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
<table>
<thead>
<tr>
<th>Content scaling level</th>
<th>Defining relations</th>
<th>POINT</th>
<th>LINE</th>
<th>AREA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Equivalence</td>
<td>+</td>
<td>+</td>
<td>Wholesale and retail establishments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td></td>
<td>Highway connectivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Land ownership</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Equivalence Greater than</td>
<td>Medium</td>
<td></td>
<td>Roads by degree of improvement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small</td>
<td>Large</td>
<td>Crop yield</td>
</tr>
<tr>
<td>Interval</td>
<td>Equivalence Greater than Ratio of intervals</td>
<td>+ 147</td>
<td>+ 210</td>
<td>Spot elevations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ 132</td>
<td></td>
<td>Gratitude</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ 122</td>
<td></td>
<td>Date of settlement</td>
</tr>
<tr>
<td>Ratio</td>
<td>Equivalence Greater than Ratio of intervals Ratio of scale values</td>
<td>Area proportional to population</td>
<td>Population density isol e ths</td>
<td>Value proportional to population density</td>
</tr>
</tbody>
</table>

*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**

<table>
<thead>
<tr>
<th>Type</th>
<th>Point</th>
<th>Line</th>
<th>Area</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>City</td>
<td>Road</td>
<td>Name of unit</td>
<td>Precipitation or soil type</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Large city</td>
<td>Major road</td>
<td>Rich county</td>
<td>Heavy precipitation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Good soil</td>
</tr>
<tr>
<td>Interval</td>
<td>Total population</td>
<td>Traffic flow</td>
<td>Percapita income</td>
<td>Precipitation exchange</td>
</tr>
<tr>
<td>Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MAP TYPES**

<table>
<thead>
<tr>
<th>Type</th>
<th>Point</th>
<th>Line</th>
<th>Area</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>Dot map</td>
<td>Network map</td>
<td>Colored area map</td>
<td>Freely colored map</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Symbol map</td>
<td>Ordered network map</td>
<td>Ordered colored map</td>
<td>Ordered chromatic map</td>
</tr>
<tr>
<td>Interval</td>
<td>Graduated symbol map</td>
<td>Flowmap</td>
<td>Choropleth map</td>
<td>Contour map</td>
</tr>
<tr>
<td>Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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
On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Physiological 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 E (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 1936 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 excused for another year.

For its final report (1940) the committee chose a common basis for its conclusions, directing its arguments to a concrete example of a sensory scale. This was the firm scale of loudness (K. B. Stevens and H. Davis, Human, 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 votes 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. 281).

It is plain from this and from other statements by the committee that the real issue is the meaning of measurement. That, 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 involved 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. B. Campbell (Final Report, p. 304), 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 selecting explicitly (a) the various rules for the assignment of numerals, (b) the mathematical properties (or group structures) 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 and 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 3-6=4-1, and equal ratios, as 8/4=4/2. 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
Nominal maps types
Point nominal map: Airport delays
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

1990 City Population
- 50,000-250,000
- 250,001-1,000,000
- 1,000,001-2,500,000
- 2,500,001-7,800,000
USGS National Map road symbols

Transportation Features

- Interstate
- US Route
- State Route
- Ramp
- Local Road (Primary / Private / Public / Secondary / Service)
- Forest Service Primary Route
- Forest Service Passenger Route
- Forest Service High Clearance Route
- 4WD Road
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

°Celsius
< 17.5
17.5 - 20.0
20.1 - 22.5
22.6 - 25.0
25.1 - 27.5
> 27.5

°Celsius
< 5.0
5.0 - 7.5
7.6 - 10.0
10.1 - 12.5
12.6 - 15.0
15.1 - 17.5
> 17.5
Interval point map: City size 1990 by proportional circle
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

The population of the United States...
Point ratio map: Synoptic weather NWS (3/16/2012)
Line ratio map: Wind flow vectors
Area ratio map: Choropleth

Population Density by County 2010
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

Methods of Data Classification
Foreign Born in Florida, 1990

A. Equal Intervals
B. Quantiles
C. Mean–Standard Deviation
D. Maximum Breaks
E. Natural Breaks
F. Optimal (median)
<table>
<thead>
<tr>
<th></th>
<th>Equal Interval</th>
<th>Quantiles</th>
<th>Mean SD</th>
<th>Maximum Breaks</th>
<th>Natural Breaks</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Considers distribution of data</td>
<td>P</td>
<td>P</td>
<td>G&lt;sup&gt;a&lt;/sup&gt;</td>
<td>G</td>
<td>VG&lt;sup&gt;b&lt;/sup&gt;</td>
<td>VG</td>
</tr>
<tr>
<td>along a number line</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of understanding</td>
<td>VG</td>
<td>VG</td>
<td>VG</td>
<td>VG</td>
<td>G</td>
<td>G&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of computation</td>
<td>VG</td>
<td>VG</td>
<td>VG</td>
<td>VG</td>
<td>VG</td>
<td>VG&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of understanding legend</td>
<td>VG&lt;sup&gt;e&lt;/sup&gt;</td>
<td>P&lt;sup&gt;f&lt;/sup&gt;</td>
<td>G</td>
<td>P&lt;sup&gt;f&lt;/sup&gt;</td>
<td>P&lt;sup&gt;f&lt;/sup&gt;</td>
<td>P&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legend values match range of data</td>
<td>P</td>
<td>VG</td>
<td>P</td>
<td>VG</td>
<td>VG</td>
<td>VG</td>
</tr>
<tr>
<td>in a class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acceptable for ordinal data</td>
<td>U</td>
<td>A</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assists in selecting number of</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>G</td>
<td>VG</td>
</tr>
<tr>
<td>classes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<sup>a</sup> Rating would be poor if data are not normal.

<sup>b</sup> Although breaks are subjectively determined, the results are often similar to those obtained by the optimal method.

<sup>c</sup> Only a good rating is assigned because of the fairly complex nature of the algorithm.

<sup>d</sup> The optimal method does require the use of a computer.

<sup>e</sup> Only a good rating would be appropriate if round numbers are not used.

<sup>f</sup> 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.

Copyright © 2009 Pearson Prentice Hall, Inc.
How many classes?

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

---

The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information

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...
Michigan Percent Unemployment
Critique time

Are some human lives worth more than others?

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

B. Minimize Boundary Error

C. Maximize Spatial Autocorrelation

D. Equalize Area in Each Class

Copyright © 2009 Pearson Prentice Hall, Inc.
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