Maps and Map Analysis II

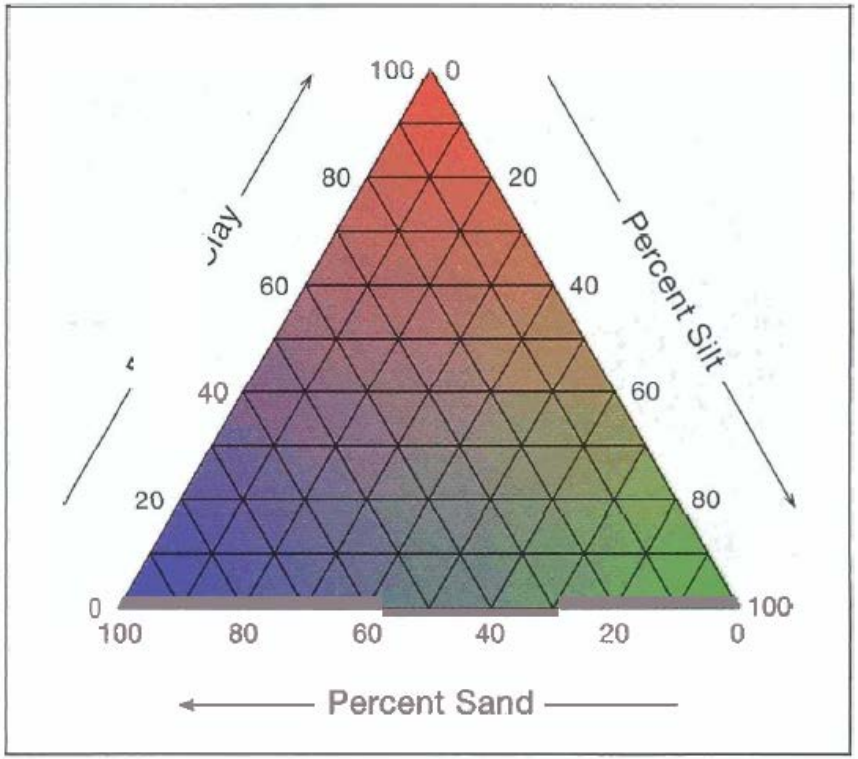
GEOG380 FA 2018

Outline

- Example analysis methods frequently used in maps
  - Bivariate mapping
  - Multivariate mapping
  - Cluster analysis

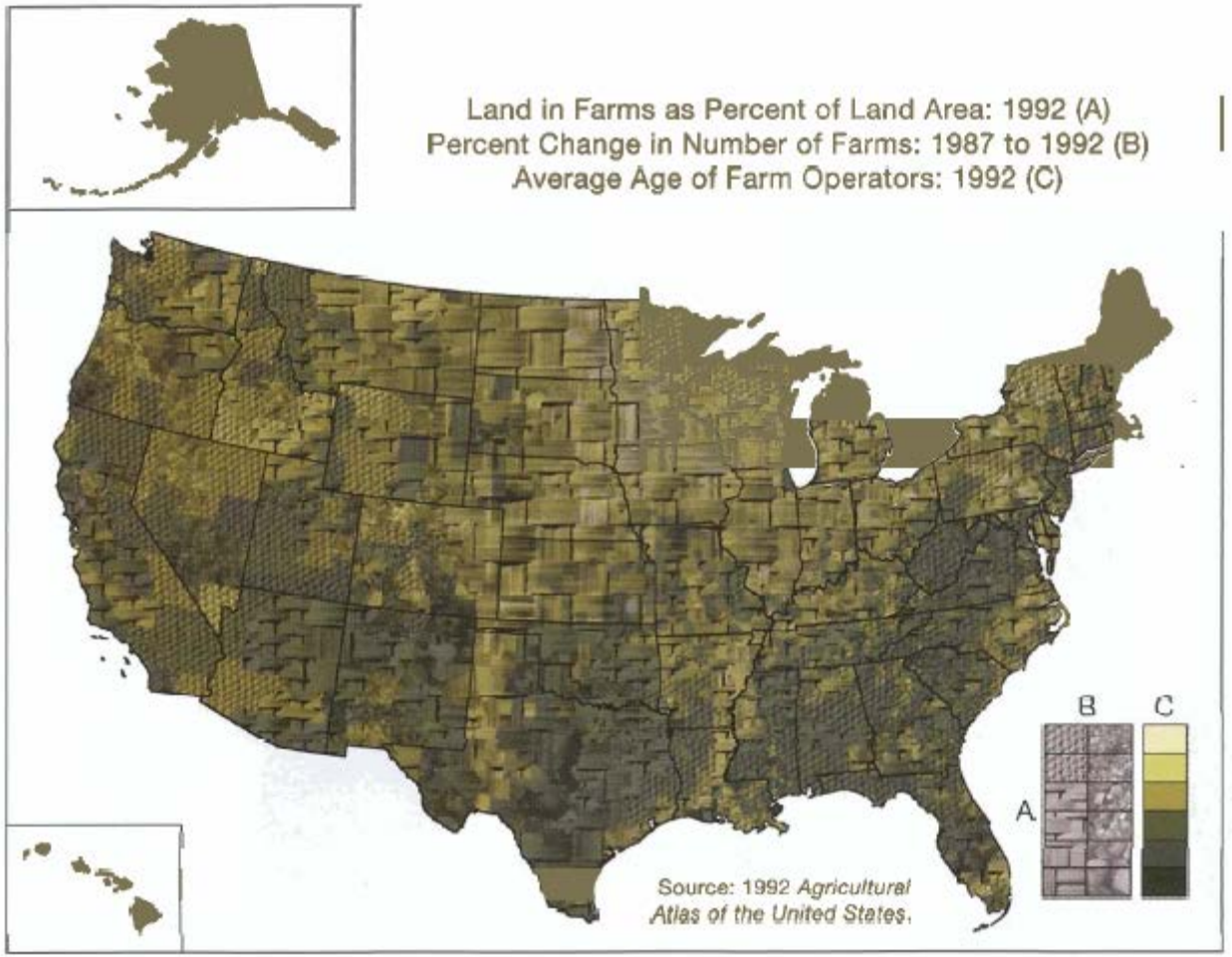
Multivariate mapping

- Combining three or more attributes on the same map
  - Three colors (e.g. RGB, CMY)
    - Restriction: three attributes should add to 100%



COLOR PLATE 18.5 An RGB color scheme for creating a trivariate choropleth map. (After Byron 1994, p. 126.)

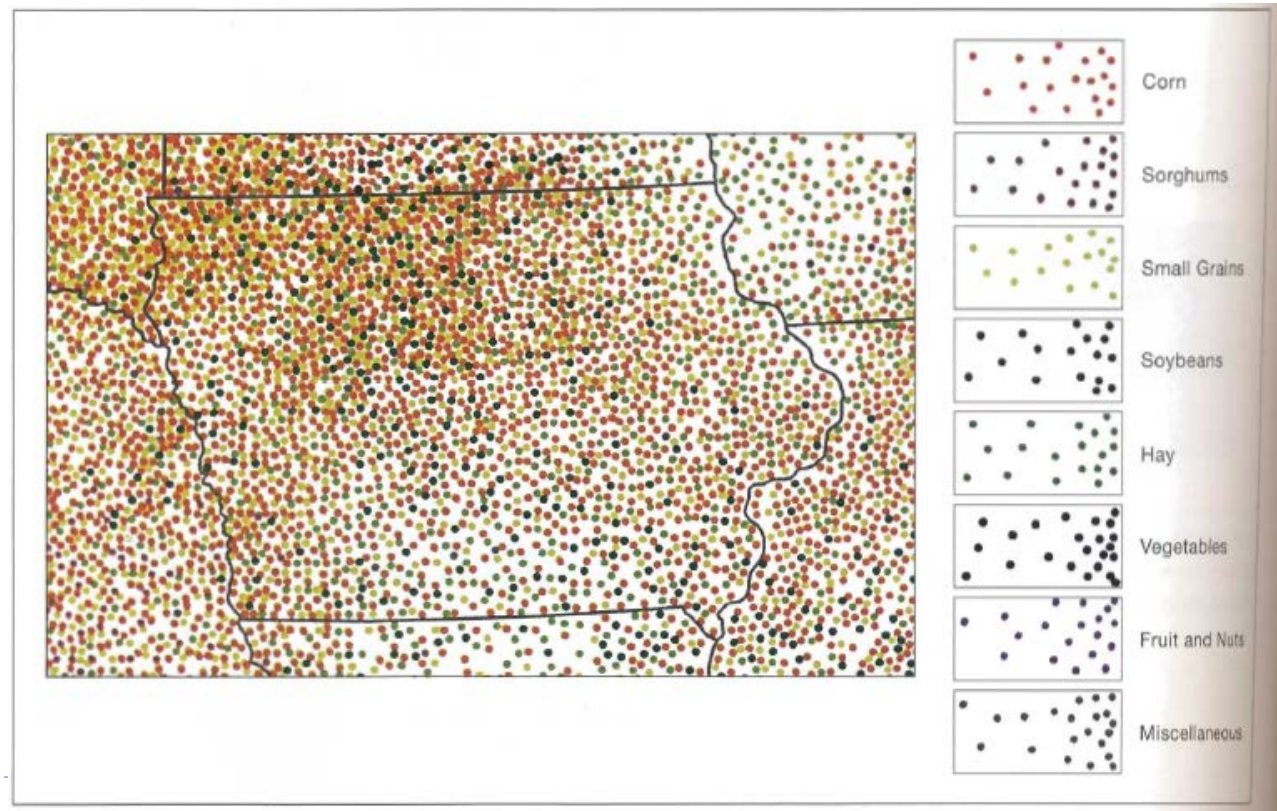
Two or three colors and textures (or patterns)
: May have some difficulty for recognition



COLOR PLATE 18.6 A trivariate choropleth map that uses pattern (or texture) for two attributes and a smooth colored tone for a third attribute. (After Interrante, V. (2000) "Harnessing natural textures for multivariate visualization." *IEEE Computer Graphics and Applications* 20, p. 9; © 2000 IEEE.)

Combining three or more attributes on the same map (cont.)

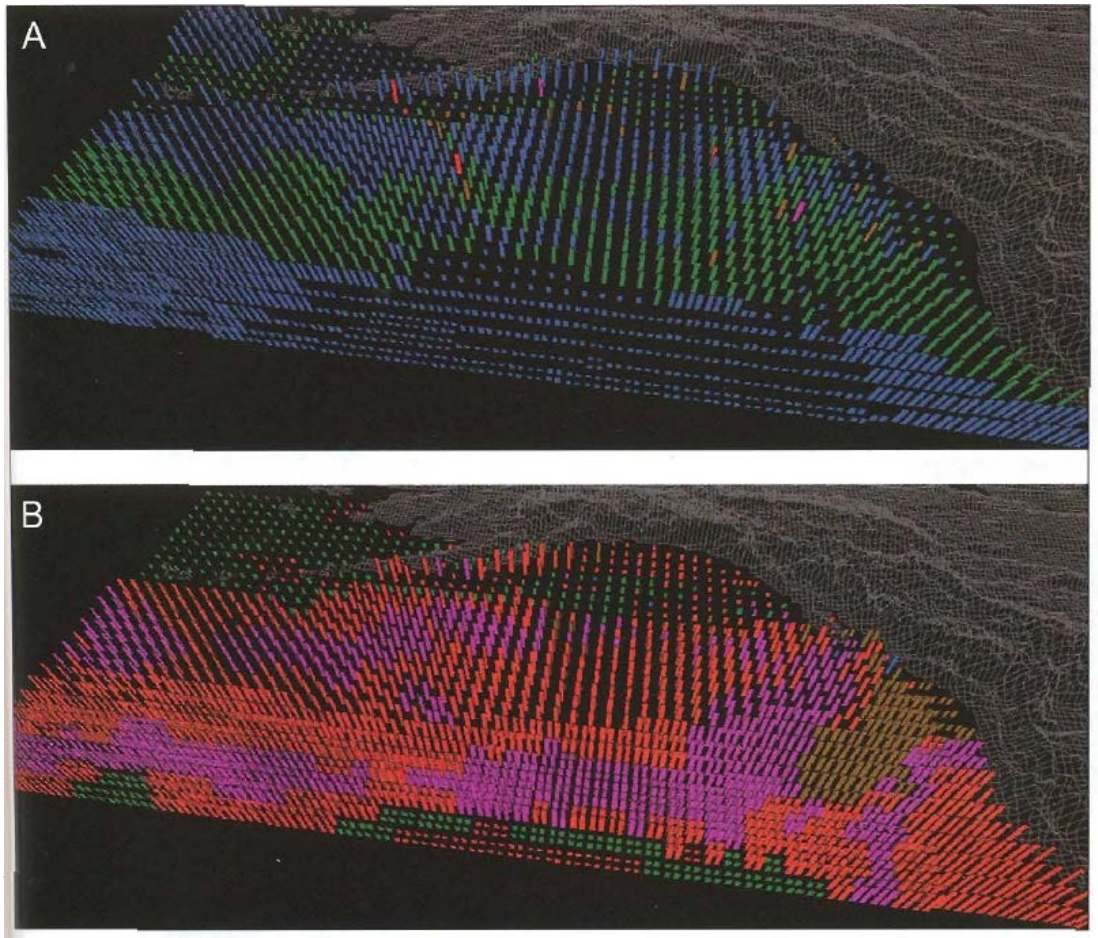
- Dots
  - Color Plate 18.7 (Plate 18)
  - Legend, density & color combination, color for highlighted attribute(s)



COLOR PLATE 18.7 A portion of a multivariate dot map constructed on the basis of pointillism. Because this map was scanned from the original map, colors shown are approximations of the actual colors. (After Jenks 1961.)

Combining three or more attributes on the same map (cont.)

- Point symbols
  - Color Plates 18.8 (Plate 19), 18.10 (Plate 20), 19.3 (Plate 22), 19.4 (Plate 22)



COLOR PLATE 18.8 Using pexels to depict multivariate data for the northern Pacific Ocean

Combining three or more attributes on the same map (cont.)

- Point symbols
  - Pie chart, glyphs, bars, etc.
  - Combination of different types of symbols (Figure 18.14, p.342)

Q. How are patterns similar or different between the areas in circle?

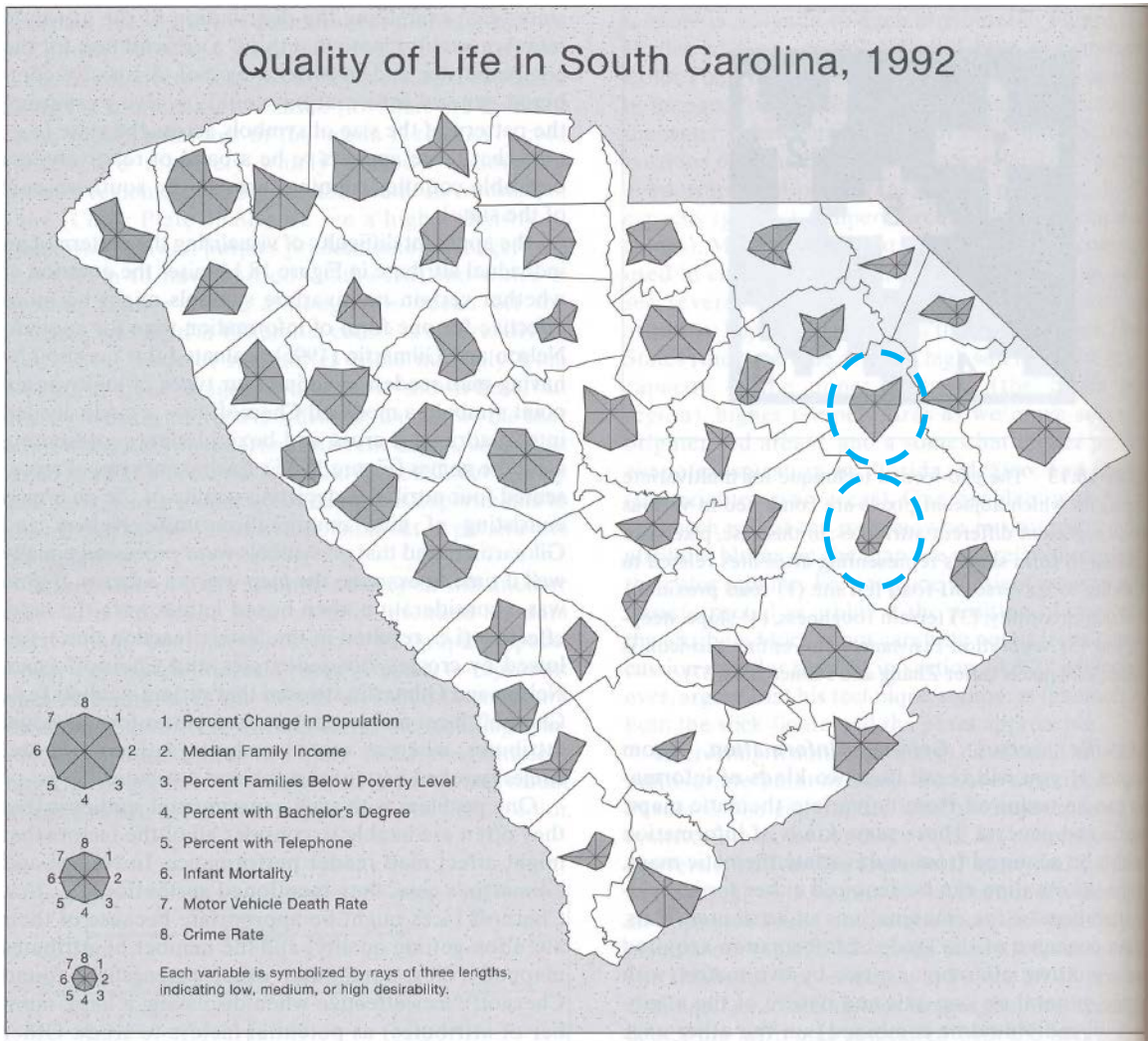


FIGURE 18.14 A multivariate map based on a combination of the star and snowflake symbols shown in Figure 18.8. (D
source: South Carolina State Budget and Control Board 1994.)

Cluster analysis
: Analysis methods frequently used in maps

- Shows which are similar and how much similar in the data by measuring numeric values
- Purposes (Romesburg 1984)
  - To create a research question
    - *E.g., Are there certain counties that have similar scores on one or more of multiple attributes in New York State (Table 18.3)?*
  - To create a research hypothesis that answers a question
    - *E.g., Does some counties have a high percentage of African Americans, a high infant mortality rate, and a decreasing population?*
  - To test a research hypothesis for the question to decide whether it should be confirmed or disproved
    - *E.g., Create a map of cluster analysis of the data such as Table 18.3*

General procedure of cluster analysis

- Standardize the dataset
  - Unit of the numeric values between variables matters
- Compute initial resemblance between the data
  - E.g. Std. dev., and etc
- Cluster the data
  - E.g. Dendrogram (next slide)
- Roughly determine an appropriate number of clusters
- Interpret the clusters
- Make a map of the resulting clusters
- Determine whether the clusters are appropriate
  - E.g., Resemblance matrix (next next slide)

Examples

Are there certain counties that have similar scores on one or more of multiple attributes in Table 18.3?

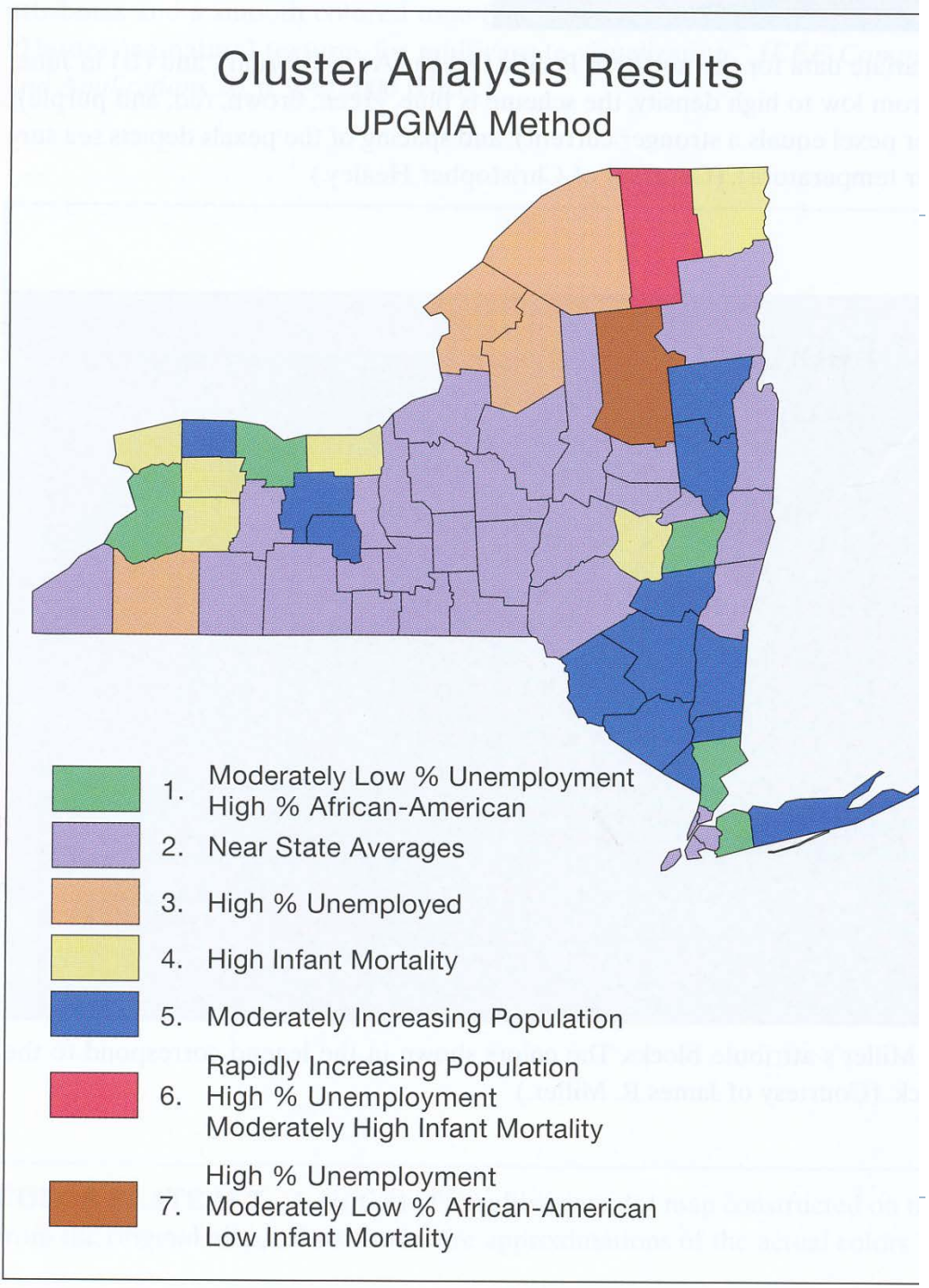
How can you know?

→ *Using cluster analysis might help*

(Slocum et al. 2009)

TABLE 18.3 Cluster analysis data for New York State counties

County	% Population Change (1990–2000)	% Unemployed	% African-American	Infant Mortality Rate (Deaths per 1,000 Live Births)
Albany	0.6	2.8	11.1	6.1
Allegany	−1.1	6.7	0.7	5.4
Broome	−5.5	3.3	3.3	8.1
Cattaraugus	−0.3	6.4	1.1	9.6
Cayuga	−0.4	4.5	4.0	6.4
Chautauqu	−1.5	4.8	2.2	8.0
Chemung	−4.3	4.8	5.8	4.9
Chenango	−0.7	4.8	0.8	4.9
Clinton	−7.1	5.2	3.6	15.2
Columbia	0.2	2.9	4.5	6.0
Cortland	−0.7	5.9	0.9	3.6
Delaware	1.5	4.8	1.2	8.0
Dutchess	8.0	3.1	9.3	5.0
Erie	−1.9	4.8	13.0	8.2
Essex	4.6	6.6	2.8	5.2
Franklin	9.9	7.6	6.6	10.7
Fulton	1.6	5.8	1.8	4.6
Genesee	0.5	4.9	2.1	14.6
Greene	7.7	5.0	5.5	4.1
Hamilton	1.9	8.2	0.4	0.0
Herkimer	−2.1	5.0	0.5	8.6
Jefferson	0.7	8.2	5.8	5.2
Lewis	0.6	7.8	0.4	9.0
Livingston	3.1	4.6	3.0	8.5
Madison	0.4	4.4	1.3	4.8
Monroe	3.0	3.8	13.7	8.2
Montgomery	−4.4	5.8	1.2	3.4
Nassau	3.6	2.7	10.1	4.8
Niagara	−0.4	5.9	6.1	10.6
Oneida	−6.1	3.8	5.7	5.9
Onondaga	−2.3	3.5	9.4	6.7
Ontario	5.4	3.7	2.1	6.8
Orange	11.0	3.1	8.1	6.5
Orleans	5.6	5.3	7.3	5.4
Oswego	0.5	6.3	0.6	4.8
Otsego	2.1	4.7	1.7	5.1
Putnam	14.1	2.5	1.6	8.2
Rensselaer	−1.2	3.8	4.7	10.7
Rockland	8.0	3.0	11.0	4.3
Saratoga	10.7	3.2	1.4	4.8
Schenectady	−1.8	3.5	6.8	6.2
Schoharie	−0.8	4.8	1.3	14.9
Schuyler	3.0	5.5	1.5	4.6
Seneca	−1.0	4.9	2.3	5.4
Steuben	−0.4	4.9	1.4	5.9
St. Lawrence	0.0	8.0	2.4	6.8
Suffolk	7.4	3.2	6.9	6.2
Sullivan	6.8	5.0	8.5	4.8



- The research question
  - *Are there certain counties that have similar scores on one or more of multiple attributes in Table 18.3?*
- The research hypothesis that answers the question
  - *Are there some counties that have a high percentage of African Americans, a high infant mortality rate, and a decreasing population?*
- Test of the research hypothesis to decide whether it should be confirmed or disproved

Color Plate 18.11 (Plate 20)
Results of clustering the New York State data in Table 18.3 using the UPGMA (Unweighted Pair Group Method with Arithmetic mean, or average linkage) method. Numbers within the legend represent cluster numbers.
Slocum et al. (2009)

Clustering data using a Dendrogram

- A tree-like structure that illustrates the resemblance coefficient values at which various clusters combined

Shorter Euclidean distances mean higher resemblance between the observations

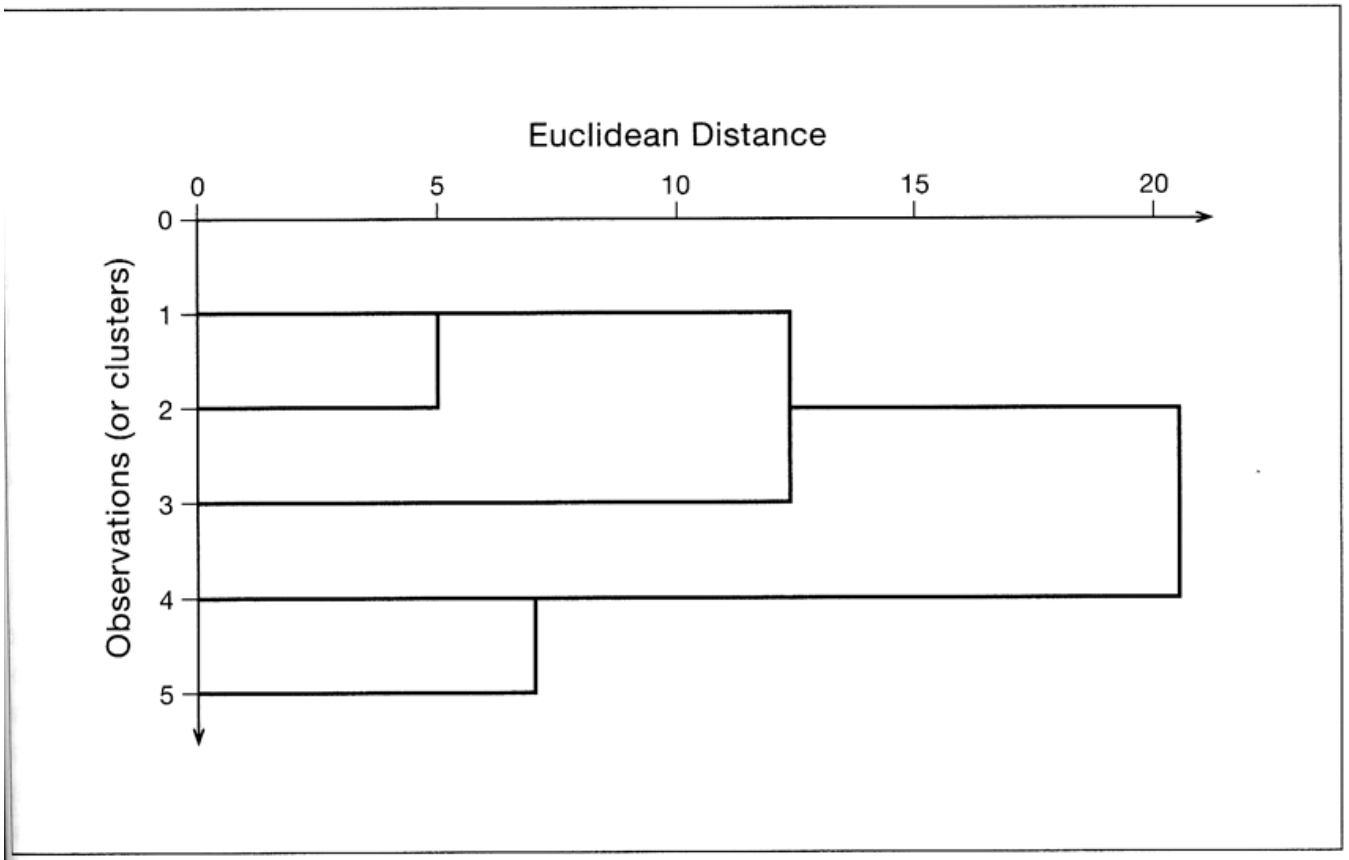


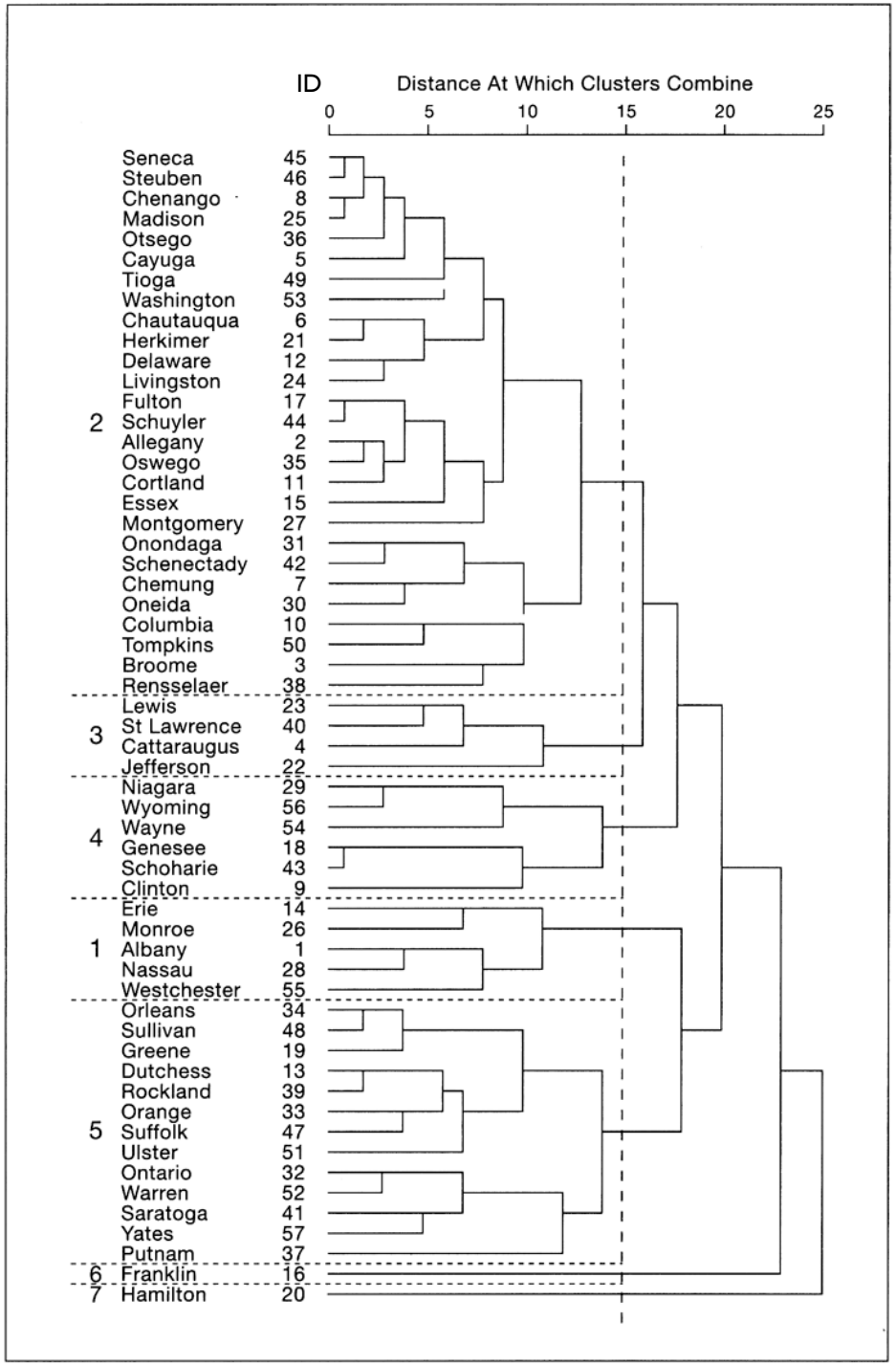
FIGURE 18.19 Dendrogram for the hypothetical data shown in Figure 18.18A using Euclidean distance as the resemblance coefficient.

Slocum et al. (2009)

Another example

- Figure 18.20
  Dendrogram of the New York State data in Table 18.3 (in a previous slide) using the UPGMA (Unweighted Pair Group Method with Arithmetic mean) method

  - How many clusters exist at the distance of clusters combination of "15"?



Slocum et al. (2009)

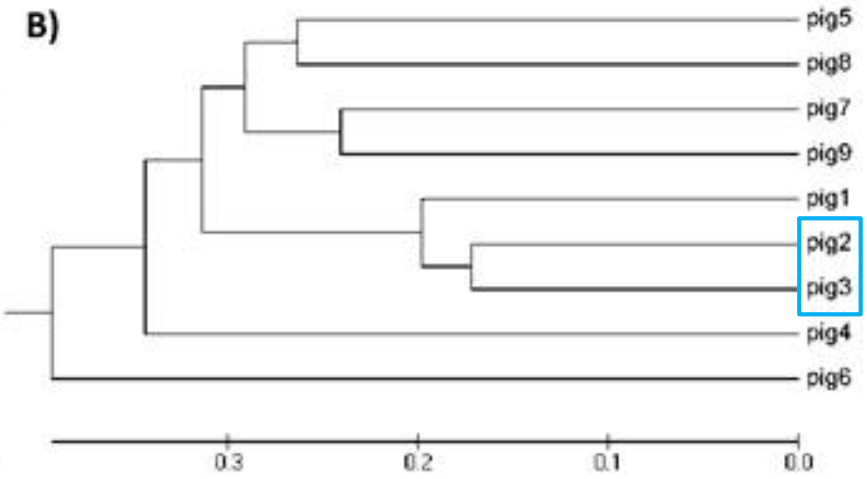
Resemblance matrix

- Degree of similarity between entities

A)

2D-immunoblots	Dice (Czekanowski or Sorenson) Measure								
	1	2	3	4	5	6	7	8	9
1	100	65.5	55.4	25.6	45.8	24.2	44.4	54	31.6
2		100	65.6	31.6	34	6.3	31.8	45.2	32.4
3			100	29.2	28.1	9.5	29.6	41.7	29.8
4				100	12.9	25	28.6	43.5	47.6
5					100	24	43.2	47.3	33.3
6						100	36.4	20	26.7
7							100	46.2	51.9
8								100	44.4
9									100

B)



Q. What do the two value mean?

Source: www.biolsci.org

Until next time…

- Readings
  - Ch. 18 & Ch. 3
- Lab1
- Questions?