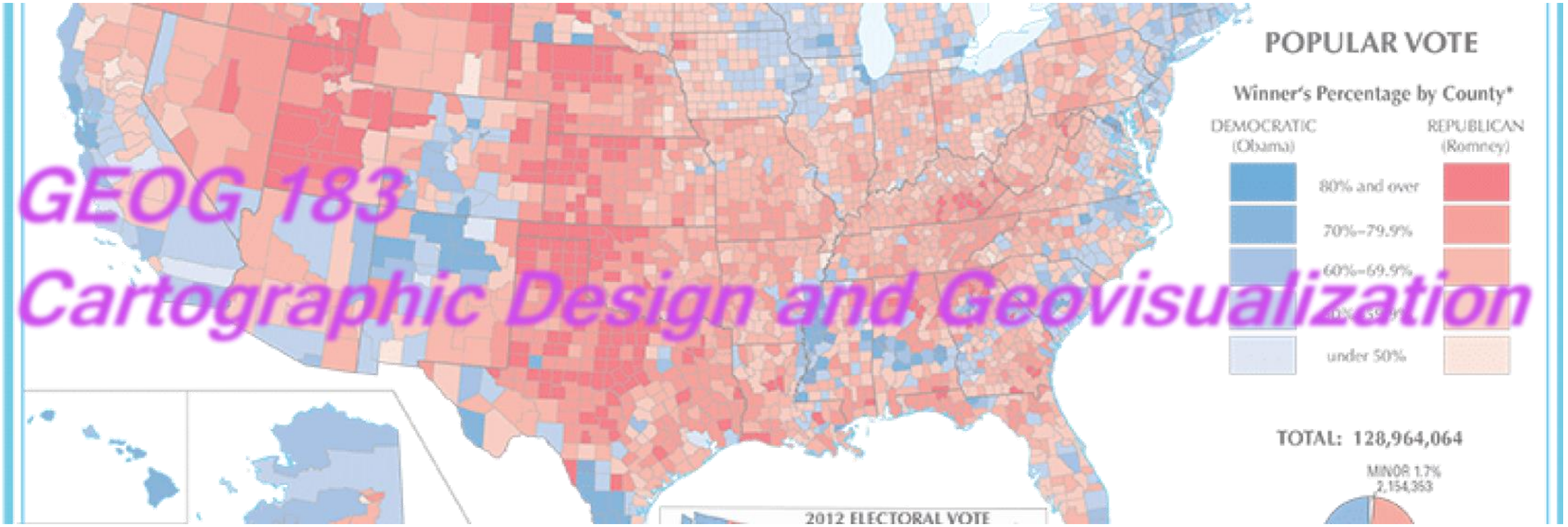


Geog183: Cartographic Design and Geovisualization Winter Quarter 2020

Lecture 6: Map types and Data types

Data types

- Data dimension: Point, Line, Area, Volume (Text)
- Data continuity: Discrete, Point, Polygon: Continuous
- Stevens data level: Nominal, Ordinal, Interval, Ratio
- Often involve classification and normalization before suitable for mapping
- E.g. Collect data at points, count points in polygon, normalize by area, then classify for choropleth map

Robinson & Sale

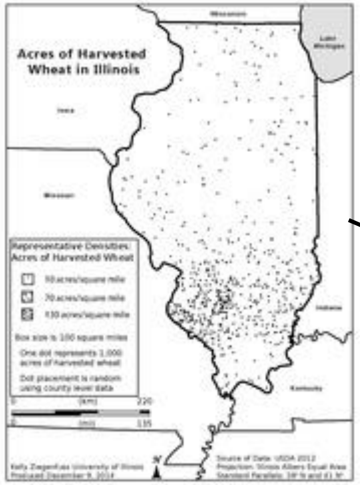
Content scaling level	Defining relations	FORM OF CARTOGRAPHIC SYMBOL		
		POINT	**LINE**	**AREA**
Nominal	Equivalence	Wholesale and retail establishments	Highway connectivity	Land ownership
Ordinal	Equivalence Greater than	Population centers	Roads by degree of improvement	Crop yield
Interval	Equivalence Greater than Ratio of intervals	Spot elevations	Graticule	Date of settlement
Ratio	Equivalence Greater than Ratio of intervals Ratio of scale values	Area proportional to population	Population density isopleths	Value proportional to population density

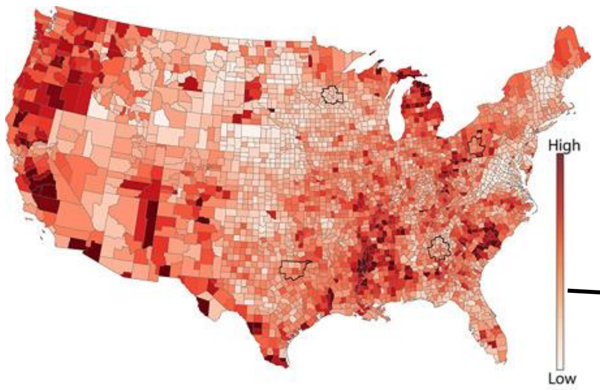
Figure 10.1 Classification by scaling and dimension. (After Robinson and Sale, *Elements of Cartography, 3d ed.* ,© 1969, by John Wiley & Sons, Inc. Used with permission.)§

Map Types

- Also divide by data types
- Data type determines what map types are suitable
- Set of map (and geoviz) types includes the standard, plus new and evolving methods
- Over time, methods went from 2D to 3D to 4D
- Slocum starts with dot, choropleth, isopleth and proportional symbol

Unwin's Classification





DATA TYPES

	Point	Line	Area	Volume
Nominal	City	Road	Name of unit	Precipitation or soil type
Ordinal	Large city	Major road	Rich county	Heavy precipitation Good soil
Interval Ratio	Total population	Traffic flow	Per capita income	Precip.in mm Cation exchange

MAP TYPES

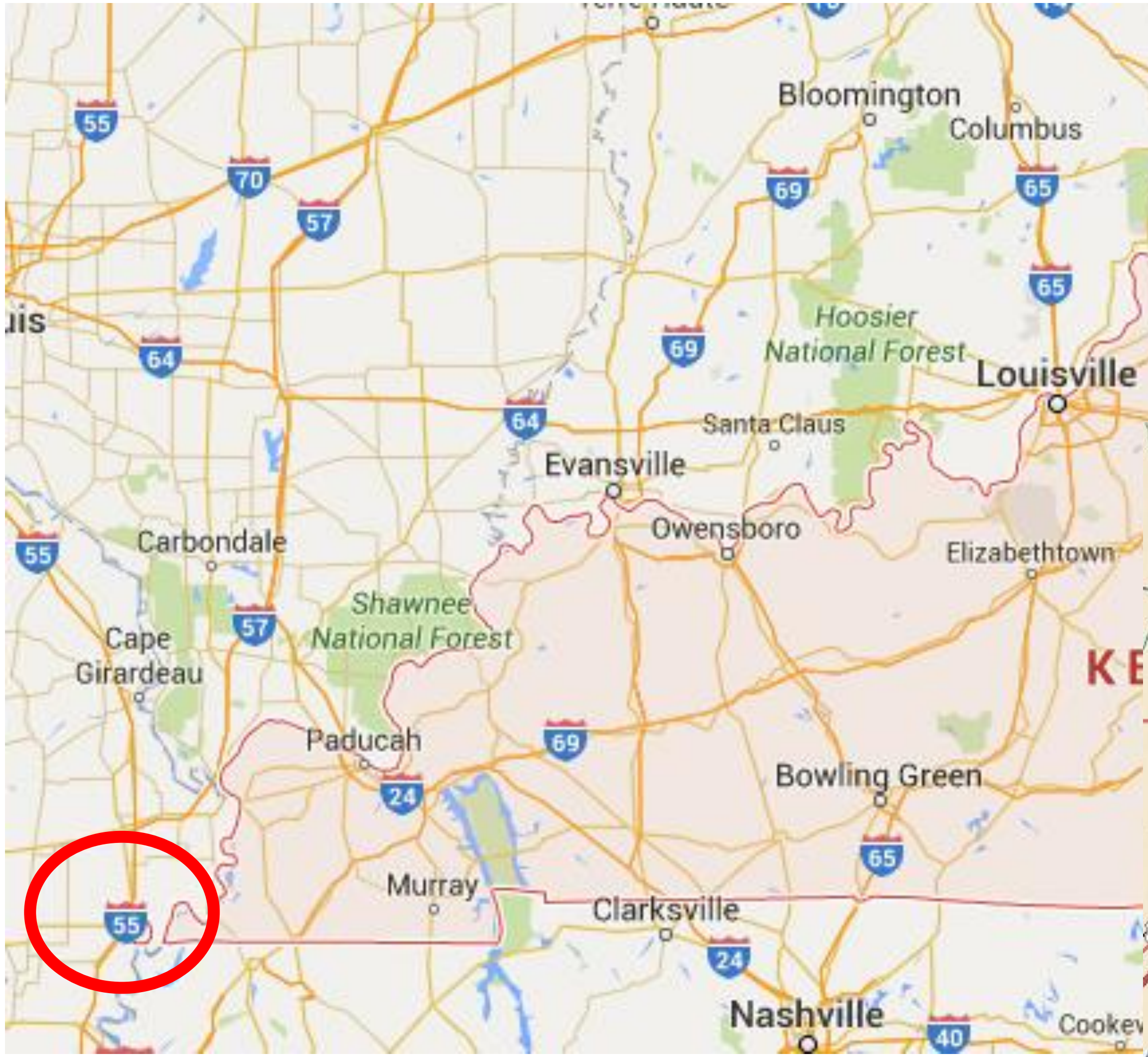
	Point	Line	Area	Volume
Nominal	Dot map	Network map	Colored area map	Freely colored map
Ordinal	Symbol map	Ordered network map	Ordered colored map	Ordered chromatic map
Interval Ratio	Graduated symbol map	Flow map	Choropleth map	Contour map

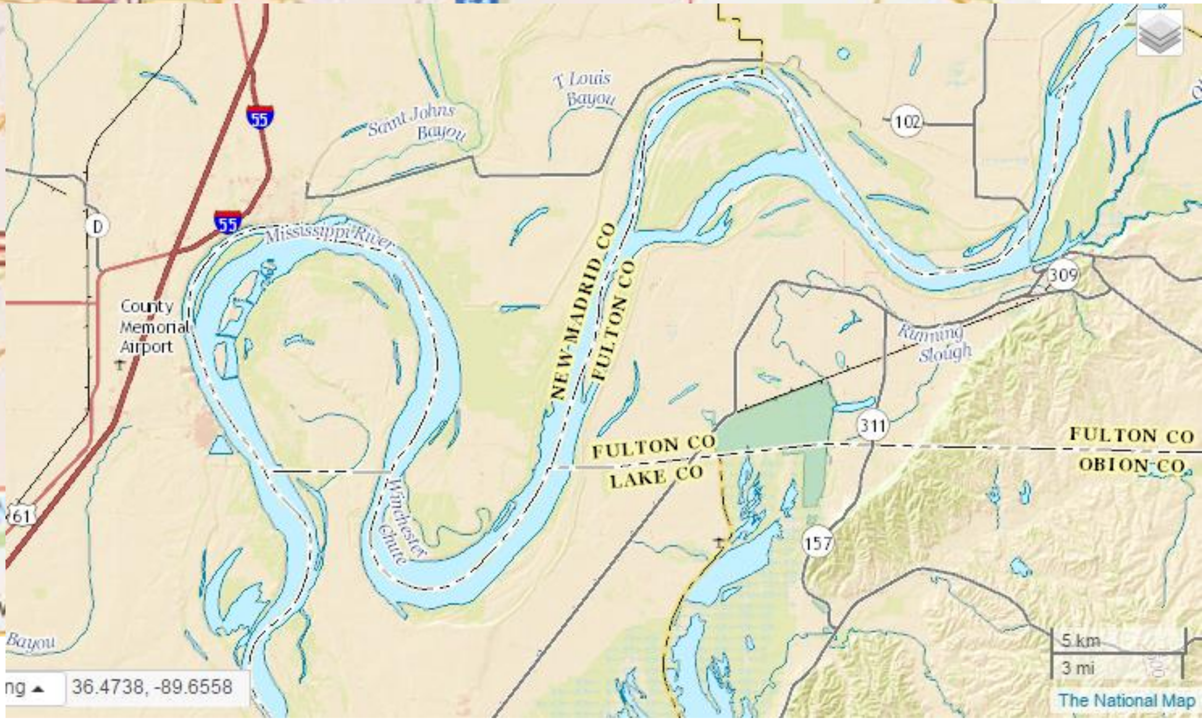
Figure 10.3 Map data and map types. (After Unwin, 1981.)§

Discrete versus Continuous

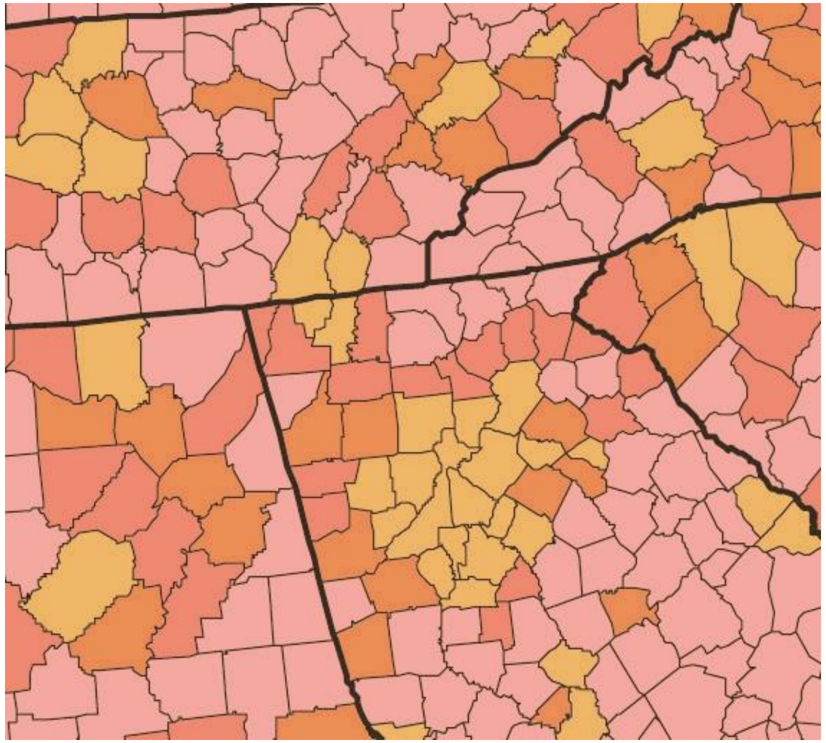
- Much geographic data relates to specific points, lines and areas
- Values are uniform within and different among
- Good example is choropleth mapping
- Choropleth means value assigned to place
- Units, especially areas, are often merely the way that data are aggregated (e.g. counties, census tracts)

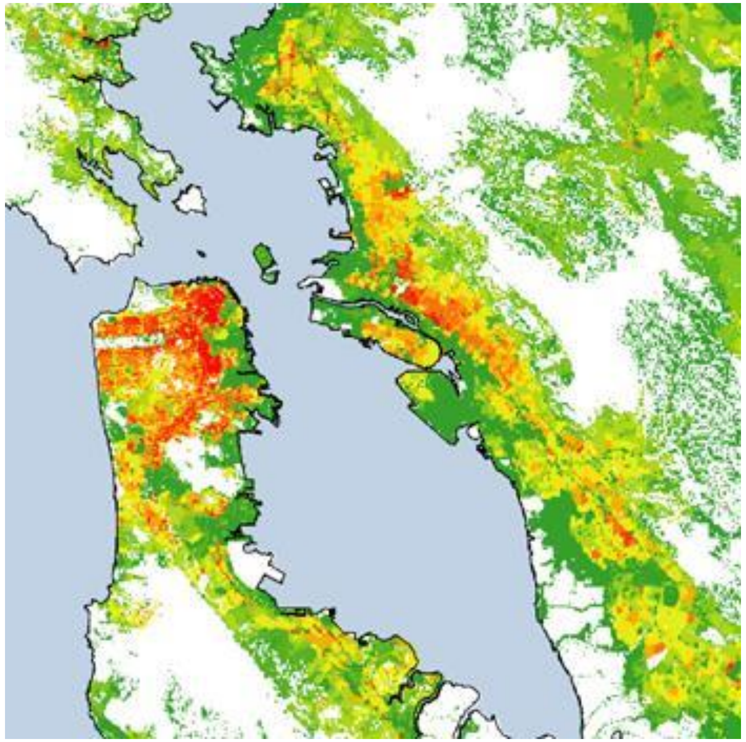
Detail matters!



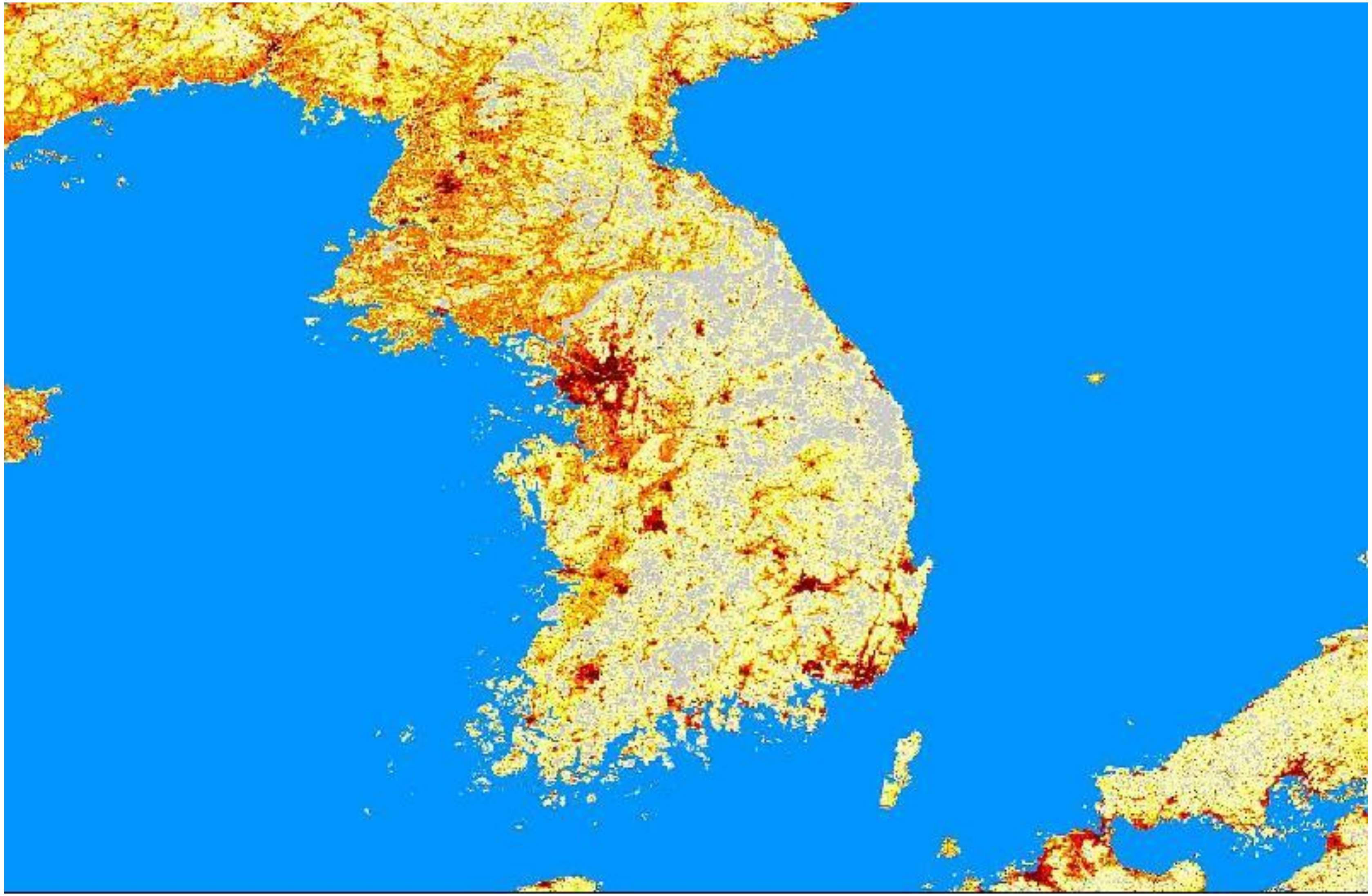


Discrete: Choropleth
But, can relax zones and add other data, e.g. land use making a dasymetric map





Dasymetric map: Landscan population density

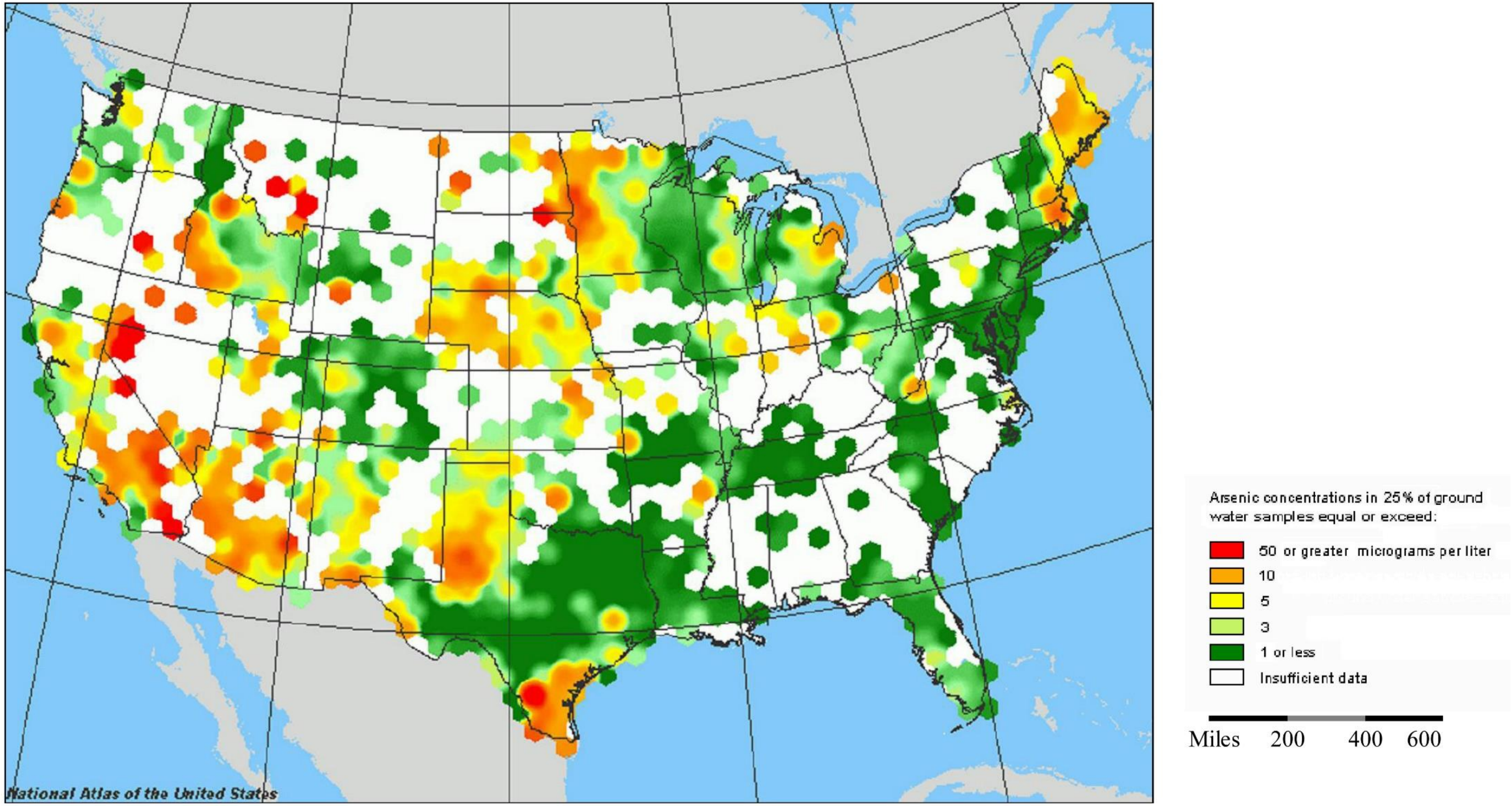


Continuous

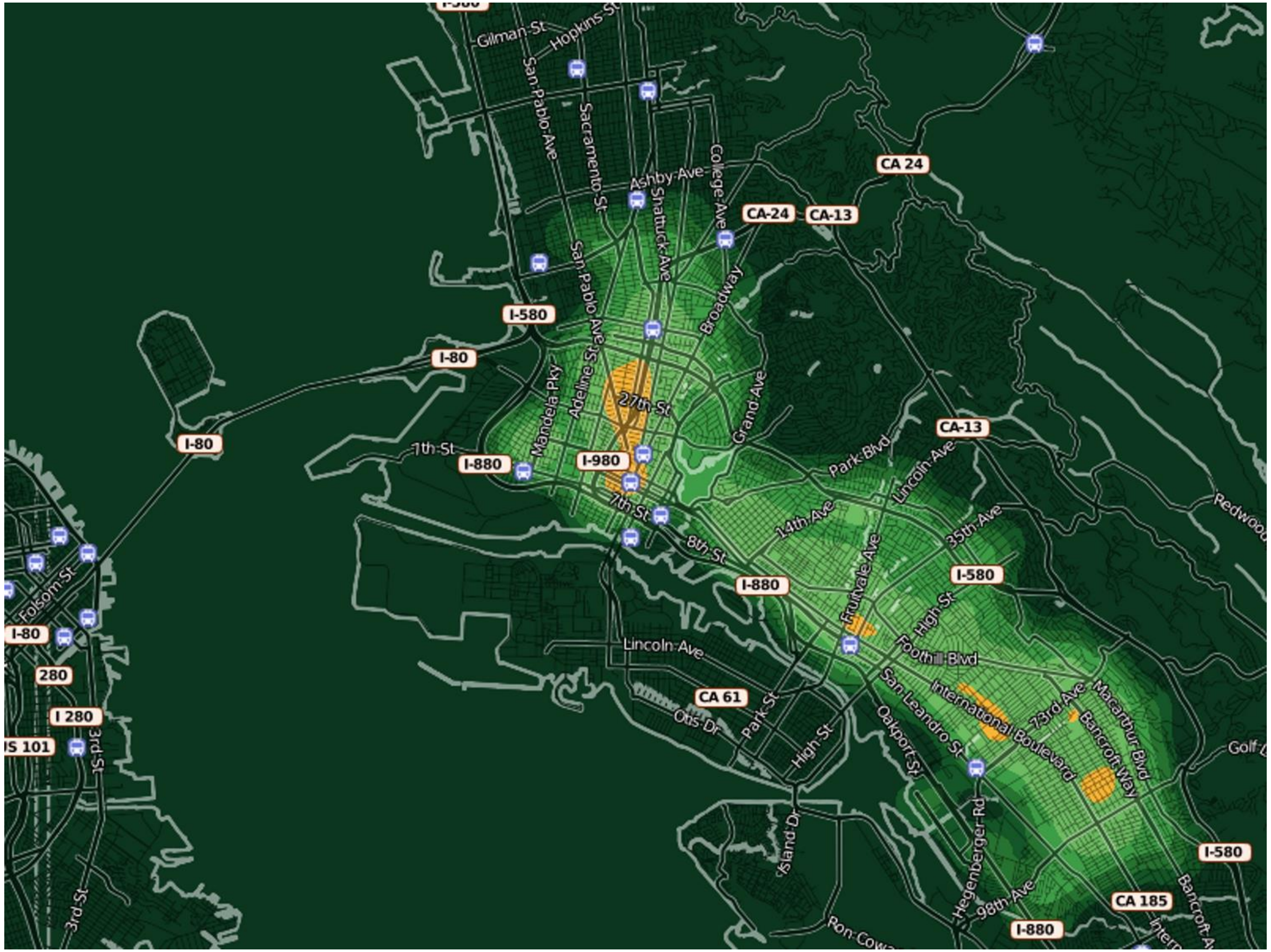
- Some geographic variables are measureable anywhere in space
- Examples: air temperature and pressure
- Creates a surface or field
- Can treat the surface like topography, and use many of the isoline and other methods for terrain
- Color and isolines often favored

Continuous image map: Biomass in tons/hectare 2000

Continuous: Hexagonally sampled arsenic in groundwater



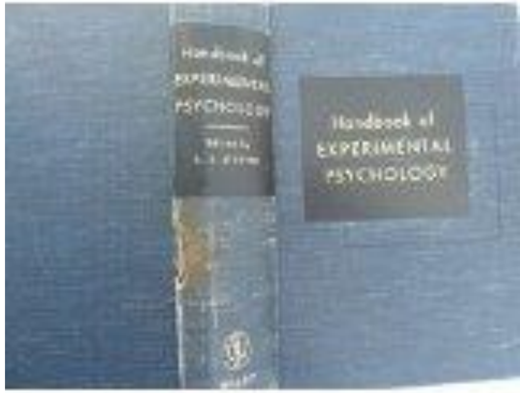
Continuous: Heat map of crime in Oakland, California 2012
Uses kernel density function



Data levels or levels of measurement

Stanley Smith Stevens
(1906-1973)

- American psychologist best known as the founder of Harvard's Psycho-Acoustic Laboratory
- Credited with Stevens' power law
- Milestone textbook, the 1400+ page "Handbook of Experimental Psychology" (1951)
- Founding organizer of the Psychonomic Society
- 1946 he introduced a theory of levels of measurement often used by statisticians



SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues "came out by that same door as they went in," and in order to have another try at agreement, the committee begged to be continued for another year.

For its final report (1940) the committee chose a common bone for its contentions, directing its arguments at a concrete example of a sensory scale. This was the Sone scale of loudness (S. S. Stevens and H. Davis. *Hearing.* New York: Wiley, 1938), which purports to measure the subjective magnitude of an auditory sensation against a scale having the formal properties of other basic scales, such as those used to measure length and weight. Again the 19 members of the committee came out by the routes they entered, and their views ranged widely between two extremes. One member submitted "that any law purporting to express a quantitative relation between sensation intensity and stimulus intensity is not merely false but is in fact meaningless unless and until a meaning can be given to the concept of addition as applied to sensation" (Final Report, p. 245).

It is plain from this and from other statements by the committee that the real issue is the meaning of measurement. This, to be sure, is a semantic issue, but one susceptible of orderly discussion. Perhaps agreement can better be achieved if we recognize that measurement exists in a variety of forms and that scales of measurement fall into certain definite classes. These classes are determined both by the empirical operations invoked in the process of "measuring" and by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A Classification of Scales of Measurement

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The problem then becomes that of making explicit (a) the various rules for the assignment of numerals, (b) the mathematical properties (or group structure) of the resulting scales, and (c) the statistical operations applicable to measurements made with each type of scale.

Scales are possible in the first place only because there is a certain isomorphism between what we can do with the aspects of objects and the properties of the numeral series. In dealing with the aspects of objects we invoke empirical operations for determining equality (classifying), for rank-ordering, and for determining when differences and when ratios between the aspects of objects are equal. The conventional series of numerals yields to analogous operations: We can identify the members of a numeral series and classify them. We know their order as given by convention. We can determine equal differences, as $8-6=4-2$, and equal ratios, as $8/4=6/3$. The isomorphism between these properties of the numeral series and certain empirical operations which we perform with objects permits the use of the series as a *model* to represent aspects of the empirical world.

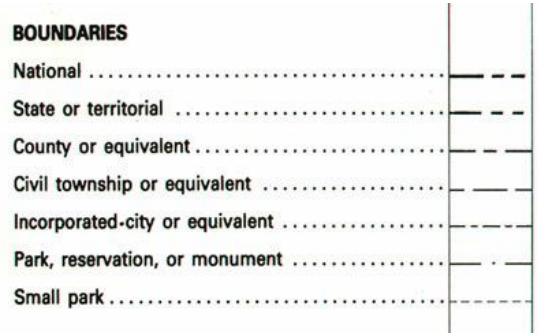
The type of scale achieved depends upon the character of the basic empirical operations performed. These operations are limited ordinarily by the nature of the thing being scaled and by our choice of procedures, but, once selected, the operations determine

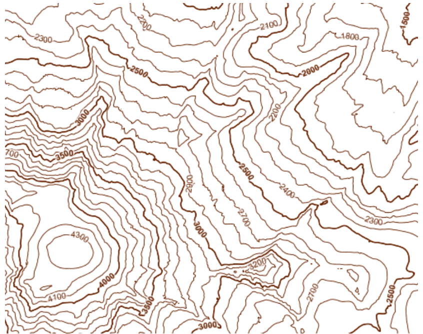
Quantitative Data Have “Levels” of Measurement

Stevens (1946)

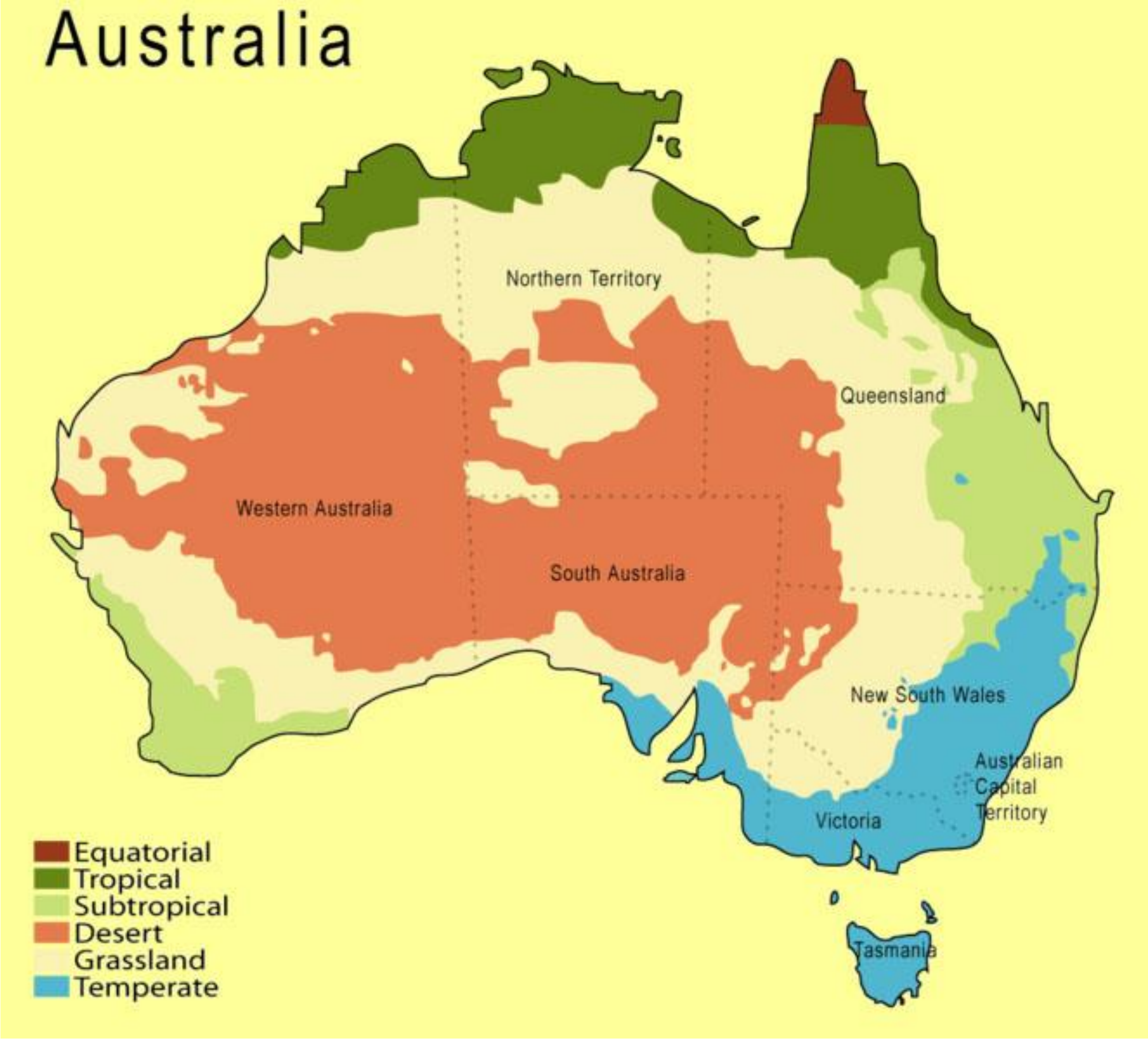
- Nominal: Has name or class only
- Ordinal: Has rank only
- Interval: Has value on arbitrary scale (e.g. Fahrenheit)
- Ratio: Has value on scale with absolute zero value (e.g. Kelvin)
- Different mathematical operations on variables are possible, depending on the level at which a variable is measured. (e.g. Forest + Agriculture = ?)
- In statistics the kinds of descriptive statistics and significance tests that are appropriate depend on the level of measurement of the variables concerned

New York City





Nominal maps types



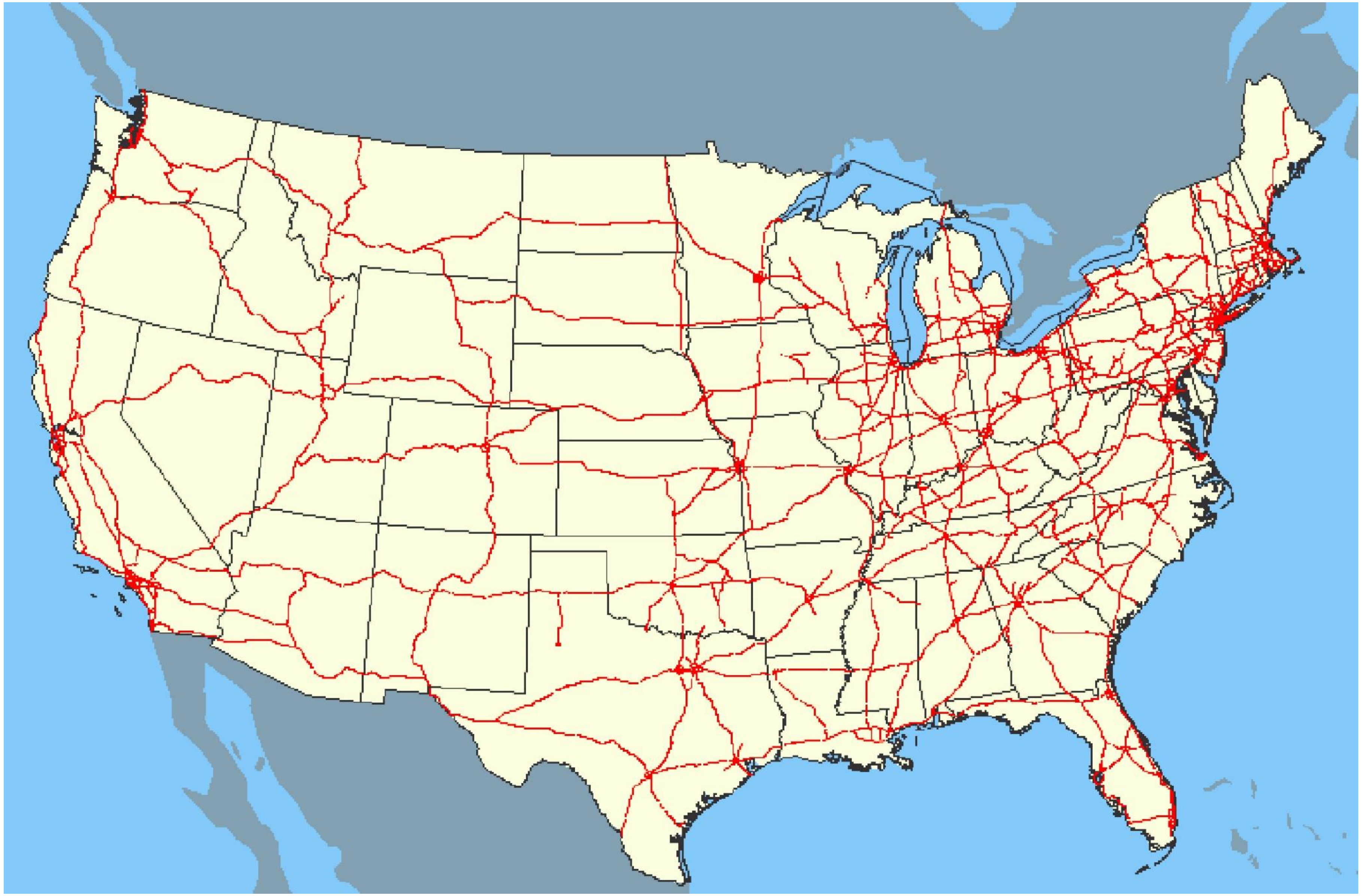
Point nominal map: Airport delays

The status information provided on this site indicates general airport conditions; it is not flight-specific. Check with your airline to determine if your flight is affected.Information on wait times at security checkpoints.

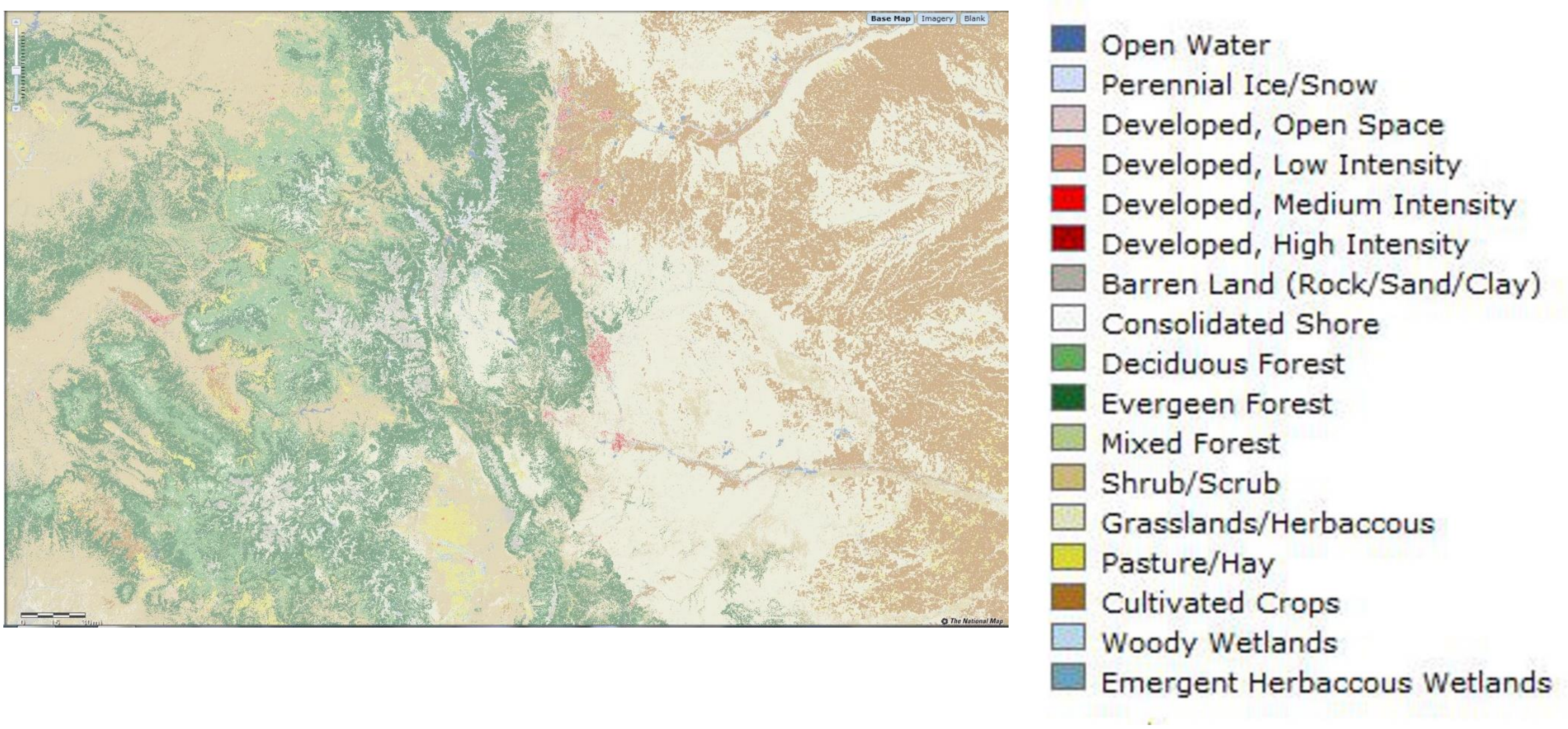
Legend

- General Arrival/Departure delays are 15 minutes or less.
- Departures are experiencing taxi delays of 16 to 45 minutes and/or arrivals are experiencing airborne holding delays of 16 to 45 minutes.
- Traffic destined to this airport is being delayed at its departure point. Check your departure airport to see if your flight may be affected.
- Departures are experiencing taxi delays greater than 45 minutes and/or arrivals are experiencing airborne holding delays greater than 45 minutes.
- This denotes a closed airport!

Line nominal map: Interstates and major highways



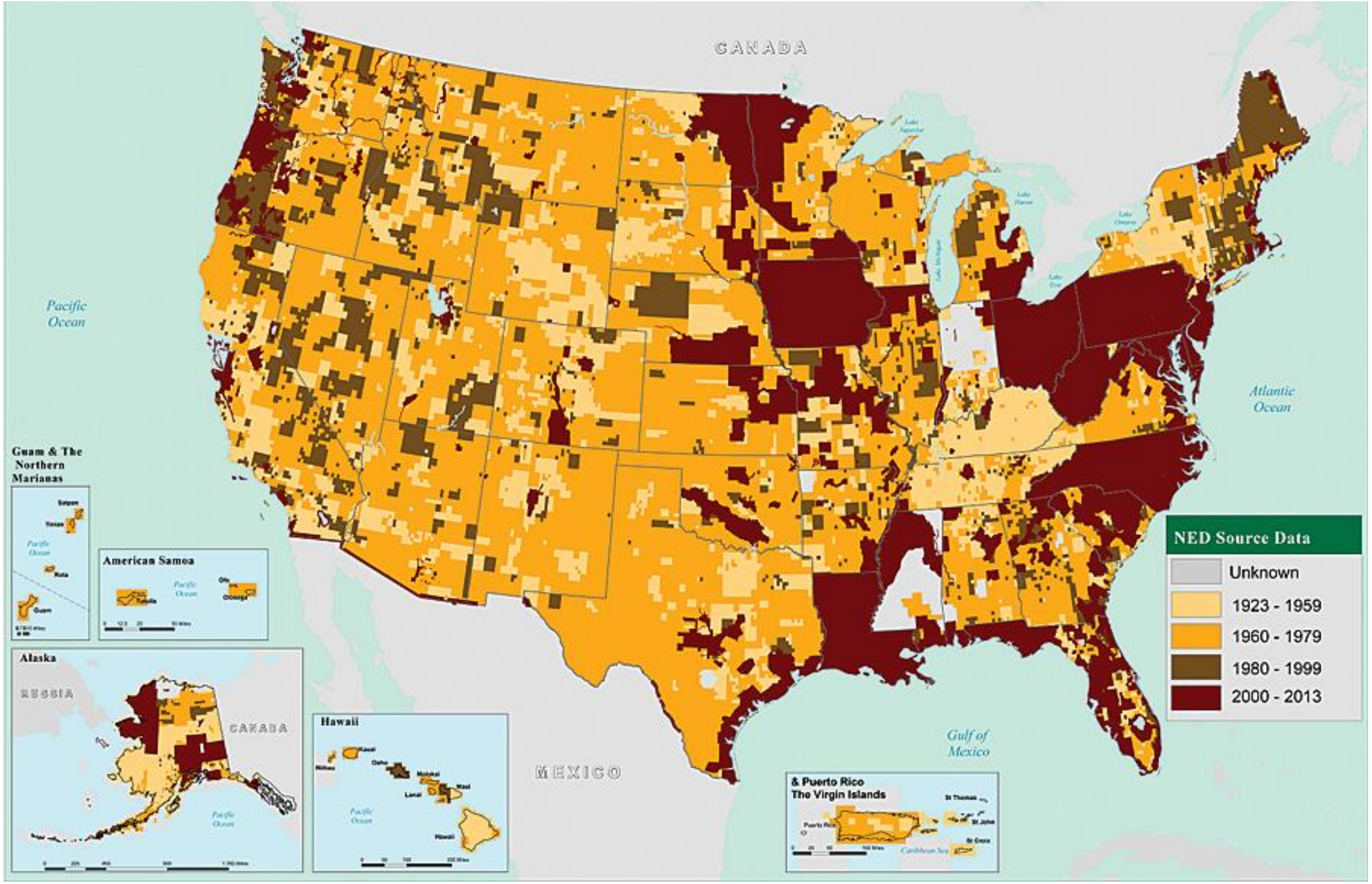
Area nominal map: Land use



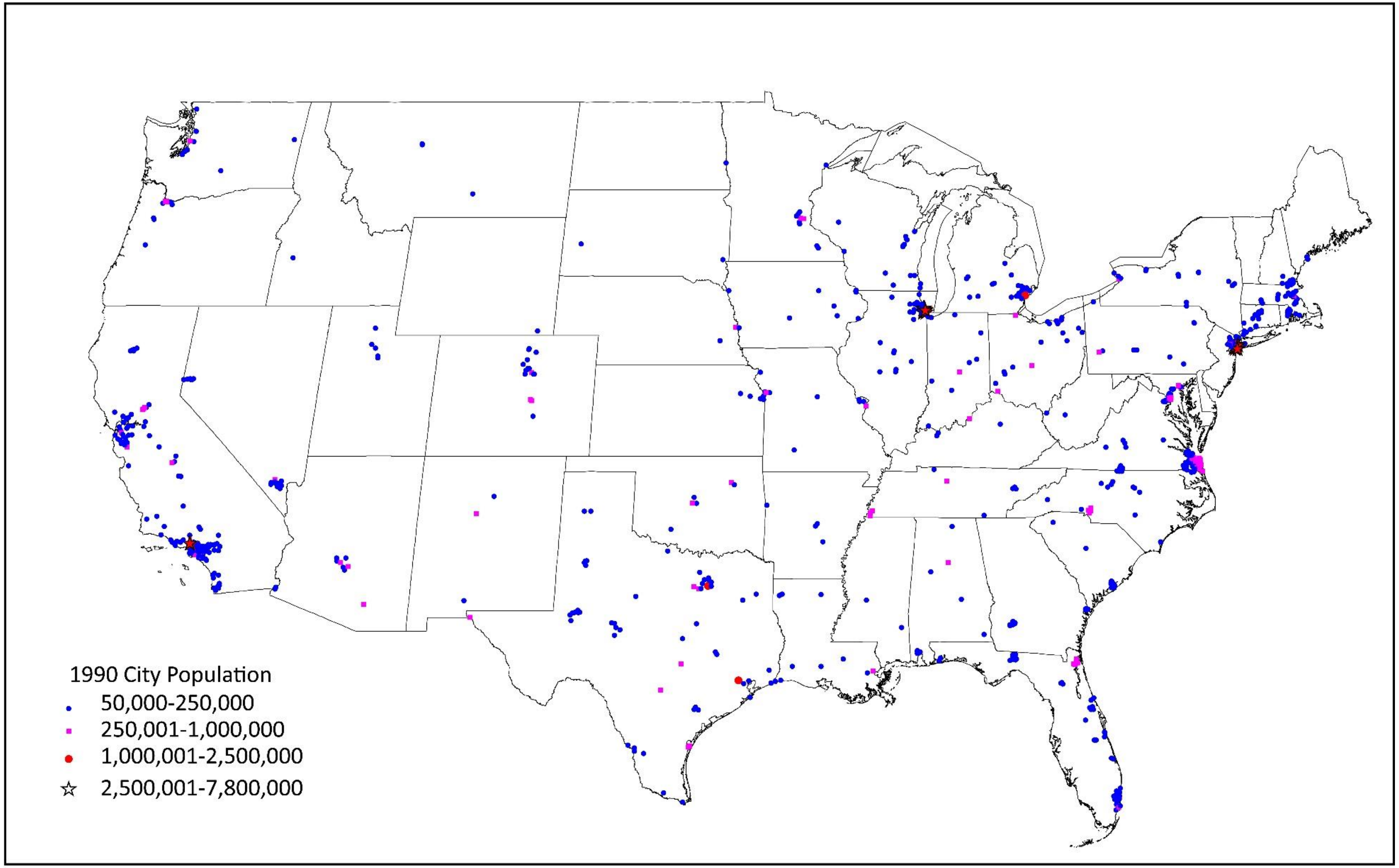
Nominal Data

- Relates to name or existence of a class
- Place names and legends important
- Point: labels at locations
- Lines: Network shown with symbols
- Areas: Classes shown by color and pattern
- Simplest data level, no real quantitative analysis possible

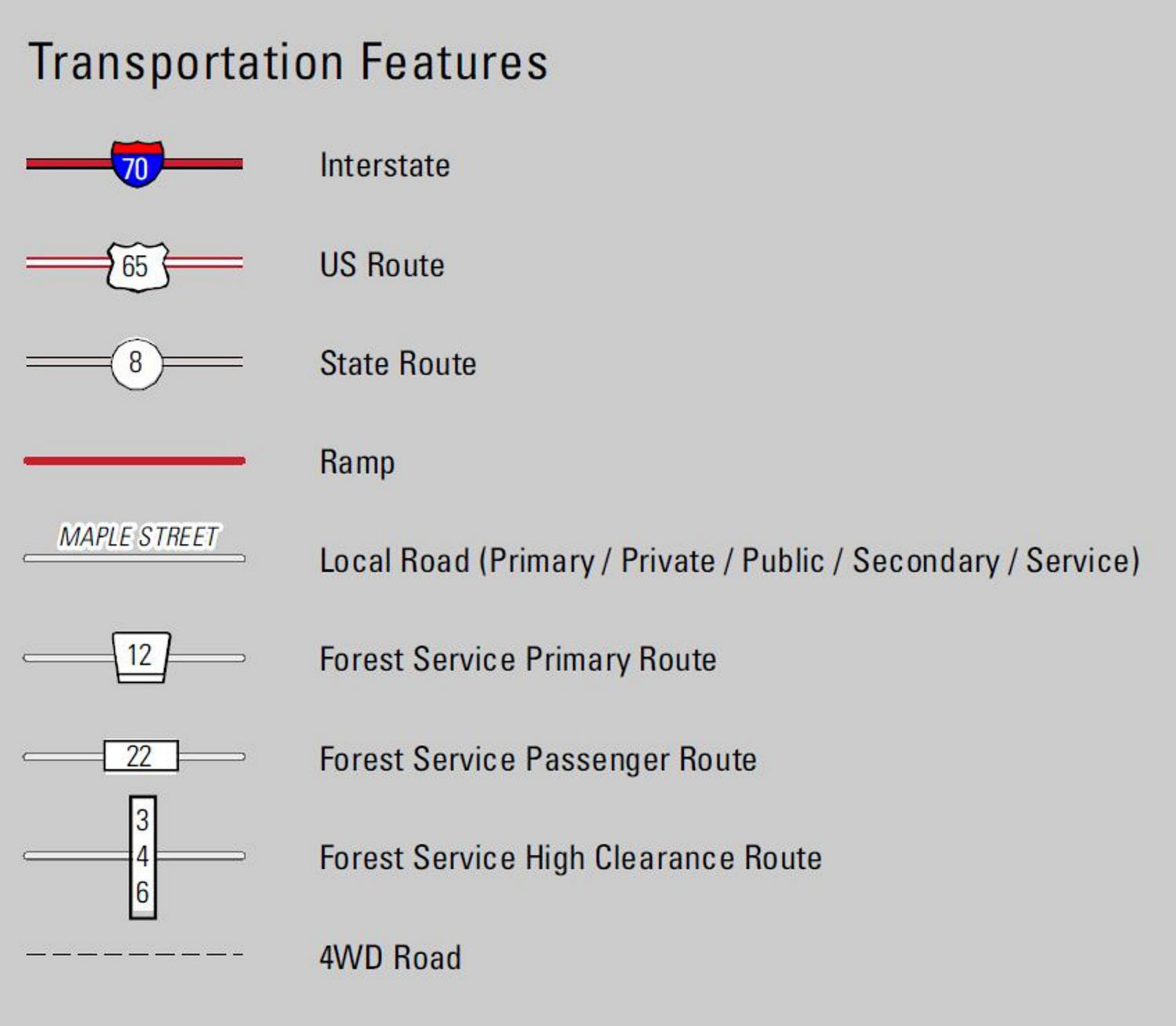
Ordinal map types



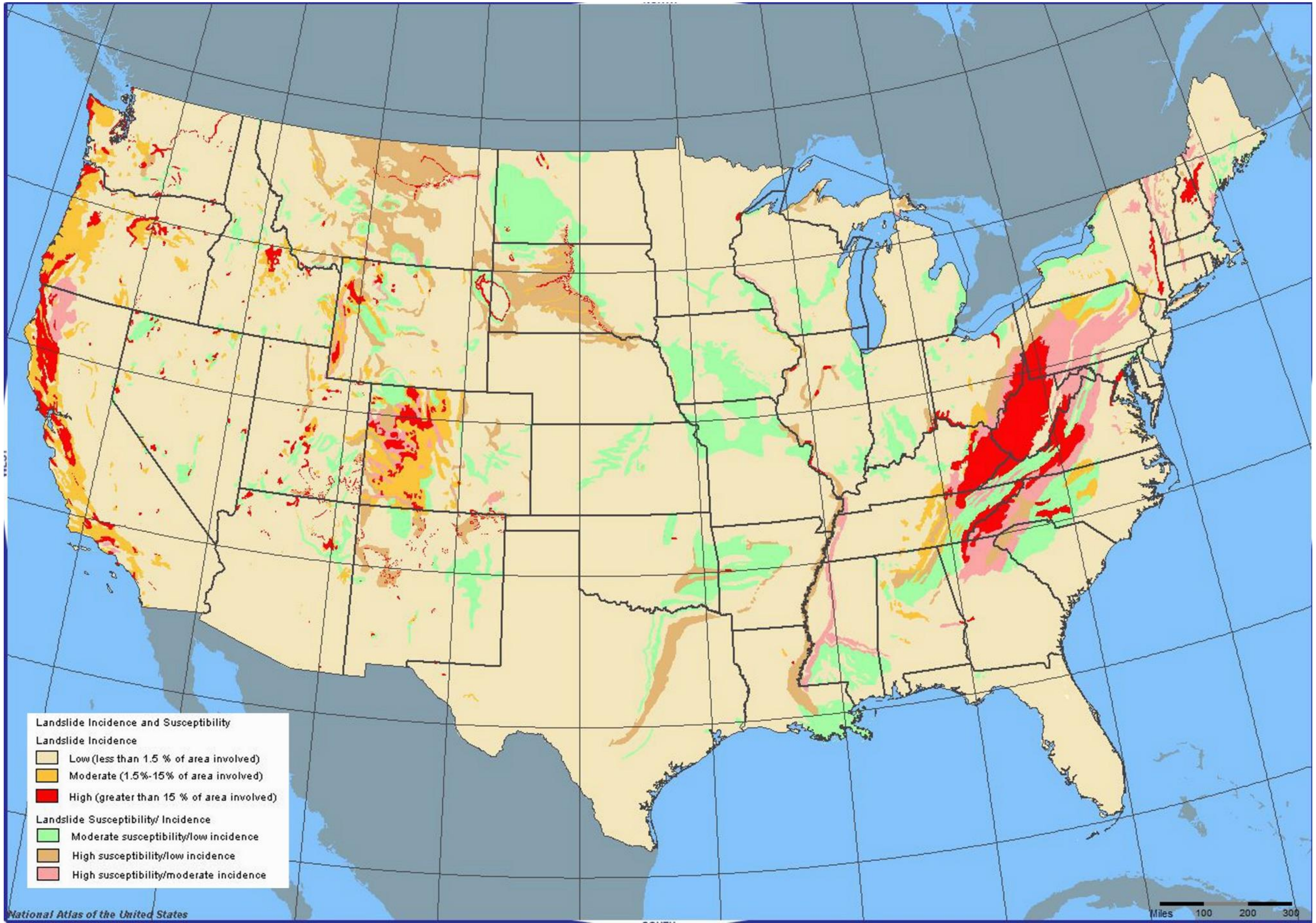
Ordinal point map: US Cities by population



USGS National Map road symbols



Ordinal area map: Likelihood and incidence of landslides

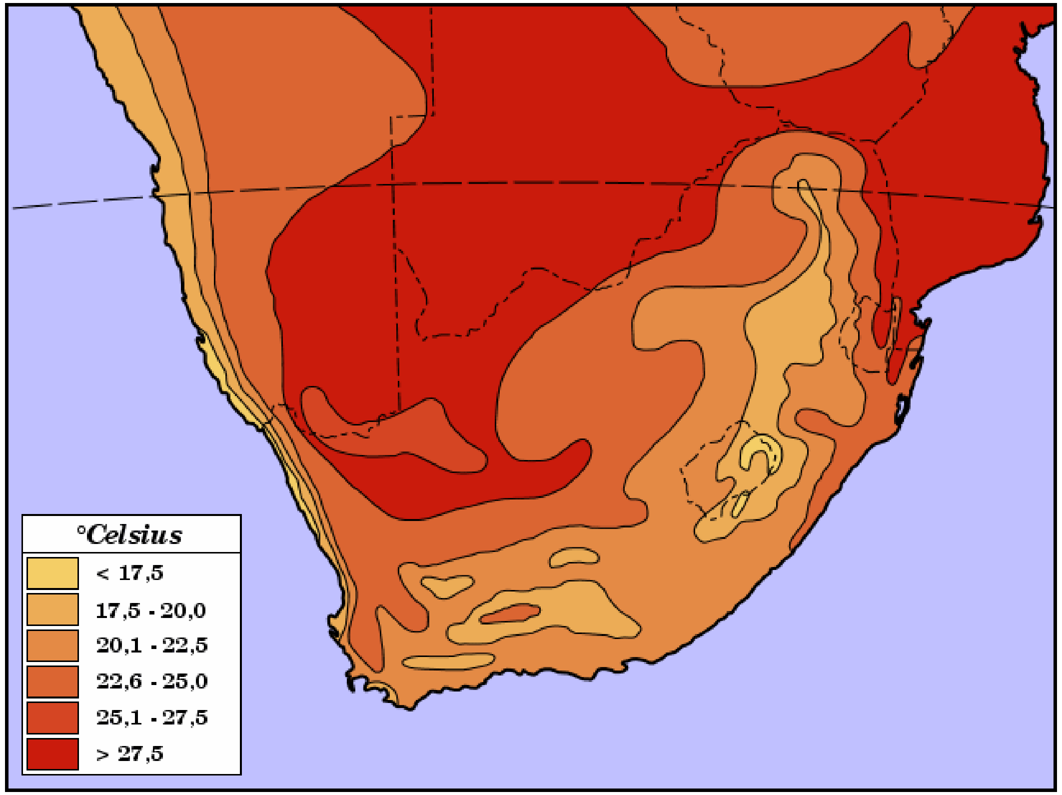


Ordinal

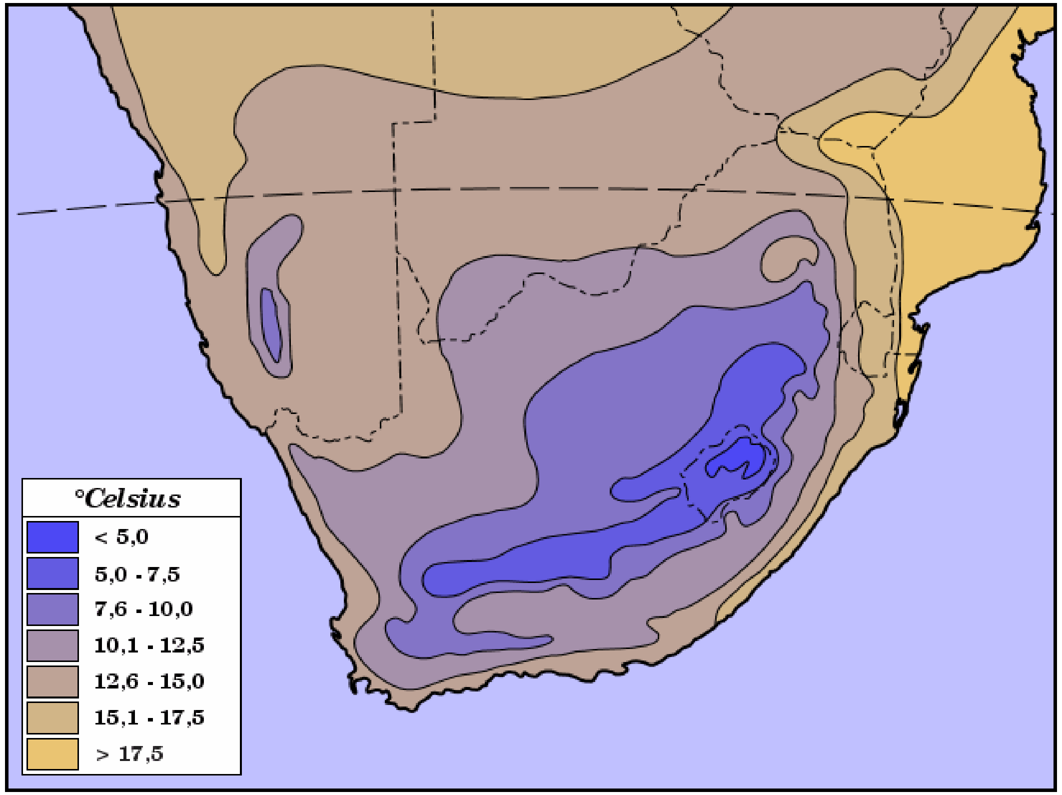
- Ordinal involves ranking
- One class or feature “above” or “below” another
- Point: Use symbols size, shape and color
- Line: Different symbols, line weights, colors
- Area: Color, pattern. Legends often high, medium, low or similar

Interval map types

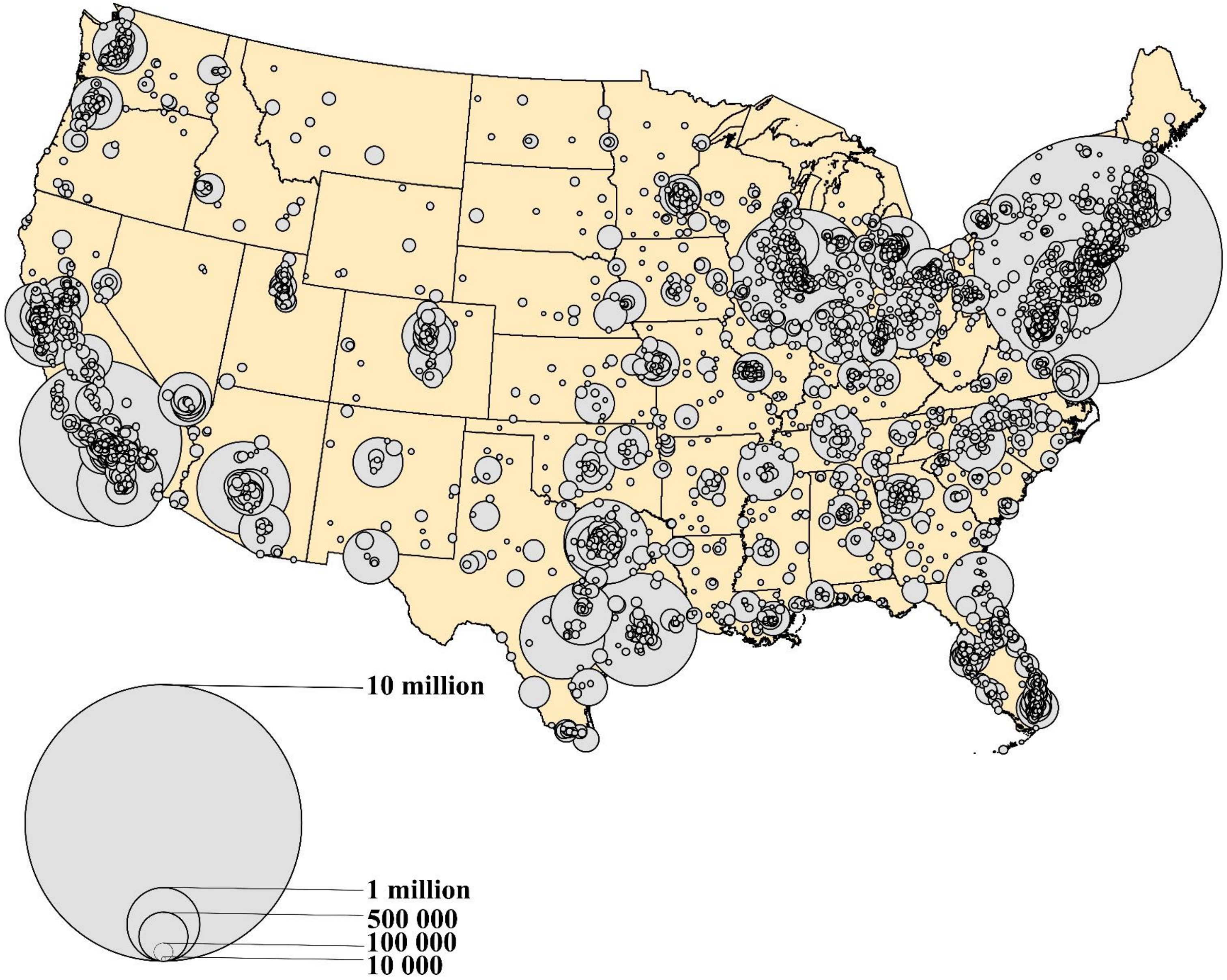
January



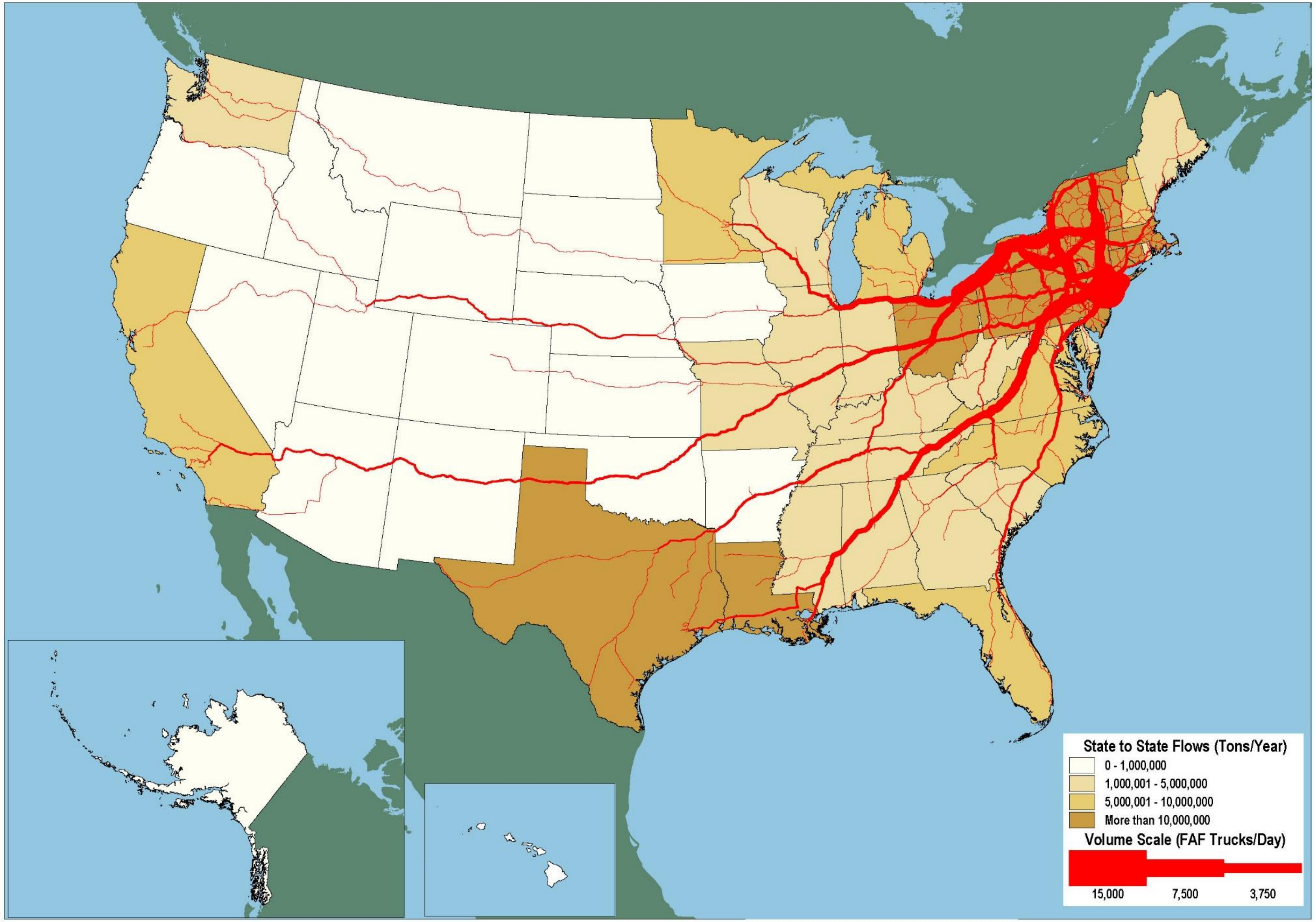
June



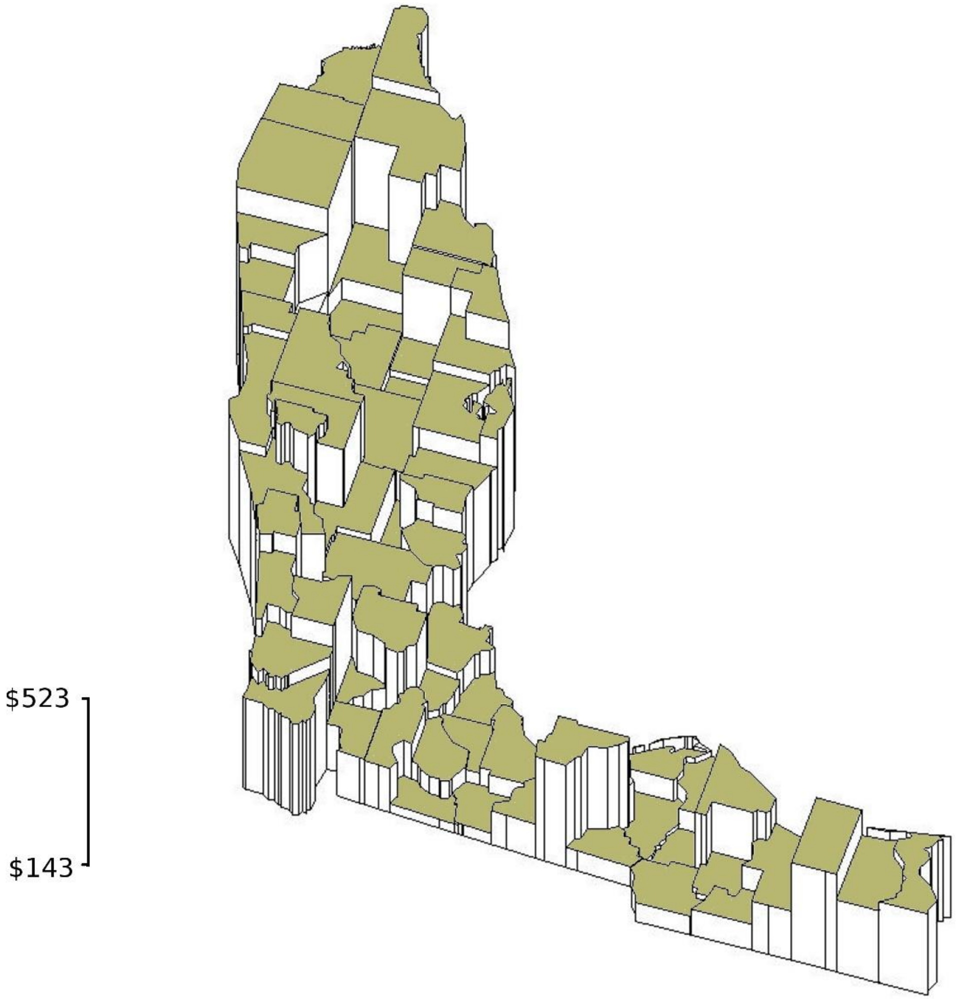
Interval point map: City size 1990 by proportional circle



Interval line map: Truck traffic to/from New York City 2007



Interval area map: Median gross rents in Florida 1990 (Prism map)



Interval

Numerical data, but on an arbitrary scale

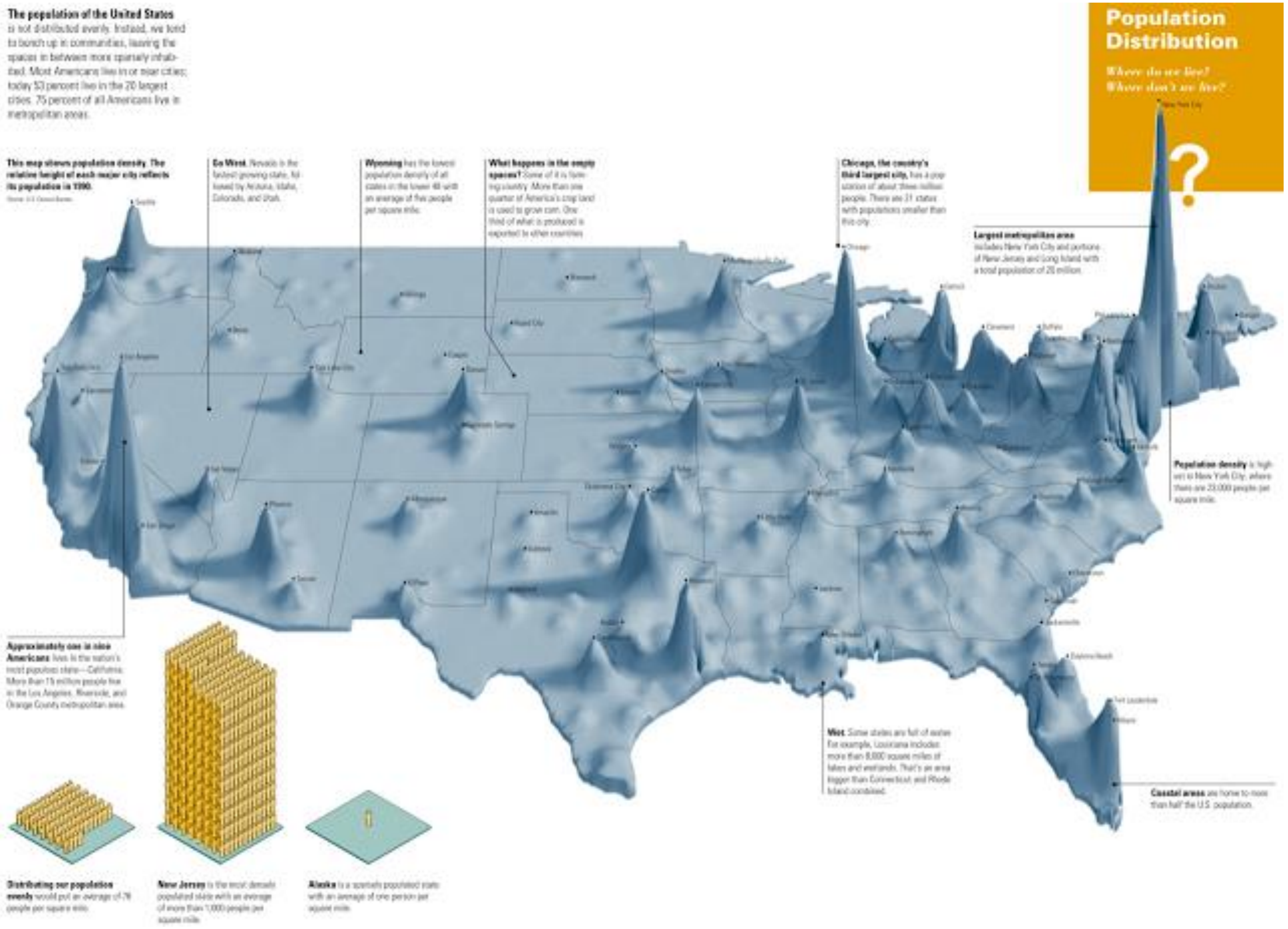
Often reflect “counts” e.g. total population

Point: Proportional symbol, usually geometric object, varies in size, sometimes classed

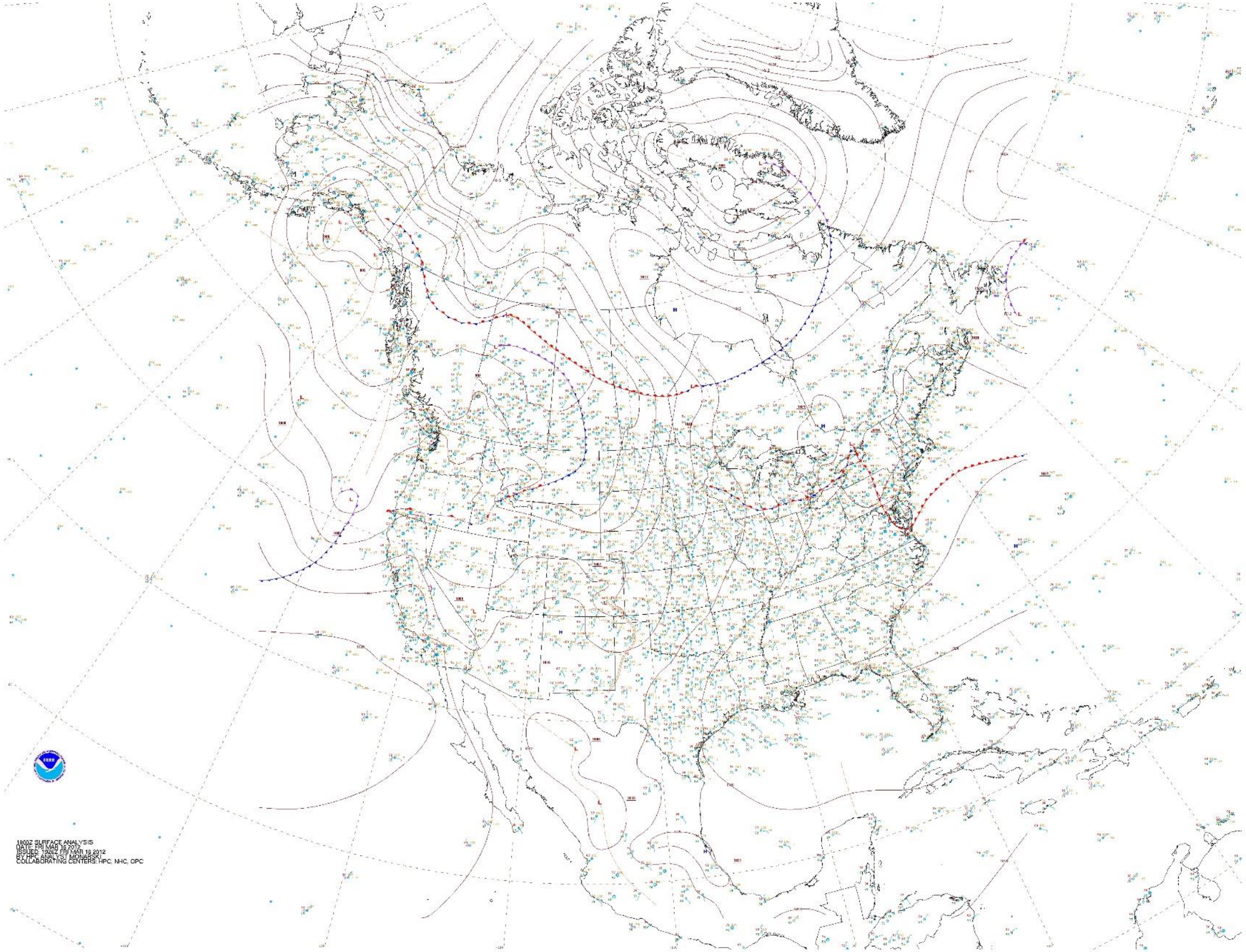
Line: Flow map, line width proportional to value

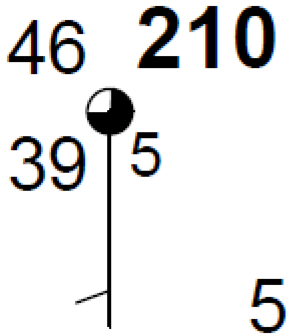
Area: Prism map, shaded map, choropleth

Ratio map types

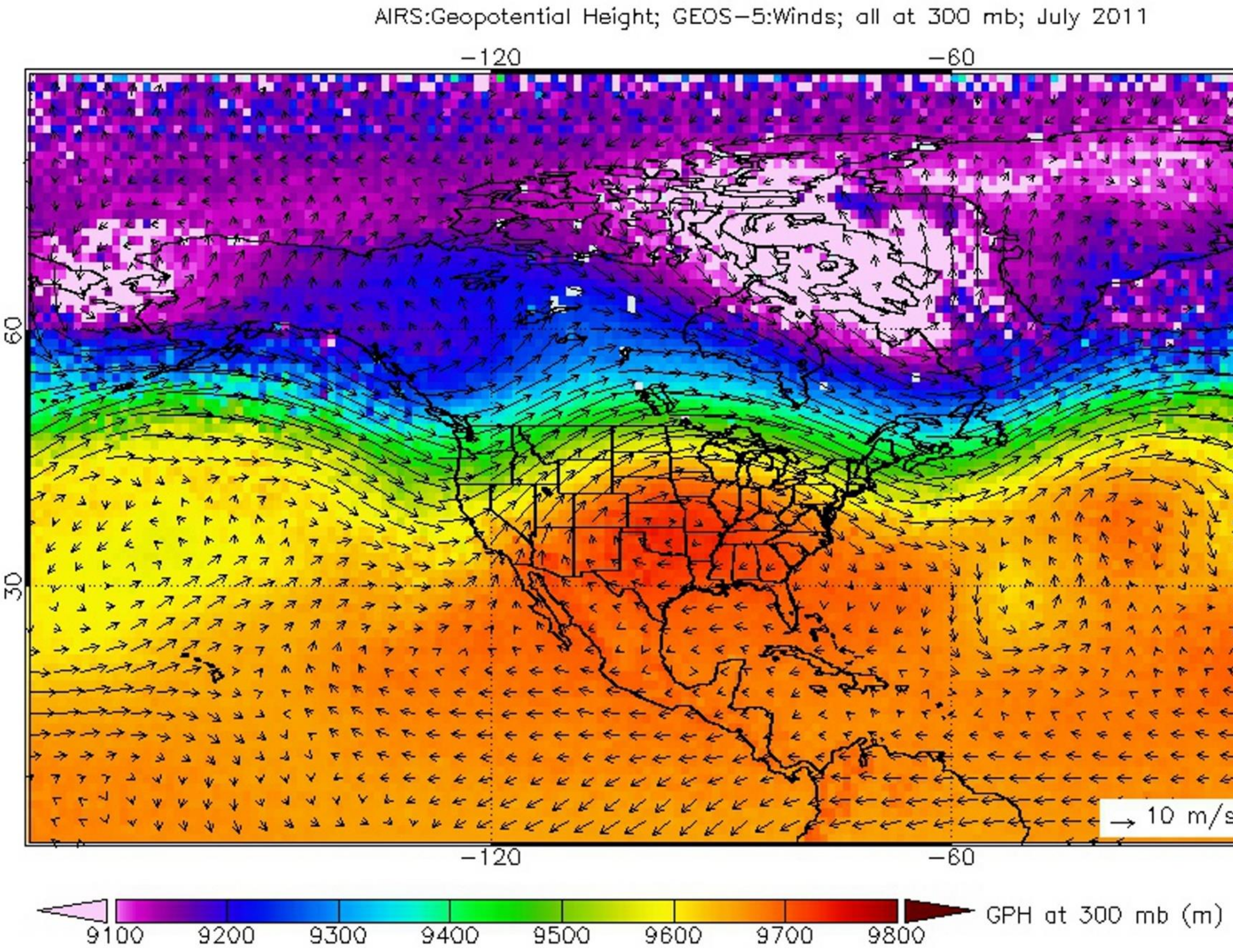


Point ratio map: Synoptic weather NWS (3/16/2012)

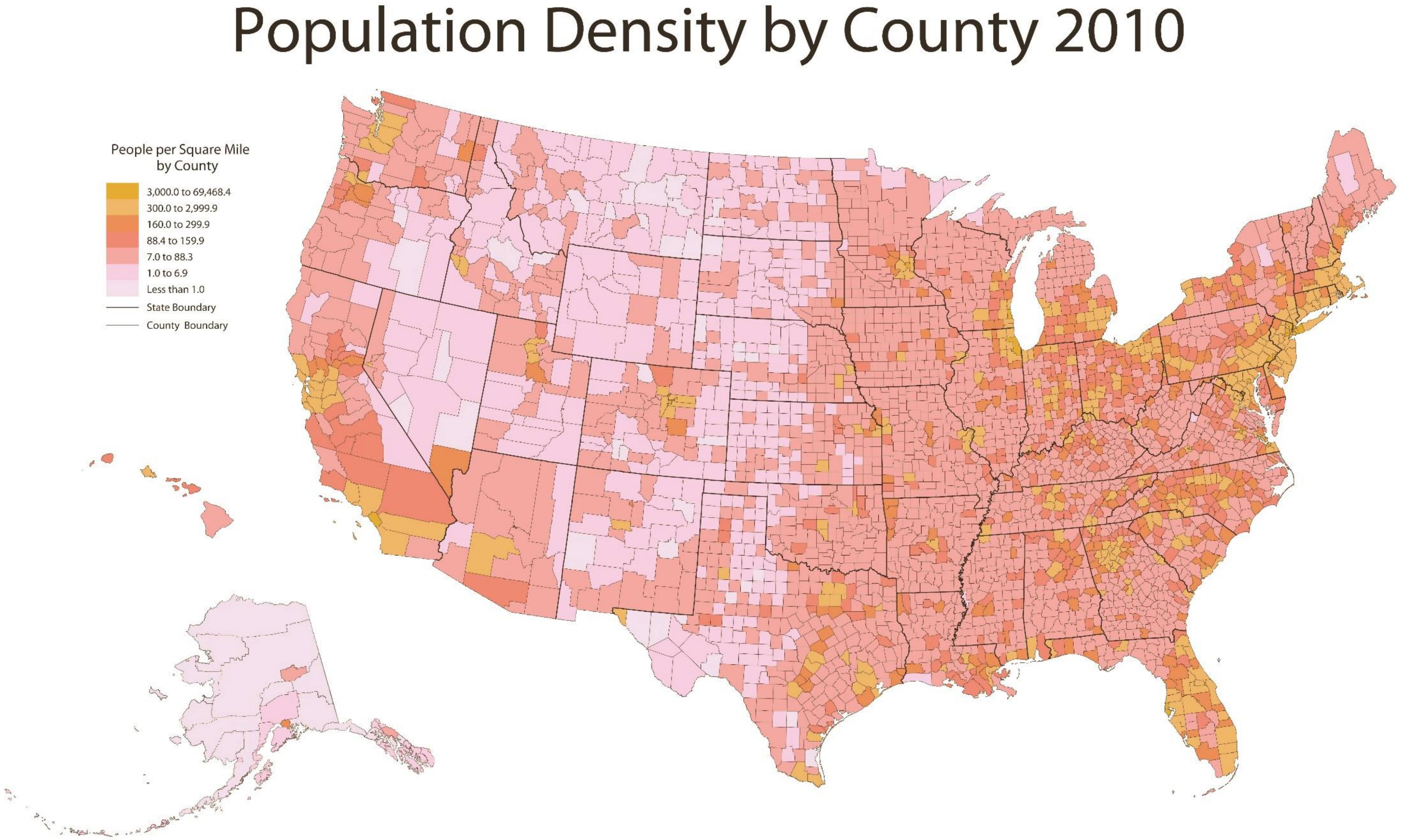




Line ratio map: Wind flow vectors



Area ratio map: Choropleth



Ratio data

- Numerical data value on a scale with an absolute zero
- Can be physical absolute (e.g. wind speed) or ratio of two numbers (people per square mile)
- Cartographic methods similar to interval
- Point: Compound point symbol with encoded data
- Line: Vectors, isolines
- Area: Choropleth and other methods, e.g. dasymetric

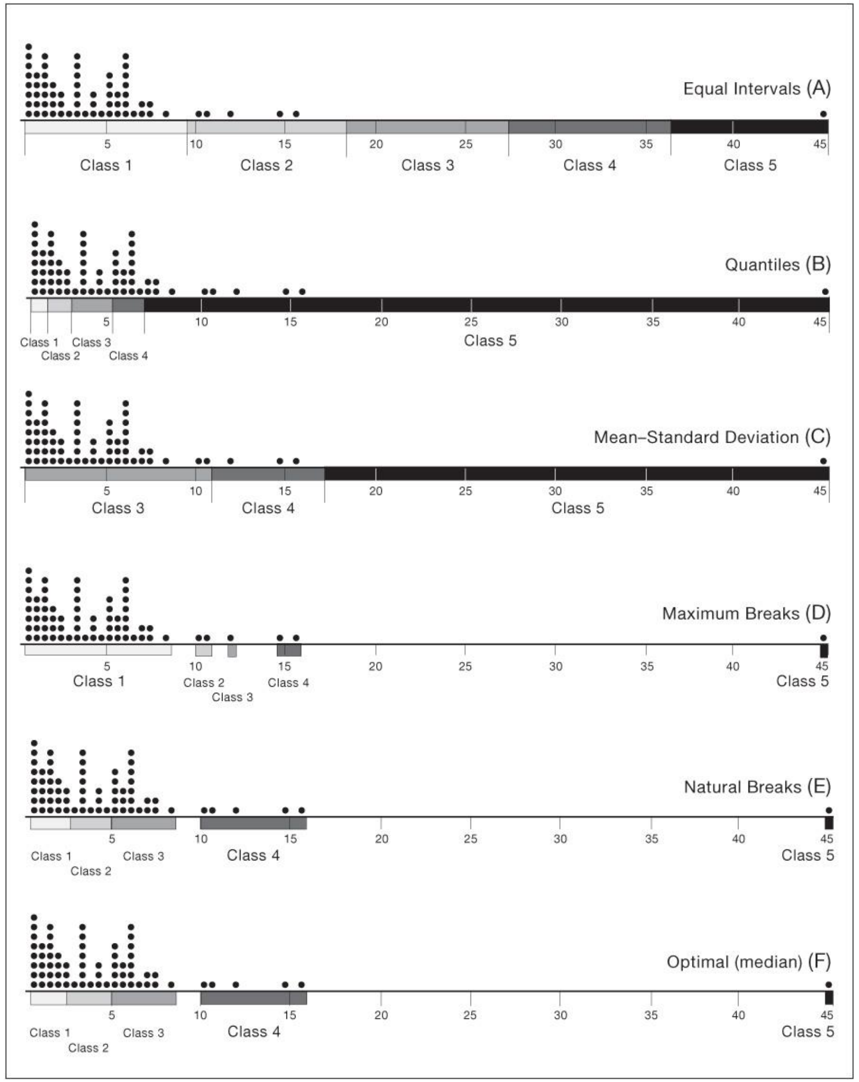
Classification

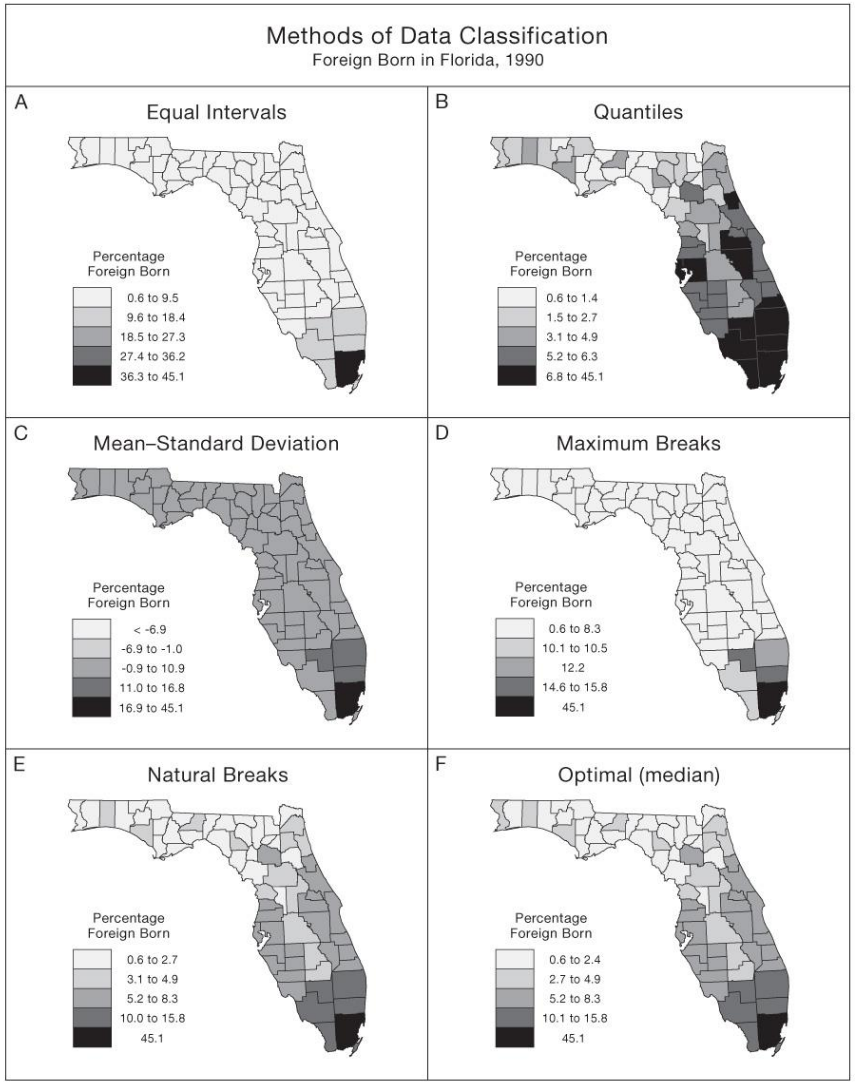
Slocum Ch. 5

Six common methods

- Equal intervals
- Quantiles
- Mean-standard deviation
- Natural breaks
- Optimal
    - Jenks-Caspall
    - Fisher-Jenks

Same distribution, different maps





Which method?

	Equal Interval	Quantiles	Mean SD	Maximum Breaks	Natural Breaks	Optimal
Considers distribution of data along a number line	P	P	G[a]	G	VG[b]	VG
Ease of understanding concept	VG	VG	VG	VG	G	G[c]
Ease of computation	VG	VG	VG	VG	VG	VG[d]
Ease of understanding legend	VG[e]	P[f]	G	P[f]	P[f]	P[f]
Legend values match range of data in a class	P	VG	P	VG	VG	VG
Acceptable for ordinal data	U	A	U	U	U	U
Assists in selecting number of classes	P	P	P	P	G	VG

P = Poor   G = Good   VG = Very Good   A = Acceptable   U = Unacceptable

[a] Rating would be poor if data are not normal.

[b] Although breaks are subjectively determined, the results are often similar to those obtained by the optimal method.

[c] Only a good rating is assigned because of the fairly complex nature of the algorithm.

[d] The optimal method does require the use of a computer.

[e] Only a good rating would be appropriate if round numbers are not used.

[f] Using rounded values may produce a good rating; some data distributions may mimic an equal interval map, thus producing a good or very good rating.

How many classes?

The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information[1]

George A. Miller (1956)

Harvard University

First published in *Psychological Review*, *63*, 81-97.

My problem is that I have been persecuted by an integer. For seven years this number has followed me around, has intruded in my most private data, and has assaulted me from the pages of our most public journals. This number assumes a variety of disguises, being sometimes a little larger and sometimes a little smaller than usual, but never changing so much as to be unrecognizable. The persistence with which this number plagues me is far more than a random accident. There is, to quote a famous senator, a design behind it, some pattern governing its appearances. Either there really is something unusual about the number or else I am suffering from delusions of persecution.

I shall begin my case history by telling you about some experiments that tested how accurately people can assign numbers to the magnitudes of various aspects of a stimulus. In the traditional language of psychology these would be called experiments in absolute judgment. Historical accident, however, has decreed that they should have another name. We now call them experiments on the capacity of people to transmit information. Since these experiments would not have been done without the appearance of information theory on the psychological scene, and since the results are analyzed in terms of the concepts of information theory, I shall have to preface my discussion with a few remarks about this theory.

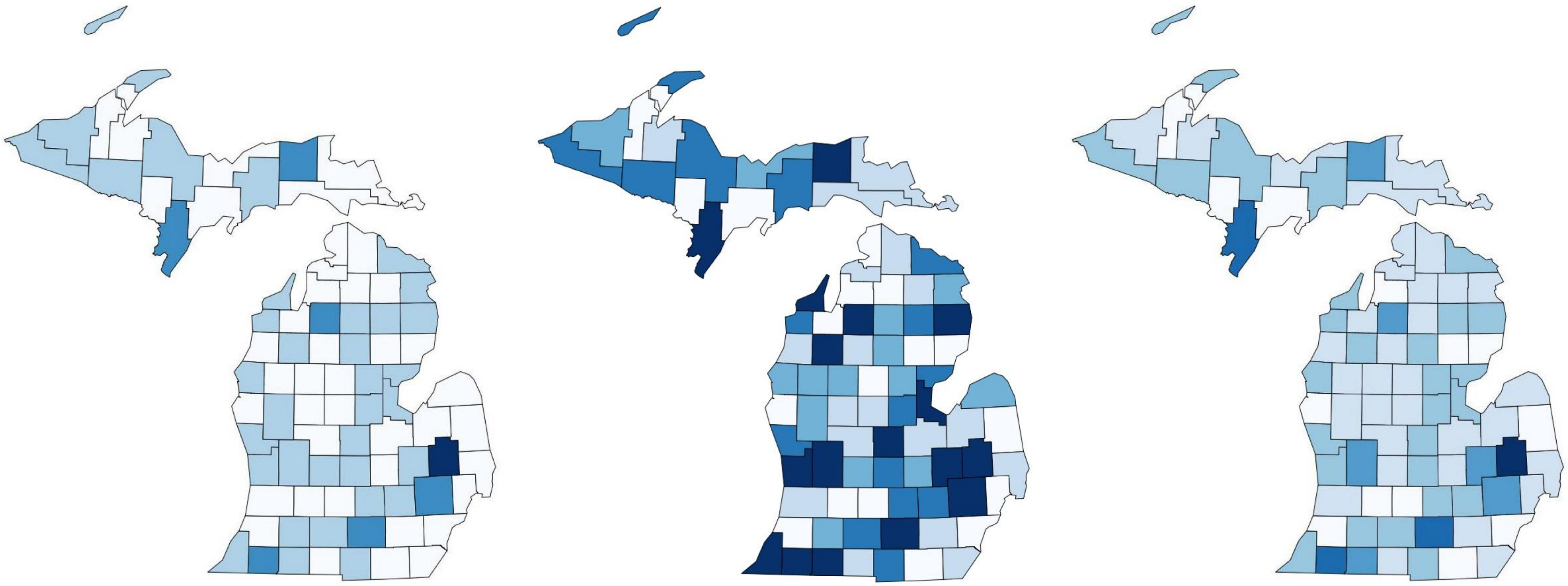
Information Measurement

The "amount of information" is exactly the same concept that we have talked about for years under the name of "variance." The equations are different, but if we hold tight to the idea that anything that increases the variance also increases the amount of information we cannot go far astray.

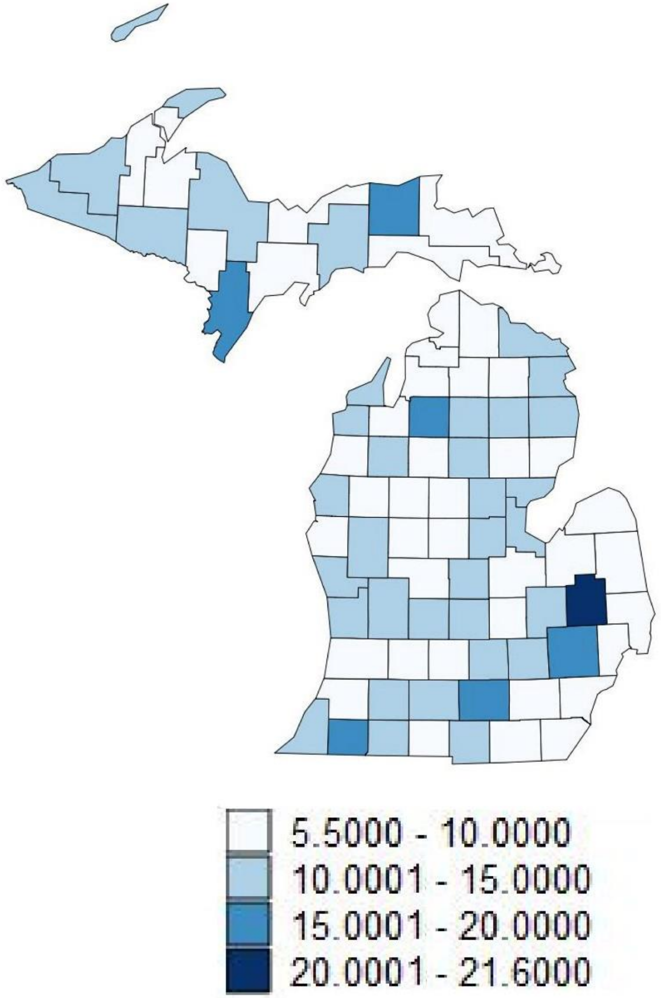
The advantages of this new way of talking about variance are simple enough. Variance is always stated in terms of the unit of measurement - inches, pounds, volts, etc. - whereas the amount of information is a dimensionless quantity. Since the information in a discrete statistical distribution does not depend upon the unit of measurement, we can extend the concept to situations where we have no metric and we would not ordinarily think of using [p. 82] the variance. And it also enables us to compare results obtained in quite different

- Are classes related? Sequenced?
- Nominal class vs. value
- George Miller (1956)
  - Short term memory capacity
  - 7 +/- 2
  - Five a safe bet!

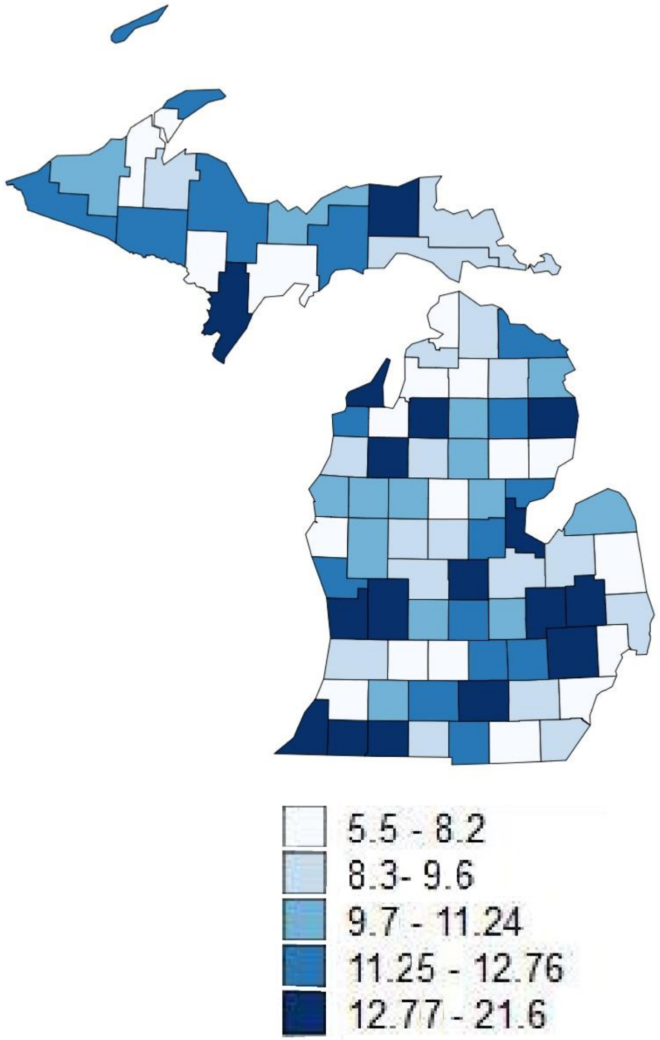
Michigan Percent Unemployment



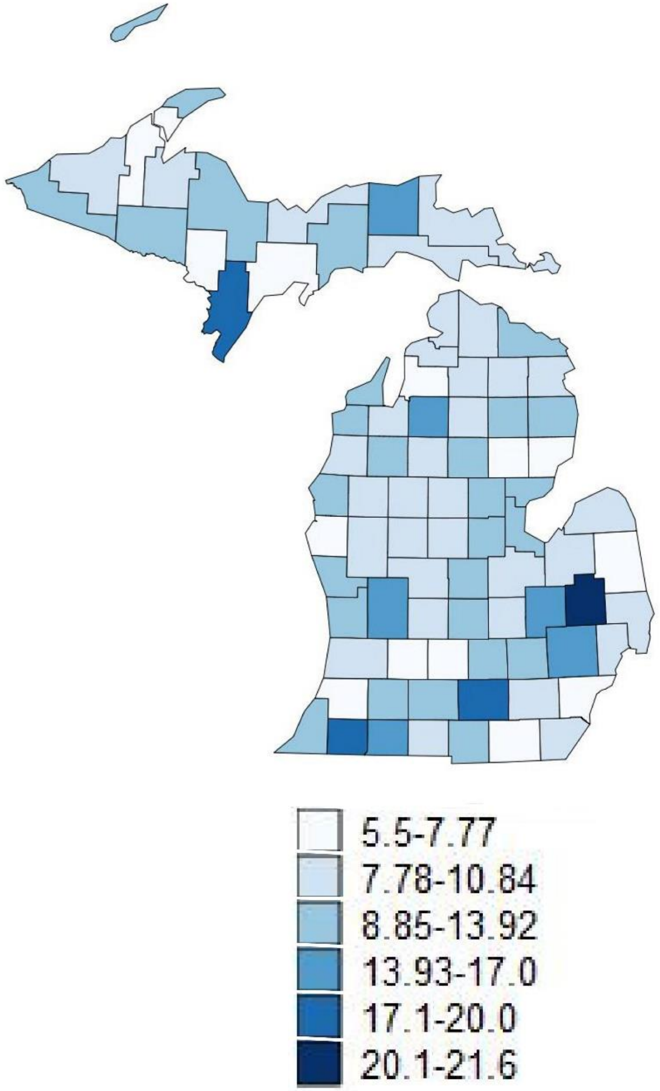
Michigan Percent Unemployment: March 2012



Equal interval
4 classes

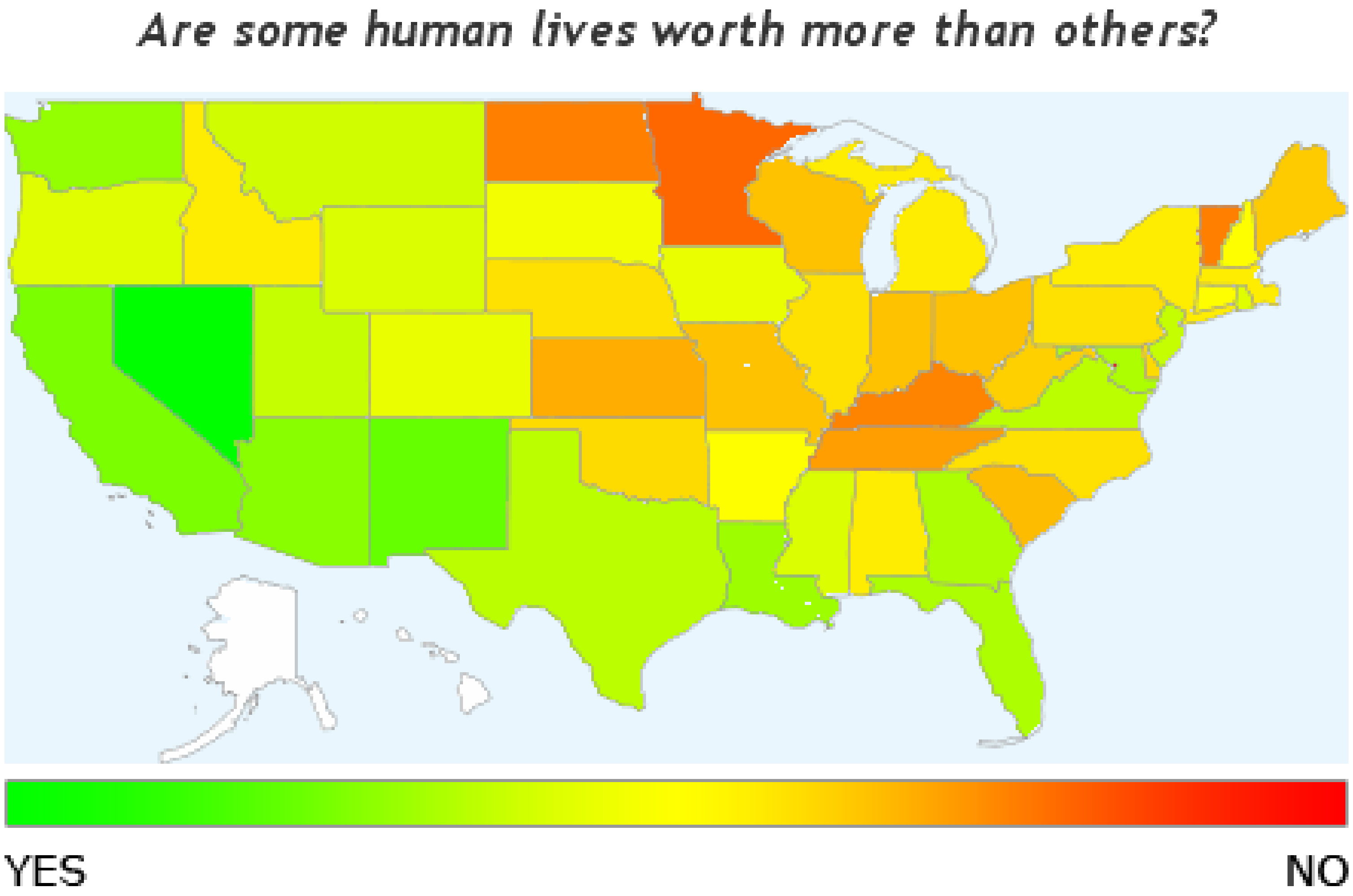


Quantile
5 classes



Standard Deviations
6 classes

Critique time



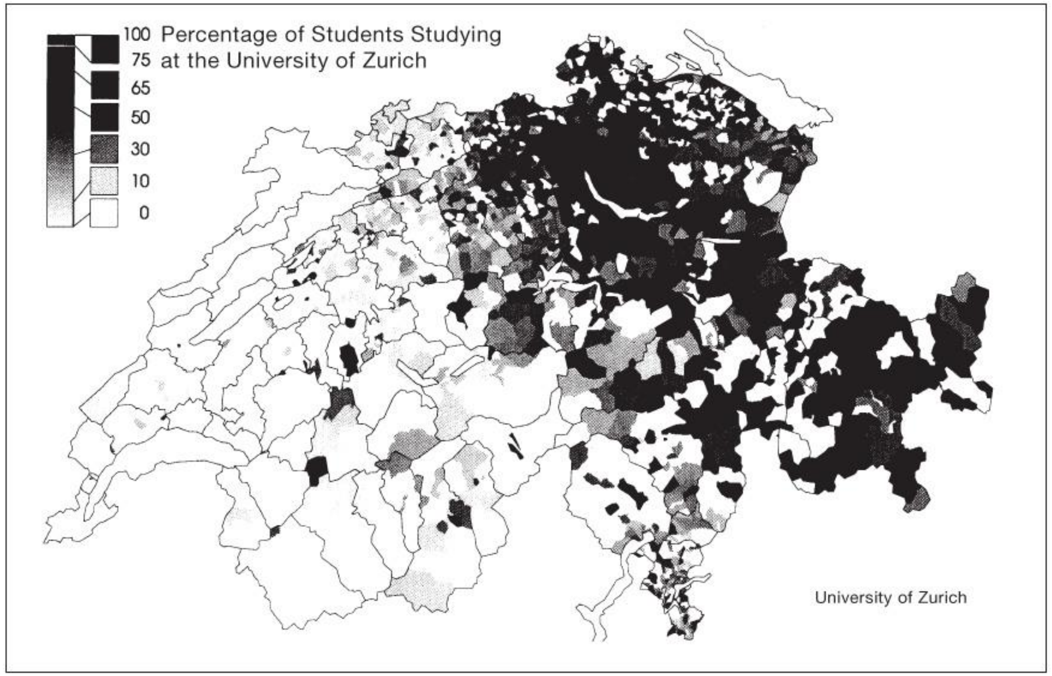
Source: OKTrends blog

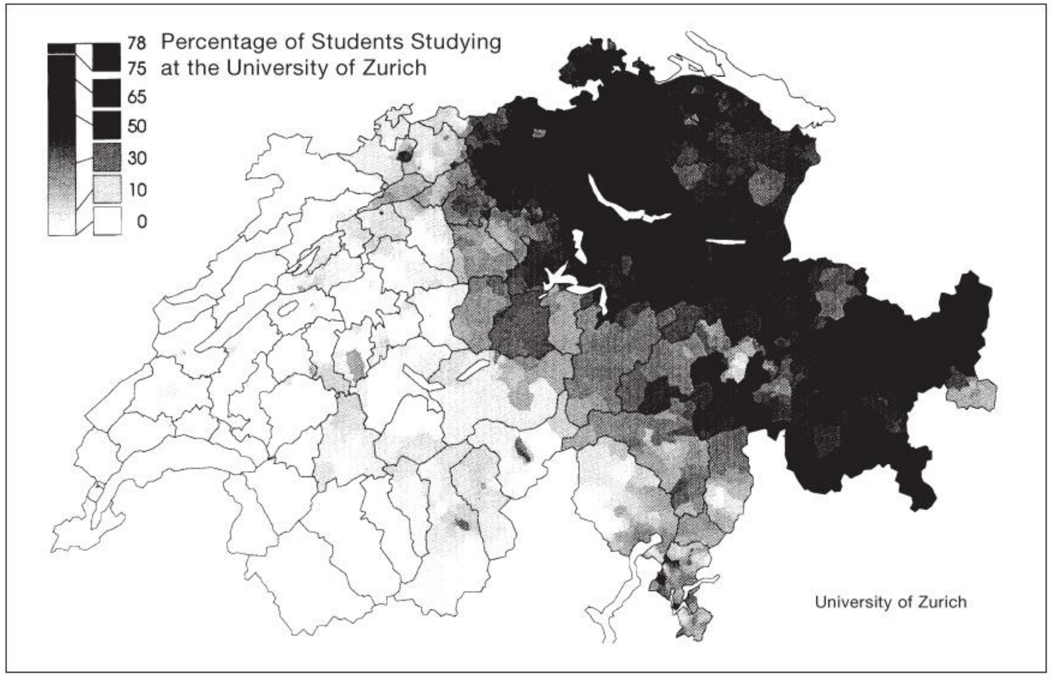
Context

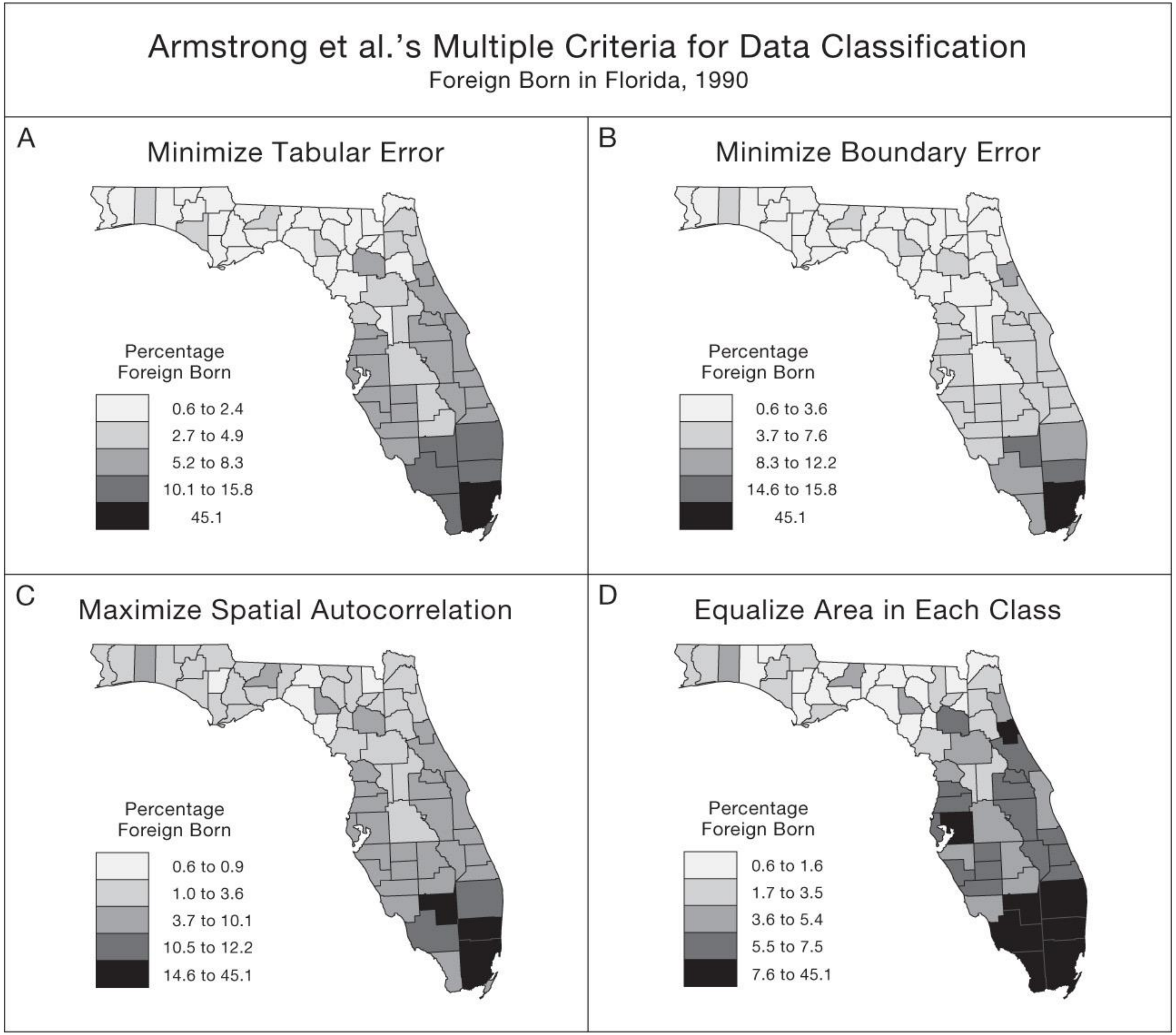
Adjust values based on surrounding values—smoothing

Proposed by Tobler (1973) and Herzog (1989)

Weighting method based on neighboring polygon values

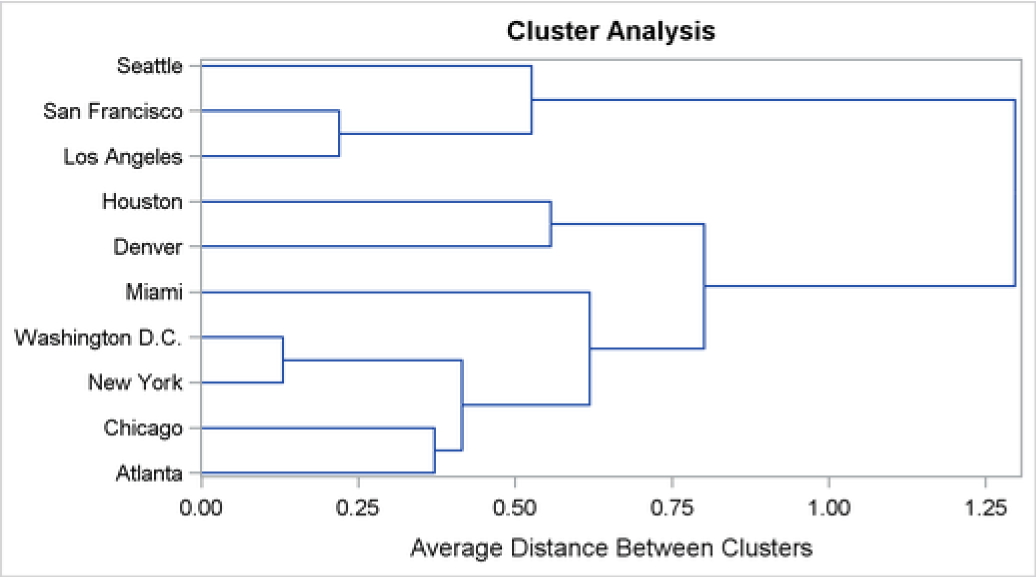


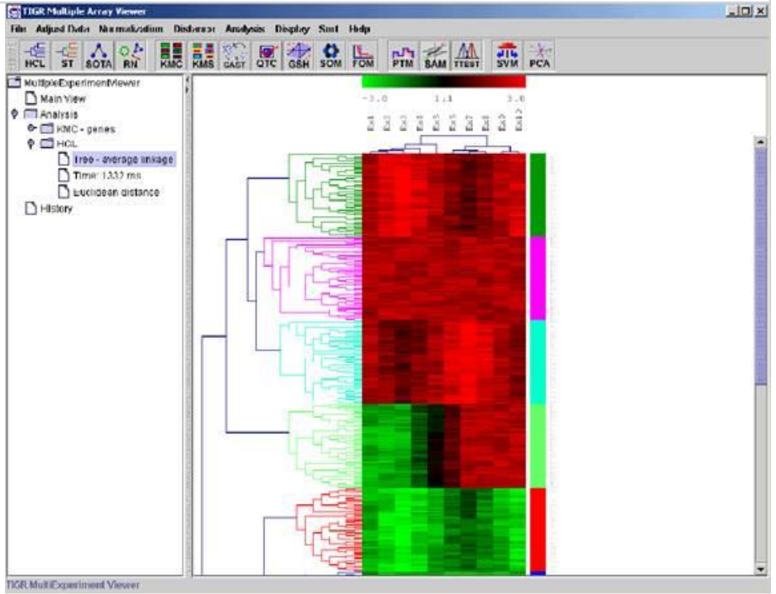


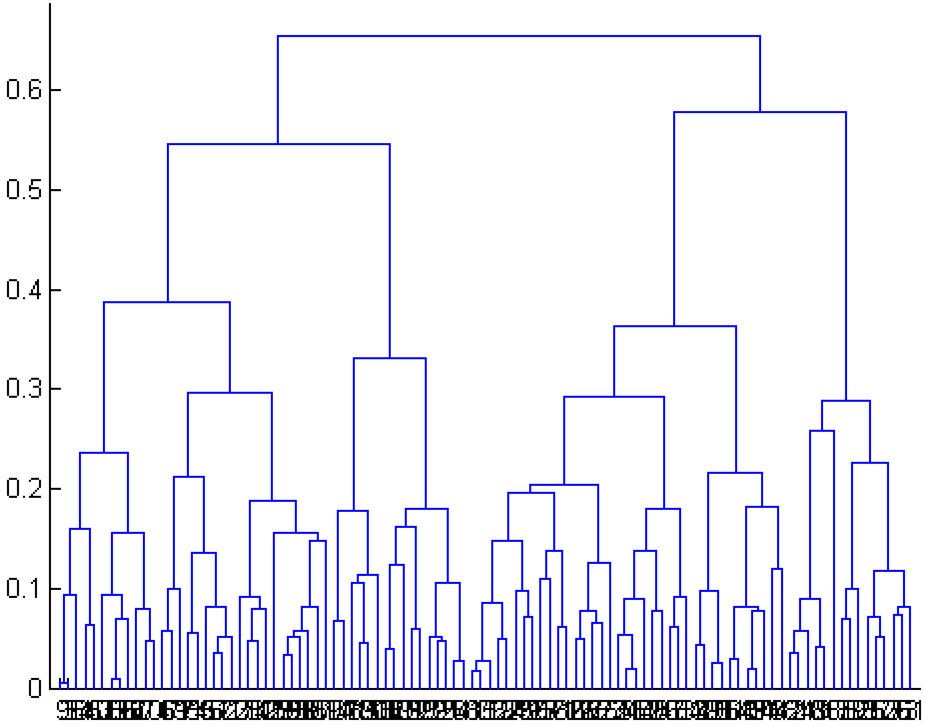


Armstrong et al.'s Multiple Criteria for Data Classification
Foreign Born in Florida, 1990
A
Minimize Tabular Error
Percentage Foreign Born
0.6 to 2.4
2.7 to 4.9
5.2 to 8.3
10.1 to 15.8
45.1
B
Minimize Boundary Error
Percentage Foreign Born
0.6 to 3.6
3.7 to 7.6
8.3 to 12.2
14.6 to 15.8
45.1
C
Maximize Spatial Autocorrelation
Percentage Foreign Born
0.6 to 0.9
1.0 to 3.6
3.7 to 10.1
10.5 to 12.2
14.6 to 45.1
D
Equalize Area in Each Class
Percentage Foreign Born
0.6 to 1.6
1.7 to 3.5
3.6 to 5.4
5.5 to 7.5
7.6 to 45.1

Clustering: Dendrogram







Summary

- Data types and map types
- Much dictated by continuity, data level and dimension of data
- Map type should be appropriate for data type
- Have covered classification methods and graphics
- Methods exist to optimize classification, statistically and visually
- Classification can impact map message, perception and interpretation