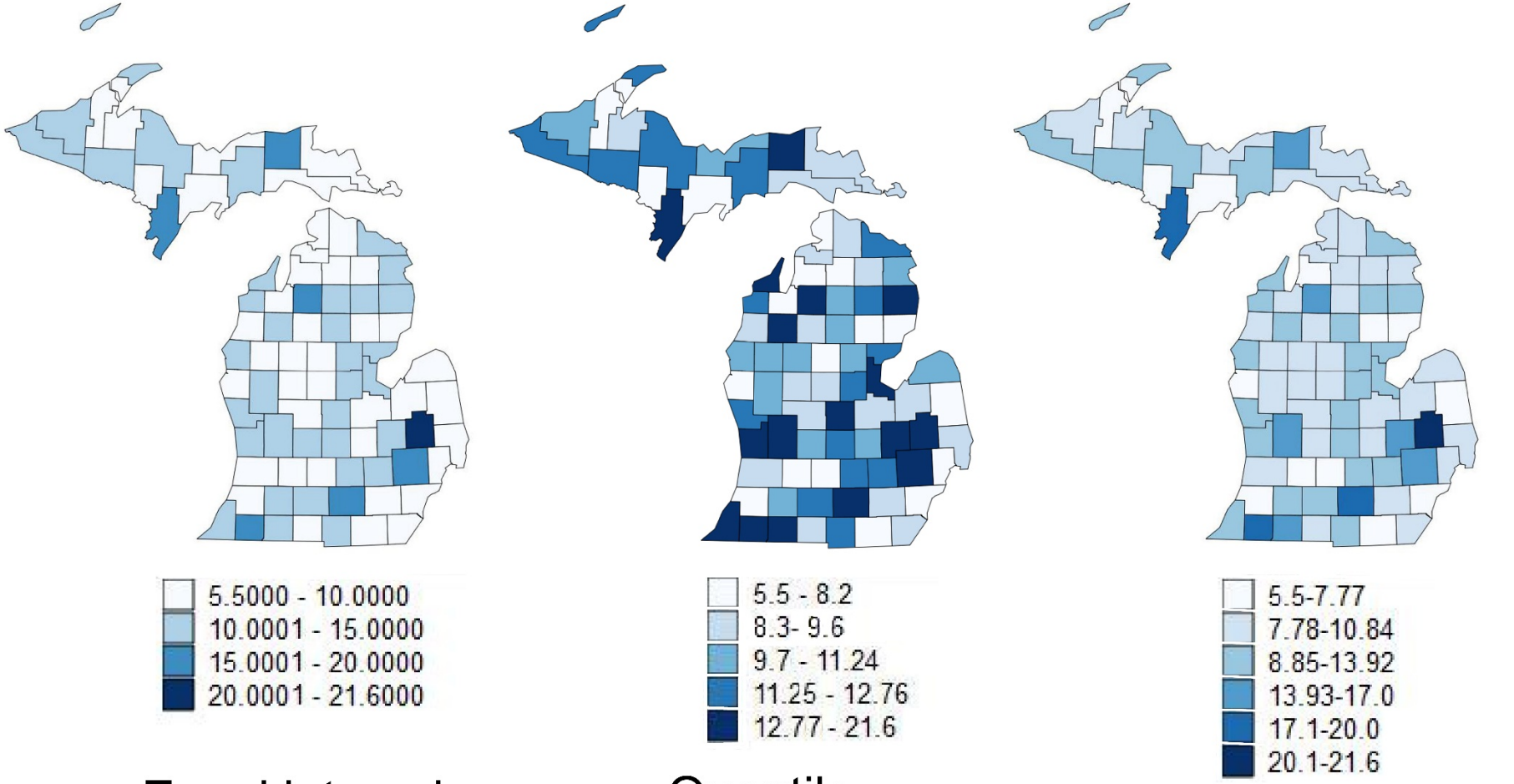
Methods for classification

- How many classes (vs. none)
- Eyeball, interactive -> Histogram
- Natural breaks
- Equal intervals
- Unequal i.e. non-linear
- Quantiles
- Standard deviations
- Jenks-Caspall analytical method



For example

Michigan Percent Unemployment: March 2012

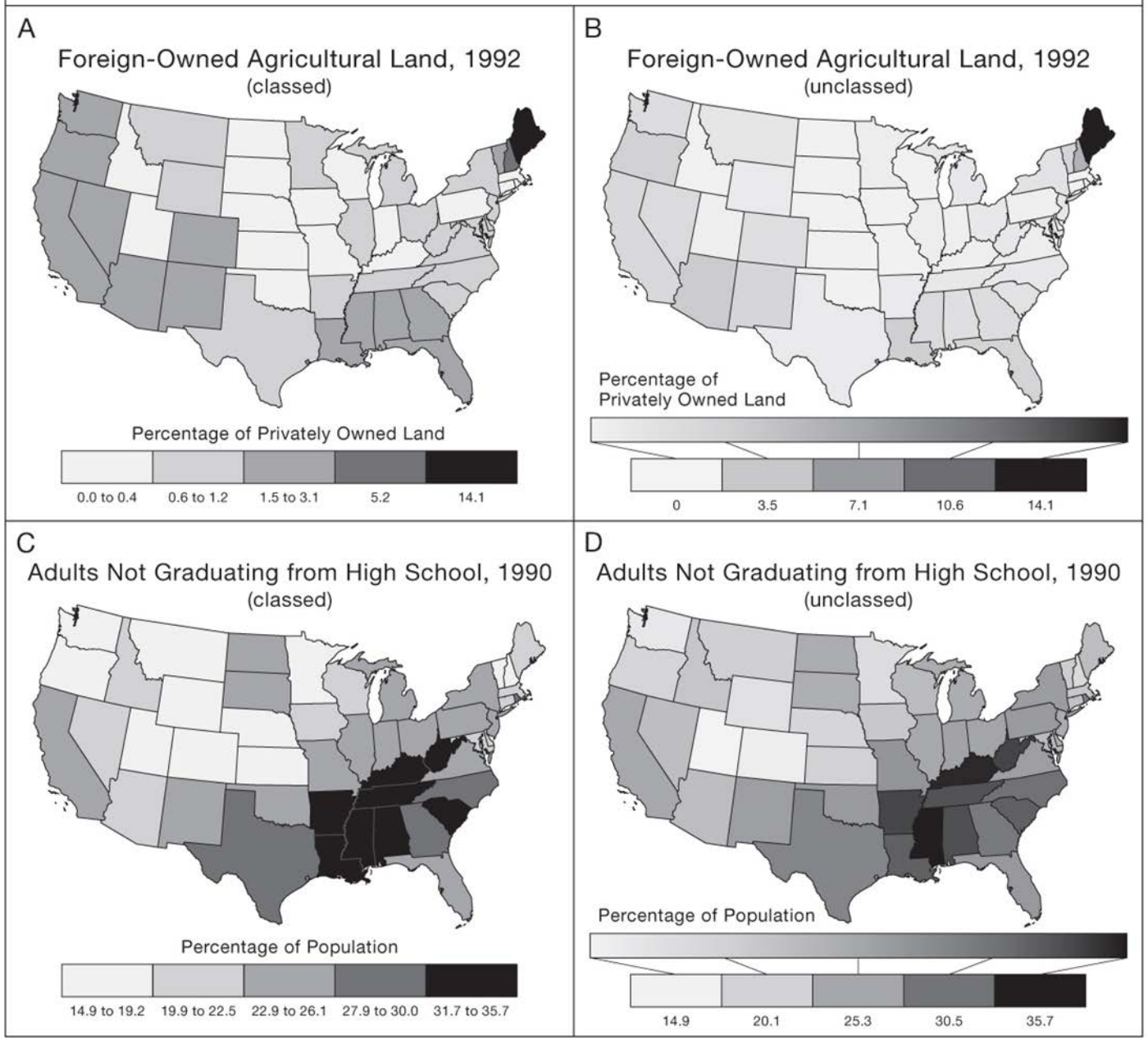


Equal interval
4 classes

Quantile
5 classes

Standard Deviations
6 classes

Classed Versus Unclassed Maps



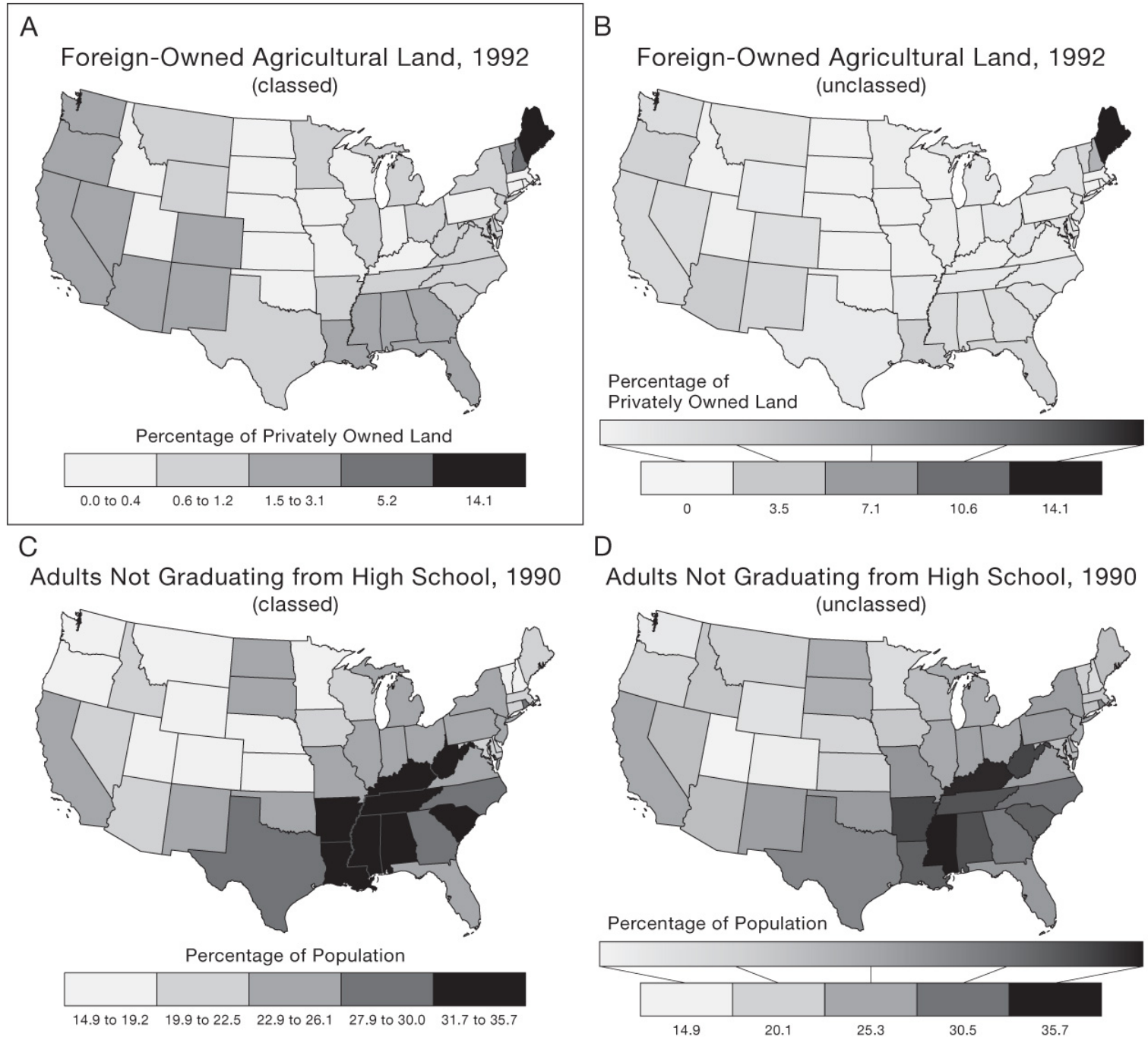
Choropleth Maps Without Class Intervals?

W. R. Tobler

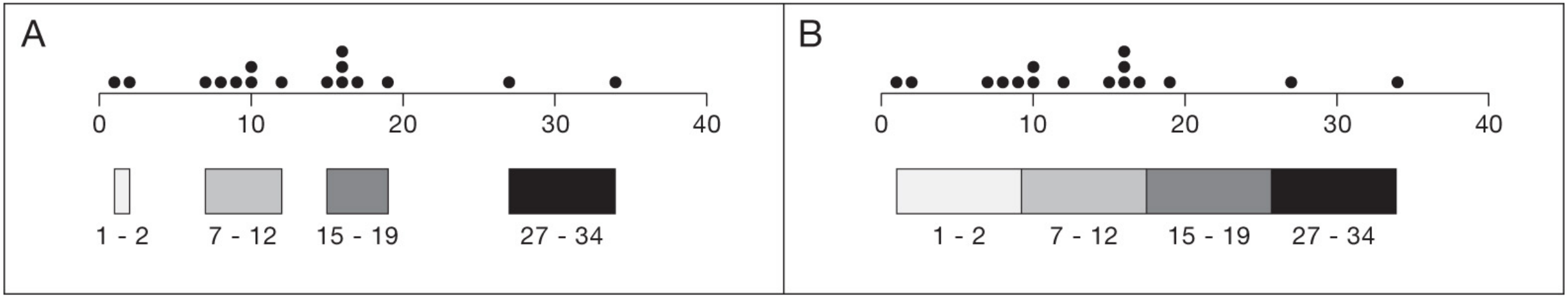
It is now technologically feasible to produce virtually continuous shades of grey by using automatic map drawing equipment. It is therefore no longer necessary for the cartographer to "quantize" data by combining values into class intervals. As a simple illustration an automatic line plotter can be programmed to draw lines virtually any distance apart (Fig. 1). Thus, one can obtain any desired destiny of inked area to white area. For example, if the geographical data, symbolized by z, are normalized to lie in the range from zero to one, then an appropriate spacing of orthogonal lines of width w is given by

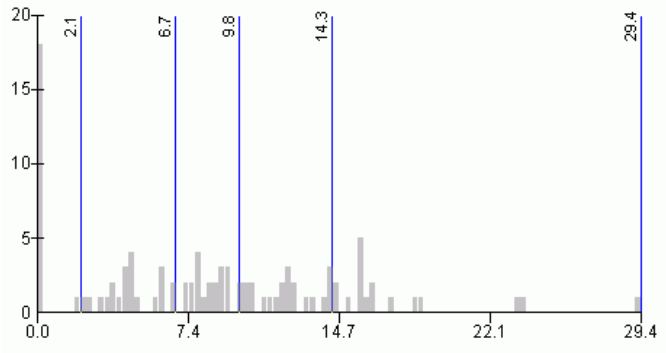
$$s = (w/z^x) \cdot [1 + (1 - z^x)^{1/2}].$$

Here an exponent ($x \approx 1.4$) of z has been chosen to approximate the nonlinear response of the human eye [*13*]. The units of the spacing s are those of w.



Histograms



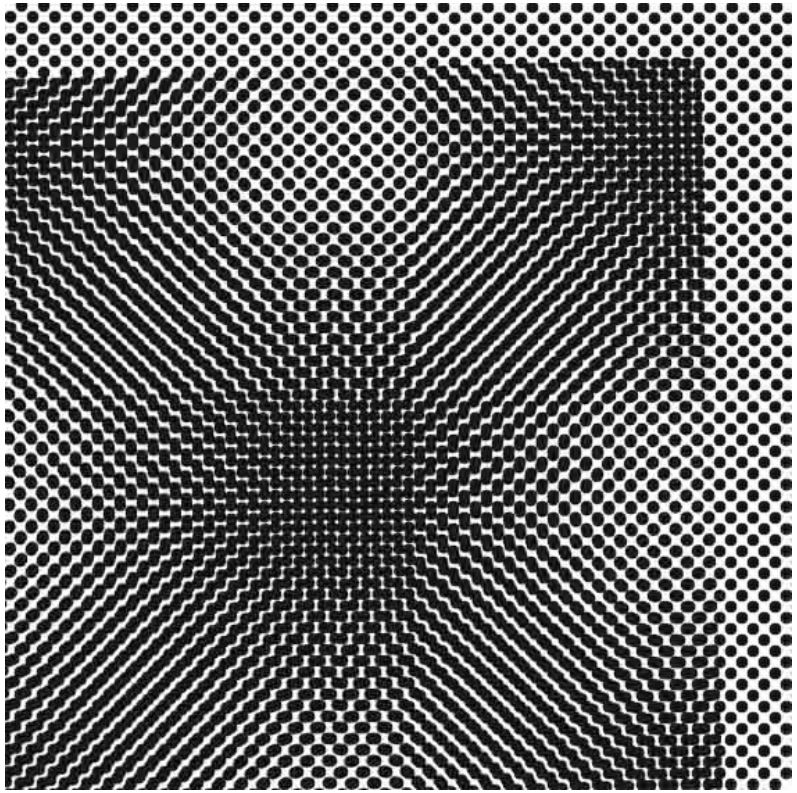


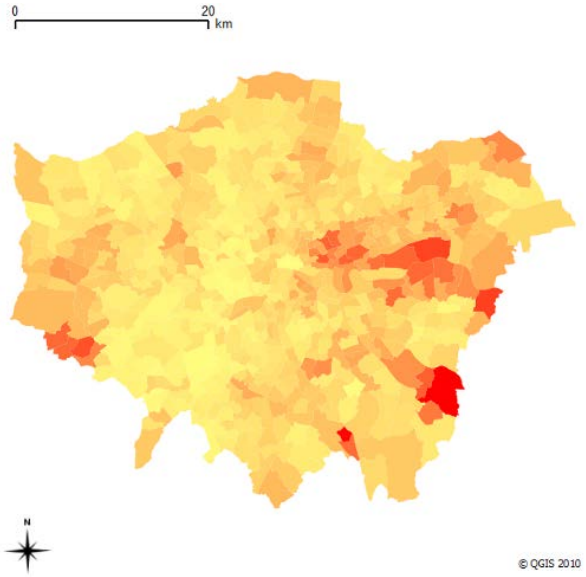
Jenks and Caspall (1971)

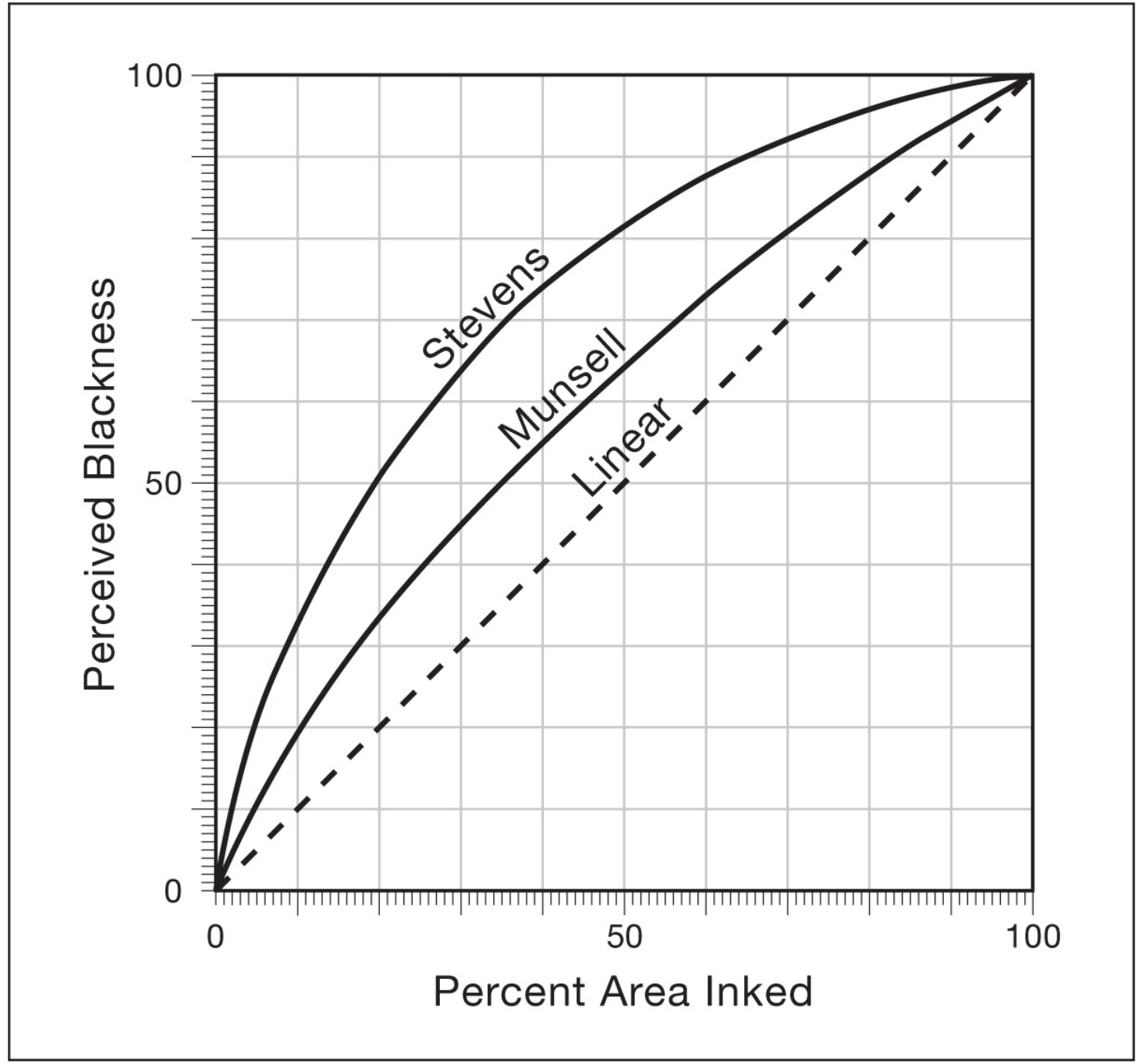
- Calculate the sum of squared deviations between classes (SDBC)
- Calculate the sum of squared deviations from the array mean (SDAM)
- Subtract the SDBC from the SDAM (SDAM-SDBC). This equals the sum of the squared deviations from the class means (SDCM)
- After inspecting each of the SDBC, a decision is made to move one unit from the class with the largest SDBC toward the class with the lowest SDBC
- New class deviations are then calculated, and the process is repeated until the sum of the within class deviations reaches a minimal value
- Final classes are uneven, but optimal

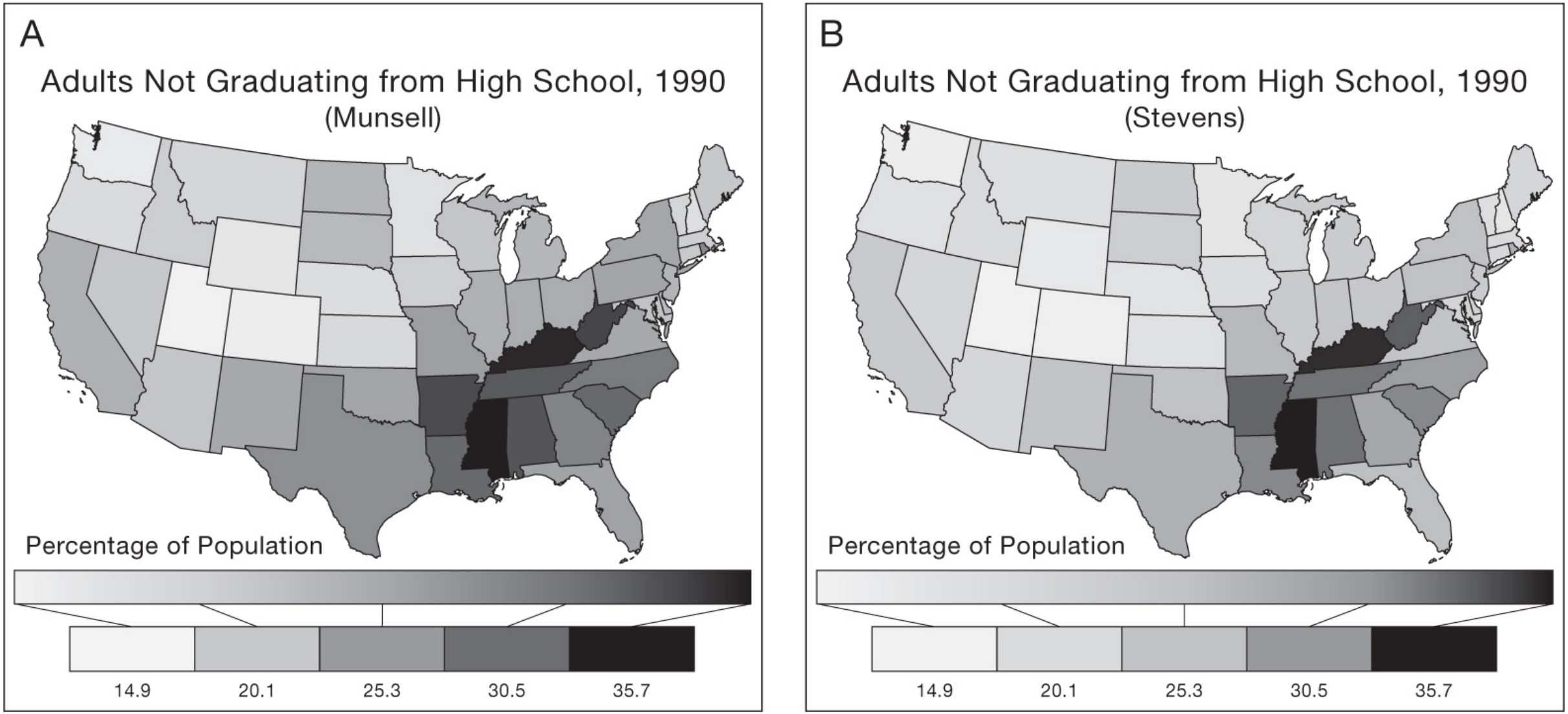
Symbolizing the choropleths

- Include areas?
- Patterns of lines (do not vary orientation)
- Dots (beware Moire patterns)
- Monochrome Shading
  - Darker is more
  - Vary density
  - Different schemas: Munsell vs. Stevens
- 2 color shading
- Color shading
  - Hue not a good variable, unless bipolar distribution
  - Use saturation or intensity

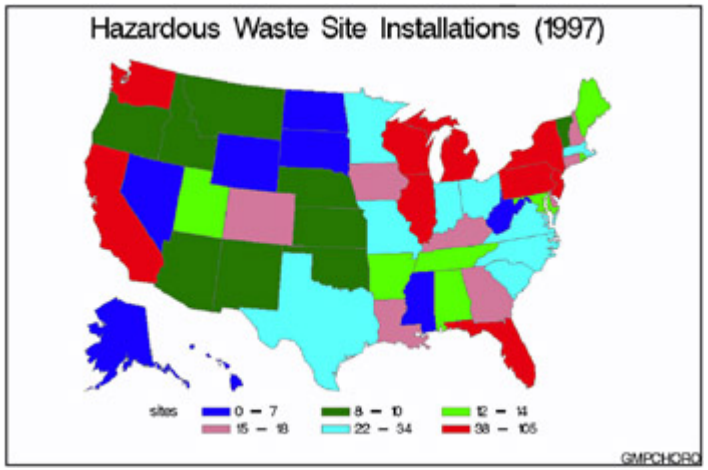


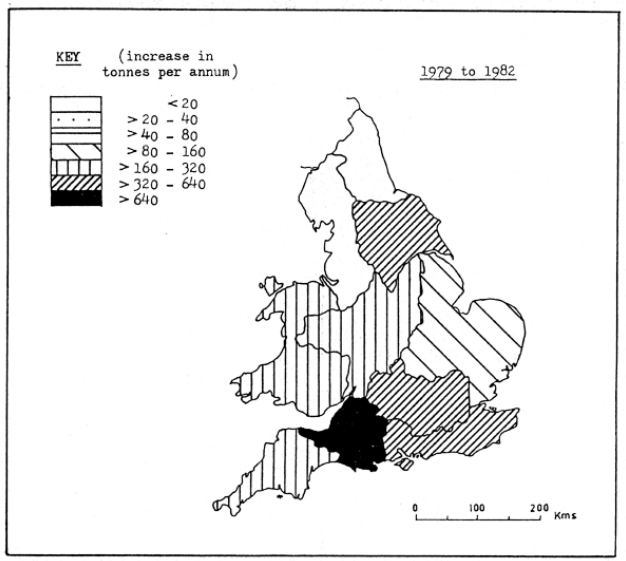


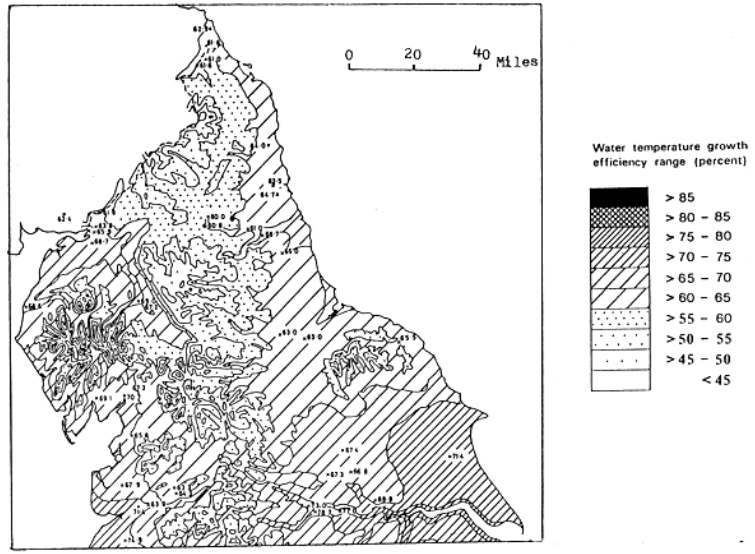


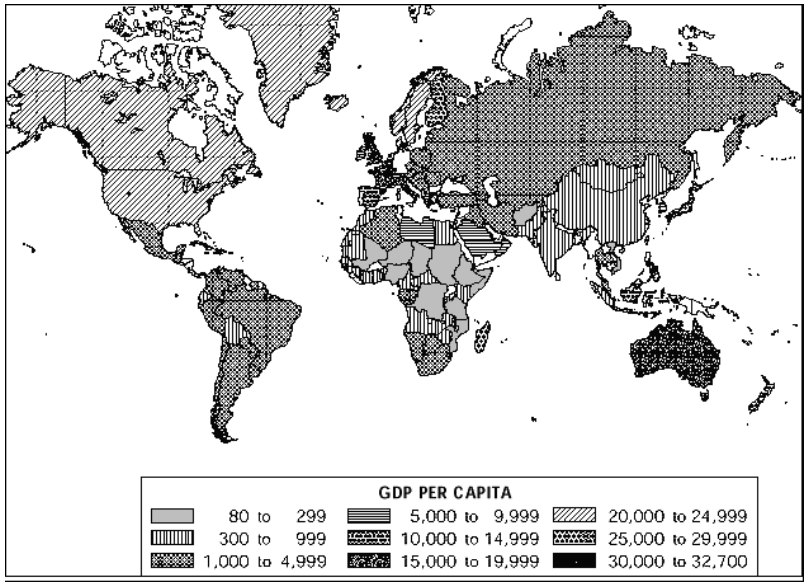


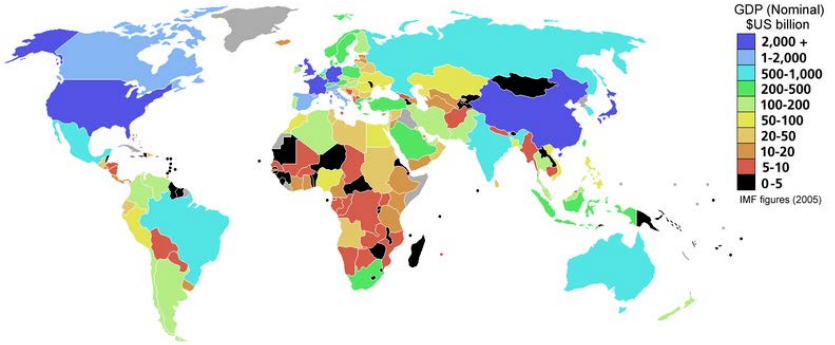
Hall of shame

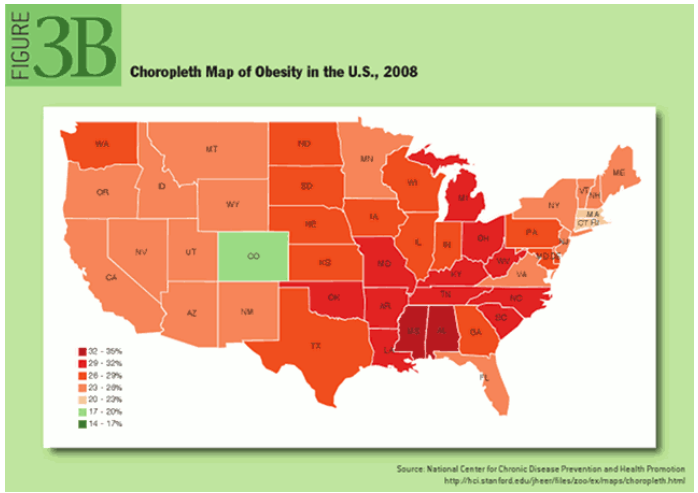




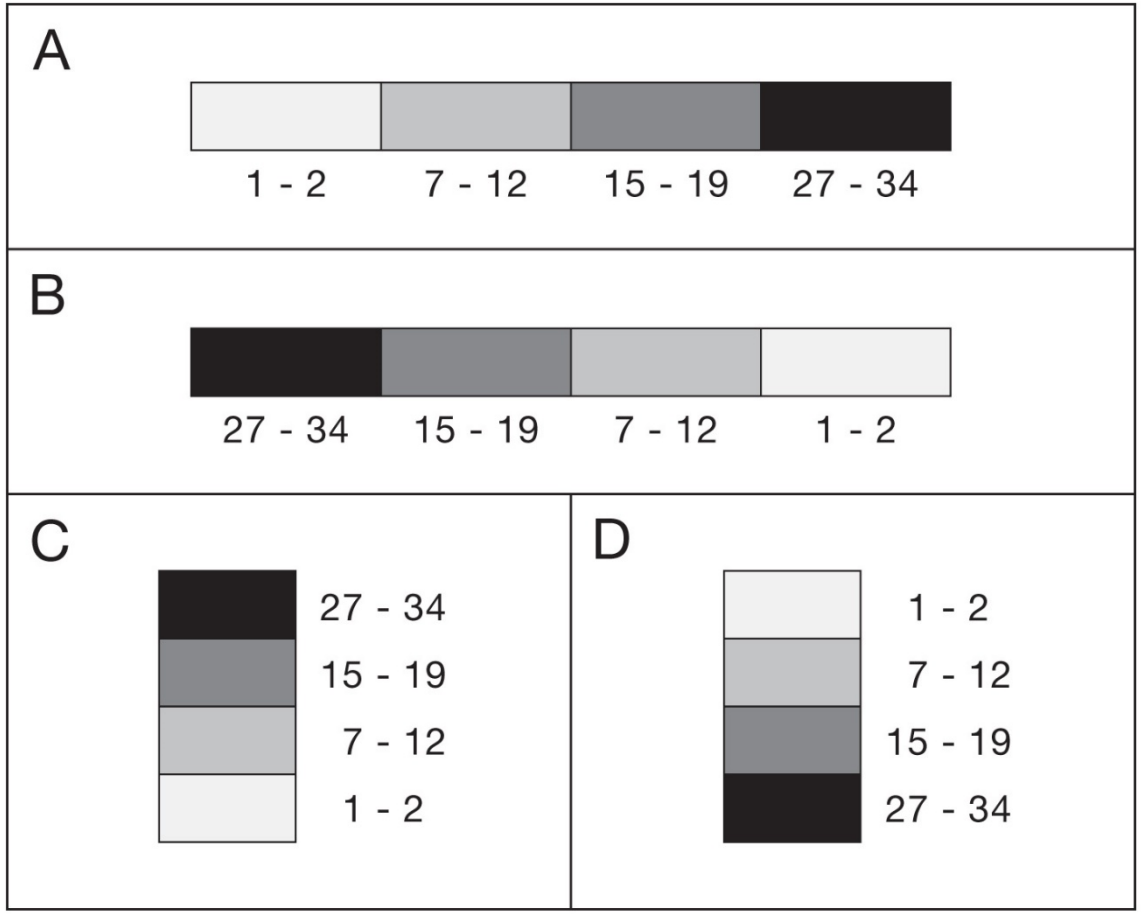


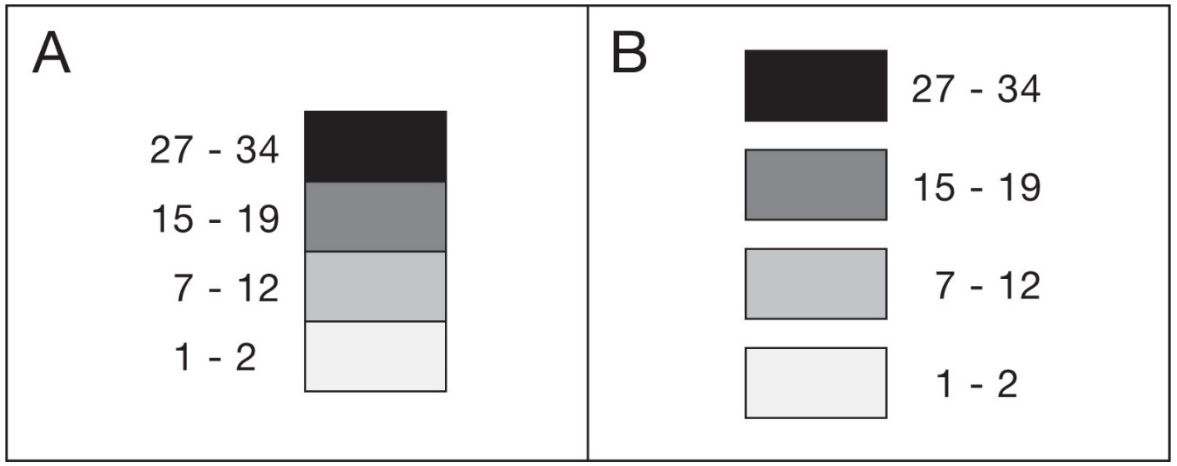




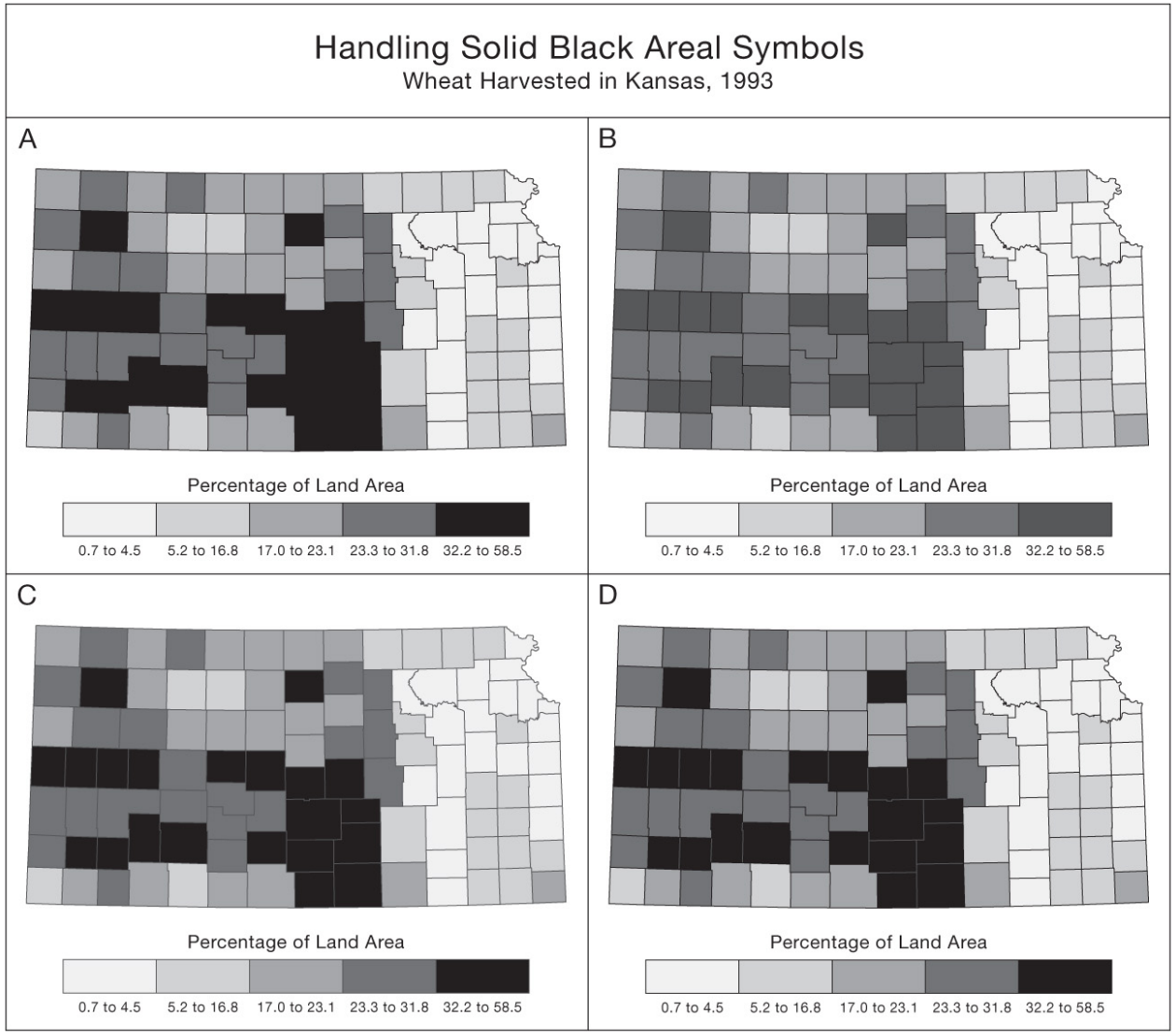


The legend





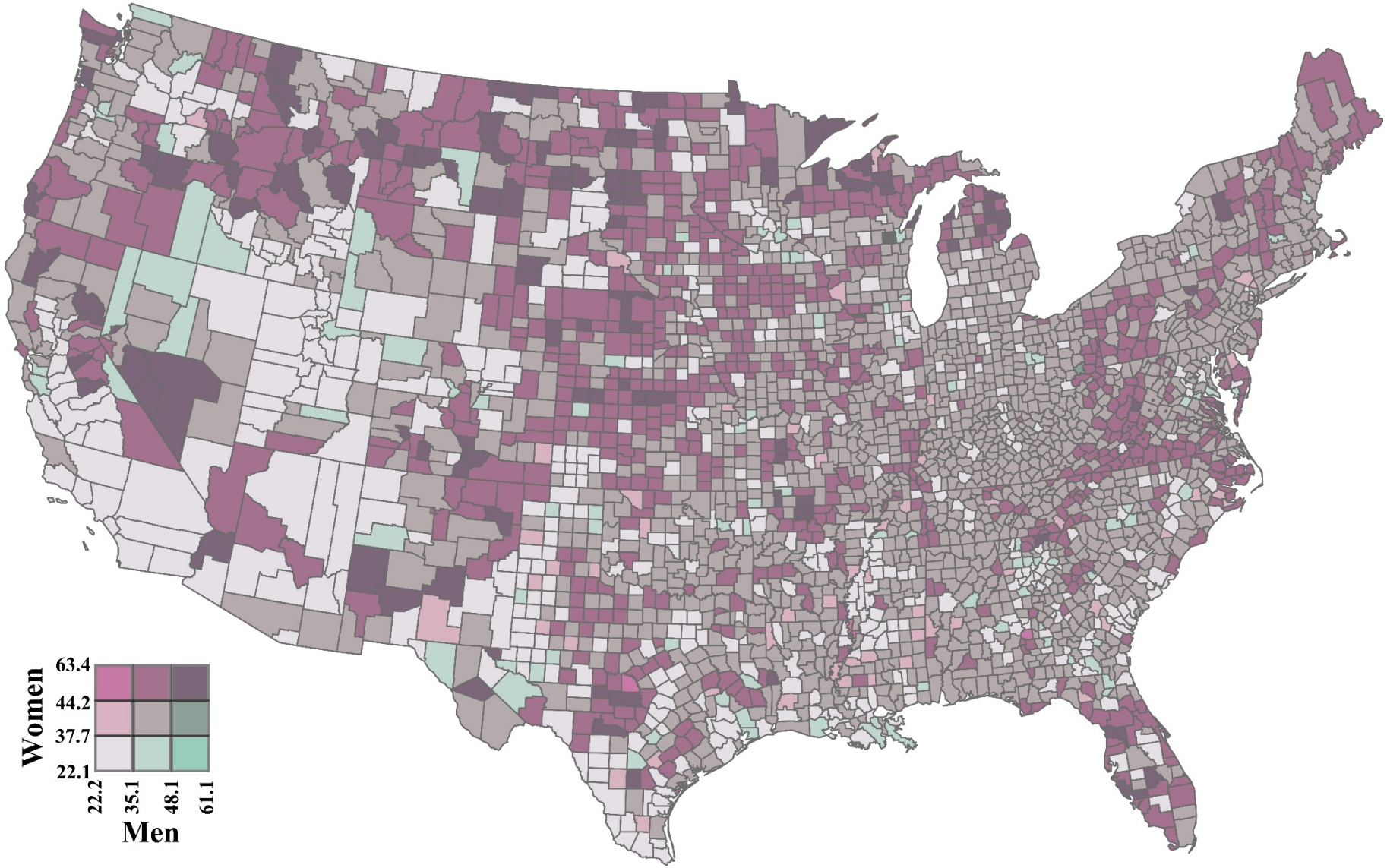
Handling the areas

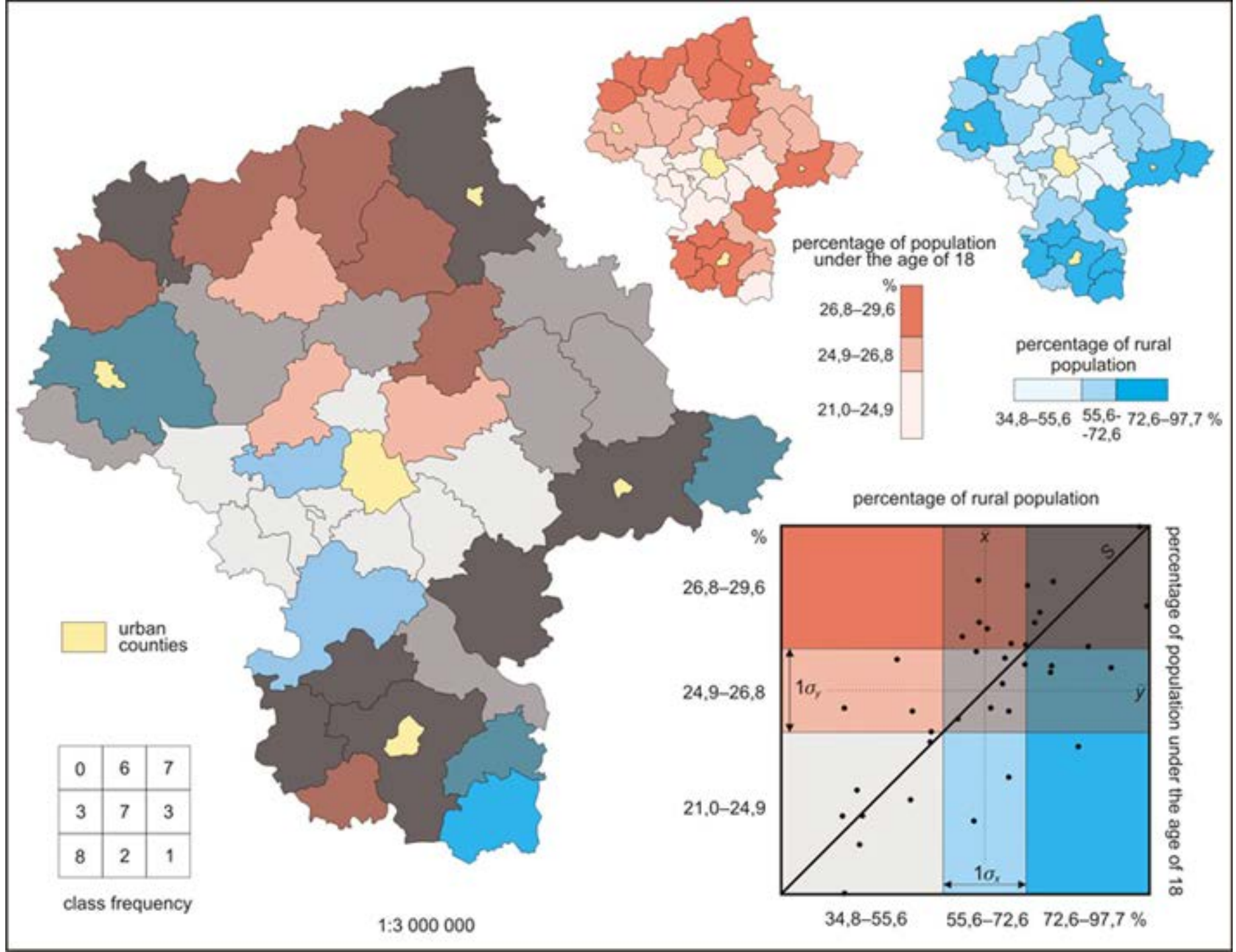


Bivariate choropleth maps

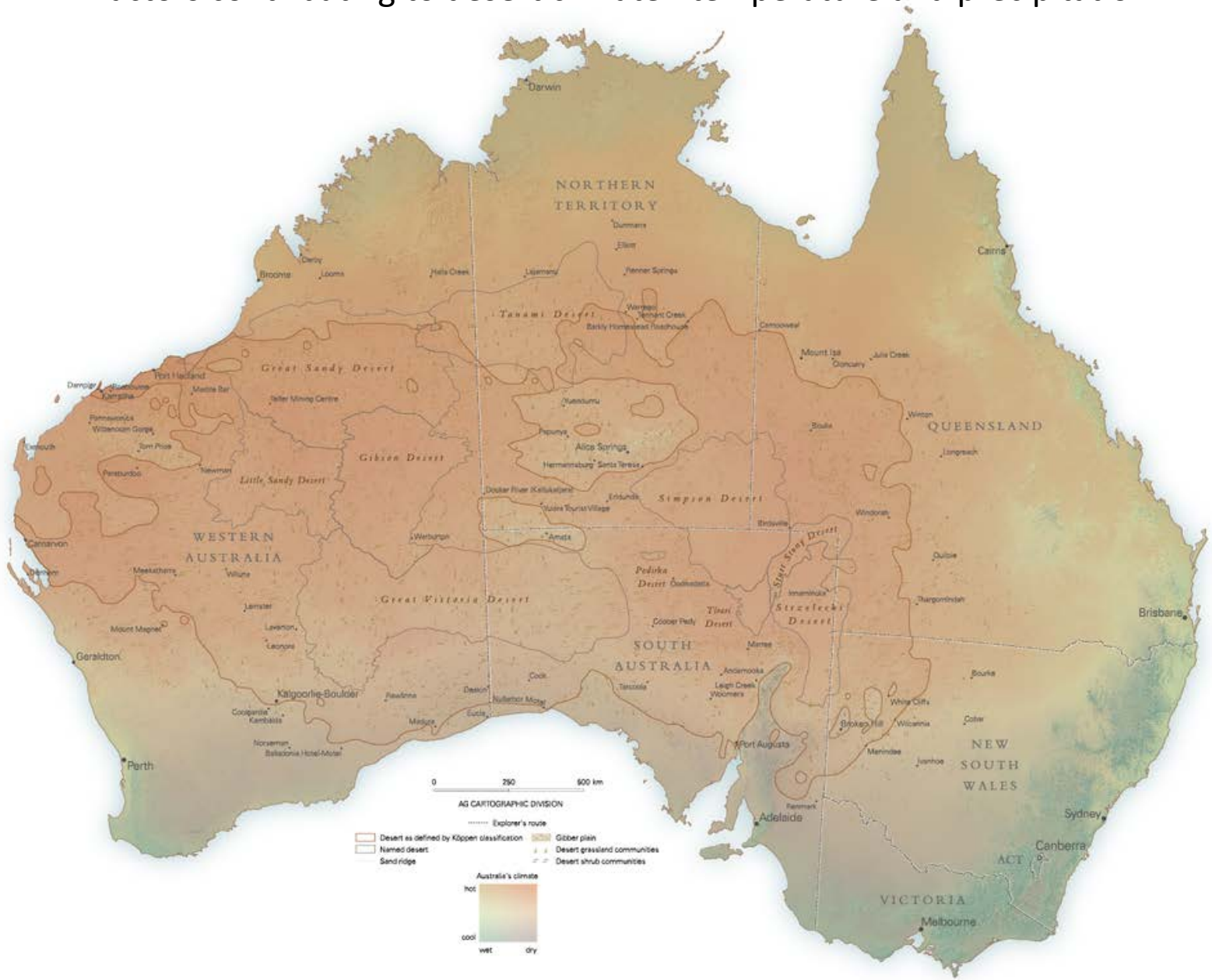
- Method for comparing two dissimilar distributions on the same map
- Popularized in the Census Atlas
- For choropleth, best with few classes
- Uses overlay legend
- Need contrasting colors
- Can do with Inkscape using automatic registration and transparency

Median Age in Years: 2010



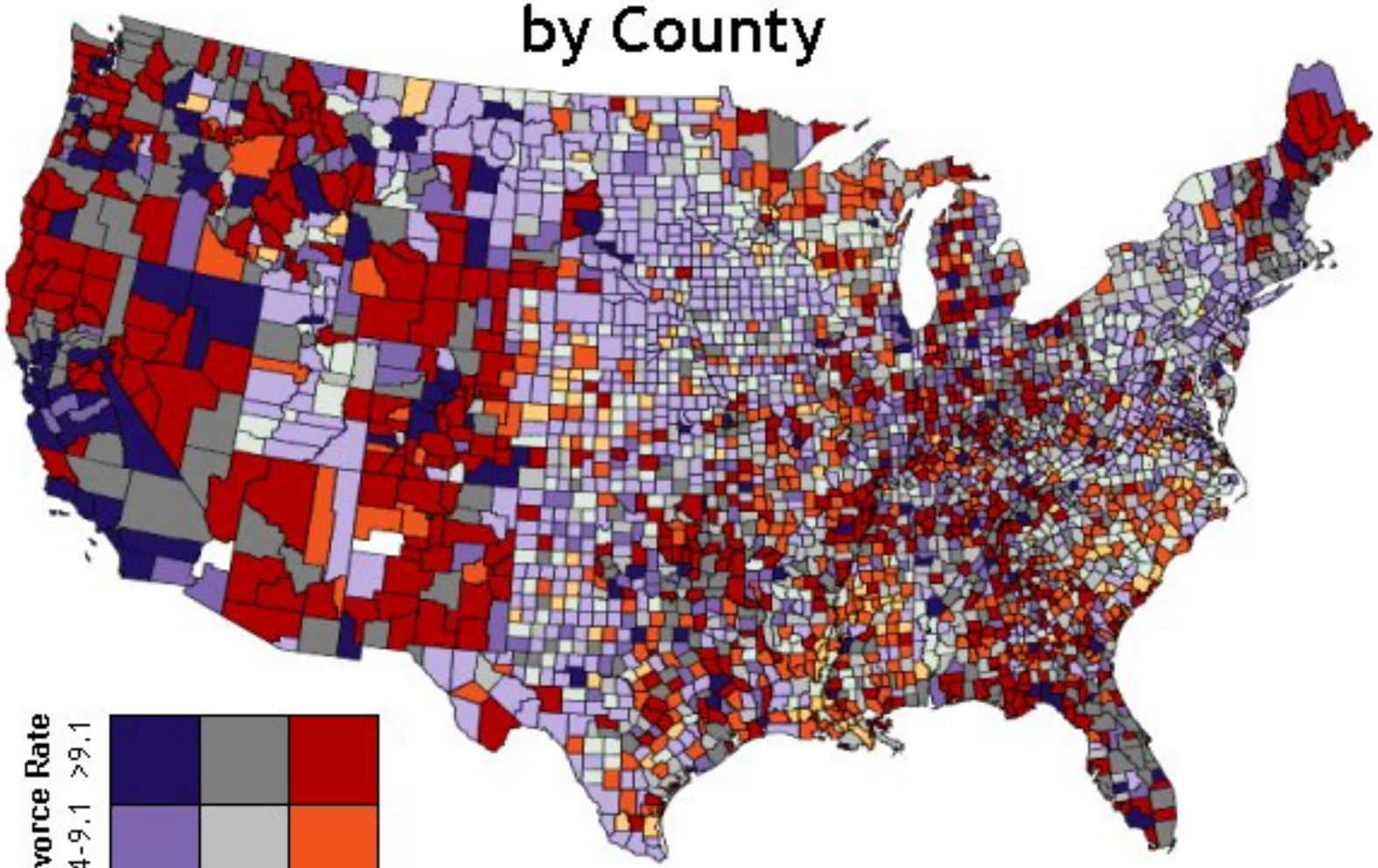

percentage of population under the age of 18
%
26,8–29,6
24,9–26,8
21,0–24,9
percentage of rural population
34,8–55,6
55,6--72,6
72,6–97,7 %
urban counties
0 6 7
3 7 3
8 2 1
class frequency
1:3 000 000
percentage of rural population
%
26,8–29,6
24,9–26,8
21,0–24,9
x
s
y
$1\sigma_y$
$1\sigma_x$
34,8–55,6
55,6–72,6
72,6–97,7 %
percentage of population under the age of 18

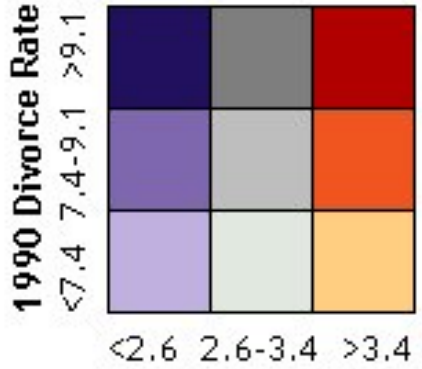
Factors contributing to desert climate - temperature and precipitation.



http://www.thecartographicdivision.com/blog

U.S. Divorce Rates by County

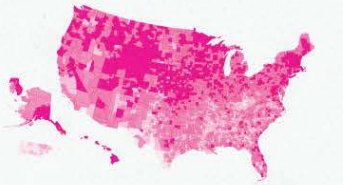




2-variable mapping scheme adapted from examples by
Dr. Cynthia A. Brewer,
see http://www.essc.psu.edu/~cbrewer/ColorSch/Schemes.html

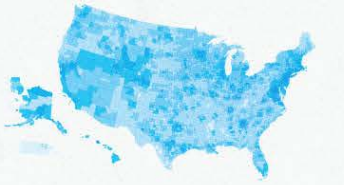
READING, WRITING, AND EARNING MONEY

The latest data from the U.S. Census's American Community Survey paints a fascinating picture of the United States at the county level. We've looked at the educational achievement and the median income of the entire nation, to see where people are going to school, where they're earning money, and if there is any correlation.



Ⓐ HIGH SCHOOL GRADUATES 65% 75% 82% 88%

Ⓑ COLLEGE GRADUATES 15% 22% 30% 40%



Ⓒ MEDIAN HOUSEHOLD INCOME $25K $40K $50K $65K

The map at right is a product of overlaying the three sets of data. The variation in hue and value has been produced from the data shown above. In general, darker counties represent a more educated, better paid population while lighter areas represent communities with fewer graduates and lower incomes.

A collaboration between GOOD and Gregory Hubacek
SOURCE: US Census

KING COUNTY, WA 81% 44% 67%

EAGLE COUNTY, CO 87% 47% 50%

HENNEPIN COUNTY, MN 69% 43% 61%

NEW YORK COUNTY, NY 84% 57% 66%

HOLMES COUNTY, OH 54% 9% 45%

BREATHITT COUNTY, KY 60% 9% 19%

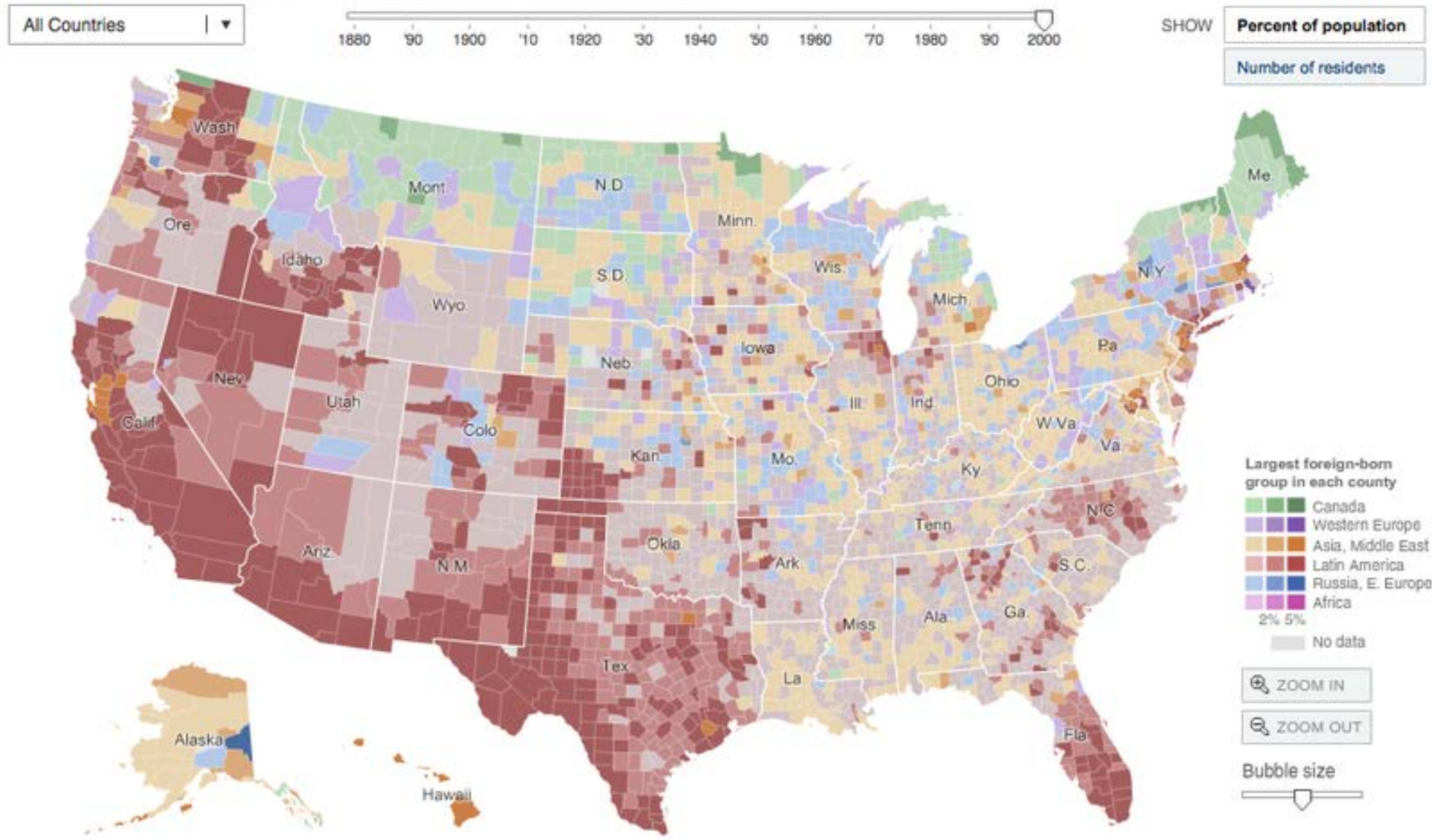
ALACHUA COUNTY, FL 81% 39% 39%

COSTILLA COUNTY, CO 73% 17% 23%

https://archive.nytimes.com/www.nytimes.com/interactive/2009/03/10/us/20090310-immigration-explorer.html

Immigration Explorer

Select a foreign-born group to see how they settled across the United States.

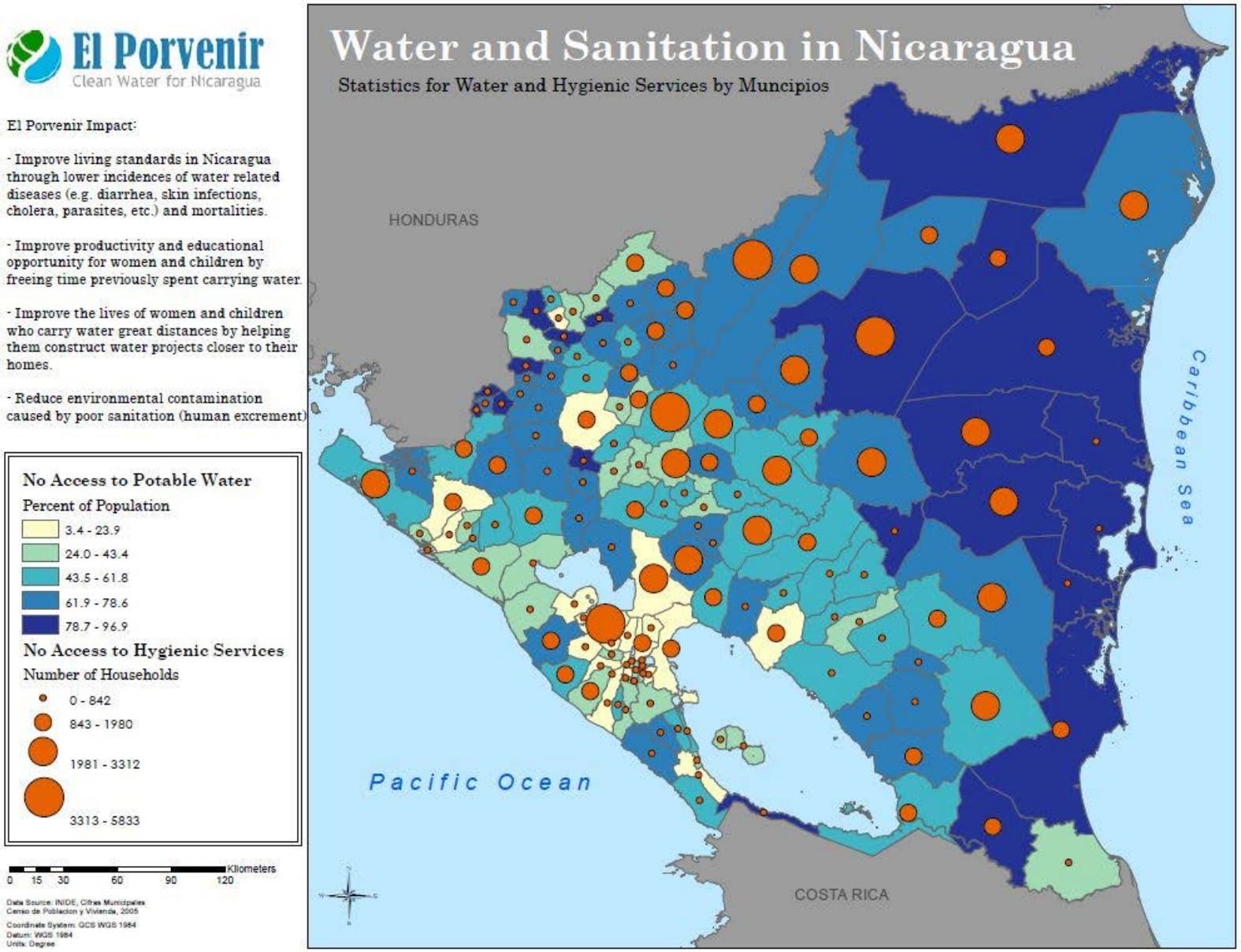


Note: Due to limitations in the Census data, foreign-born populations are not available in all areas for all years.

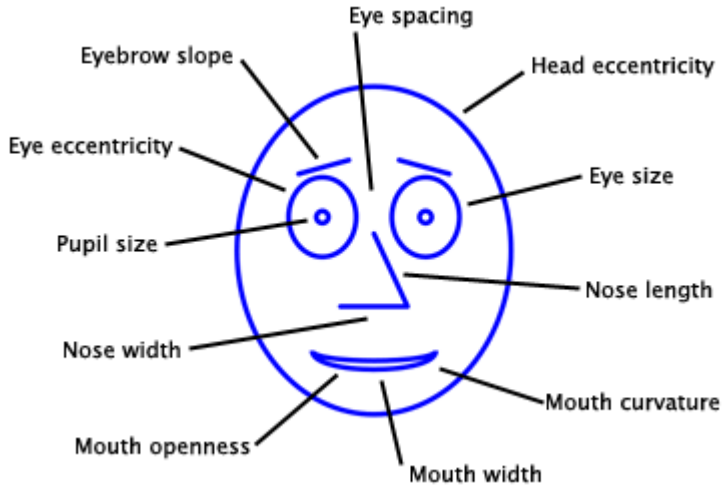
Sources: Social Explorer, www.socialexplorer.com; Minnesota Population Center; U.S. Census Bureau

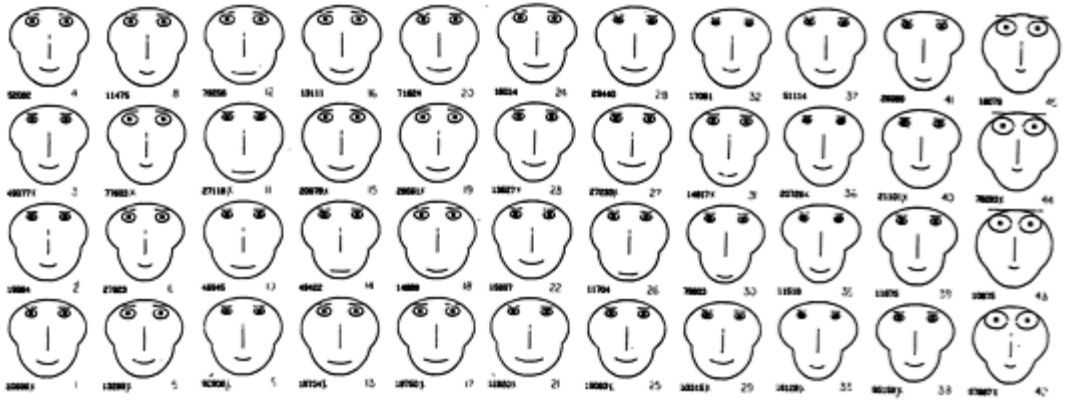
Matthew Bloch and Robert Gebeloff/The New York Times

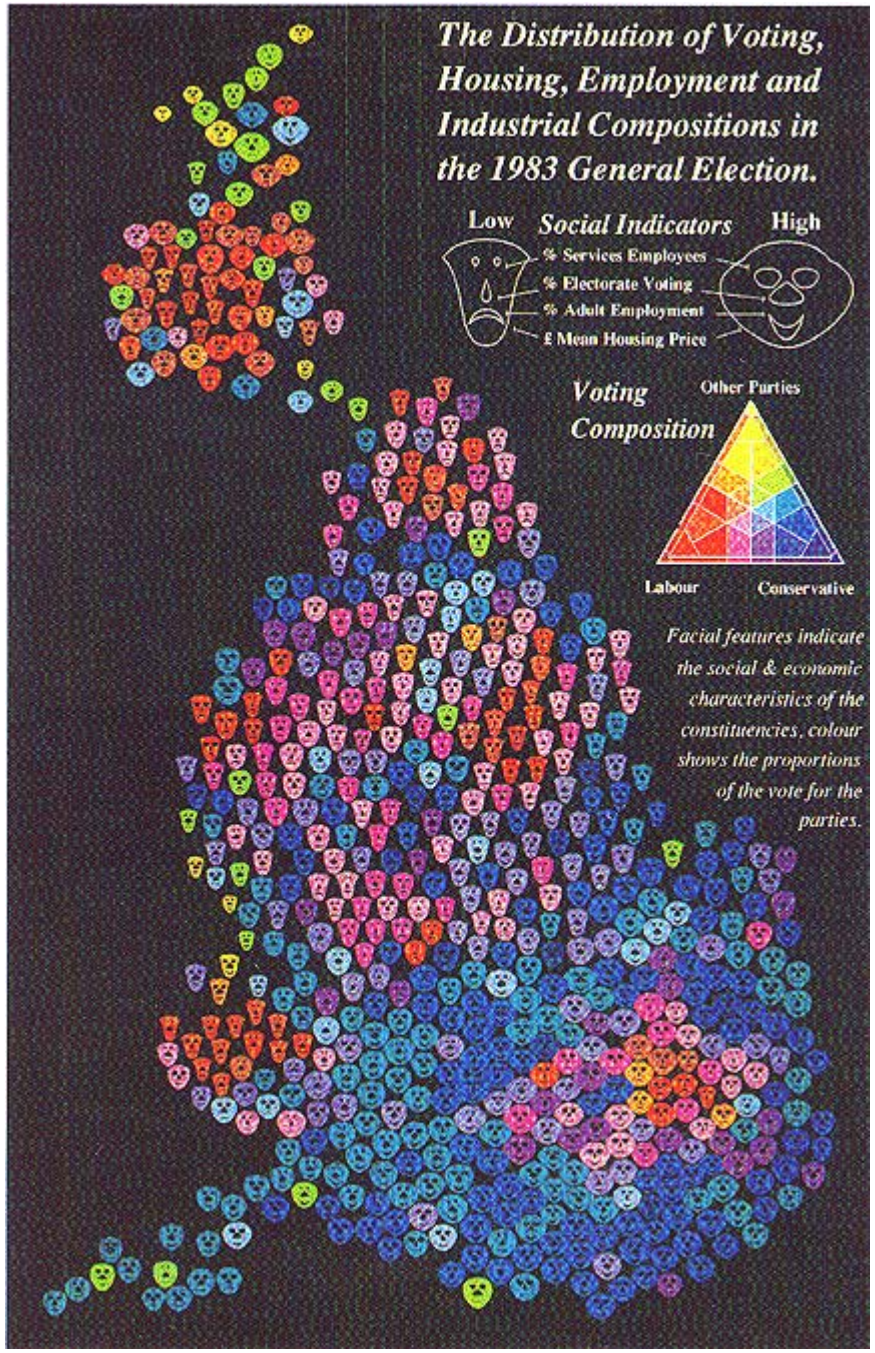
Multivariable by multimethod



Chernov Faces (Dorling, 1995)







Summary

- Map symbolization depends of data dimensions, level, scale
- Must often first standardize and classify
- Choose method or map type: choropleth, isopleth, dot, proportional symbol
- Take care with visual variables and their interactions
- Examined choropleth mapping in detail: classification critical
- Looked at bivariate and multivariate methods
- Complexity bad, simplicity good