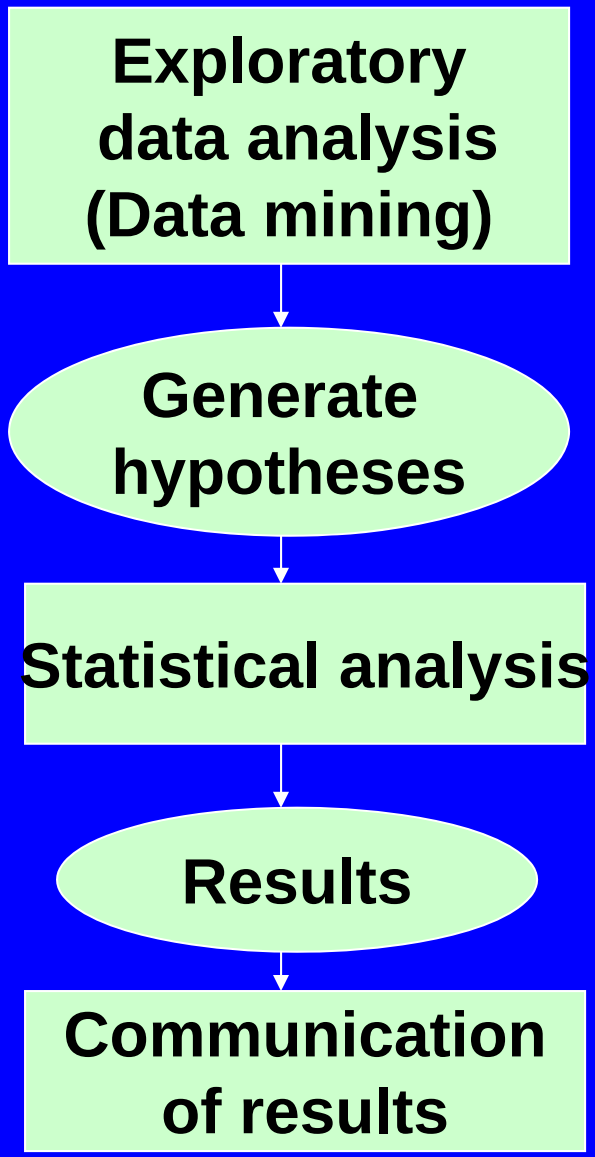
New Tools for Understanding Cancer Patterns

National Cancer Advisory Board

September 9, 2001

Linda Williams Pickle, PhD
Statistical Research and Applications Branch, SRP, NCI

A Framework for Analysis



Goals of Cancer Data Exploration

- Identifying broad geographic patterns
- Identifying clusters of unusually high rates (“hot spots”)
- Generating hypotheses about causes of these geographic patterns
- Illustrating relationships among risk factors

New visualization tools

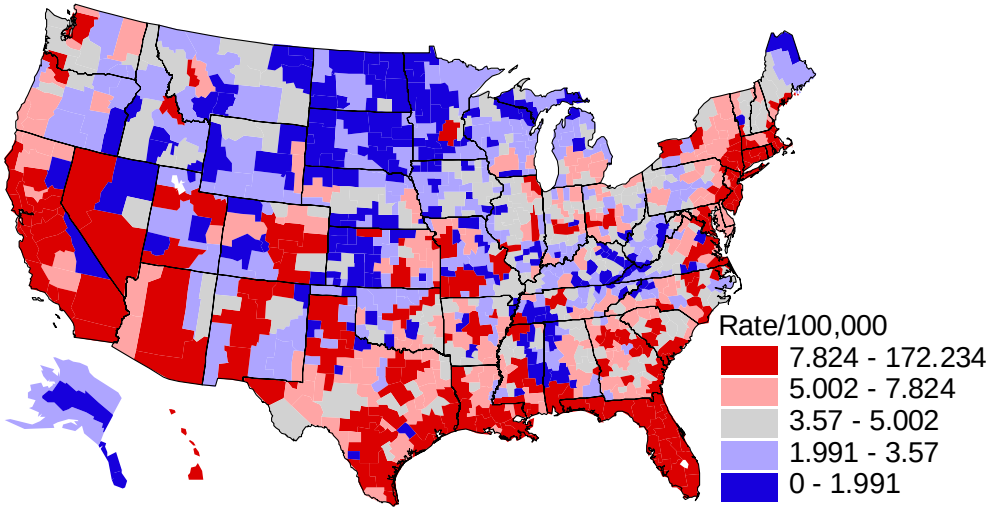
- Map smoothing methods
- Linked maps and graphs
- Maps conditioned on levels of risk factors
- Multivariate plots linked to maps
- Animation over time

Map smoothing methods

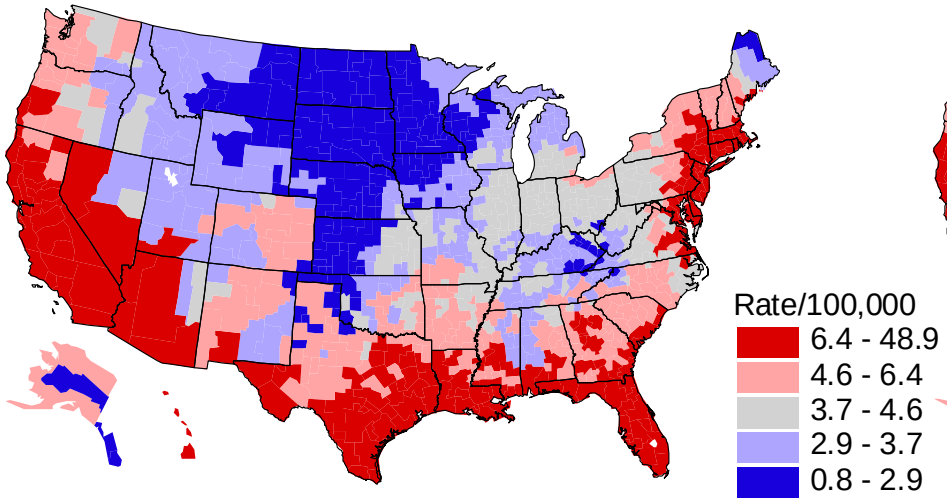
- Smoothing is a method of removing some of the variability in a quantitative map, e.g., of rates
- Maps of cancer rates for small areas, e.g., counties, can be difficult to interpret because of background “noise” (variability)
- Previous methods of smoothing, did not account for varying population size across the areas
- New methods now can incorporate weights so that very stable rates are smoothed less than rates that are more variable due to small populations

HIV mortality rates, 1988-92

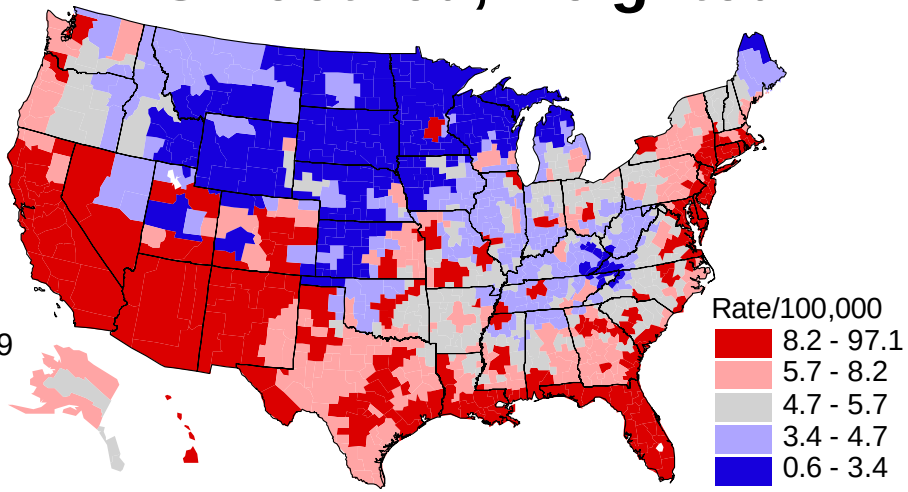
Original data:



Smoothed, unweighted



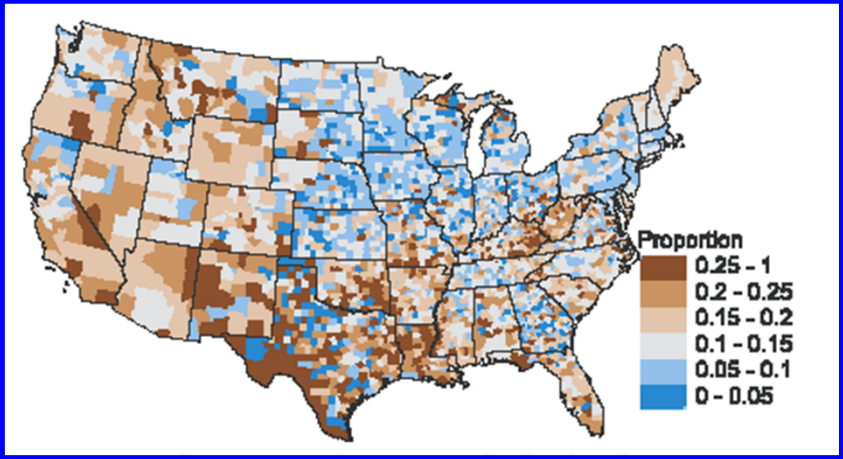
Smoothed, weighted



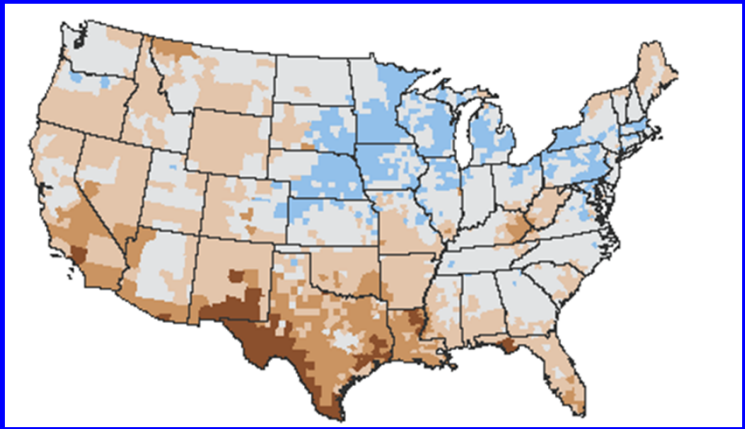
Source: Pickle et al., Atlas of United States Mortality, NCHS, 1996.

Proportion of residents without health care coverage, Behavioral Risk Factor Survey

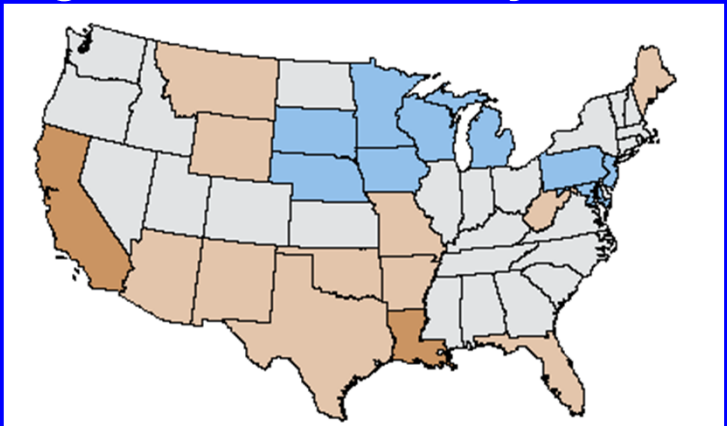
Original data, 1992-98, by county:



Smoothed county data

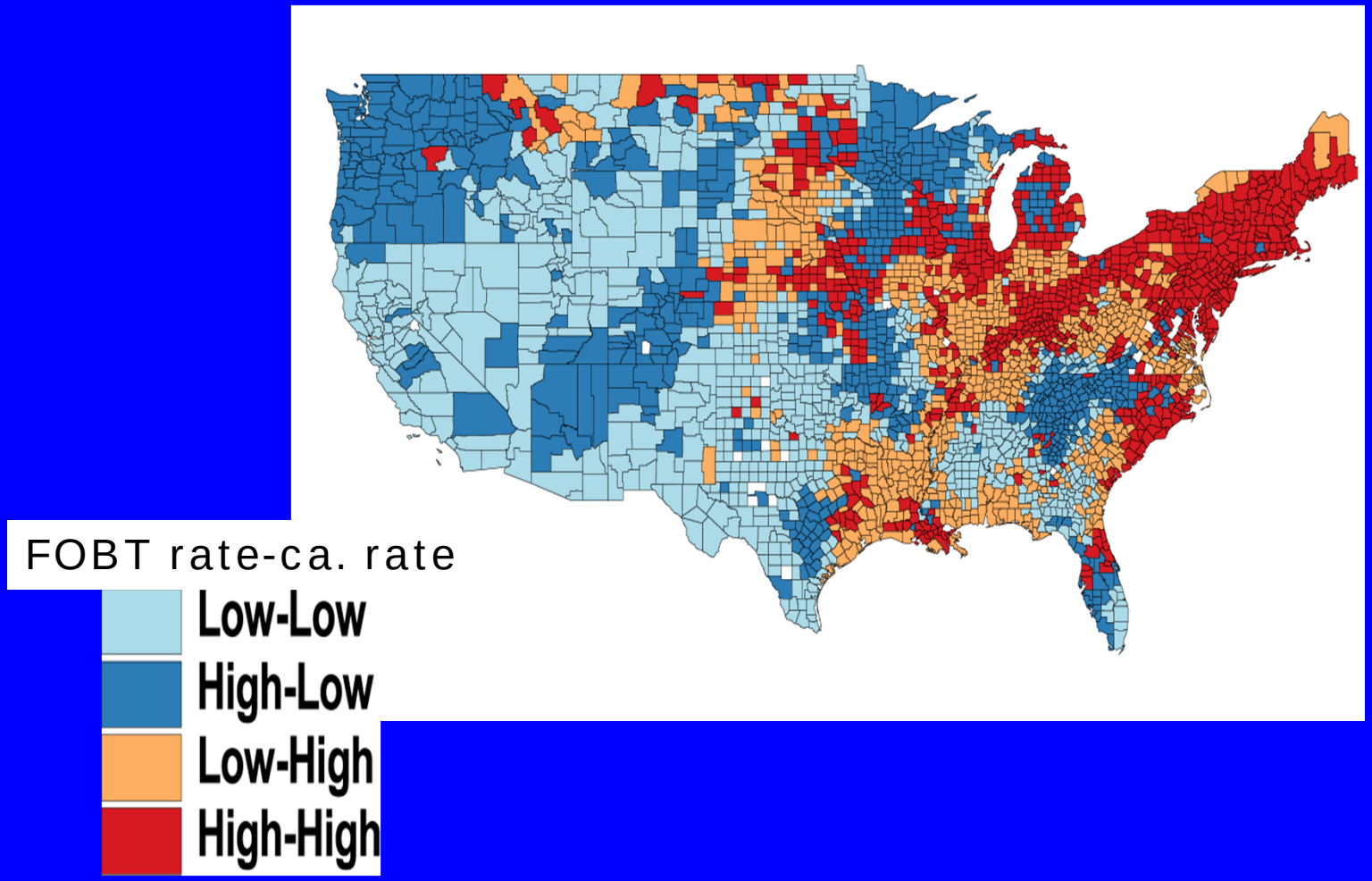


Original data, 1995, by state



Source: Pickle and Su, Am J Prev Med 2002, based on CDC BRFSS data for 1992-98

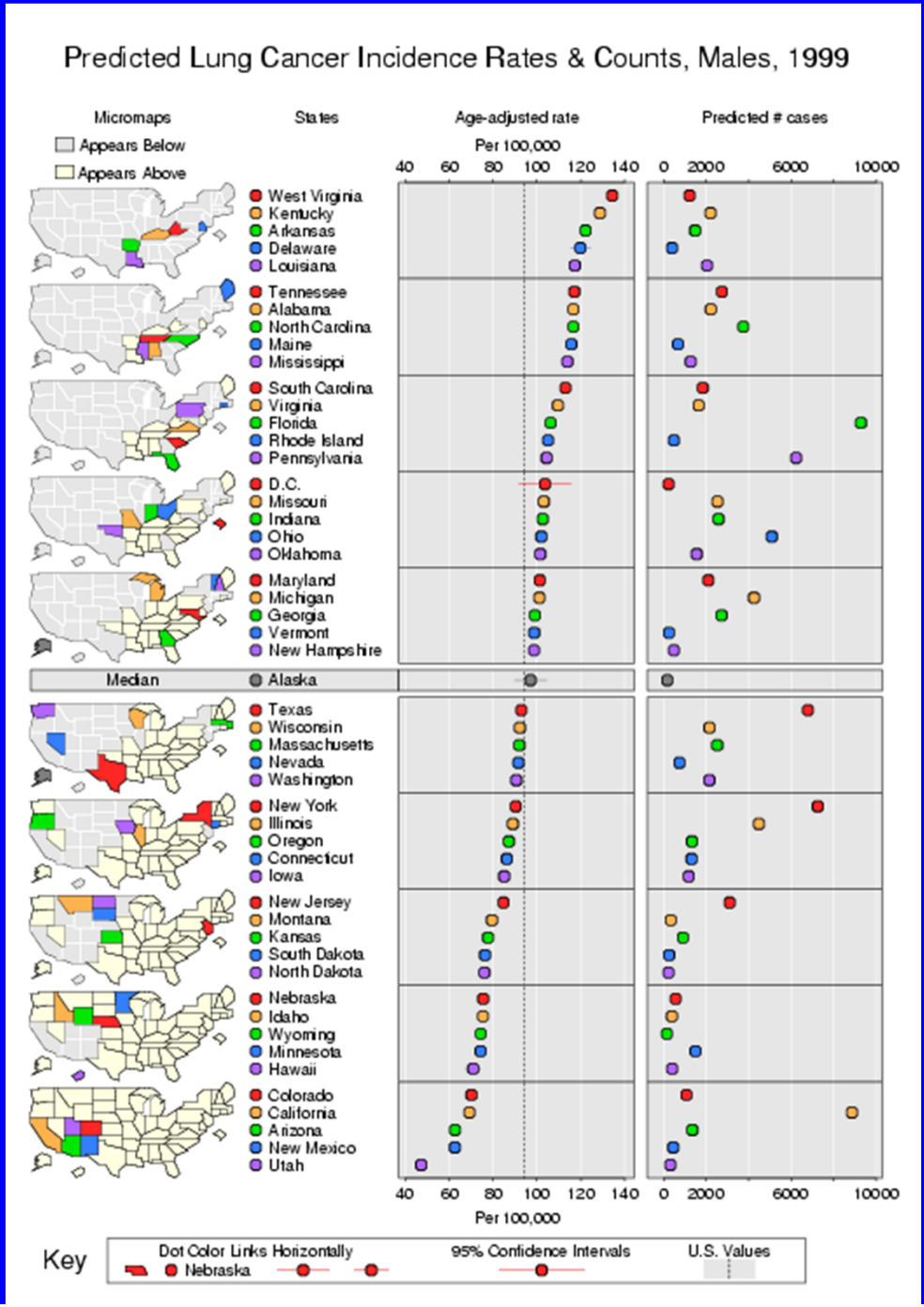
Colorectal Screening Behavior and Cancer Mortality Rates



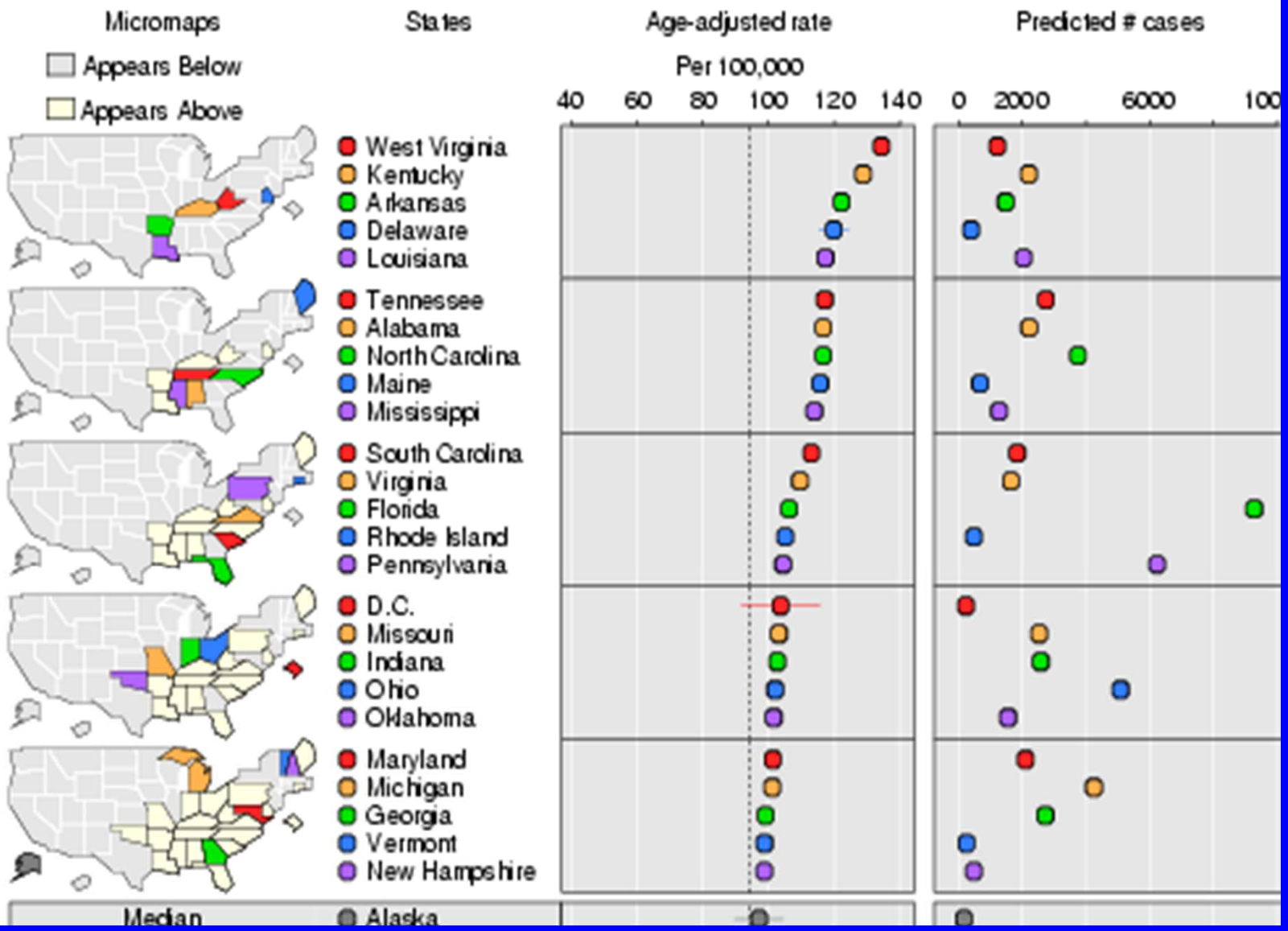
Can we judge geographic patterns and see statistical detail on the same graphic?

- "Linked micromap plot" links statistical graphs to maps, so patterns of both can be seen
- Can be static for publication or dynamic to allow user interaction over the web
- Any geographic units can be plotted – not restricted to states or counties
- Flexible types of statistical graphs

Linked Micromap Plot



Predicted Lung Cancer Incidence Rates & Counts, Males, 1999

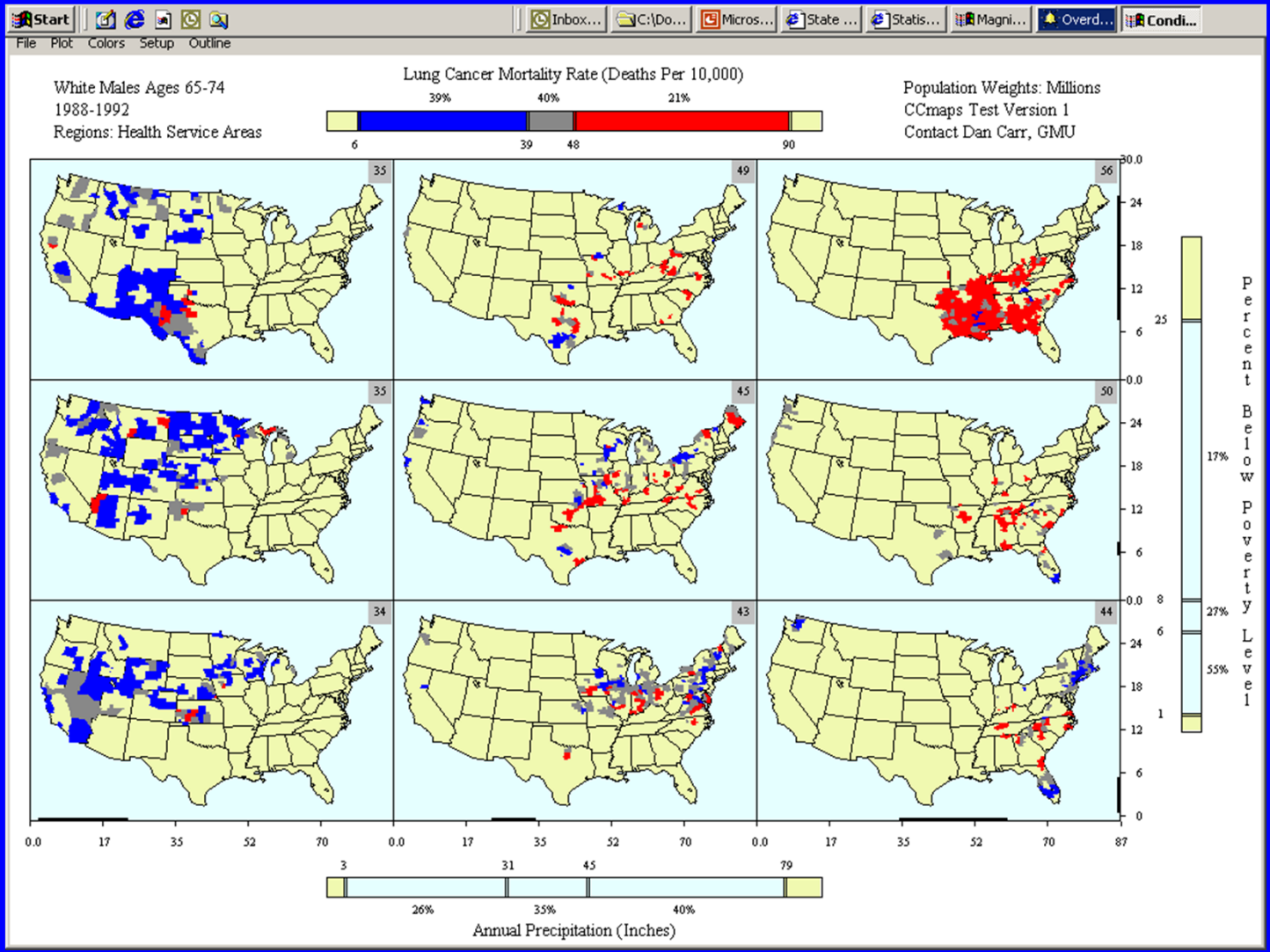


How can we explore the relationships between cancer rates and potential risk factors?

- Cancer rate maps conditioned on risk factors or confounding variables (e.g., income)
- “Conditioned choropleth (area shaded) maps” decompose a single color-coded map into strata according to 1 or 2 other factors
- Slider bars let the user choose how to stratify

Demo of conditioned maps

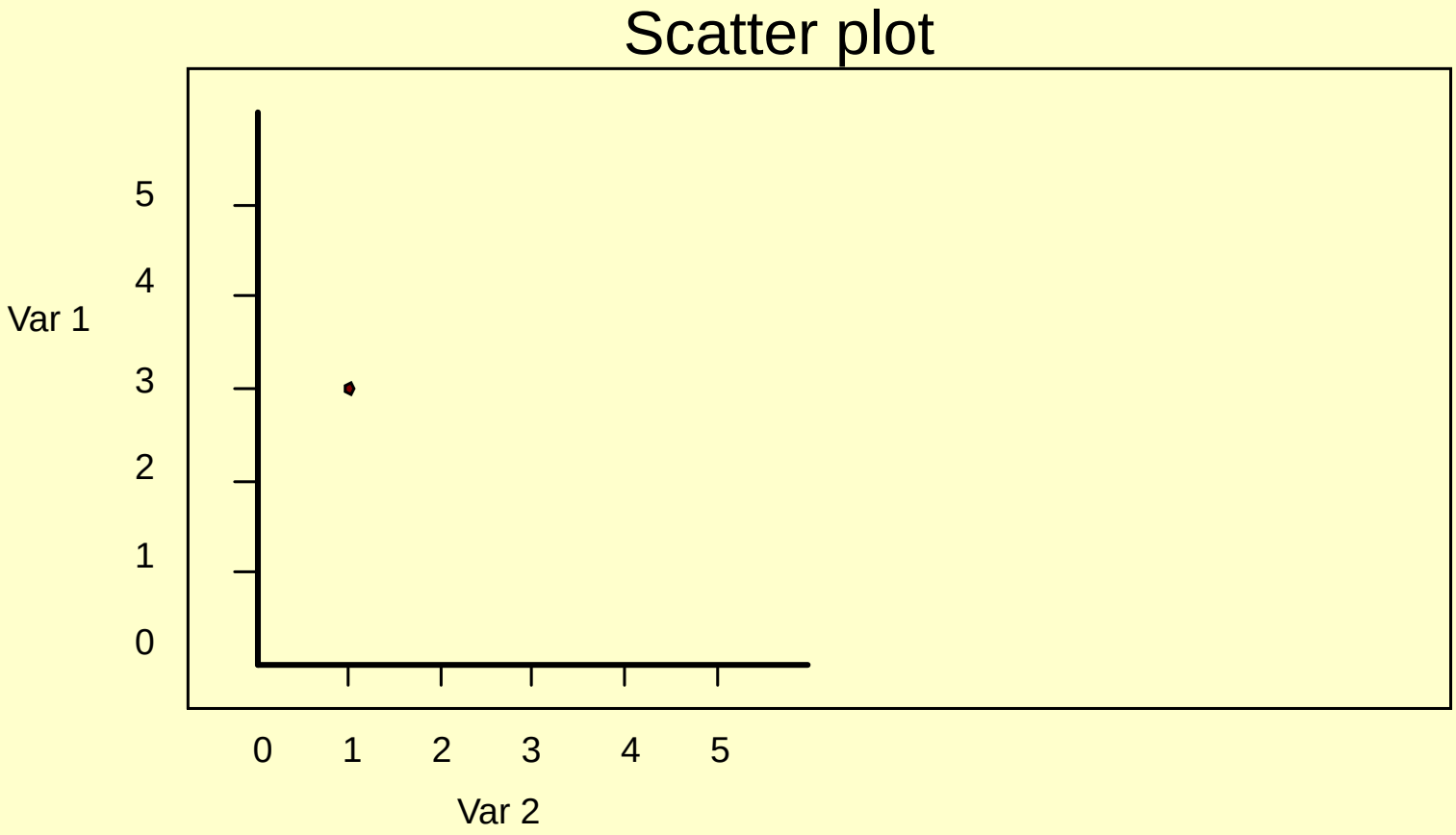
Let users control classifications (high, medium, low) with sliders

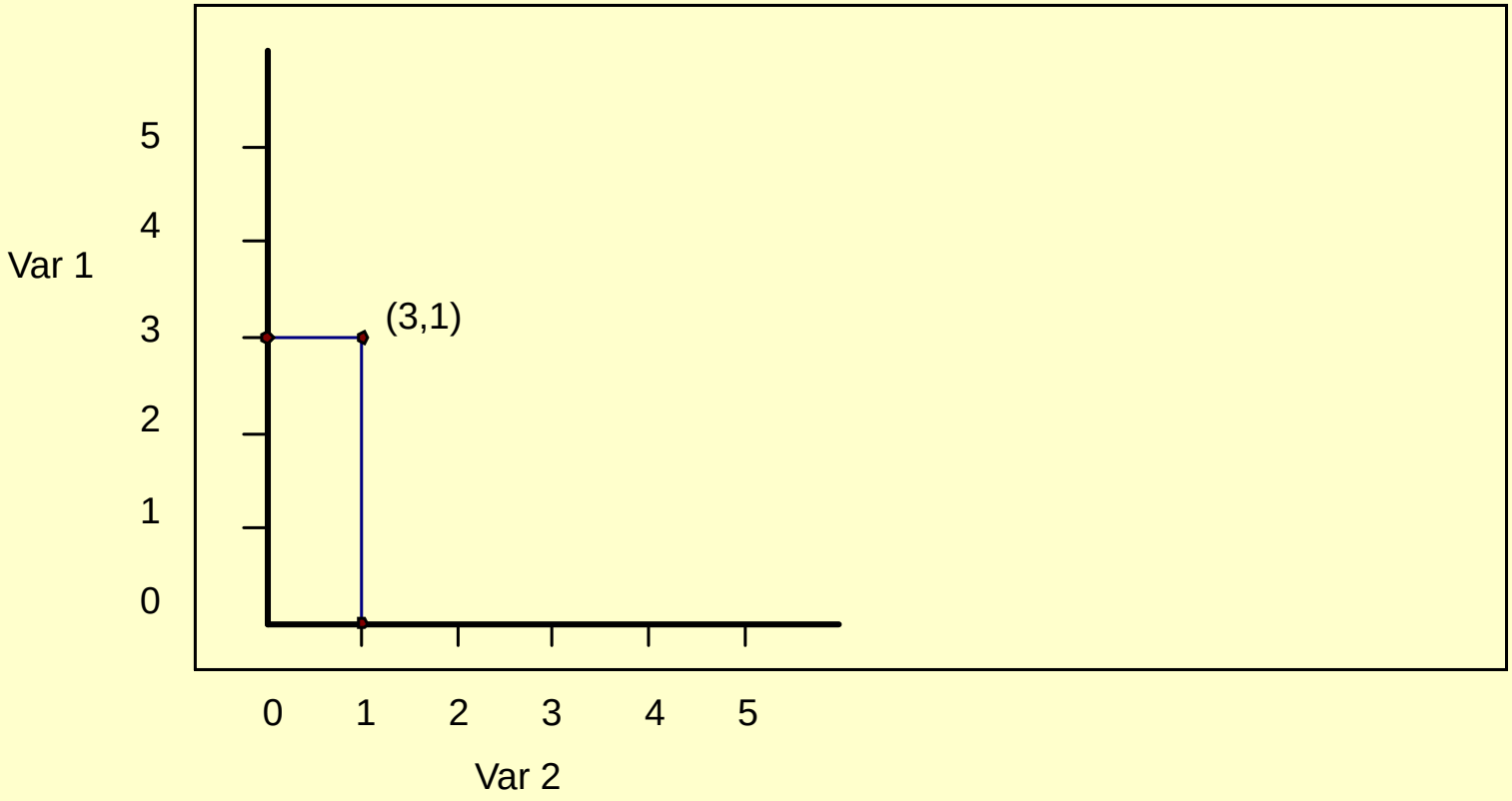


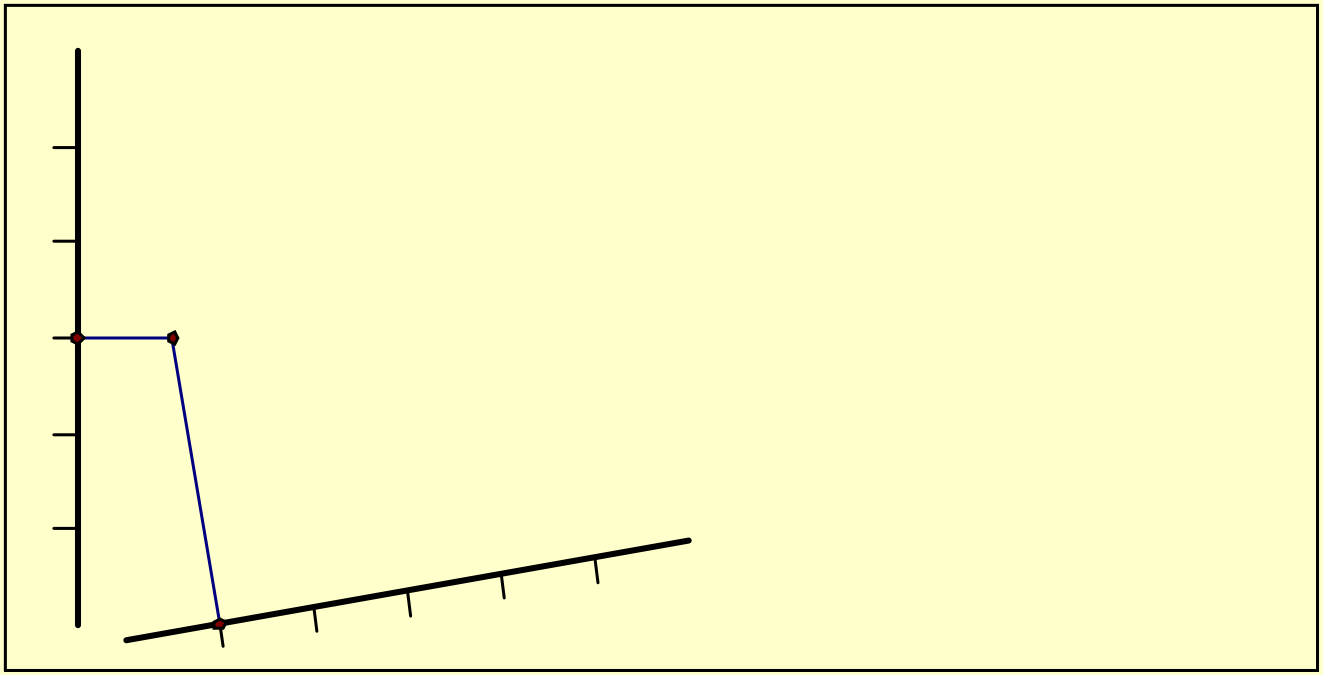
Is there a similar tool for more complex multivariate relationships among factors?

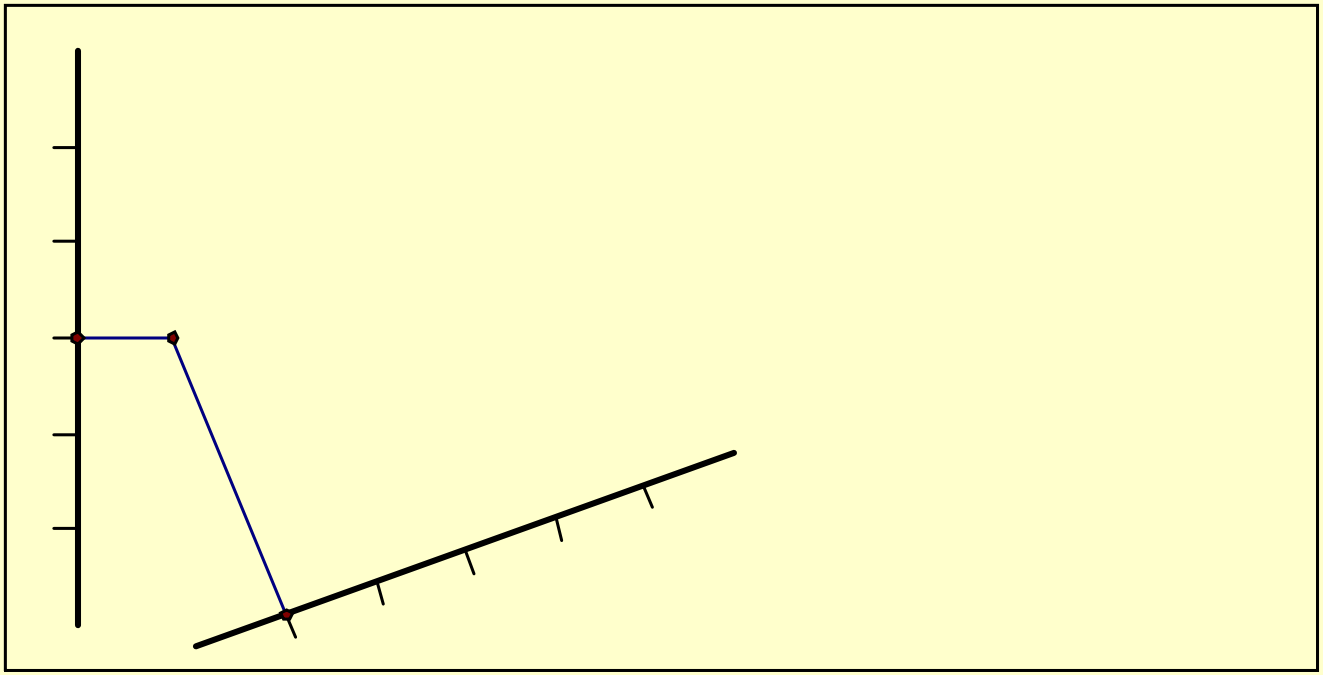
- Parallel coordinate plot linked to a map
- “Linking and brushing”: User can color-code data on plot, see where these places are on map, or vice versa
- Can examine virtually an unlimited # of factors, unlike many other types of plots
- Can add complex statistical tools behind the scenes to guide the user to interesting patterns

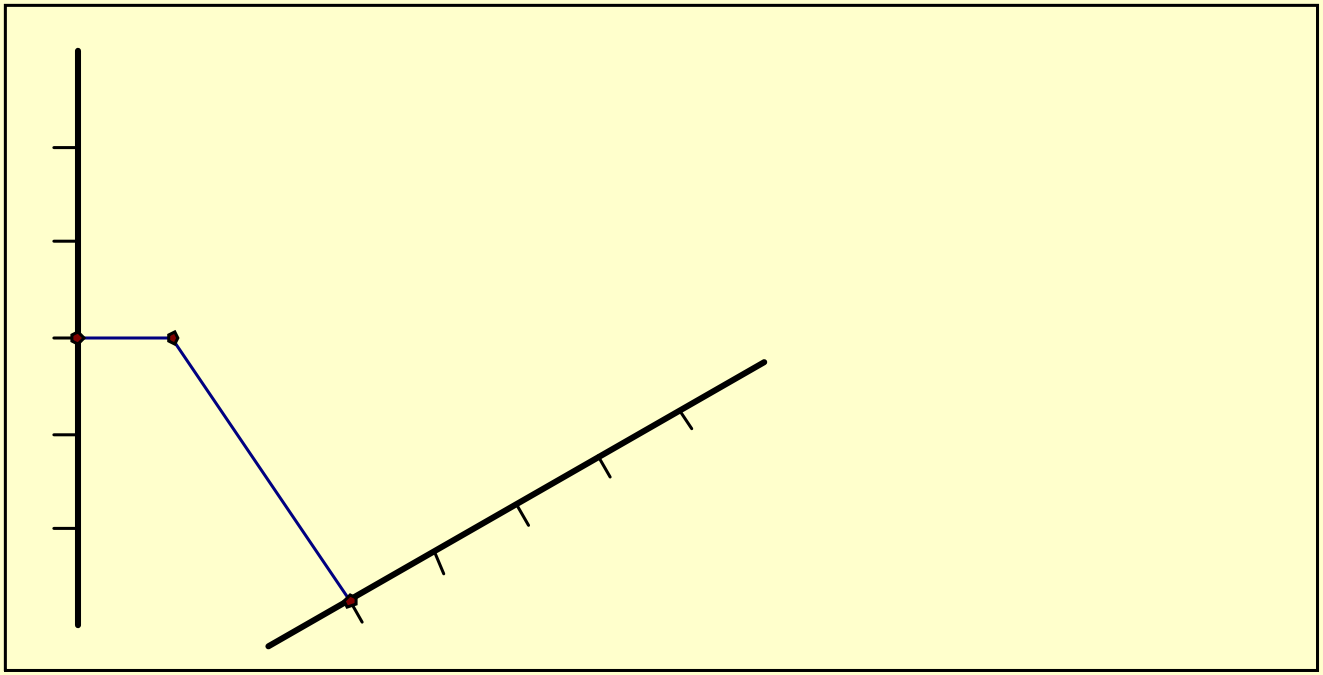
What is a parallel coordinate plot?

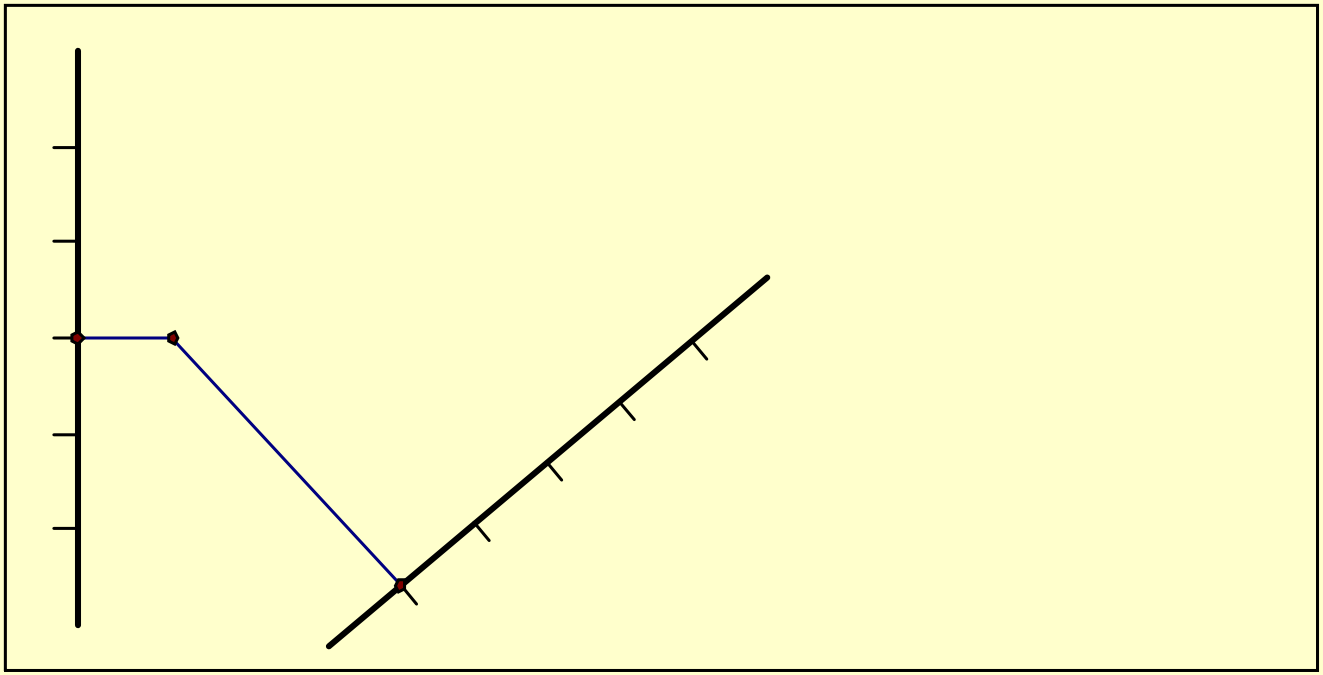


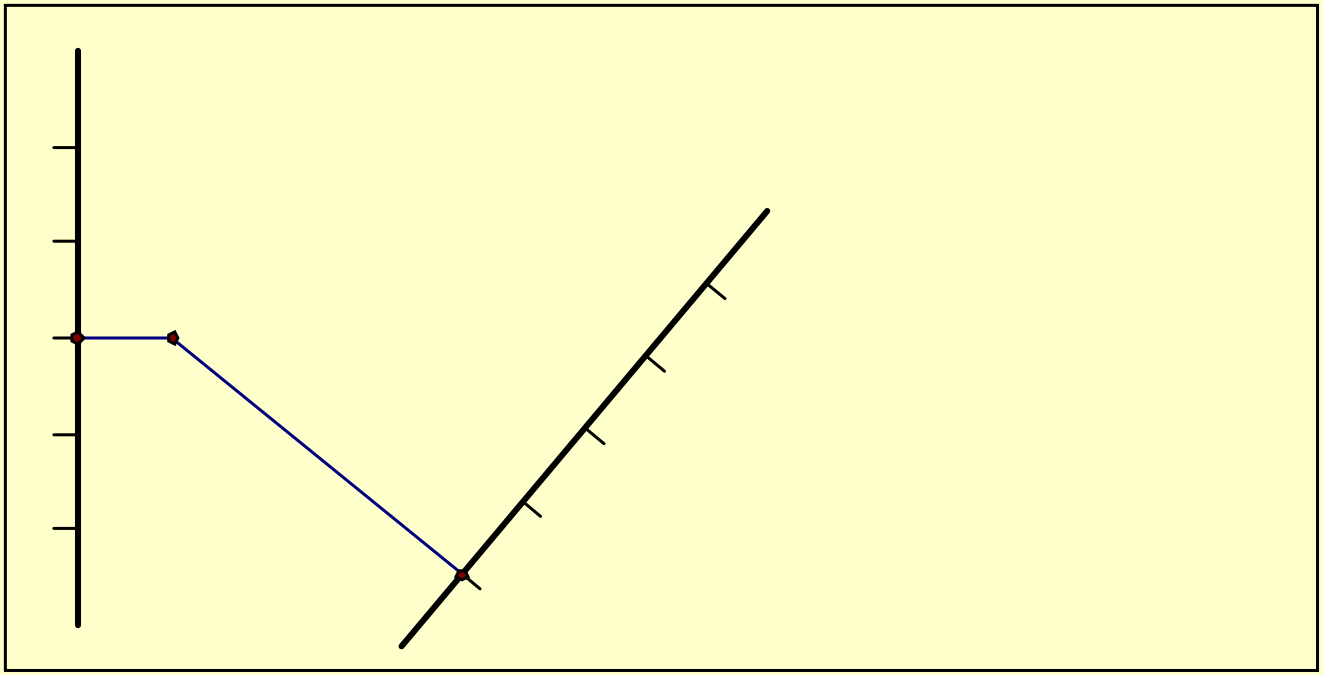

Var 1
5
4
3
2
1
0
(3,1)
0
1
2
3
4
5
Var 2

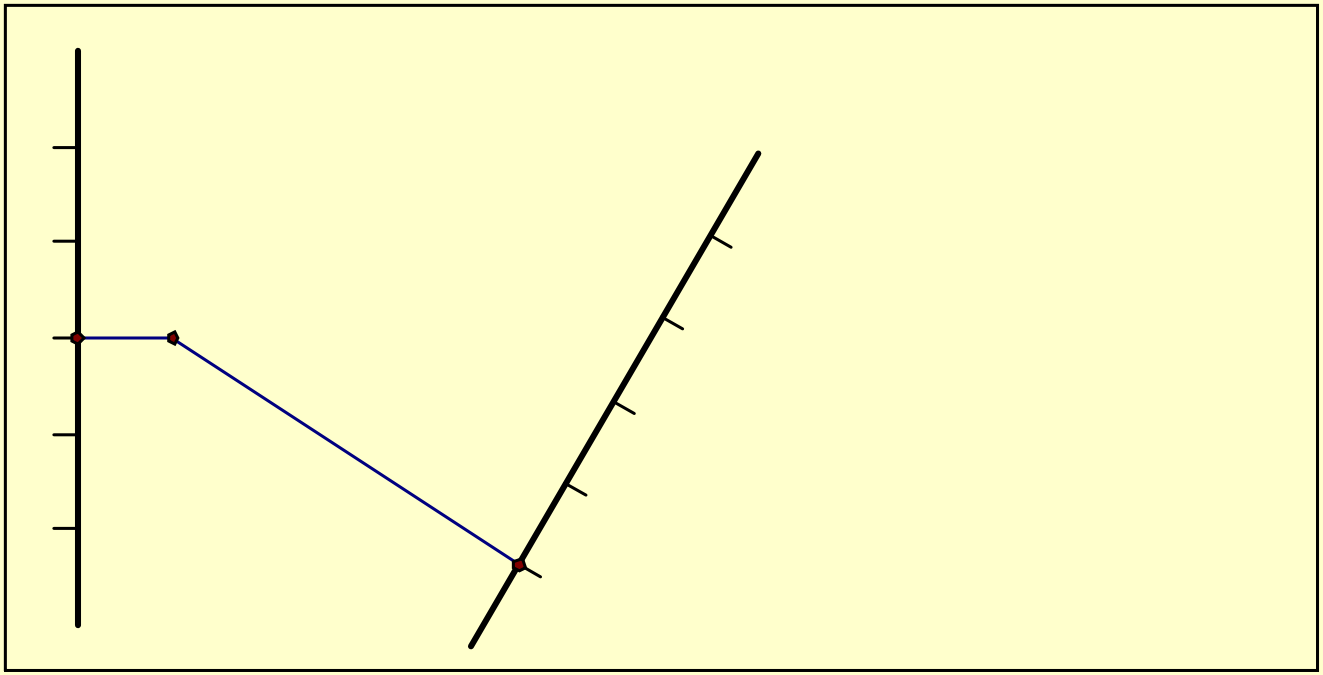


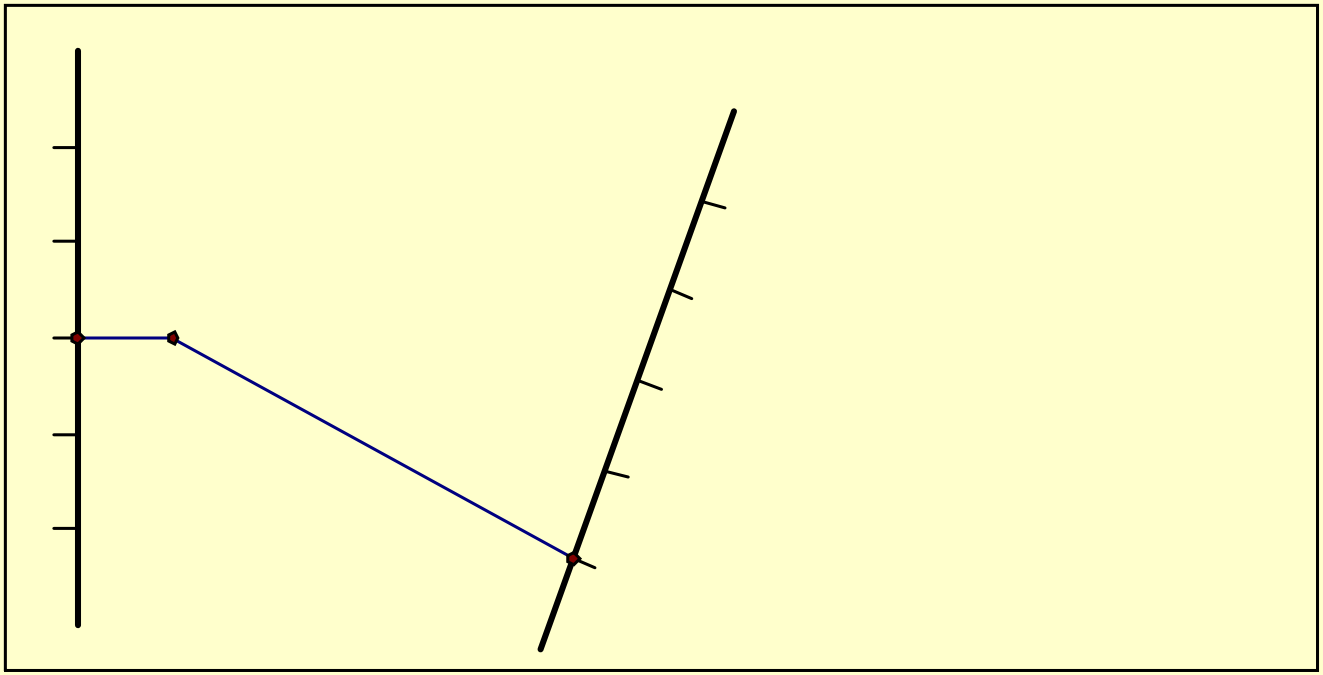


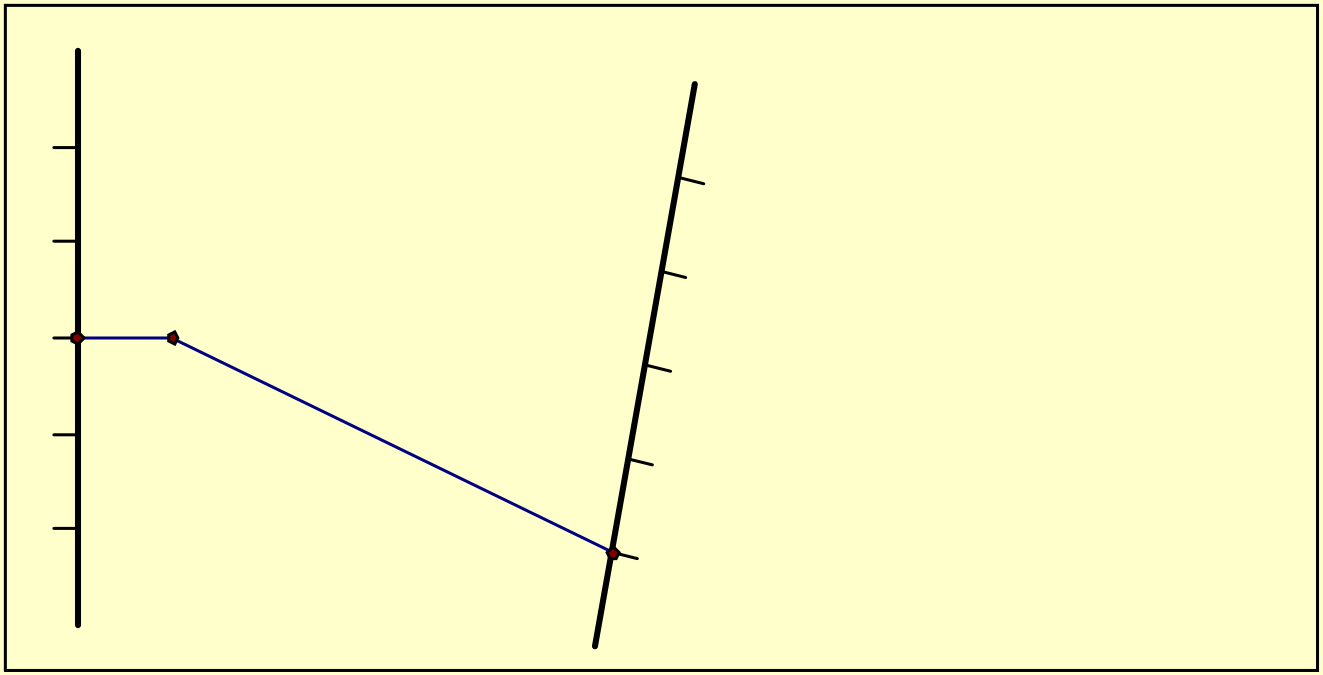


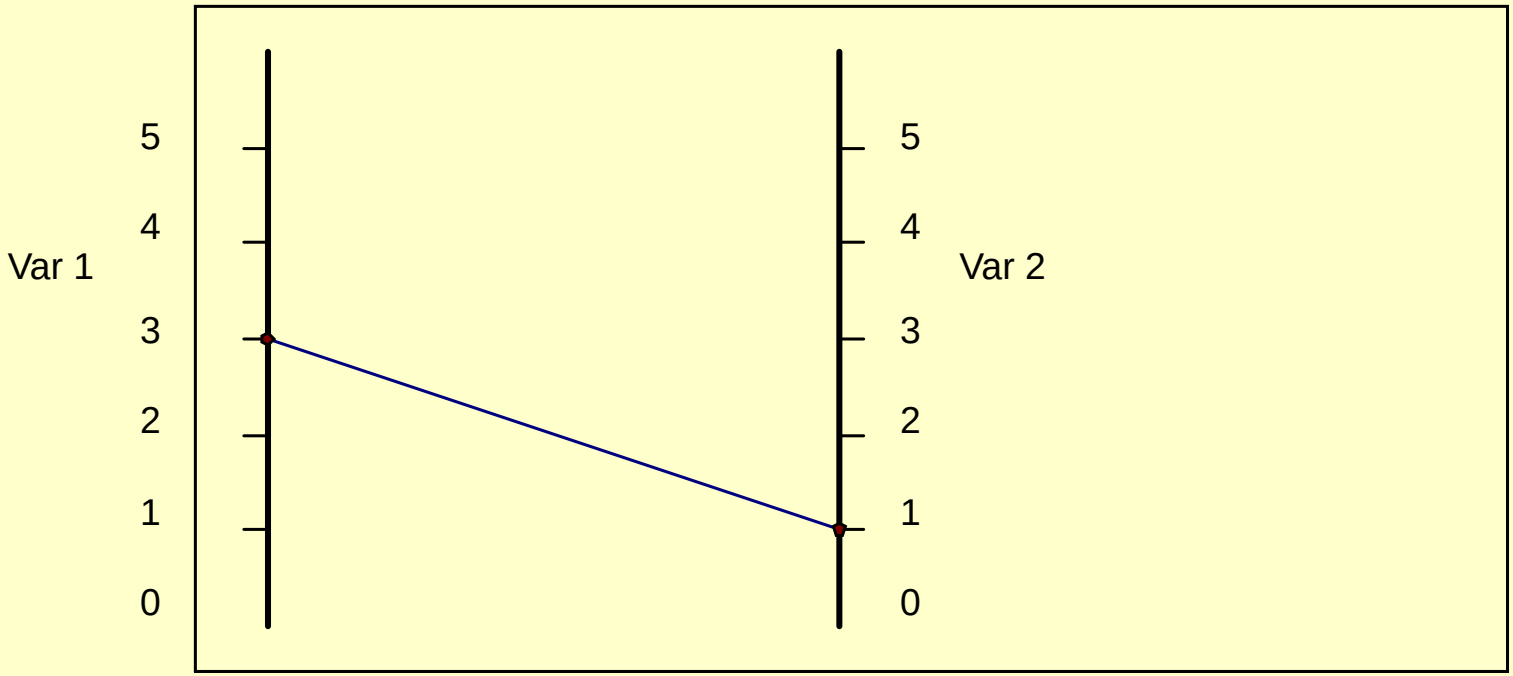


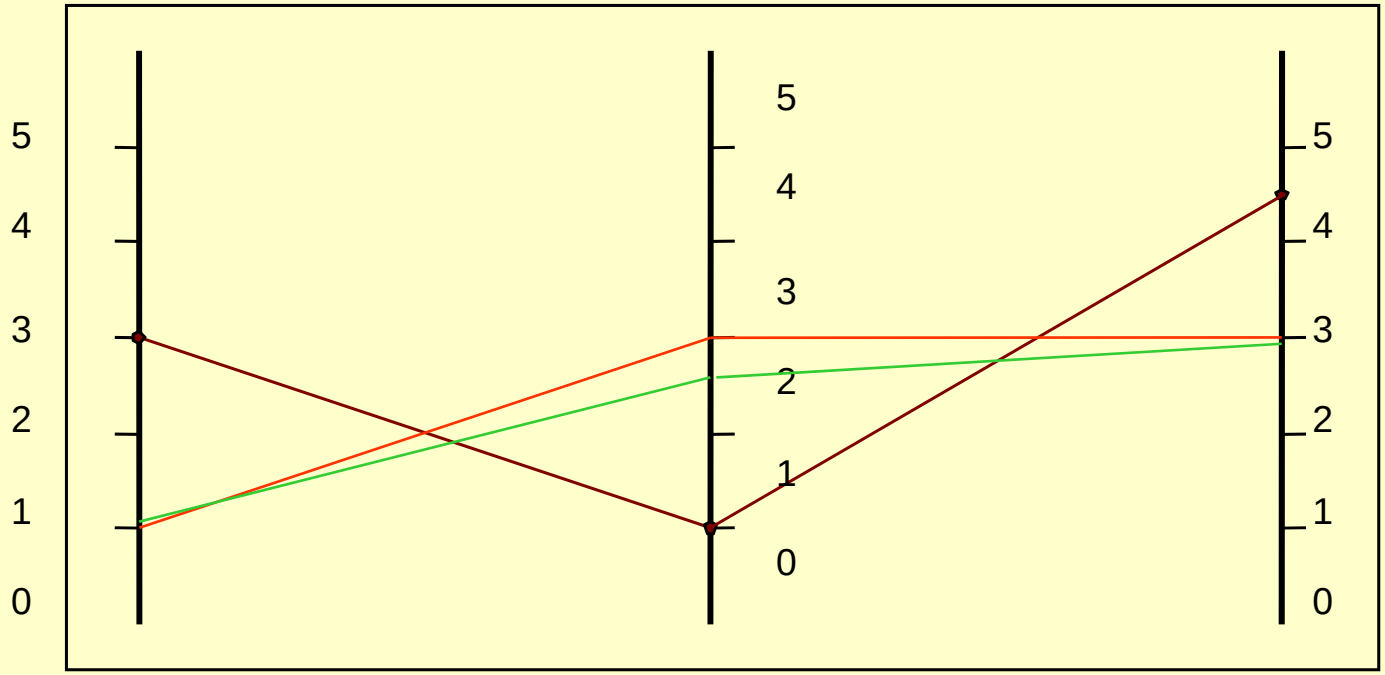




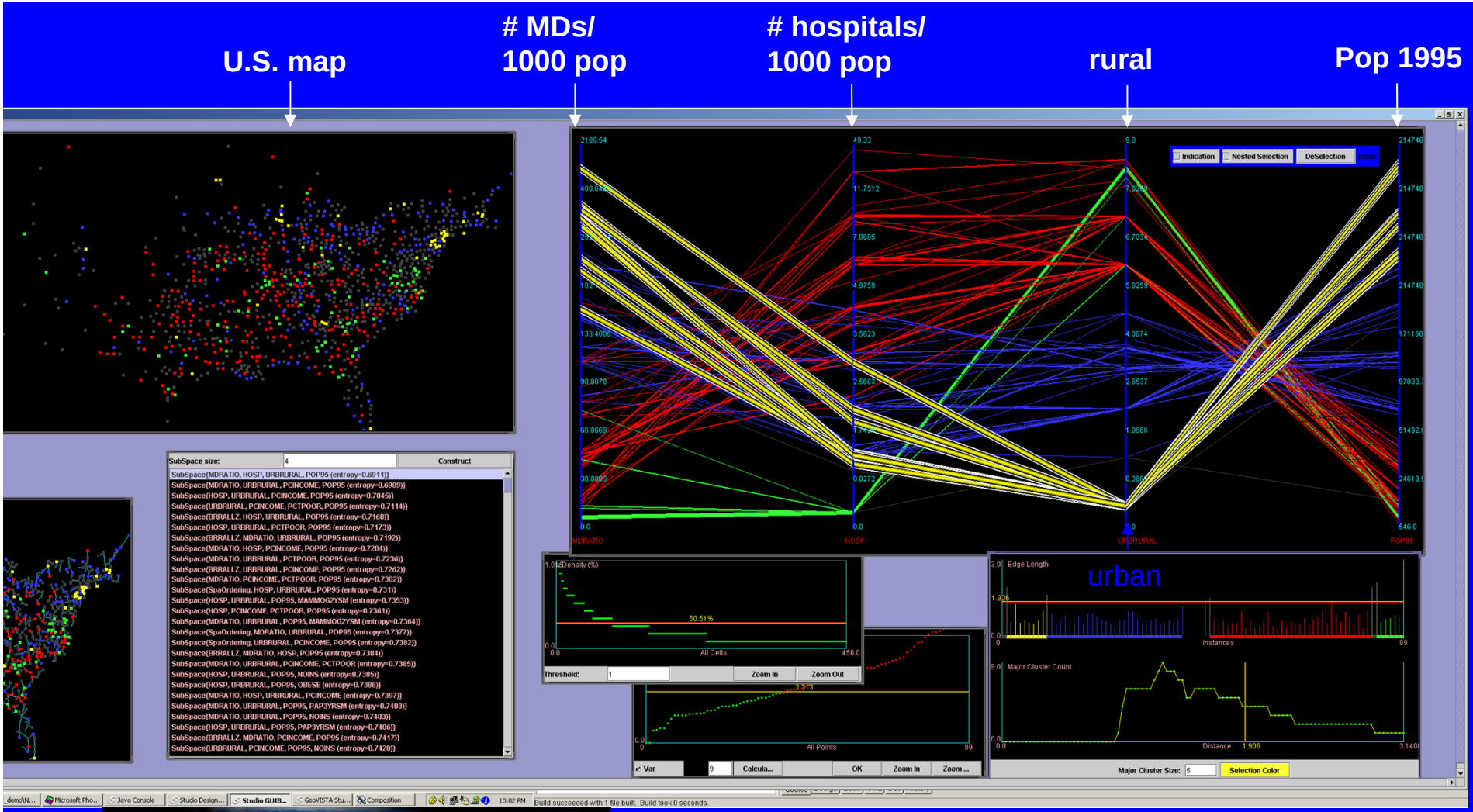





Var 1
5
4
3
2
1
0
5
4
3
2
1
0
Var 2


5
4
3
2
1
0
5
4
3
2
1
0
5
4
3
2
1
0

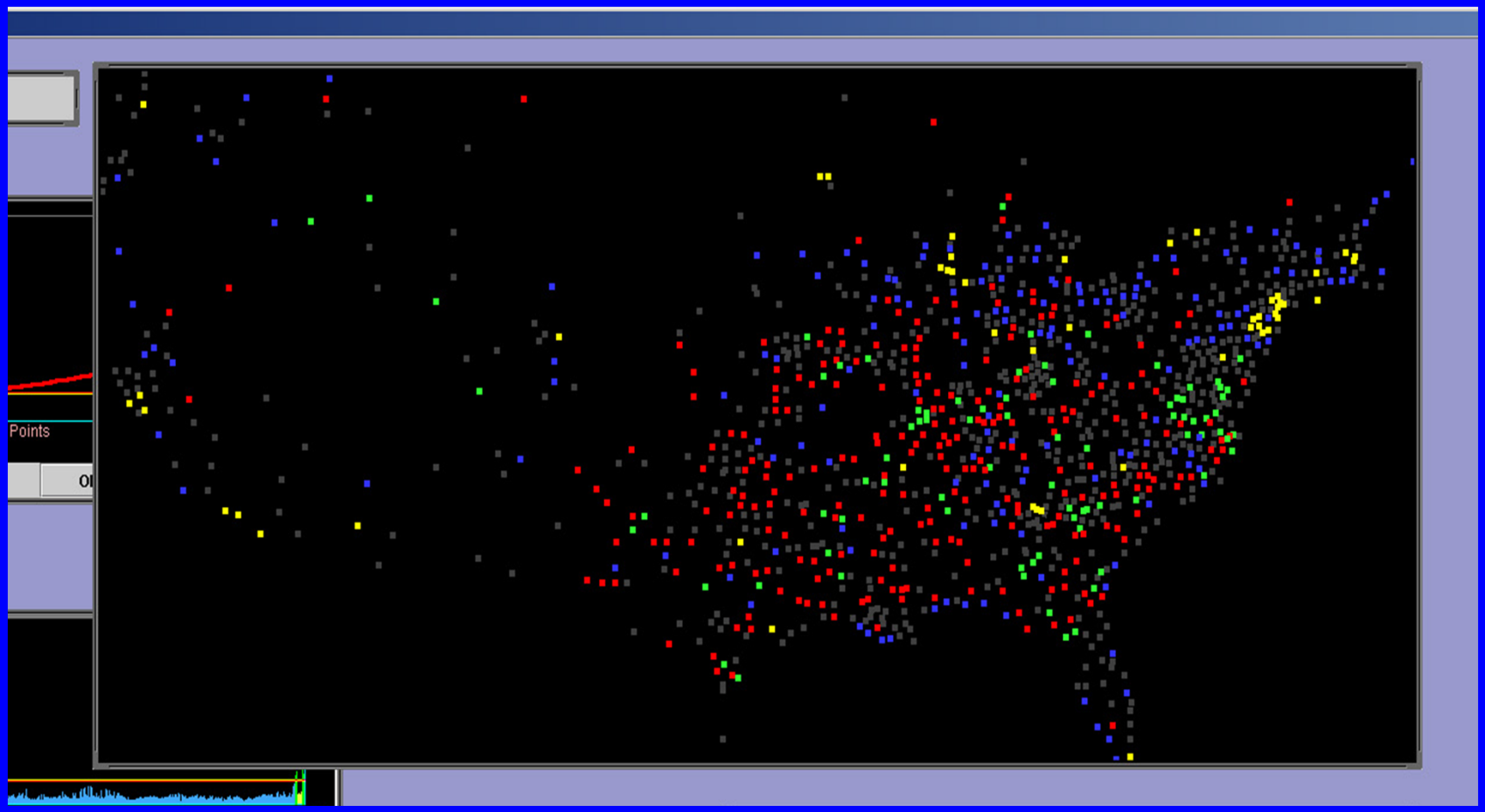
Combining statistical clustering algorithms with new visualization tools



GeoVISTA Center

Alan MacEachren, Diansheng Guo, Penn State University

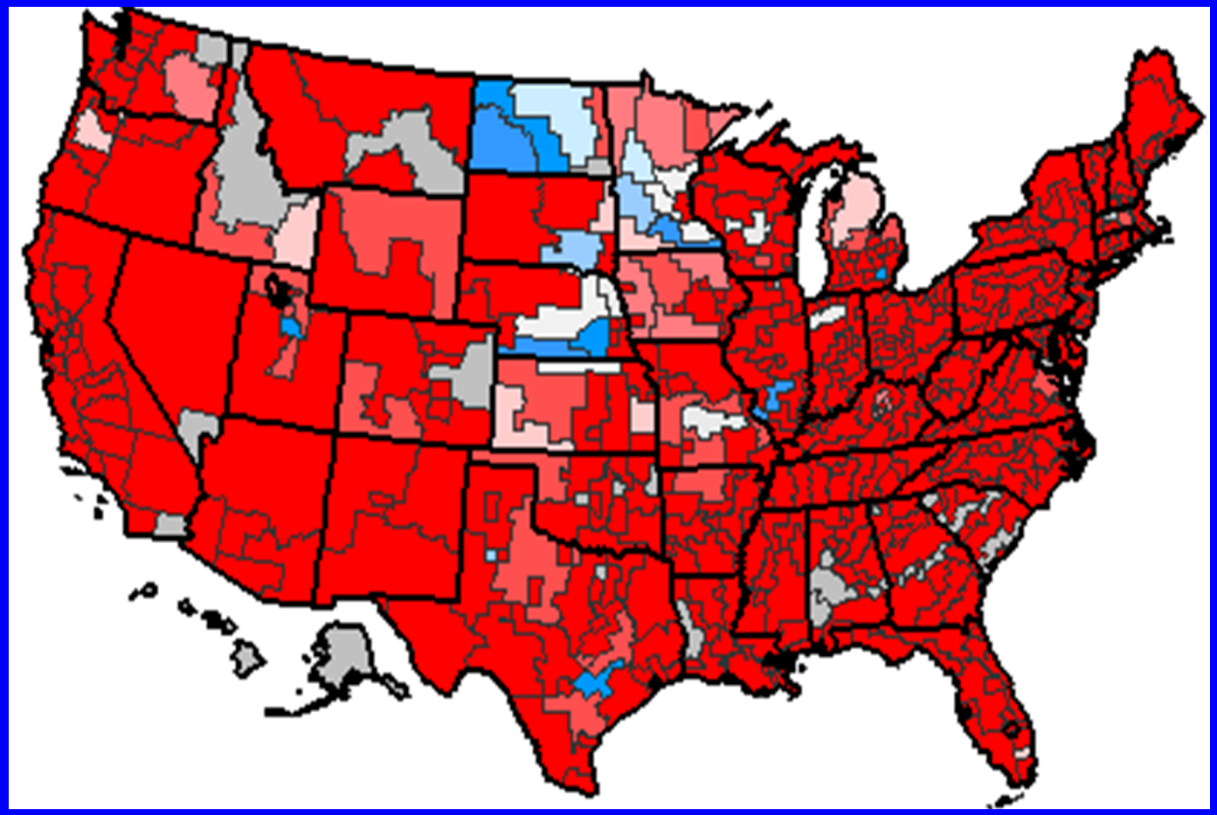
Combining statistical clustering algorithms with new visualization tools



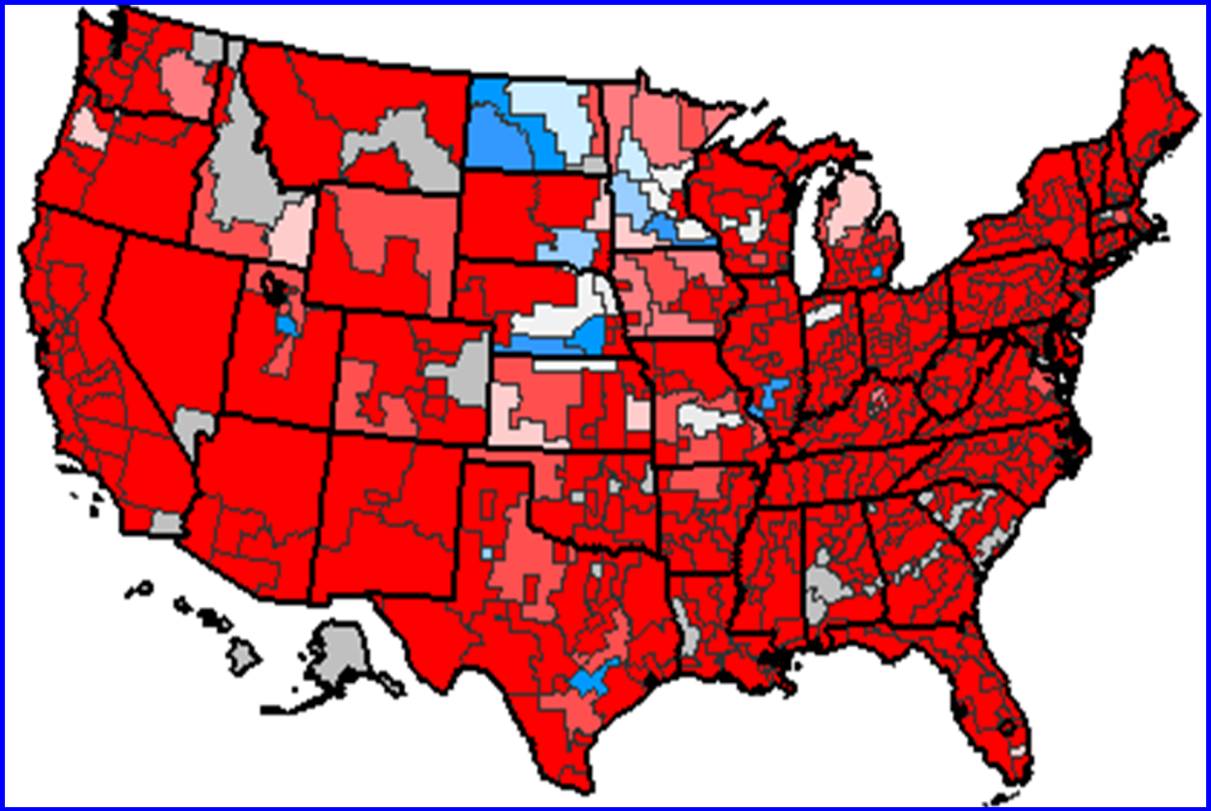
Alan MacEachren, Diansheng Guo, Penn State University

Cervix uteri cancer mortality rates, white females, 1950-94
Time trends by State Economic Area
(dark red=highest rates, dark blue=lowest, gray=sparse data)

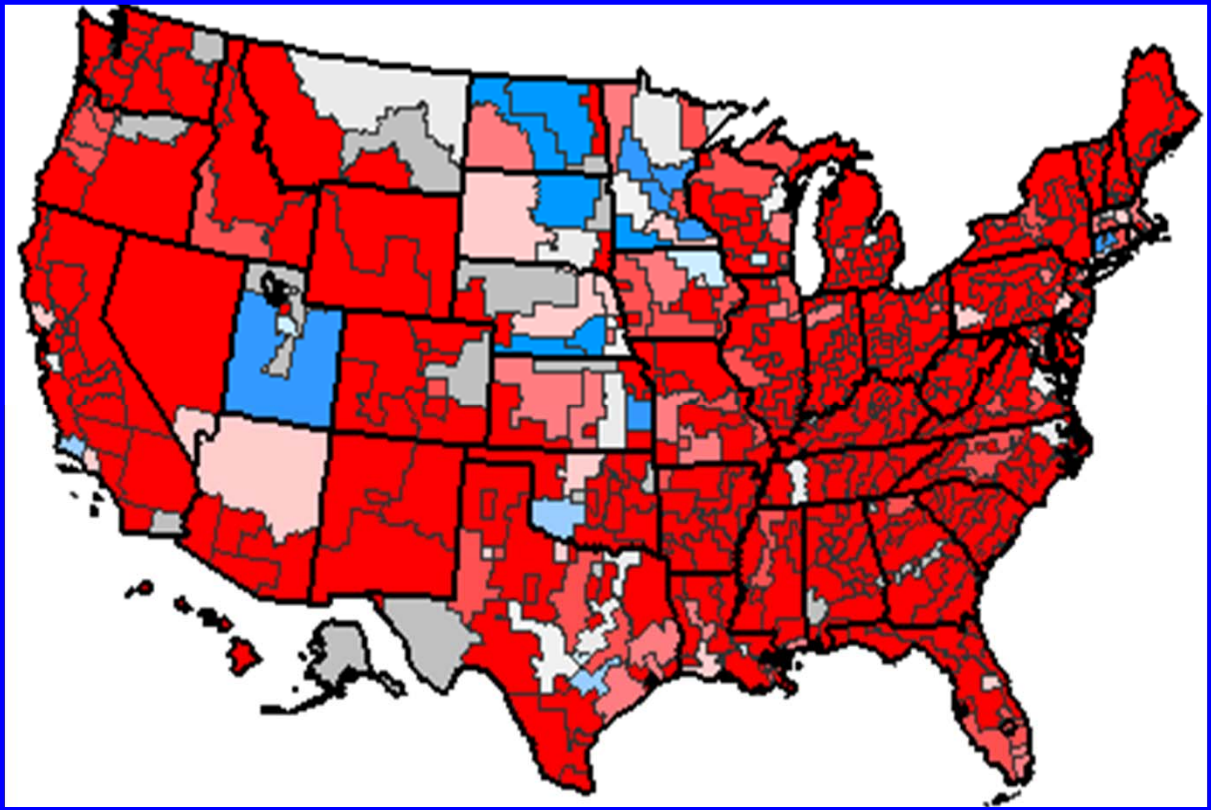
1950-54



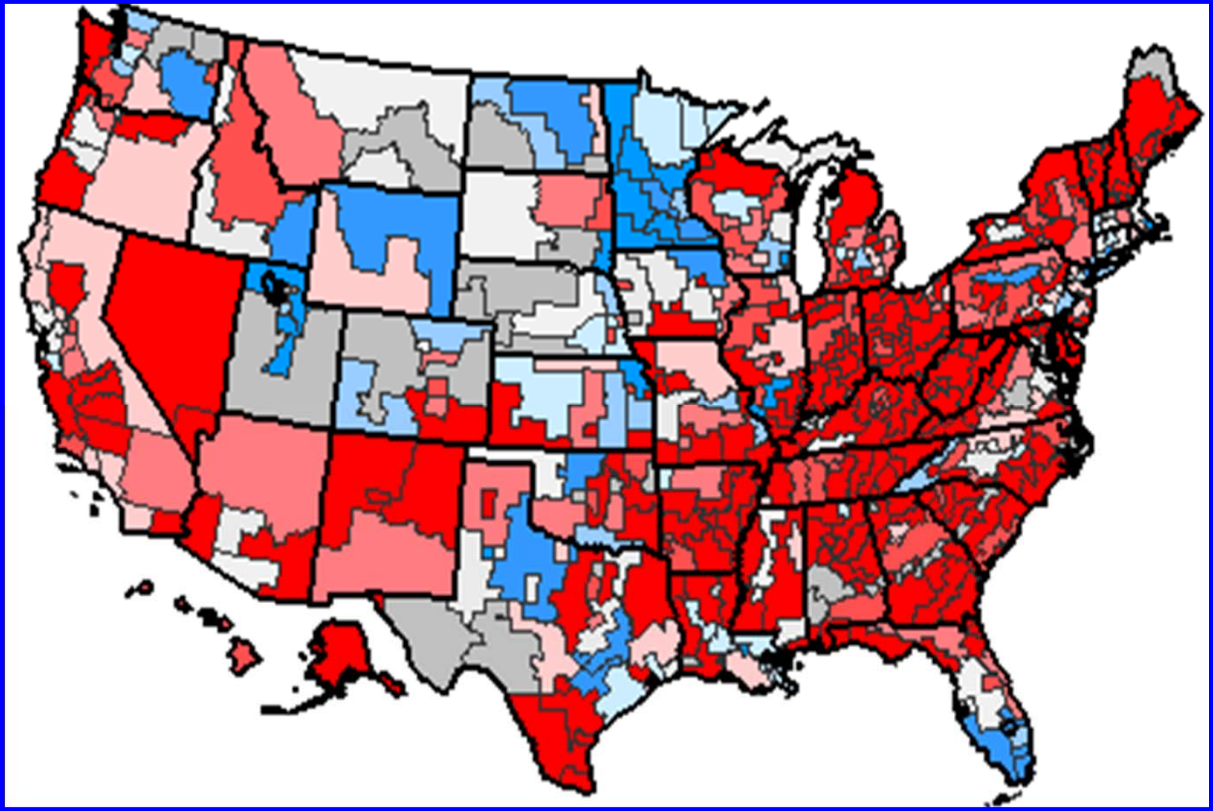
1955-59



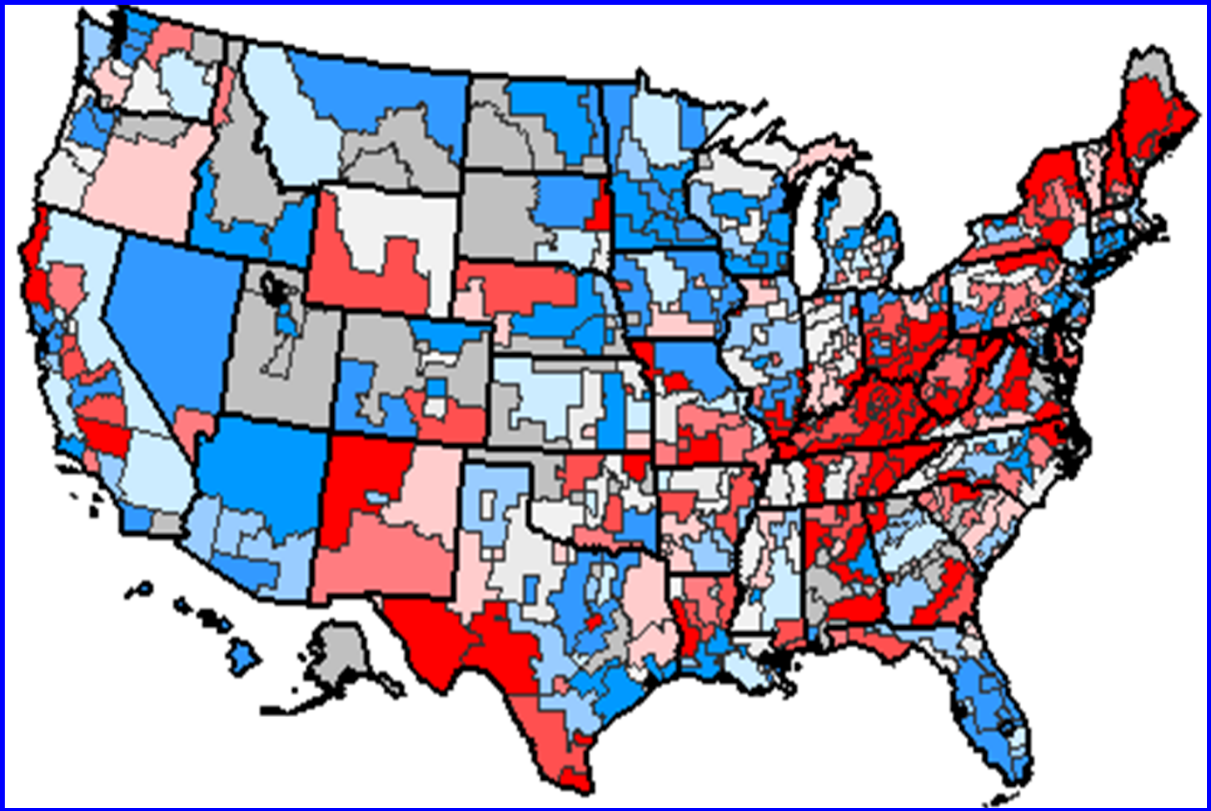
1960-64



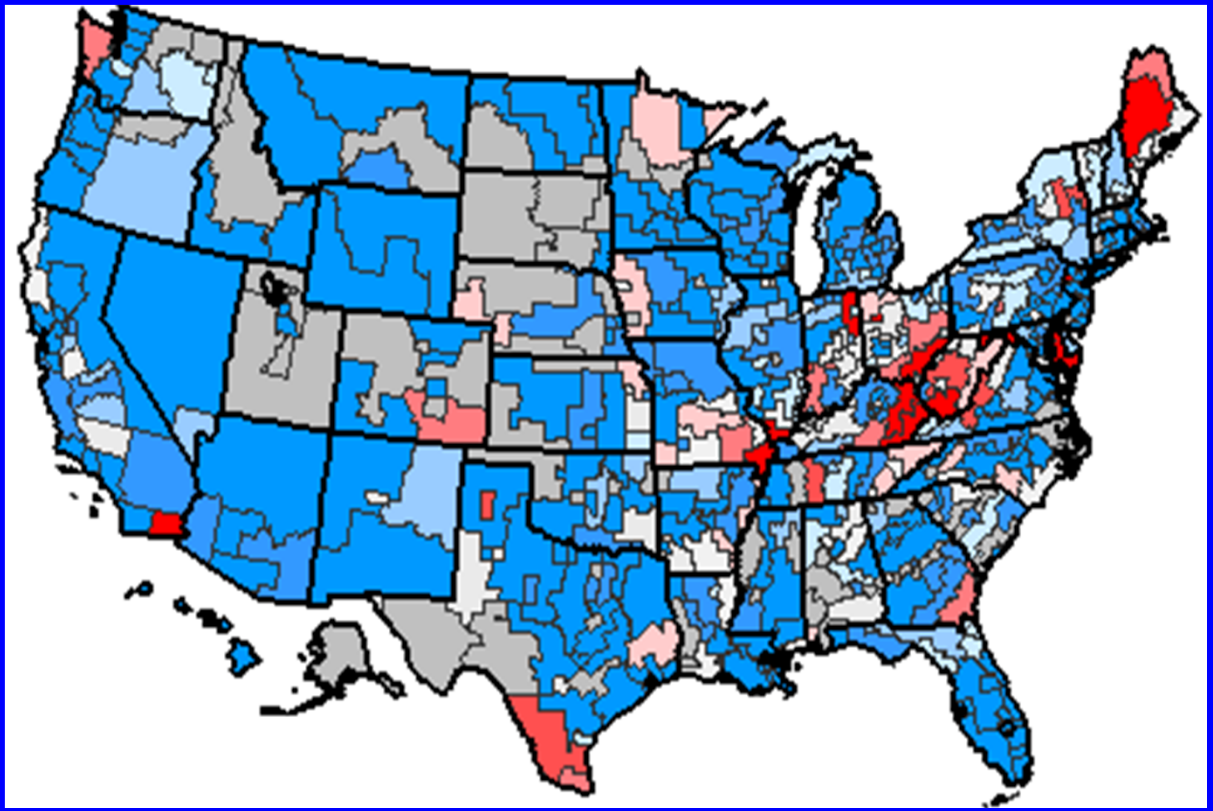
1965-69



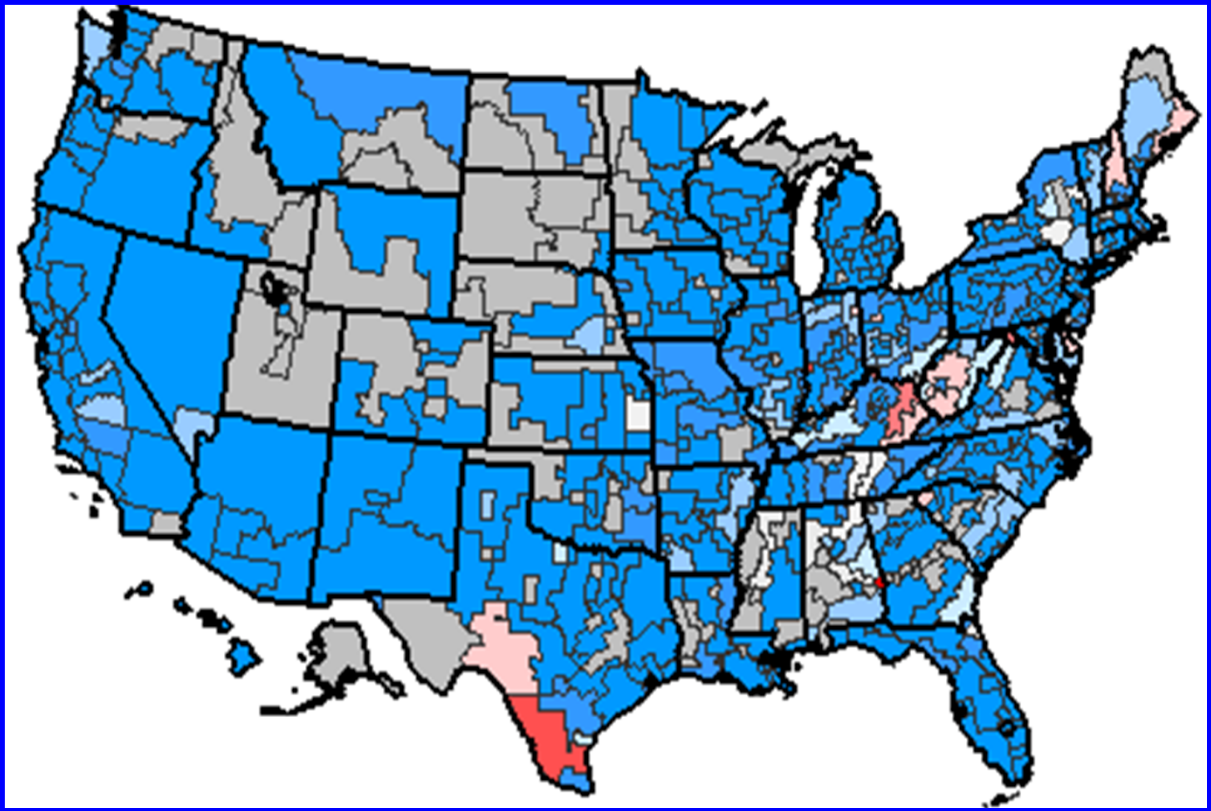
1970-74



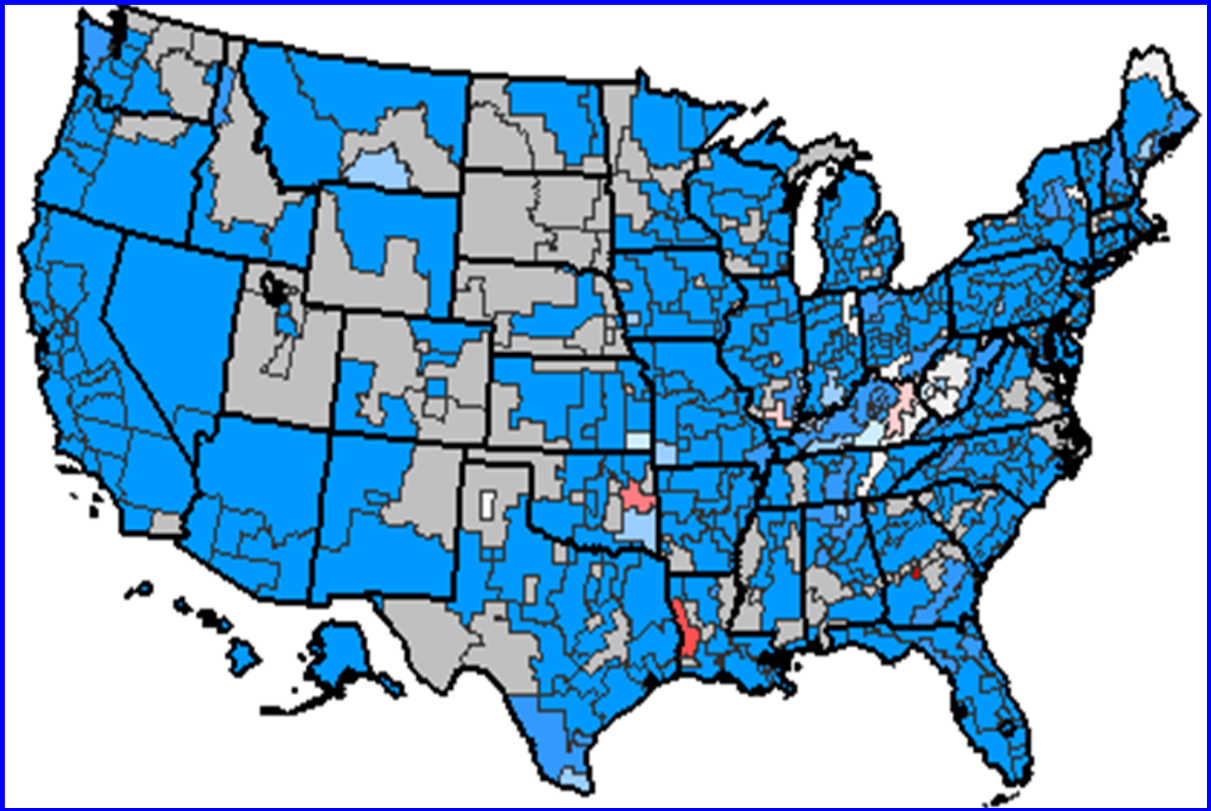
1975-79



1980-84

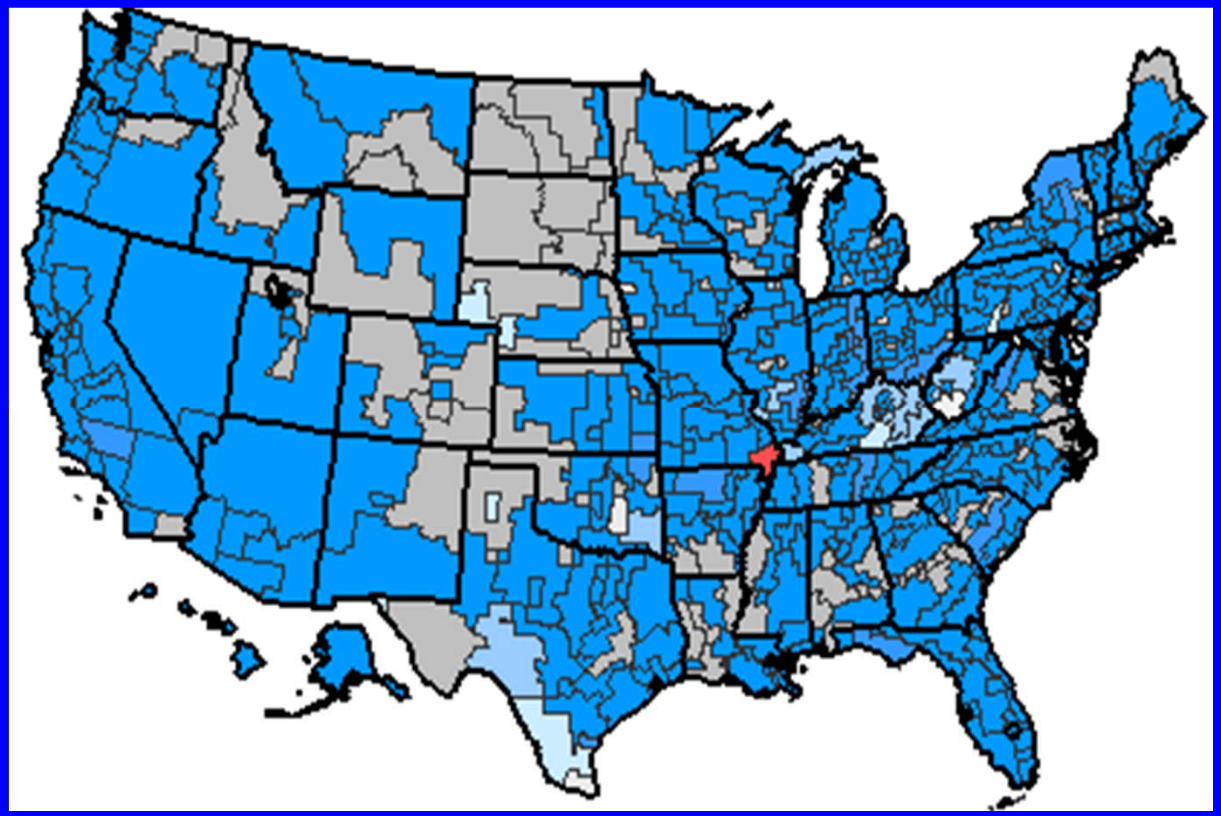


1985-89



Cervix uteri cancer mortality rates, white females, 1950-94
Time trends by State Economic Area
(dark red=highest rates, dark blue=lowest, gray=sparse data)

1990-94



Collaborators

- NCI is a partner in a Digital Government Initiative (NSF) grant to develop better visualization tools to "improve the way government serves the American people" (web site: diggov.org)

  Dan Carr, George Mason University

  Alan MacEachren, Penn State University

  David Scott, Rice University

- NCI geographic information systems grant

  Alan MacEachren, Penn State University

- NCI sabbatical

  Dan Carr, George Mason University