

FIT3179 Data Visualisation

Data Classification

Reading

- **Required reading**

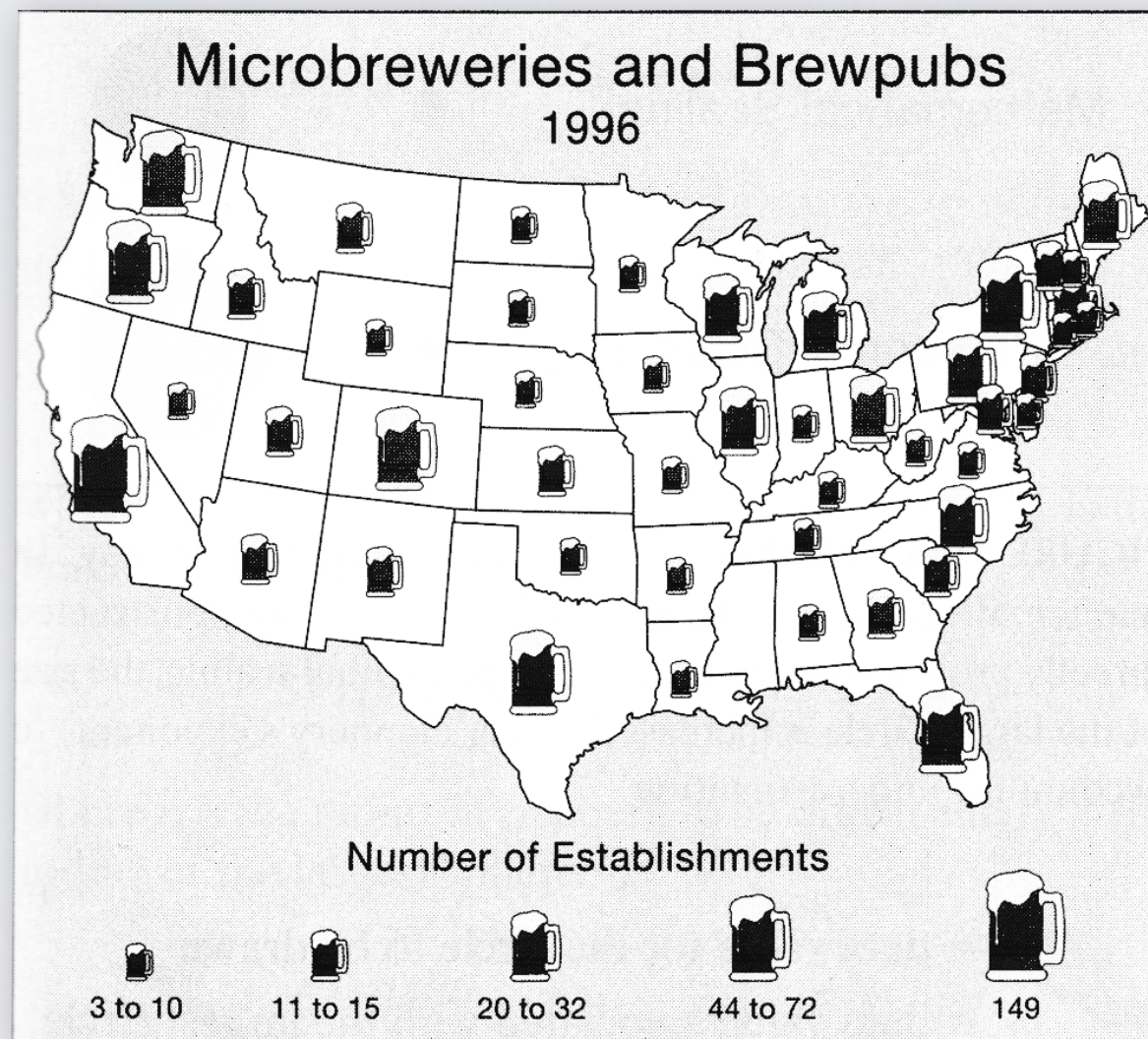
axismaps, The Basics of Data Classification, online: <http://axismaps.github.io/thematic-cartography/articles/classification.html>

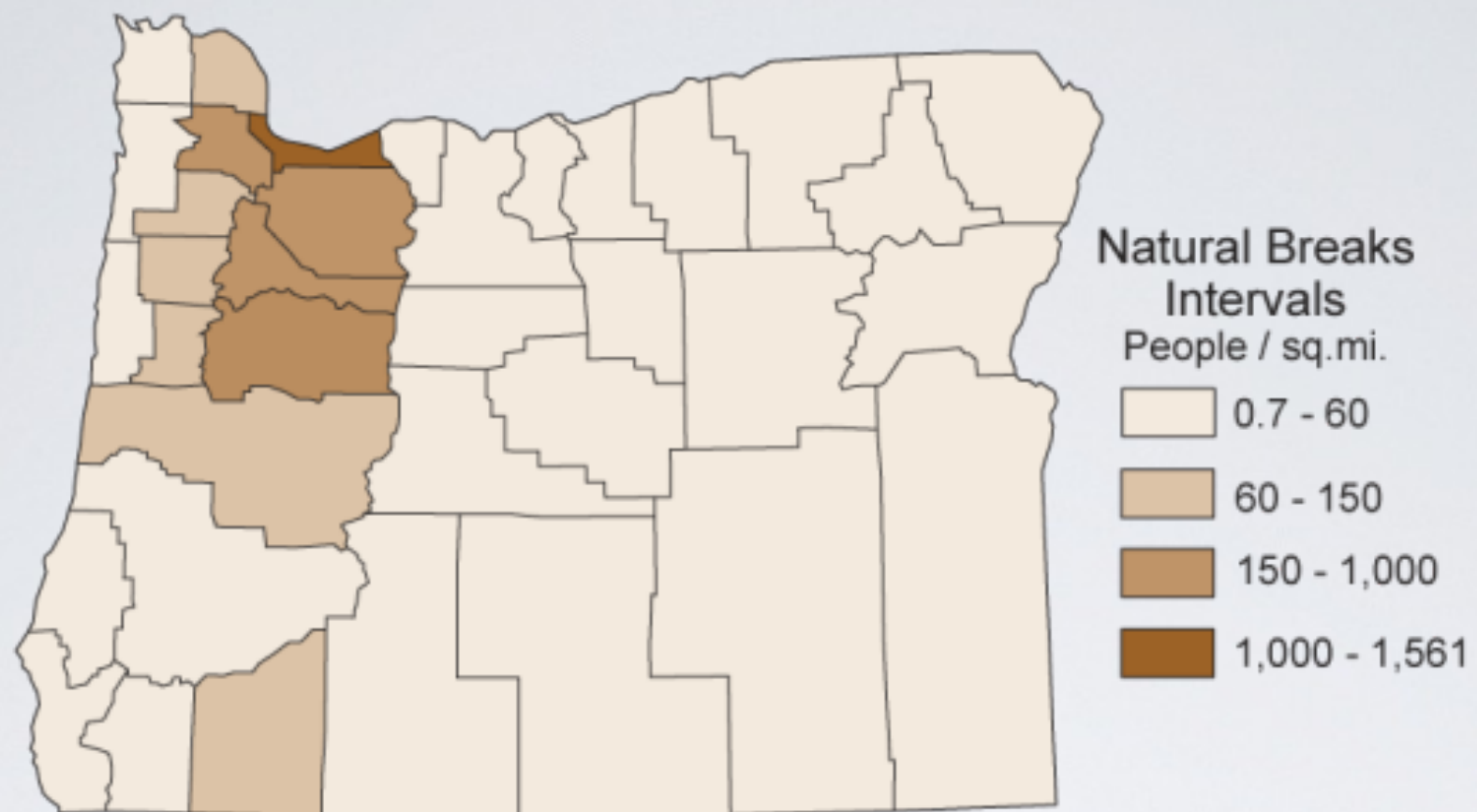
- **Recommended reading**

Slocum, T. et al. 2005. "Chapter 5: Data Classification." Thematic cartography and geographic visualization, Second Edition.

Data Classification

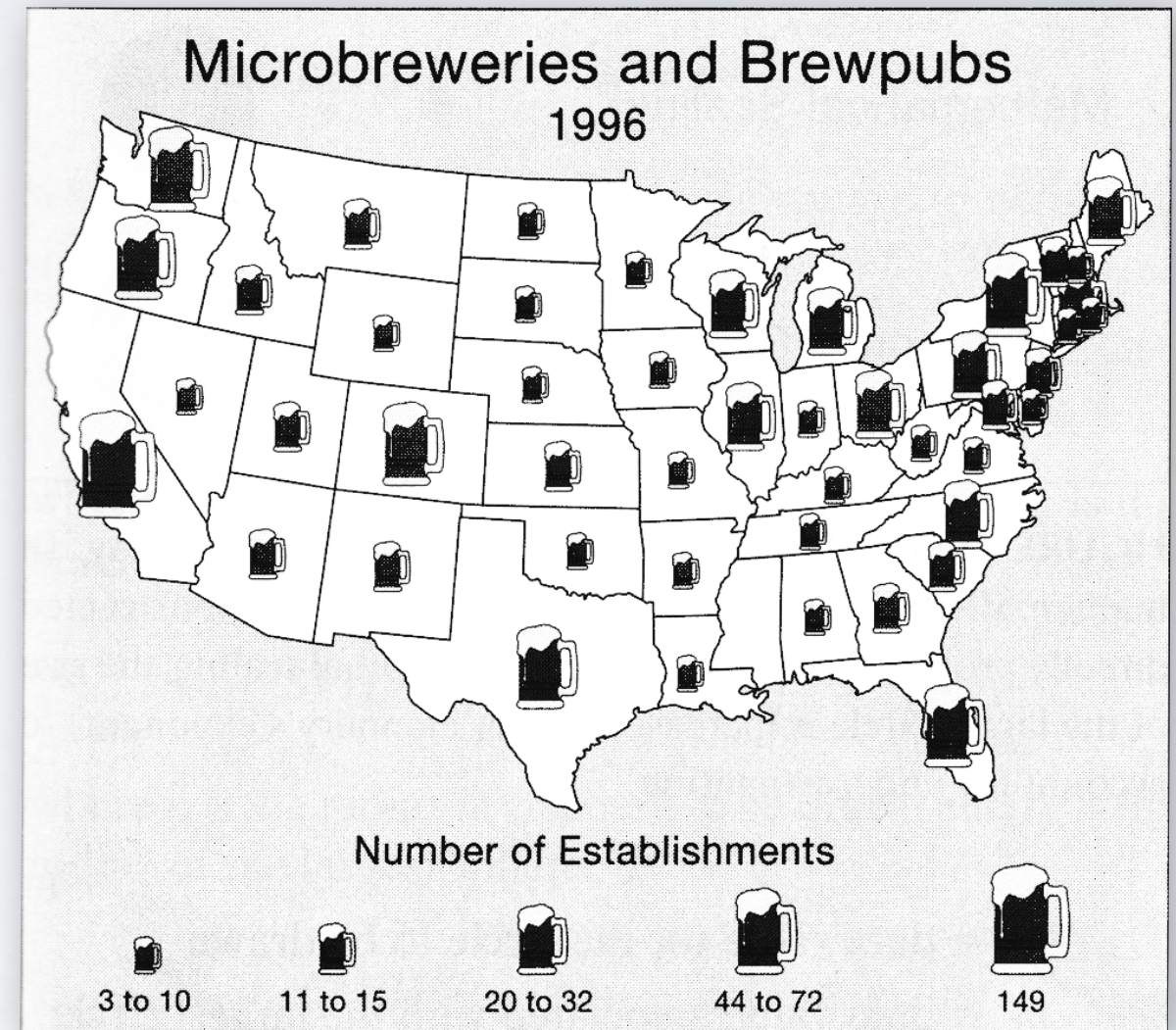
- For diagrams, choropleth maps, proportional symbols, flow lines, etc.
- For quantitative data: reduce from quantitative to ordinal





Data Classification

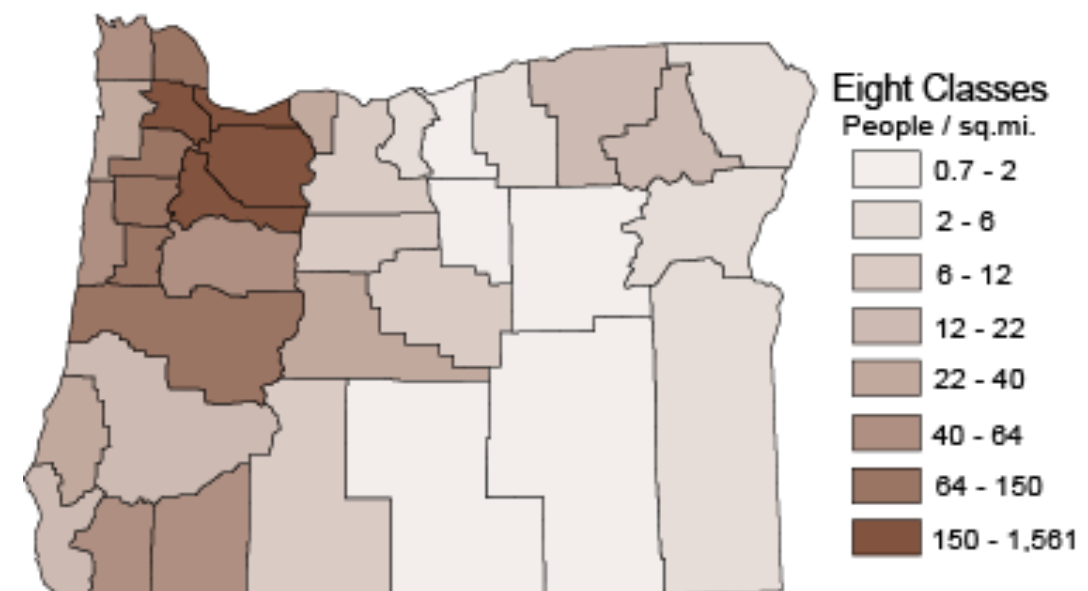
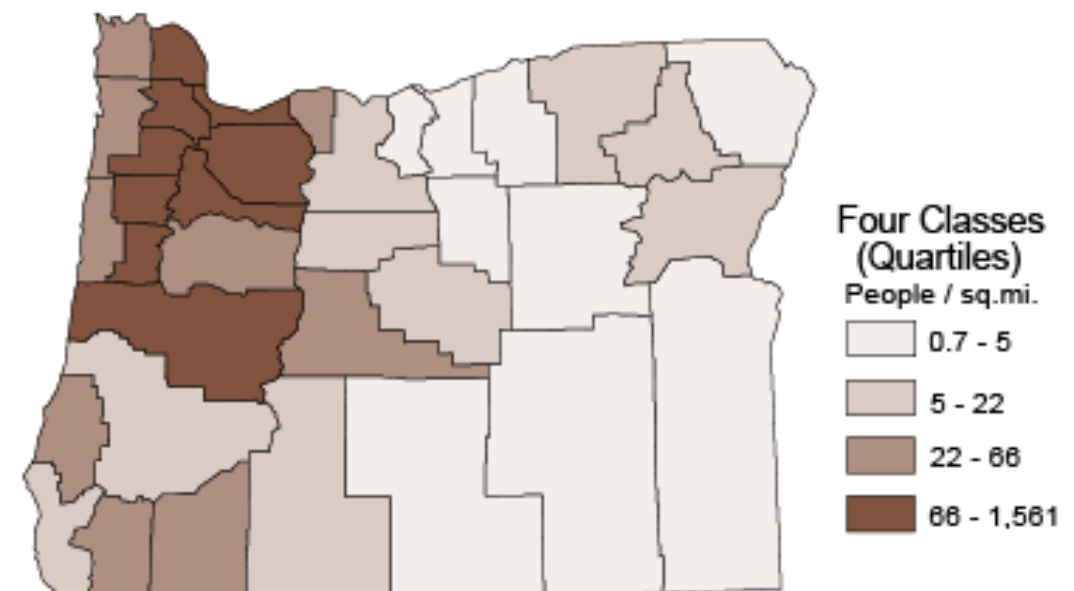
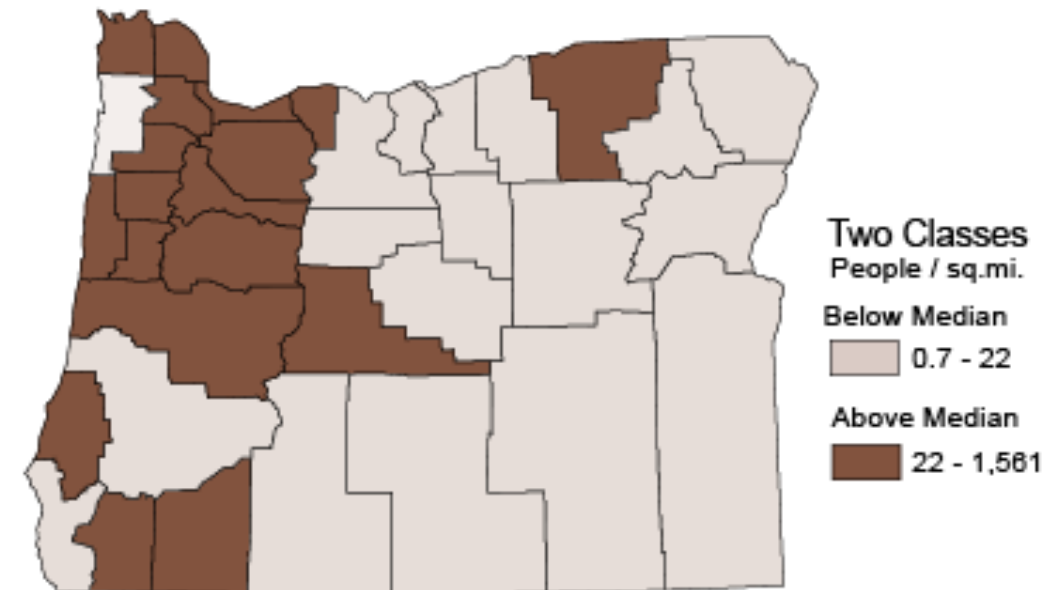
- Why classify data?
 - Simplify data to make visualisation easier to read.
 - Clarify the message.
 - Show trends.
- Two questions:
 - How many classes?
 - What class limits?



Number of Classes

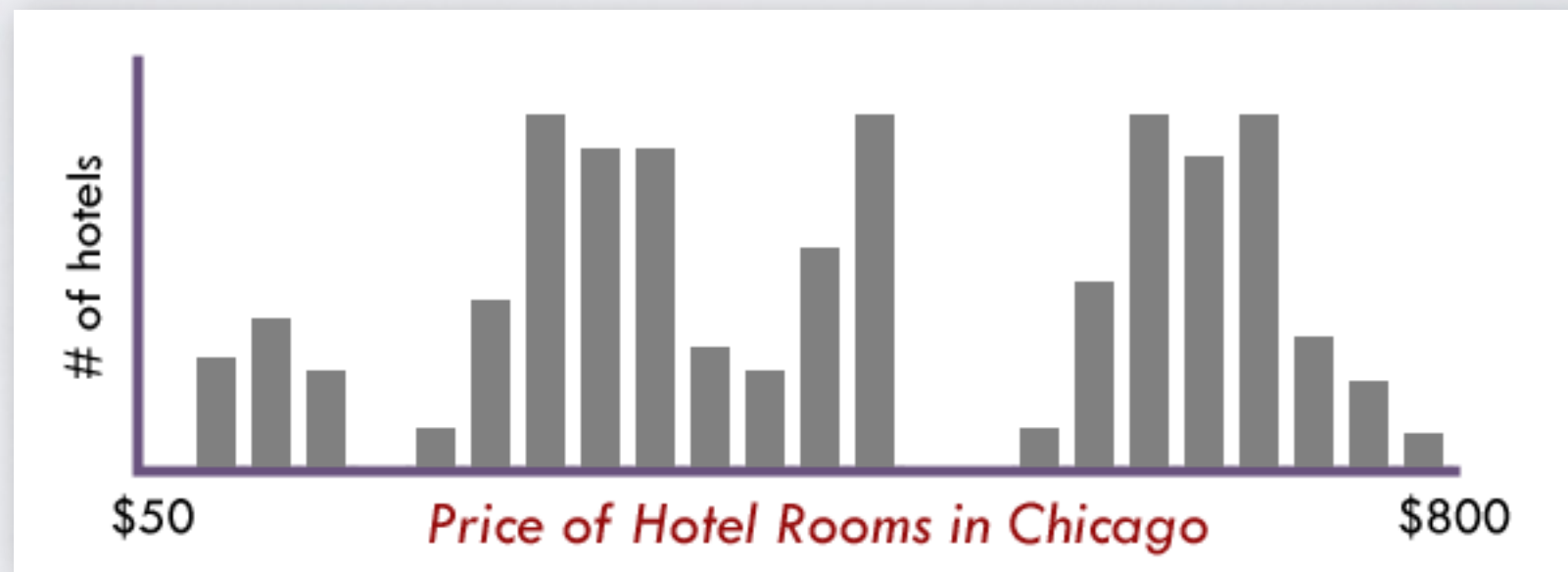
- Q: How many classes?
- A: Normally not more than 7 or 8.
The more classes, the more difficult a mark is to match with the legend.
- Fewer classes: visualisation
 - vis easier to read,
 - vis easier to remember,
 - clearer pattern,
 - but loss of details and information.

Oregon Population Density -- 2000



Data Classification Methods

- Goal: group together similar observations and split apart observations that are substantially different.
- Minimise within-group variance and maximise between-group differences
- Identify gaps in the histogram of your data

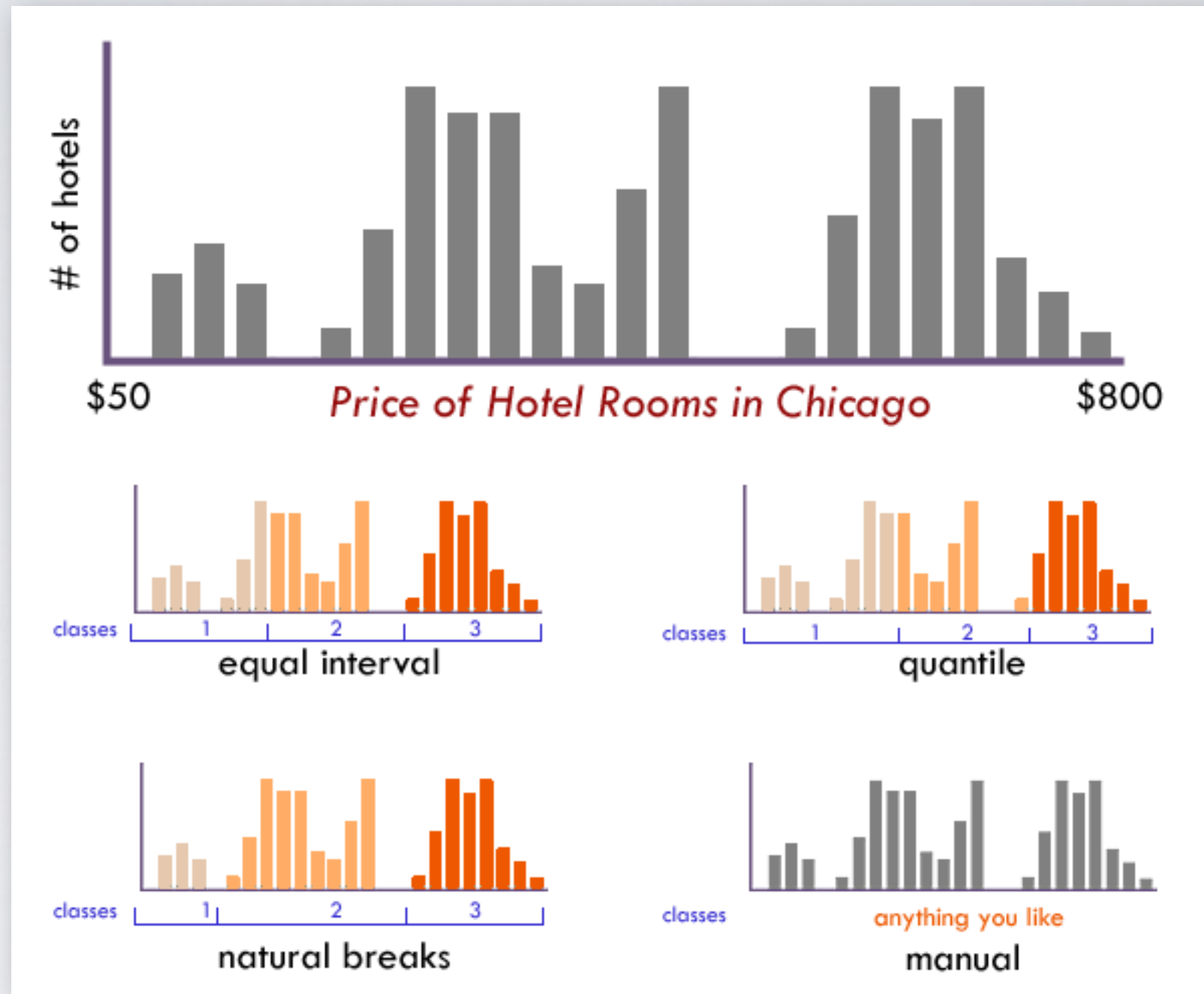


<http://axismaps.github.io/thematic-cartography/articles/classification.html>

Class Breaks: Guidelines

- Group similar values in one class
 - Show clusters and extreme values
 - Avoid empty classes if possible
 - No overlap between classes
 - Avoid gaps between classes, as they are confusing
-

Data Classification Methods

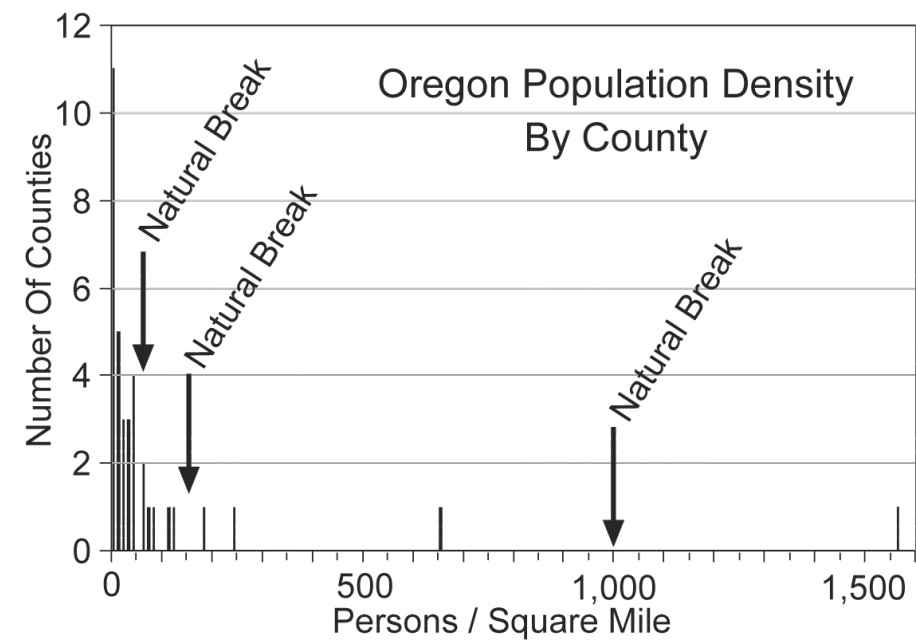
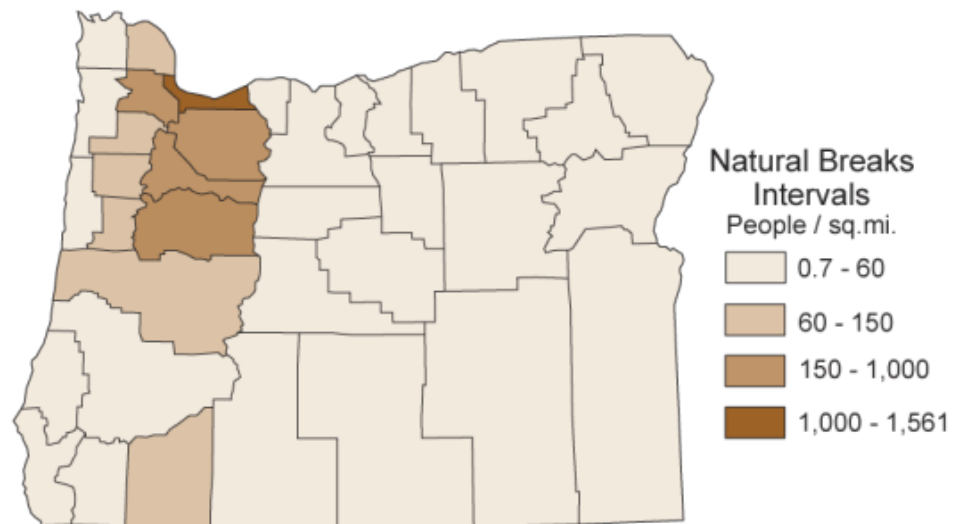
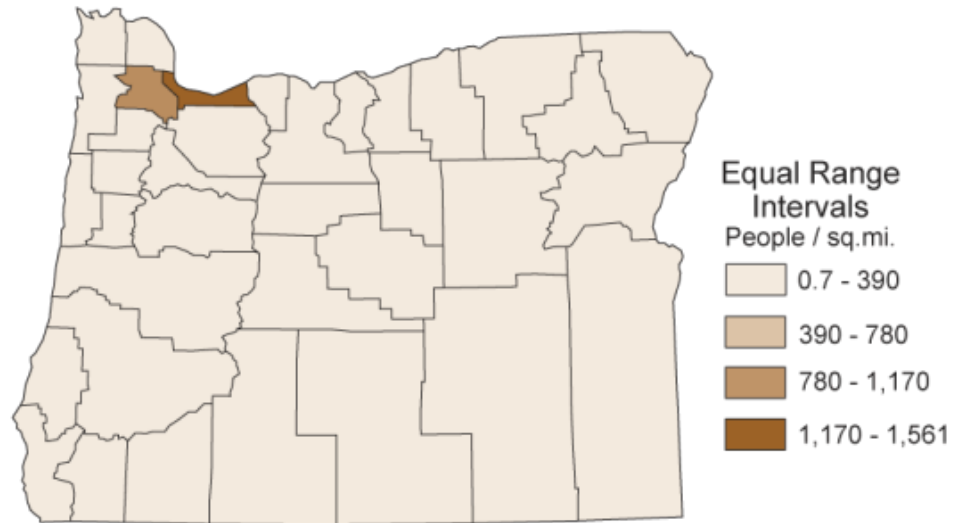


<http://axismaps.github.io/thematic-cartography/articles/classification.html>

Data Classification Methods

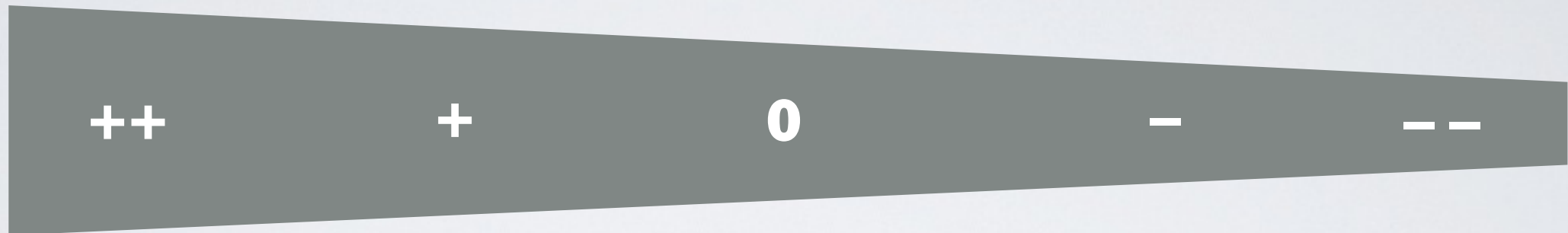
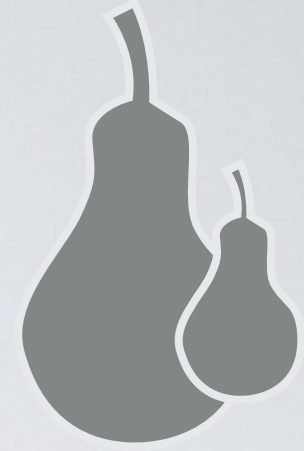
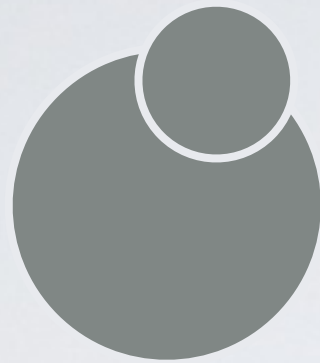
- **Equal intervals:** class limits are equidistant
 - Problem: not good for skewed data, as empty classes are likely.
 - **Quantiles:** equal number of observations in each class
 - Problem: classes can have very different ranges.
 - **Natural breaks:** minimises within-class variance and maximises between-class differences (for given number of classes). Algorithm: Jenks natural breaks optimisation (a clustering method).
 - **Manual:** adjust to “round” numbers, set class breaks at critical values (e.g. mean value, or legal threshold value). Needed when comparing multiple data sets.
-

Oregon Population Density -- 2000



Class Breaks: Guidelines

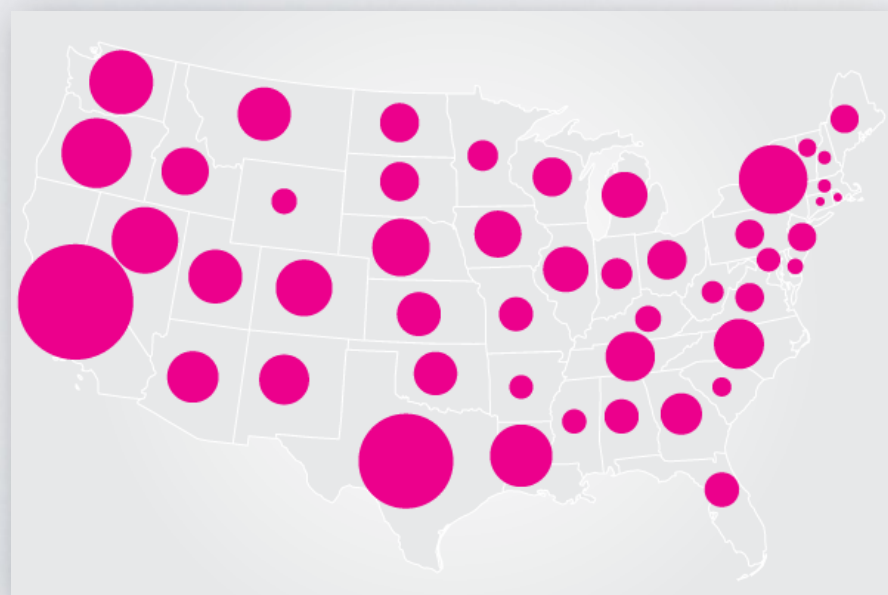
- Goal: group together similar observations and split apart observations that are substantially different.
 - Minimise within-group variance and maximise between-group differences.
 - Make the map simple to read. Limit the number of classes.
 - All mathematical classification schemes suggest class breaks at “non-round” numbers, which should be changed to reduce the cognitive load for the reader.
 - E.g., round class breaks, or change breaks to meaningful values.
-



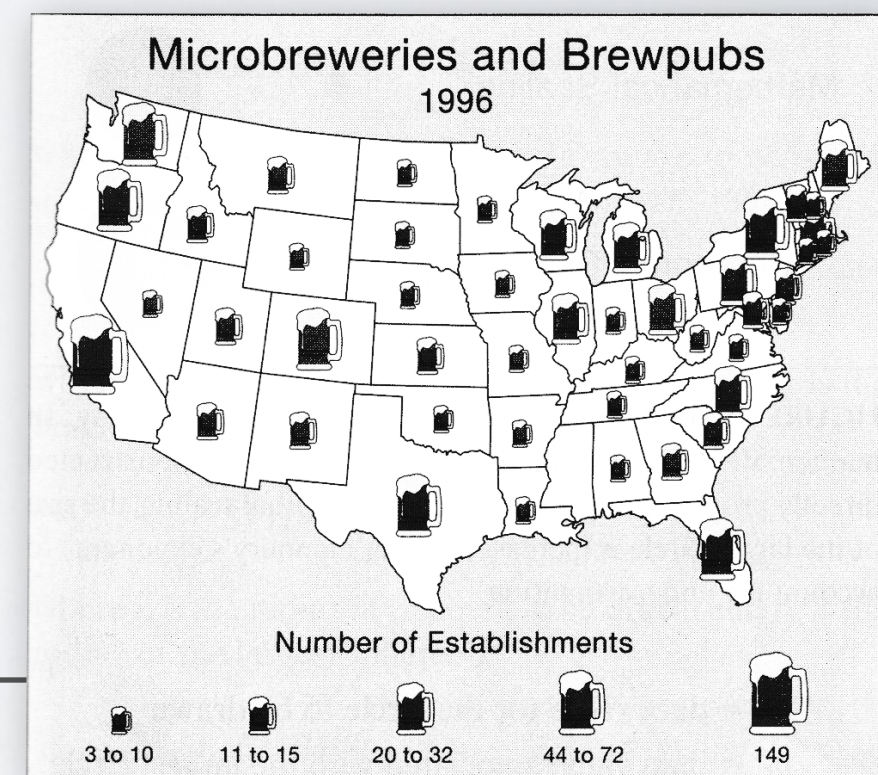
Classification optional

Classification needed

Area-proportional



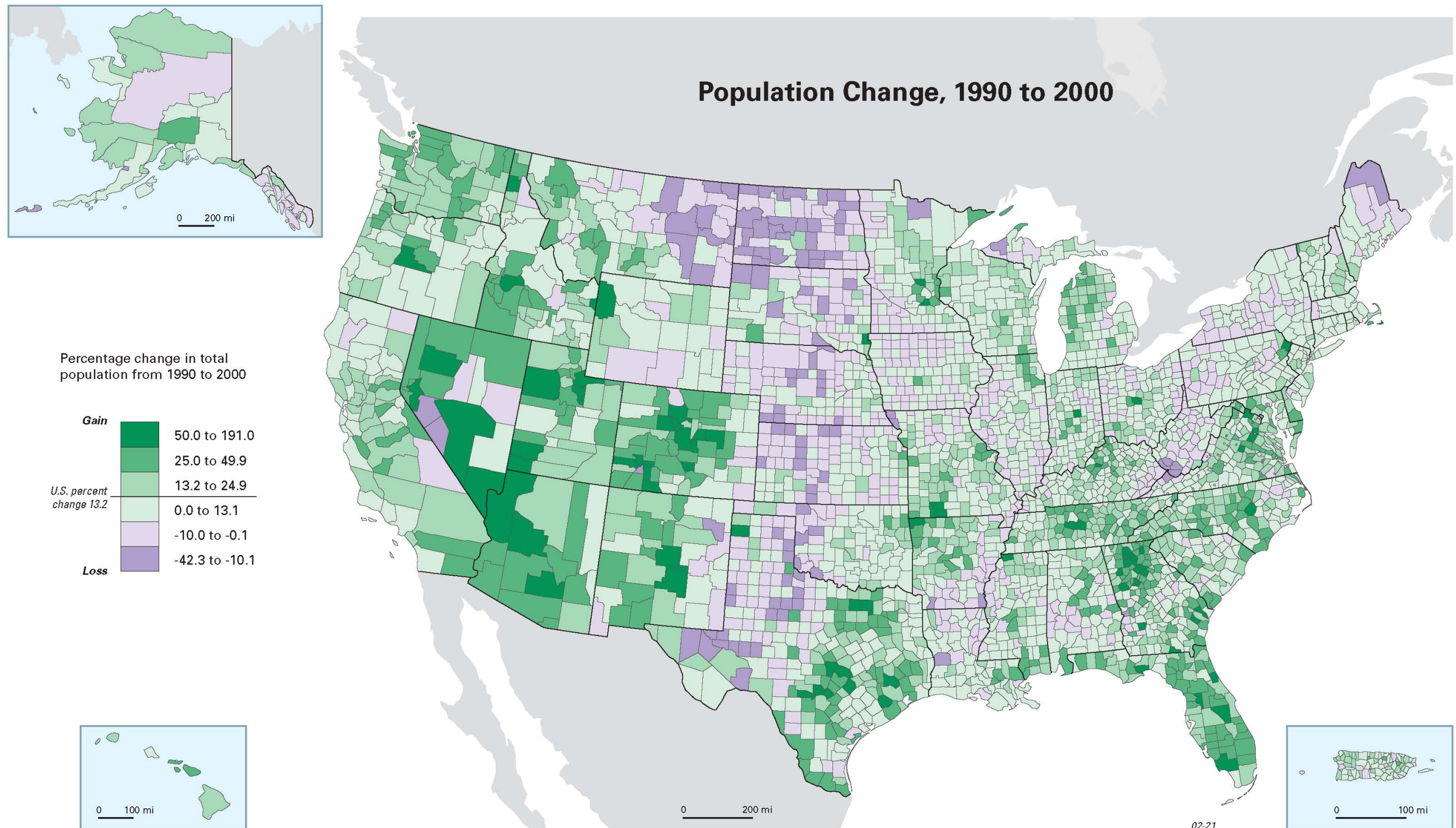
Graduated symbol



Continuous vs. Classed Representation

- Classify...
 - when values are not known accurately
 - when the data set has “natural” boundaries
 - when the order of magnitude is more important than exact values
 - to simplify the representation
 - for use with pictographic symbols

Classified Data on Choropleth Map



Classified Data for Graduated Symbol Map



Classified Data on Dot Map

