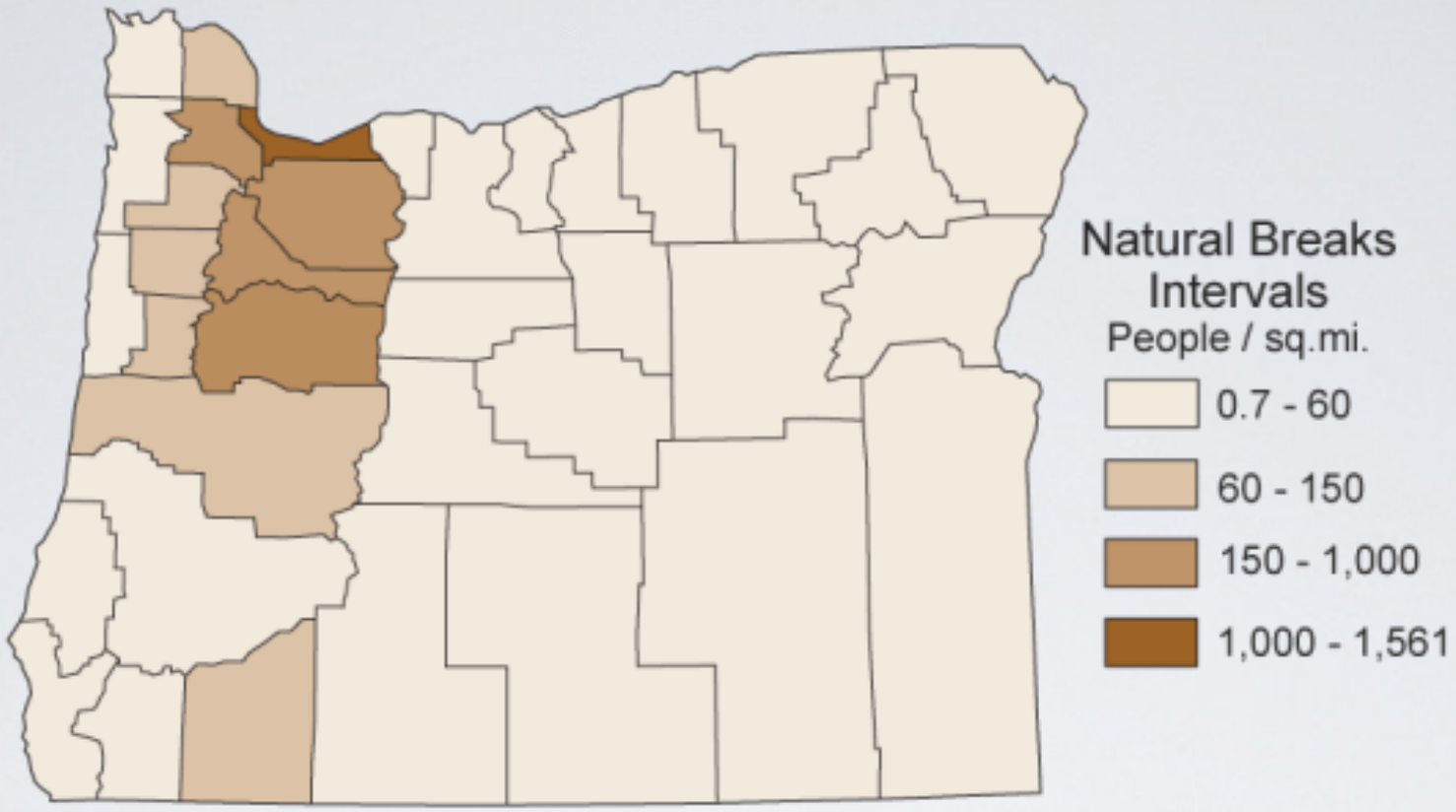
FIT3179 Data Visualisation

Data Classification

Reading

- **Required reading**
  axismaps, The Basics of Data Classification, online: https://www.axismaps.com/guide/data/data-classification/

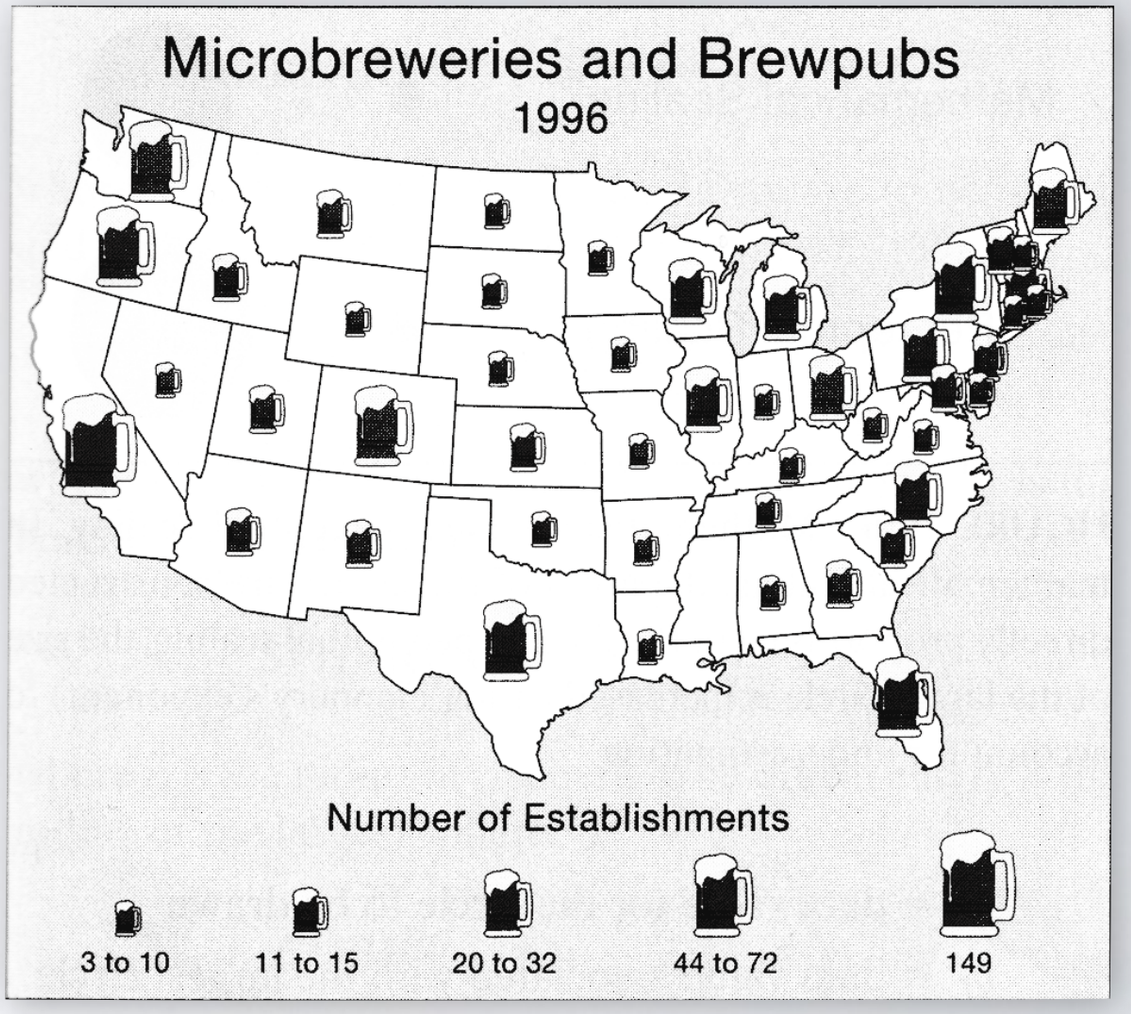
- **Optional reading**
  Slocum, T. et al. 2005. "Chapter 5: Data Classification." Thematic cartography and geographic visualization, Second Edition.


Natural Breaks
Intervals
People / sq.mi.
0.7 - 60
60 - 150
150 - 1,000
1,000 - 1,561


Unclassed
People / sq.mi.
0.7
780
1,561

Data Classification

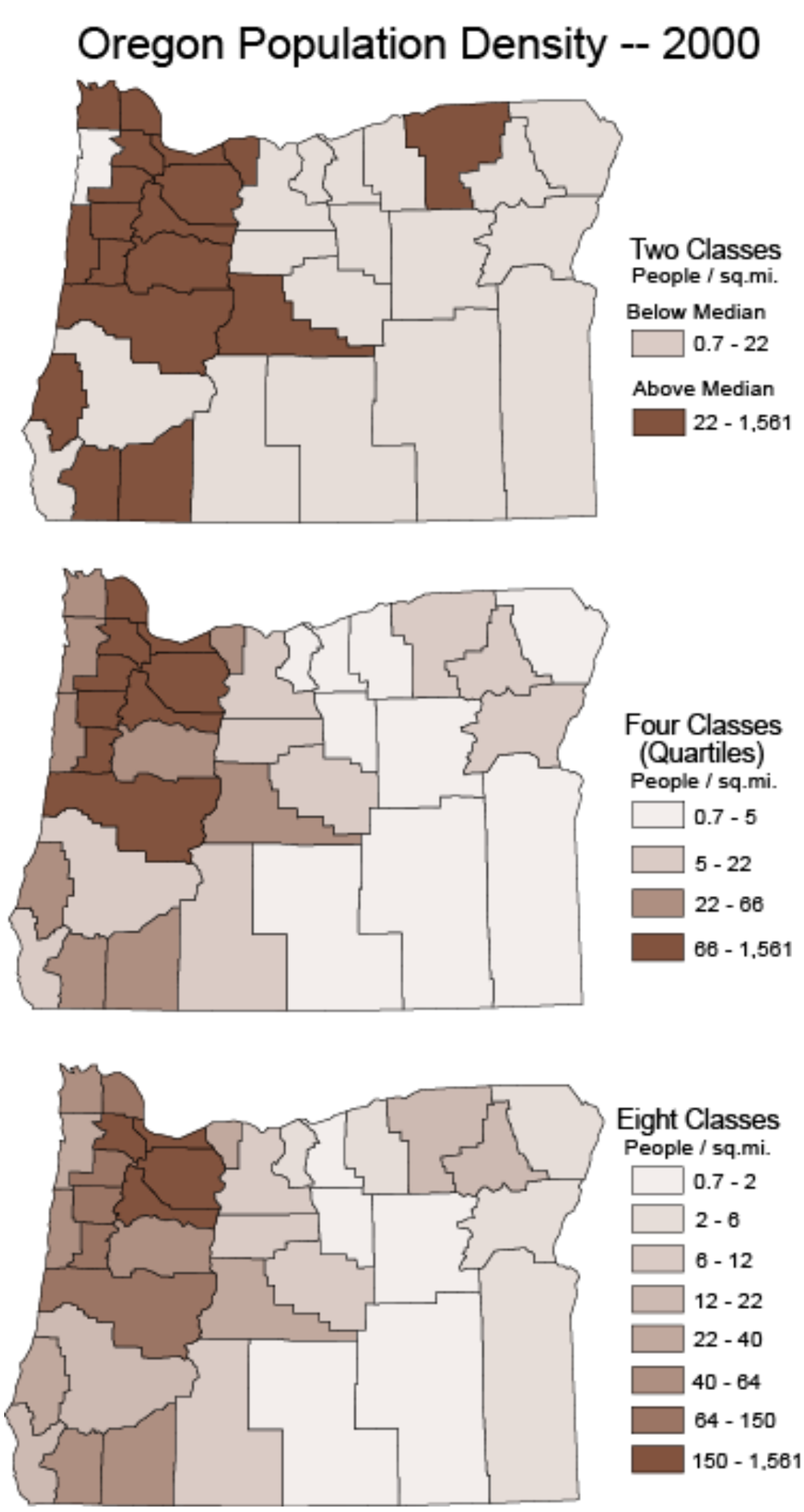
- For non-spatial diagrams and maps (choropleth maps, proportional symbols, flow lines, etc.)
- Why classify data?
  - Simplify data to make visualisation easer to read.
  - Clarify the message.
  - Show trends.
- Two questions:
  - How many classes?
  - What class limits?



Slocum et al., Thematic Cartography and Geographic Visualization, 2.ed, Fig 16.3, 2005

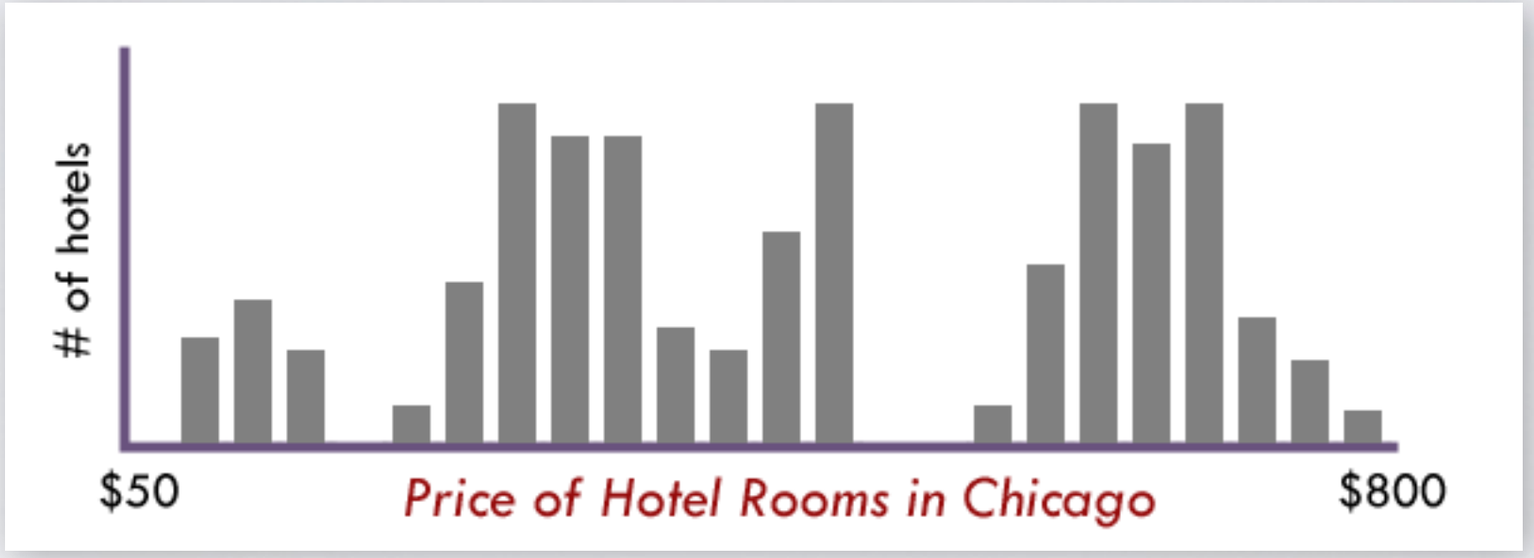
Number of Classes

- 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:
  - vis easier to read,
  - vis easier to remember,
  - clearer pattern,
  - but loss of details and information (no micro reading).



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

- 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.
- Show clusters and extreme values
- Avoid empty classes
- No overlap between classes

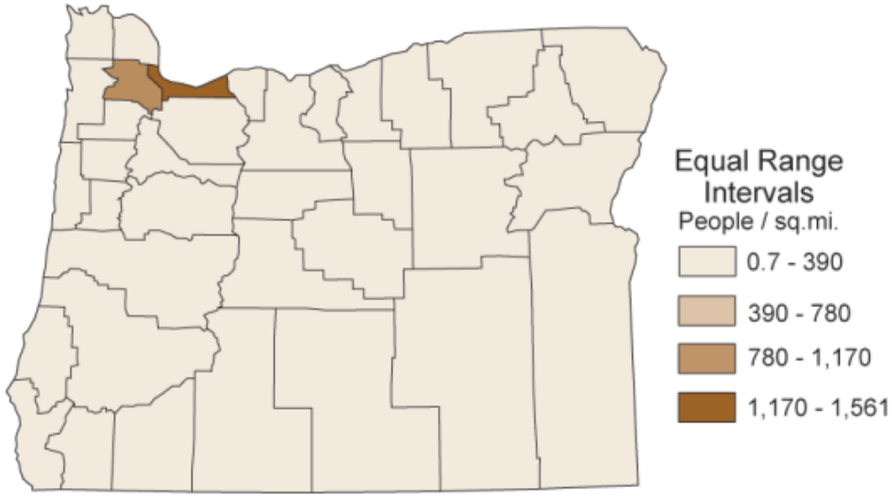
Data Classification Methods

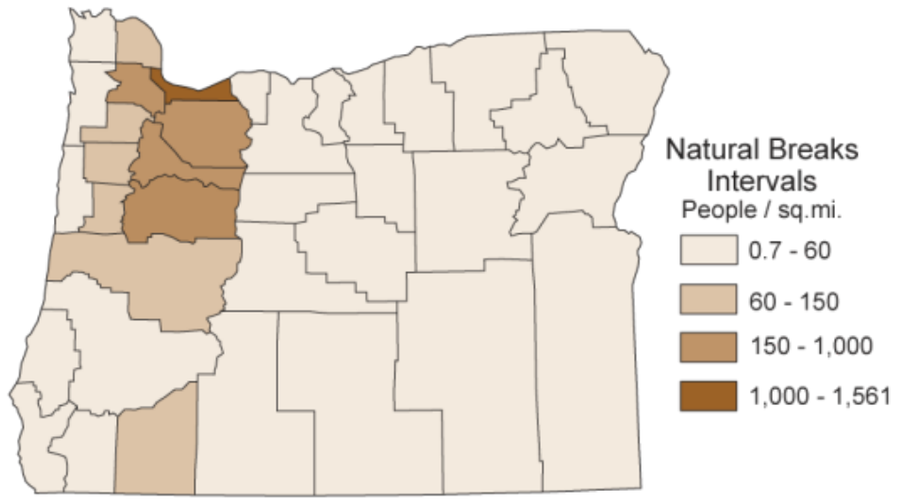
http://axismaps.github.io/thematic-cartography/articles/classification.html

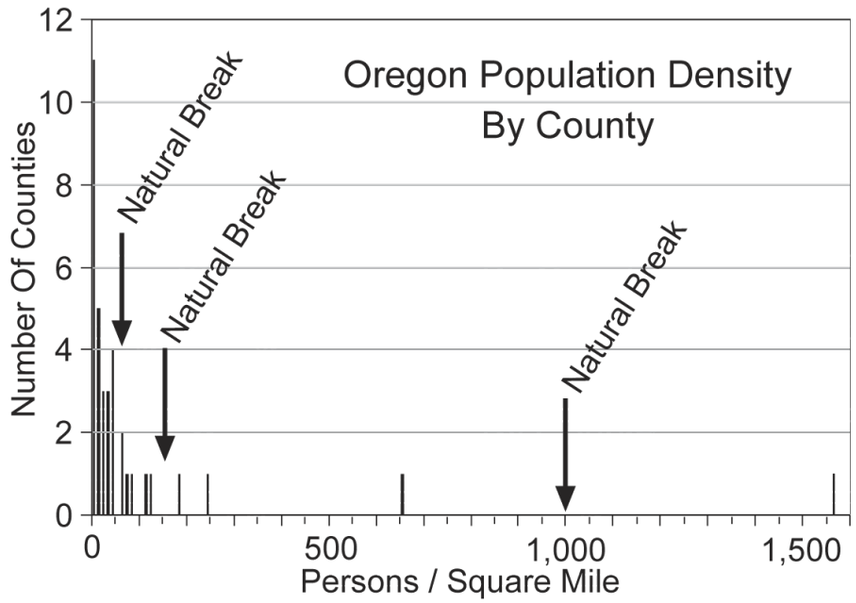
Data Classification Methods

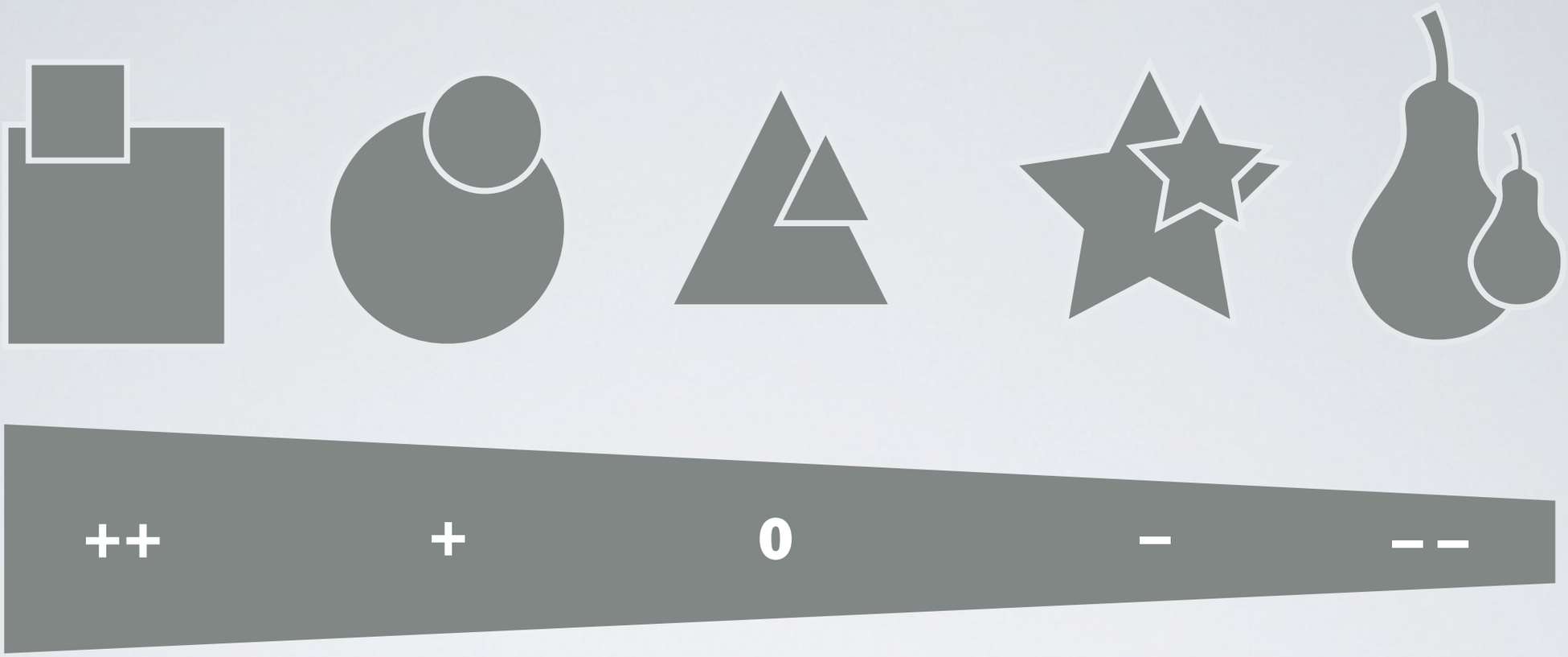
- **Equal intervals**: class limits are equidistant
  - Problem: not good for skewed data, as empty classes are likely.
- **Quantiles**: equal number of observations in each class
  - Problem: classes can have very different ranges.
- **Natural breaks**: minimises within-class variance and maximises 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



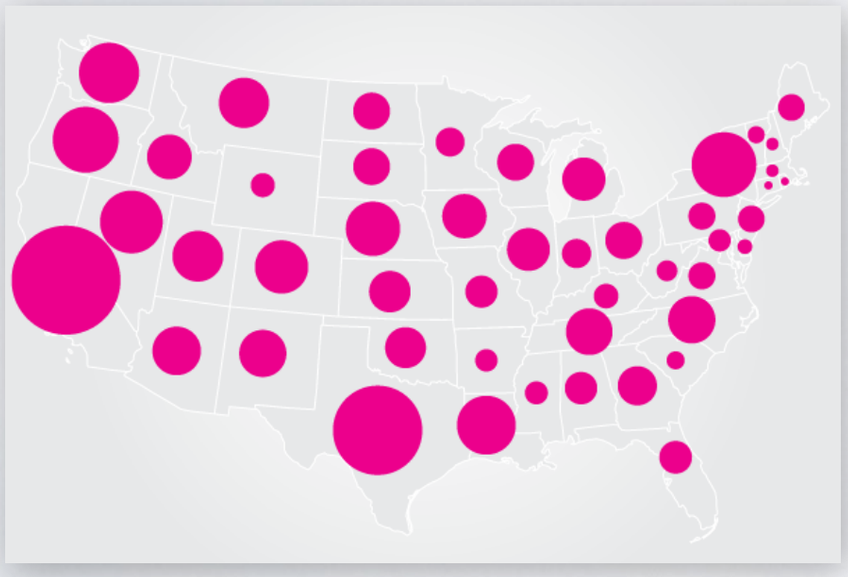




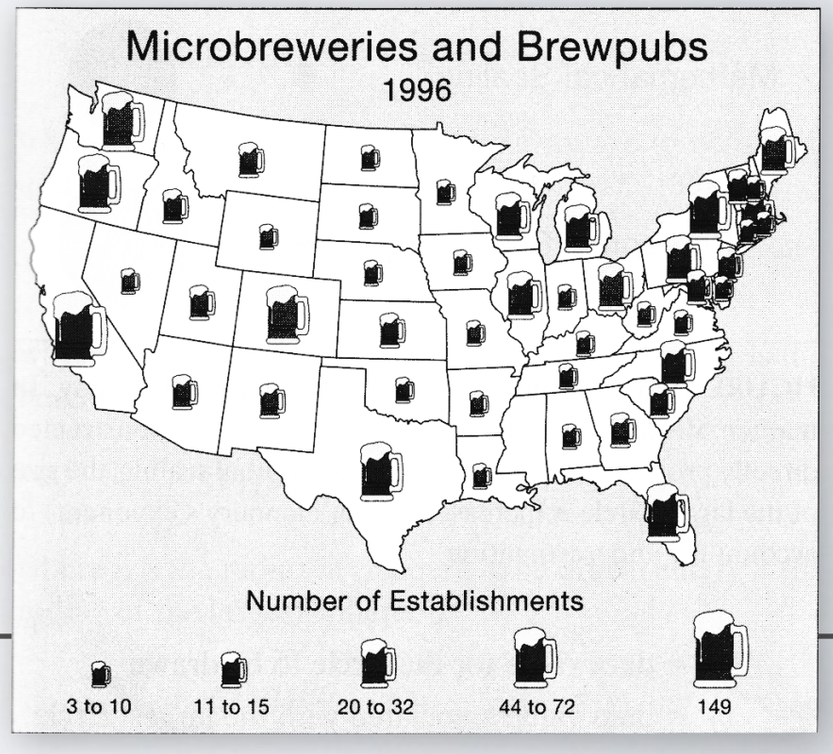


Classification optional **Classification needed**

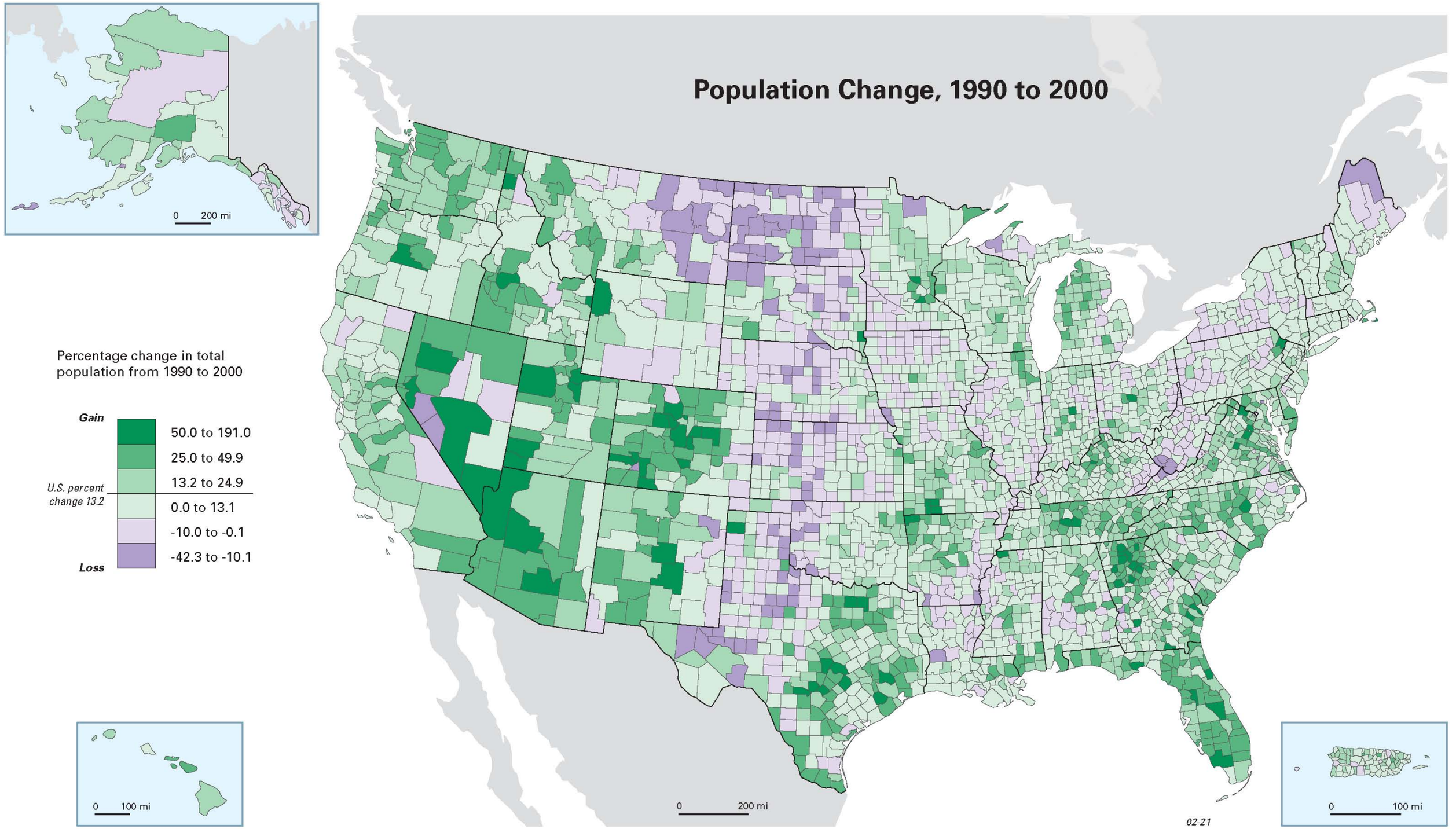
Area-proportional



Graduated symbol



Classified Data on Choropleth Map



Diverging color schema. Census Atlas of the United States, 2007

Classified Data for Graduated Symbol Map

Power Plants
Production in Megawatt

Classified Data on Dot Map