Improving the communication of geographic patterns of disease through computer-based tools

Linda Williams Pickle, Ph.D.

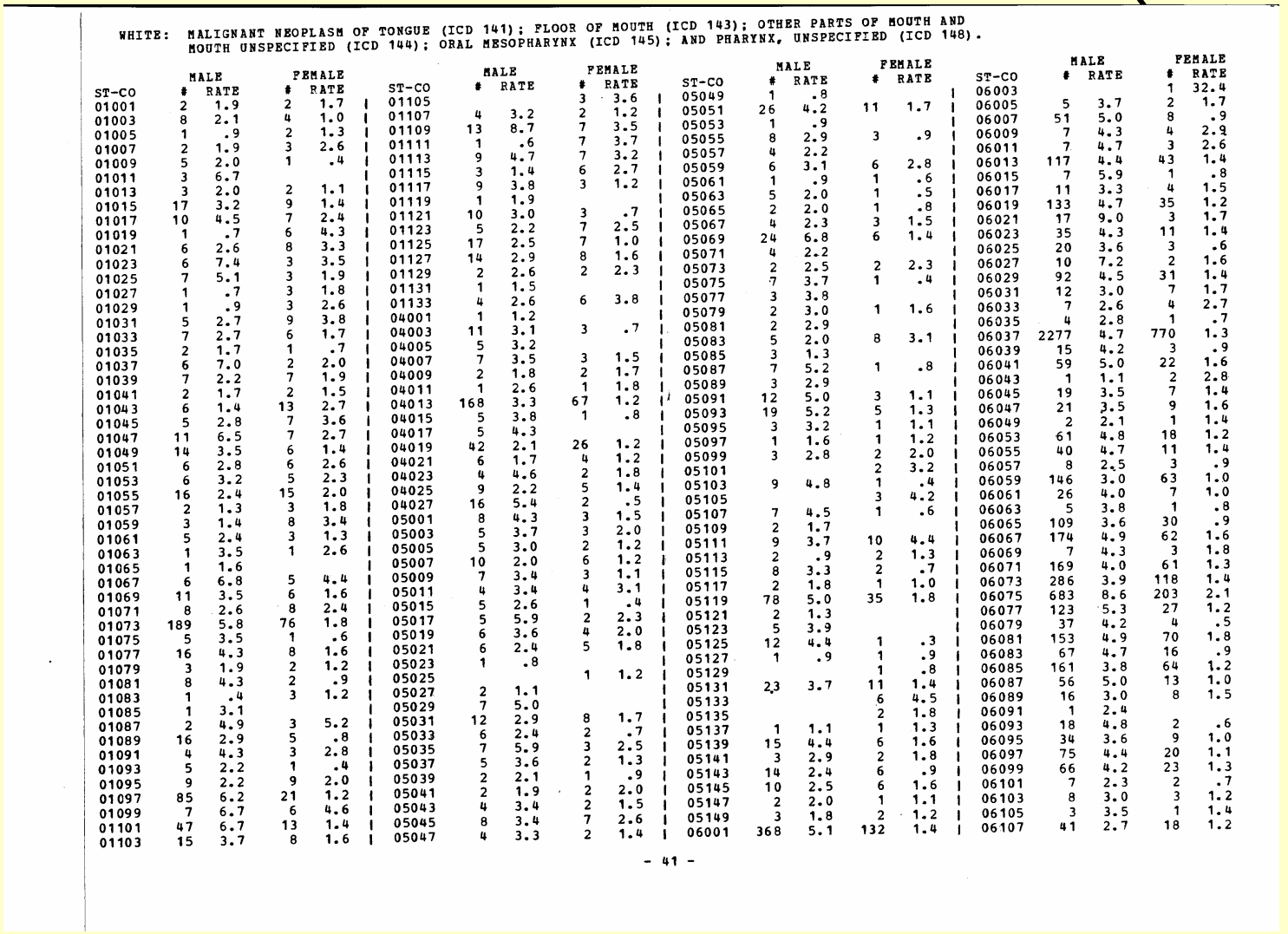
Email: PICKLEL@MAIL.NIH.GOV

ASA meeting
August 10, 2005

Outline

- Background
- Map design research at NCHS & NCI
  - Cognitive research methods
  - Basic map style
  - Legend design
  - Color choices
  - Indication of unreliable rates
  - Classification of rates into color categories
- Development of new graphical tools for communication
  - Smoothing
  - Cluster identification
  - Linked micromap plots
  - Exploratory Spatio-Temporal Analysis Tool (ESTAT)
- Communication over the web
  - NCI Cancer Atlas
  - Long Island Breast Cancer Study GIS
  - State Cancer Profiles

Mortality data by county, sex, race & cancer: Published in tabular form in 1974 (700 p.!)

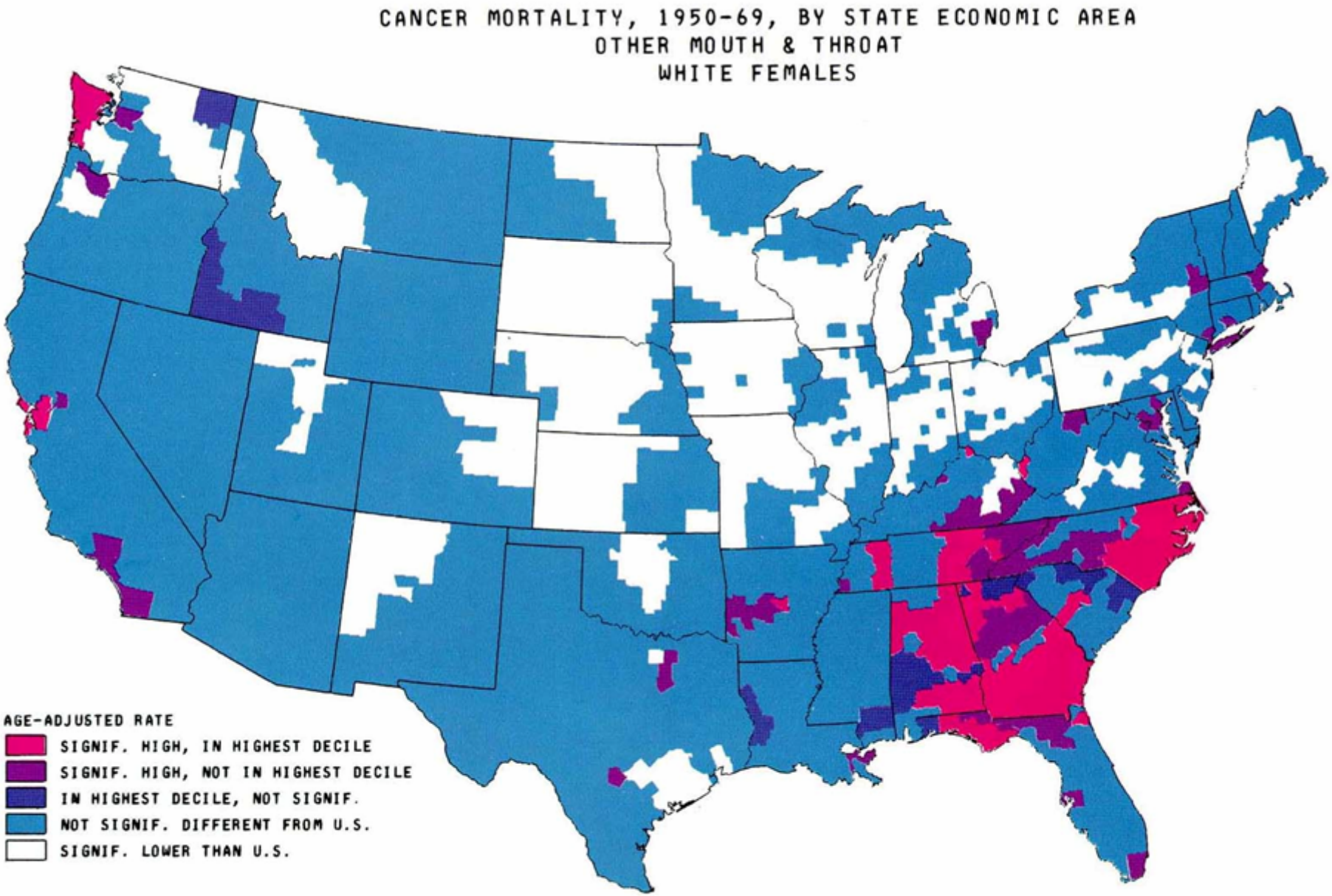


WHITE: MALIGNANT NEOPLASM OF TONGUE (ICD 141); FLOOR OF MOUTH (ICD 143); OTHER PARTS OF MOUTH AND MOUTH UNSPECIFIED (ICD 144); ORAL MESOPHARYNX (ICD 145); AND PHARYNX, UNSPECIFIED (ICD 148).

	MALE		FEMALE			MALE		FEMALE			MALE		FEMALE			MALE		FEMALE	
ST-CO	#	RATE	#	RATE	ST-CO	#	RATE	#	RATE	ST-CO	#	RATE	#	RATE	ST-CO	#	RATE	#	RATE
01001	2	1.9	2	1.7	01105			3	3.6	05049	1	.8			06003			1	32.4
01003	8	2.1	4	1.0	01107	4	3.2	2	1.2	05051	26	4.2	11	1.7	06005	5	3.7	2	1.7
01005	1	.9	2	1.3	01109	13	8.7	7	3.5	05053	1	.9			06007	51	5.0	8	.9
01007	2	1.9	3	2.6	01111	1	.6	7	3.7	05055	8	2.9	3	.9	06009	7	4.3	4	2.9
01009	5	2.0	1	.4	01113	9	4.7	7	3.2	05057	4	2.2			06011	7	4.7	3	2.6
01011	3	6.7			01115	3	1.4	6	2.7	05059	6	3.1	6	2.8	06013	117	4.4	43	1.4
01013	3	2.0	2	1.1	01117	9	3.8	3	1.2	05061	1	.9	1	.6	06015	7	5.9	1	.8
01015	17	3.2	9	1.4	01119	1	1.9			05063	5	2.0	1	.5	06017	11	3.3	4	1.5
01017	10	4.5	7	2.4	01121	10	3.0	3	.7	05065	2	2.0	1	.8	06019	133	4.7	35	1.2
01019	1	.7	6	4.3	01123	5	2.2	7	2.5	05067	4	2.3	3	1.5	06021	17	9.0	3	1.7
01021	6	2.6	8	3.3	01125	17	2.5	7	1.0	05069	24	6.8	6	1.4	06023	35	4.3	11	1.4
01023	6	7.4	3	3.5	01127	14	2.9	8	1.6	05071	4	2.2			06025	20	3.6	3	.6
01025	7	5.1	3	1.9	01129	2	2.6	2	2.3	05073	2	2.5	2	2.3	06027	10	7.2	2	1.6
01027	1	.7	3	1.8	01131	1	1.5			05075	7	3.7	1	.4	06029	92	4.5	31	1.4
01029	1	.9	3	2.6	01133	4	2.6	6	3.8	05077	3	3.8			06031	12	3.0	7	1.7
01031	5	2.7	9	3.8	04001	1	1.2			05079	2	3.0	1	1.6	06033	7	2.6	4	2.7
01033	7	2.7	6	1.7	04003	11	3.1	3	.7	05081	2	2.9			06035	4	2.8	1	.7
01035	2	1.7	1	.7	04005	5	3.2			05083	5	2.0	8	3.1	06037	2277	4.7	770	1.3
01037	6	7.0	2	2.0	04007	7	3.5	3	1.5	05085	3	1.3			06039	15	4.2	3	.9
01039	7	2.2	7	1.9	04009	2	1.8	2	1.7	05087	7	5.2	1	.8	06041	59	5.0	22	1.6
01041	2	1.7	2	1.5	04011	1	2.6	1	1.8	05089	3	2.9			06043	1	1.1	2	2.8
01043	6	1.4	13	2.7	04013	168	3.3	67	1.2	05091	12	5.0	3	1.1	06045	19	3.5	7	1.4
01045	5	2.8	7	3.6	04015	5	3.8	1	.8	05093	19	5.2	5	1.3	06047	21	3.5	9	1.6
01047	11	6.5	7	2.7	04017	5	4.3			05095	3	3.2	1	1.1	06049	2	2.1	1	1.4
01049	14	3.5	6	1.4	04019	42	2.1	26	1.2	05097	1	1.6	1	1.2	06053	61	4.8	18	1.2
01051	6	2.8	6	2.6	04021	6	1.7	4	1.2	05099	3	2.8	2	2.0	06055	40	4.7	11	1.4
01053	6	3.2	5	2.3	04023	4	4.6	2	1.8	05101			2	3.2	06057	8	2.5	3	.9
01055	16	2.4	15	2.0	04025	9	2.2	5	1.4	05103	9	4.8	1	.4	06059	146	3.0	63	1.0
01057	2	1.3	3	1.8	04027	16	5.4	2	.5	05105			3	4.2	06061	26	4.0	7	1.0
01059	3	1.4	8	3.4	05001	8	4.3	3	1.5	05107	7	4.5	1	.6	06063	5	3.8	1	.8
01061	5	2.4	3	1.3	05003	5	3.7	3	2.0	05109	2	1.7			06065	109	3.6	30	.9
01063	1	3.5	1	2.6	05005	5	3.0	2	1.2	05111	9	3.7	10	4.4	06067	174	4.9	62	1.6
01065	1	1.6			05007	10	2.0	6	1.2	05113	2	.9	2	1.3	06069	7	4.3	3	1.8
01067	6	6.8	5	4.4	05009	7	3.4	3	1.1	05115	8	3.3	2	.7	06071	169	4.0	61	1.3
01069	11	3.5	6	1.6	05011	4	3.4	4	3.1	05117	2	1.8	1	1.0	06073	286	3.9	118	1.4
01071	8	2.6	8	2.4	05015	5	2.6	1	.4	05119	78	5.0	35	1.8	06075	683	8.6	203	2.1
01073	189	5.8	76	1.8	05017	5	5.9	2	2.3	05121	2	1.3			06077	123	5.3	27	1.2
01075	5	3.5	1	.6	05019	6	3.6	4	2.0	05123	5	3.9			06079	37	4.2	4	.5
01077	16	4.3	8	1.6	05021	6	2.4	5	1.8	05125	12	4.4	1	.3	06081	153	4.9	70	1.8
01079	3	1.9	2	1.2	05023	1	.8			05127	1	.9	1	.9	06083	67	4.7	16	.9
01081	8	4.3	2	.9	05025			1	1.2	05129			1	.8	06085	161	3.8	64	1.2
01083	1	.4	3	1.2	05027	2	1.1			05131	23	3.7	11	1.4	06087	56	5.0	13	1.0
01085	1	3.1			05029	7	5.0			05133			6	4.5	06089	16	3.0	8	1.5
01087	2	4.9	3	5.2	05031	12	2.9	8	1.7	05135			2	1.8	06091	1	2.4		
01089	16	2.9	5	.8	05033	6	2.4	2	.7	05137	1	1.1	1	1.3	06093	18	4.8	2	.6
01091	4	4.3	3	2.8	05035	7	5.9	3	2.5	05139	15	4.4	6	1.6	06095	34	3.6	9	1.0
01093	5	2.2	1	.4	05037	5	3.6	2	1.3	05141	3	2.9	2	1.8	06097	75	4.4	20	1.1
01095	9	2.2	9	2.0	05039	2	2.1	1	.9	05143	14	2.4	6	.9	06099	66	4.2	23	1.3
01097	85	6.2	21	1.2	05041	2	1.9	2	2.0	05145	10	2.5	6	1.6	06101	7	2.3	2	.7
01099	7	6.7	6	4.6	05043	4	3.4	2	1.5	05147	2	2.0	1	1.1	06103	8	3.0	3	1.2
01101	47	6.7	13	1.4	05045	8	3.4	7	2.6	05149	3	1.8	2	1.2	06105	3	3.5	1	1.4
01103	15	3.7	8	1.6	05047	4	3.3	2	1.4	06001	368	5.1	132	1.4	06107	41	2.7	18	1.2

- 41 -

Source: Mason & McKay, U.S. Cancer Mortality by County: 1950-1969, DHEW Publ. No.(NIH) 74-615, 1974.



Source: Mason et al., *Atlas of Cancer Mortality for U.S. Counties*, NCI, 1975.

Cognitive Research Methods

- Focus groups
- Designed experiments
  - Focused on a single map element
  - Random order of maps seen
  - Subjects required to answer several types of questions about each map
  - Statistical analysis of % errors
- Think-aloud: “tell me what you are doing”

Statistical map reading tasks

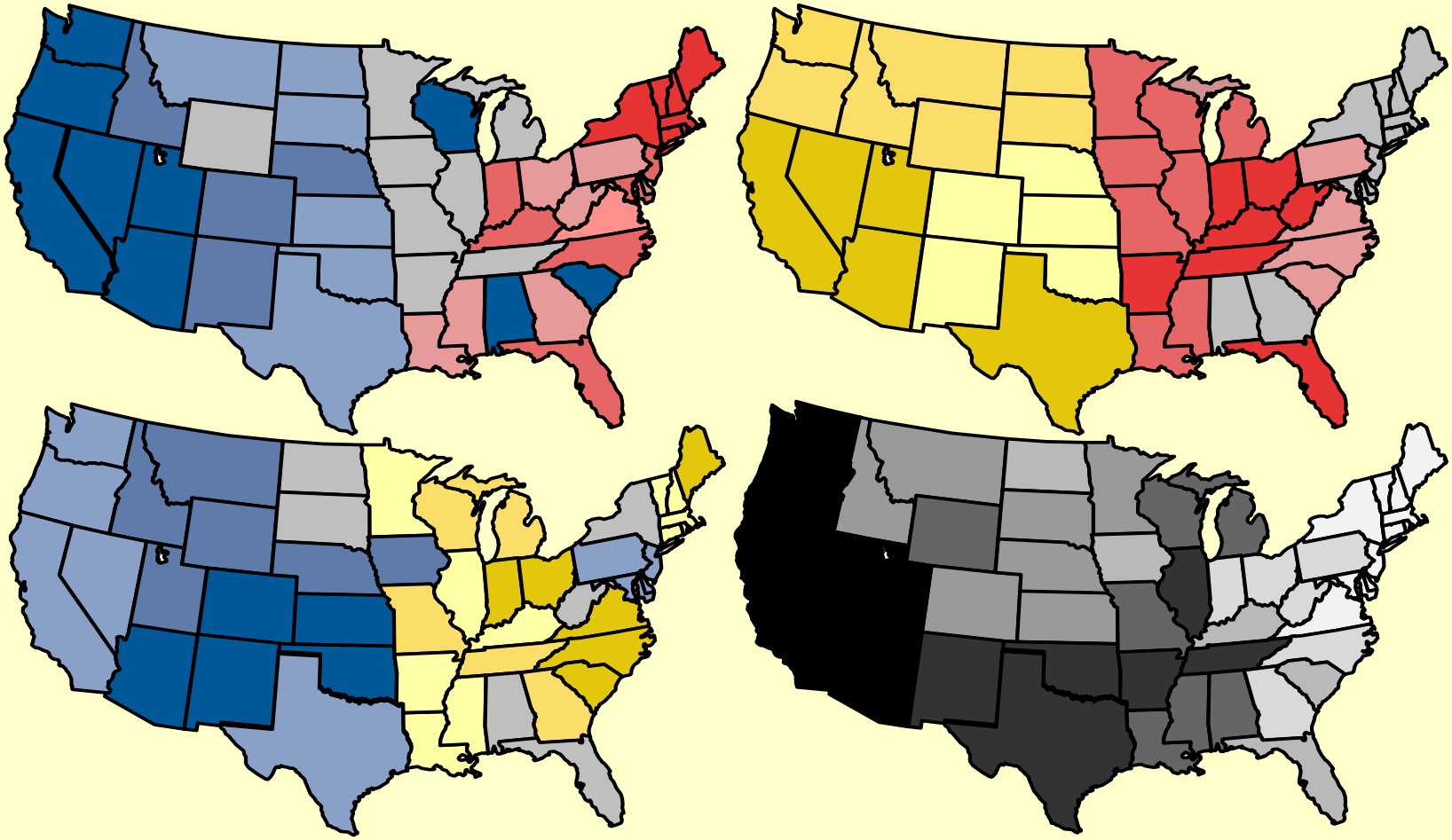
- Rate readout - what is approx. rate?
- Pattern recognition -clusters? outliers?
  -regional patterns?
- Pattern comparisons - compare maps

Target audience: Epidemiologists,
public health professionals

Results of early studies

- Choropleth (area-shaded) maps preferred & used most accurately by epidemiologists
- Legend: standard vertical fixed-box style
- Colors:
  - Very distinct colors best for rate readout (Hastie 1995)
  - Color gradient best for pattern recognition (Lewandowsky 1995)
  - Double-ended (diverging) scale combines gradients of 2 distinct hues; further tested for both types of questions
  - Color conventions (expectations) matter: darker or warmer color used for higher rates

What do you expect?
Do color conventions matter?

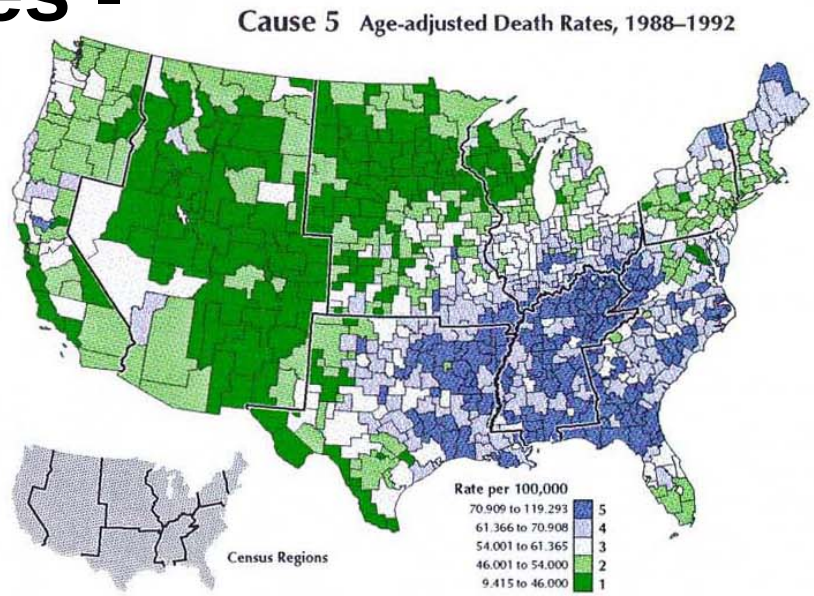


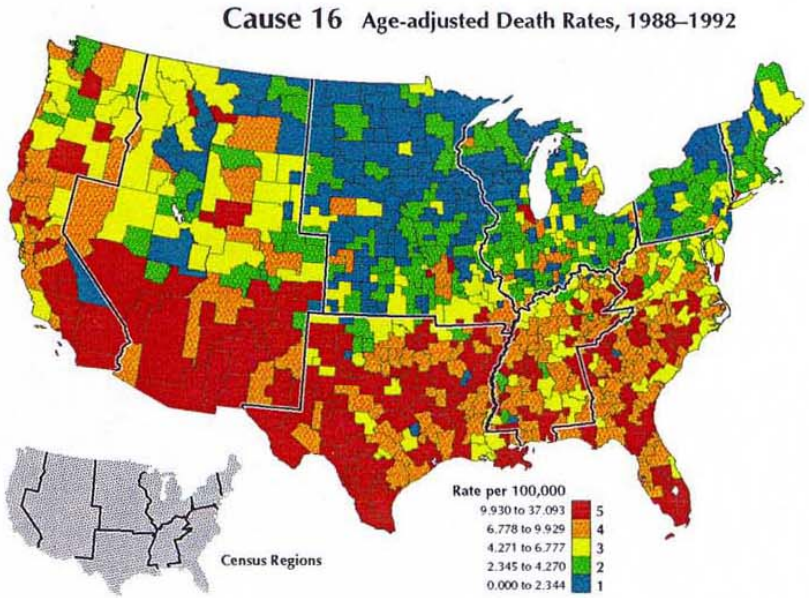
Source: Carswell M, in Pickle & Herrmann, eds., *Cognitive Aspects of Statistical Mapping*,1995

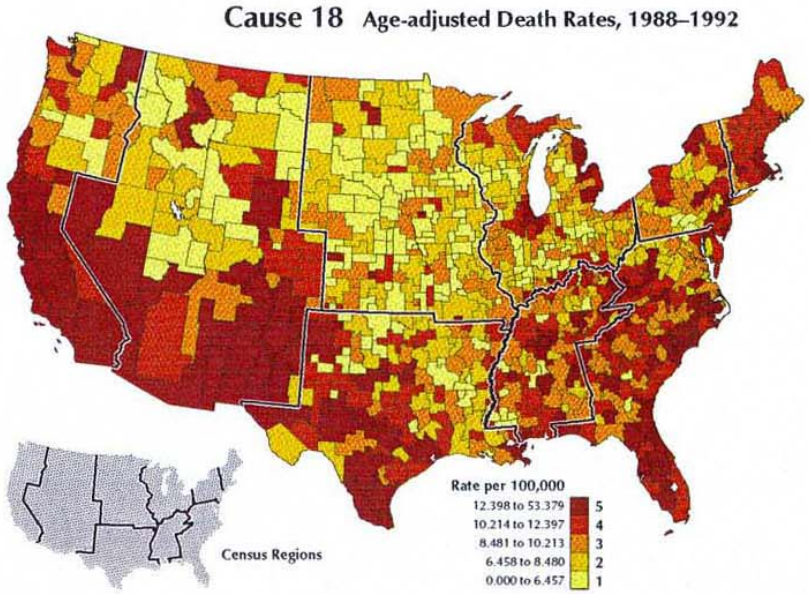
Evaluating color schemes - Sample test maps

Figure 3. Example Quarter-Scale Test Maps

Mapped variables 5, 16, and 18 are shown with the five-class Purple/Green diverging, Spectral, and Red-Yellow sequential schemes.

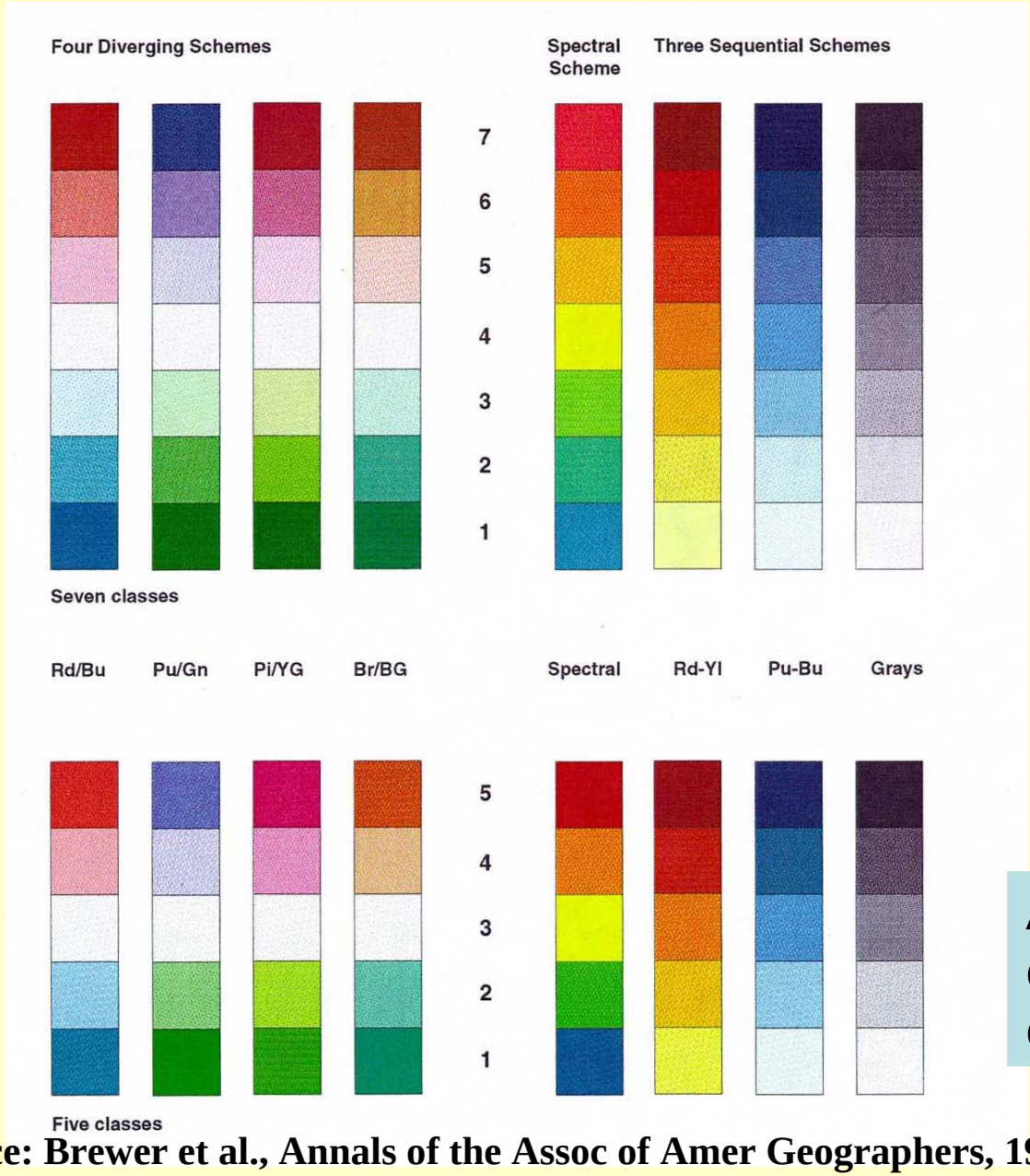






Source: Brewer et al., Annals of the Assoc of Amer Geographers, 1997.

Color schemes tested

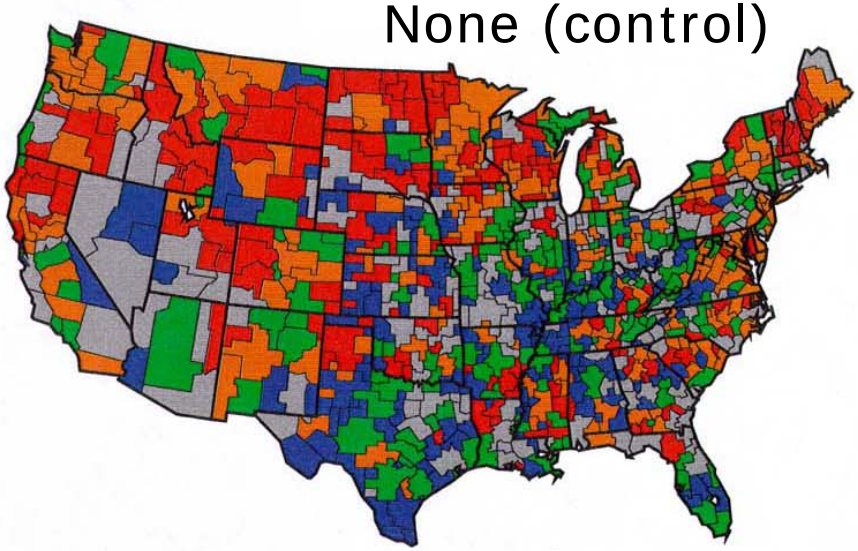


A new web tool for choosing colors: colorbrewer.org

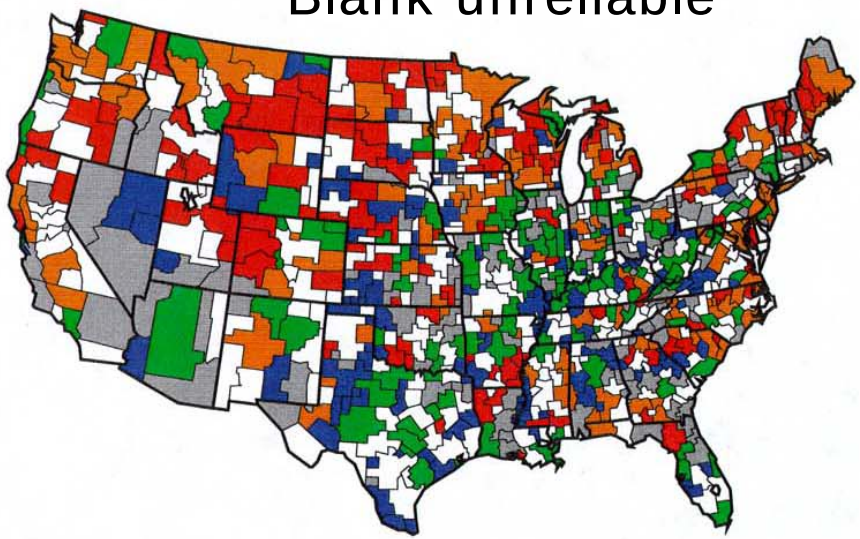
Source: Brewer et al., Annals of the Assoc of Amer Geographers, 1997.

Reliability Representation (study #1)

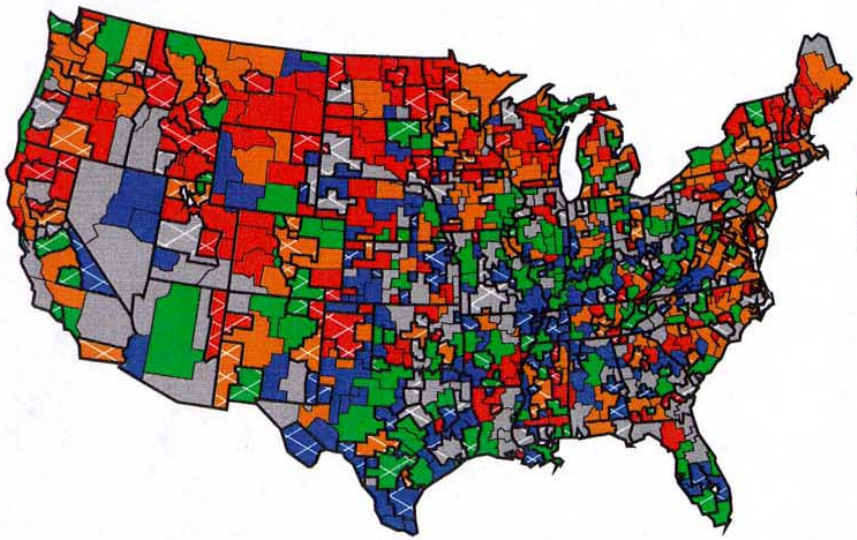
None (control)



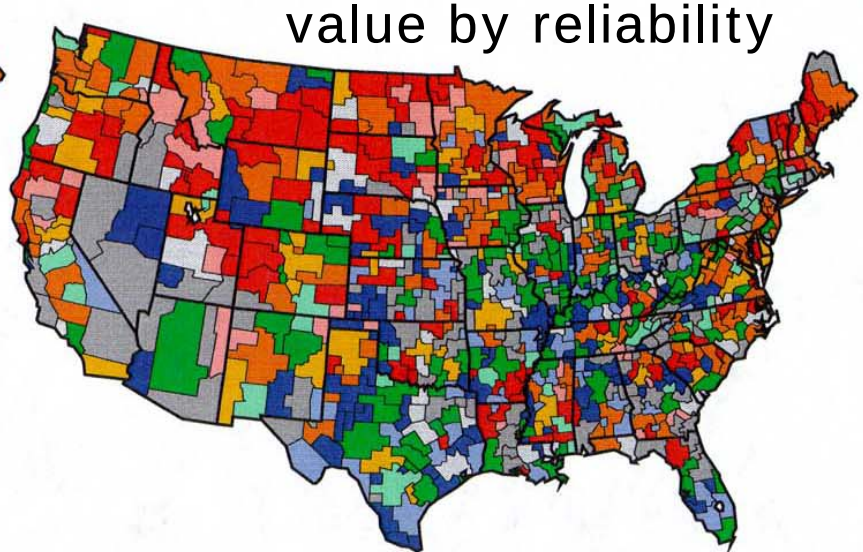
Blank unreliable



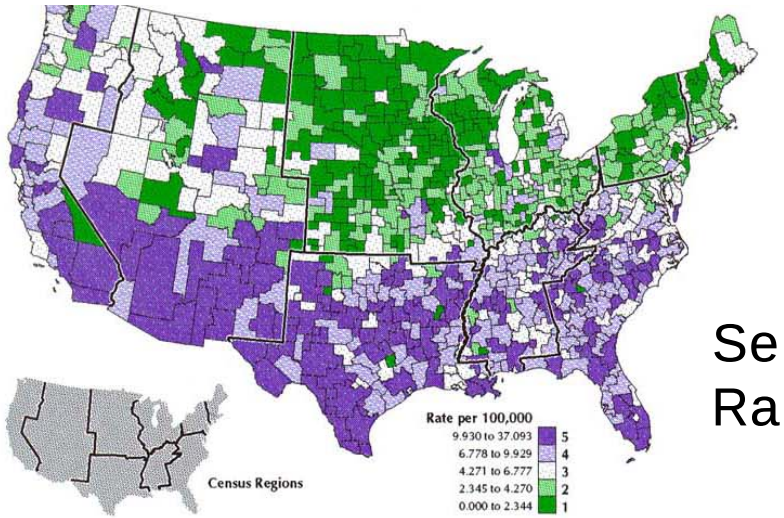
Hatch unreliable

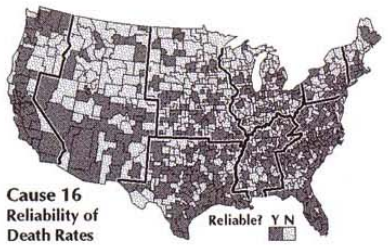


Bivariate color scheme – value by reliability



Reliability Representation (Study #2)



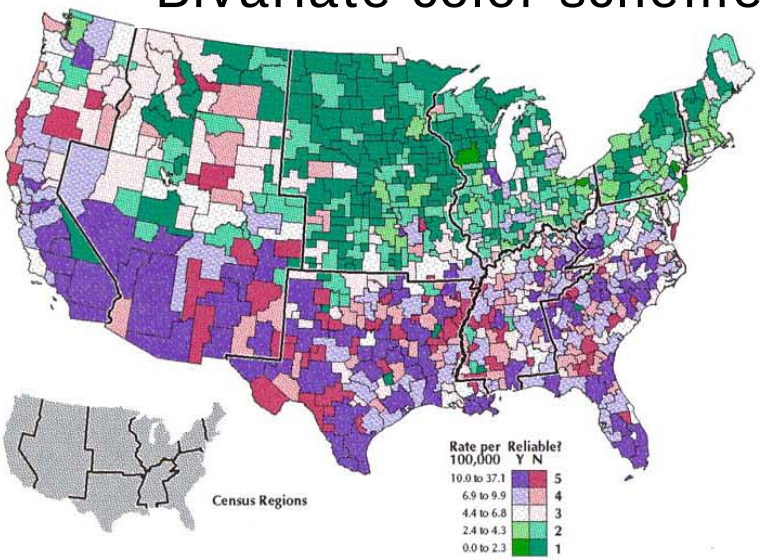


Sample Test Maps:

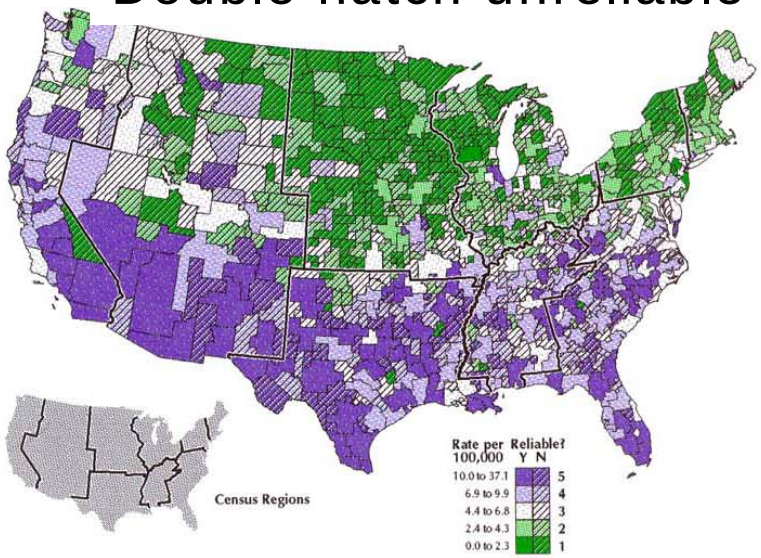
- cause: 16
- scale: quarter
- color scheme: purple-green
- reliability schemes:
  - map pairs
  - color change
  - texture overlay

Separate maps for
Rate and reliability

Bivariate color scheme



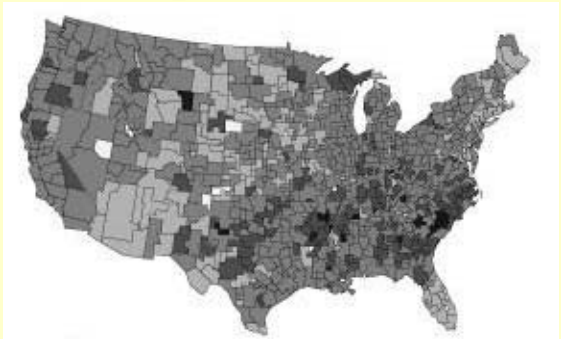
Double hatch unreliable



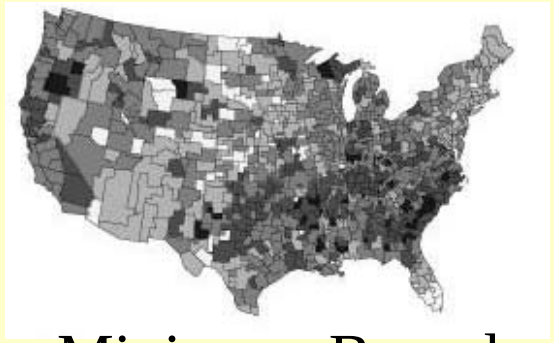
Source: MacEachren et al., Environment & Planning A, 1998.

Cutpoint Methods Tested

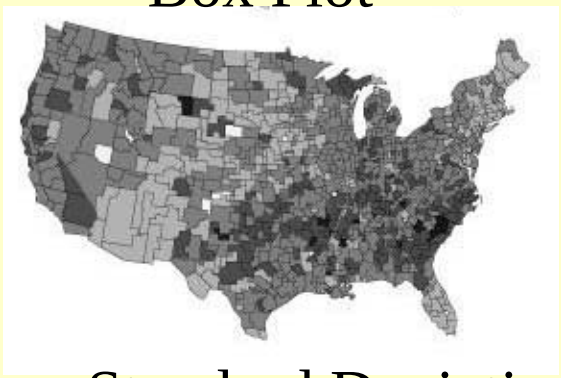
Equal Width



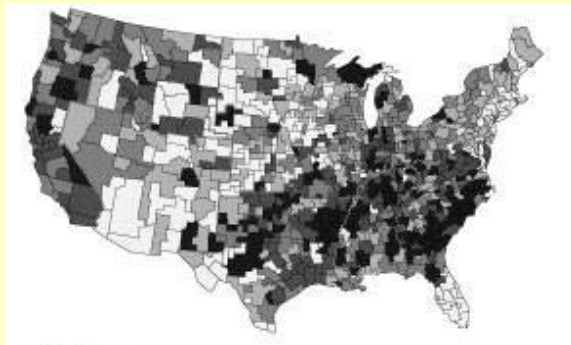
Natural Breaks (Jenks)



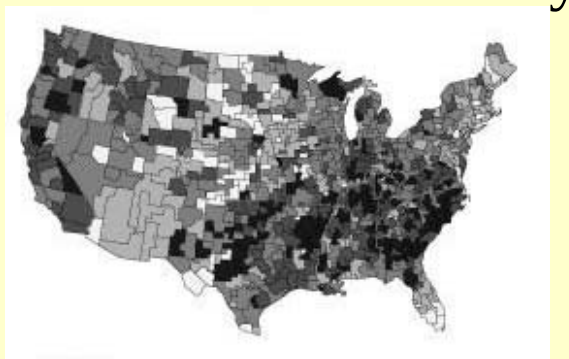
Box Plot



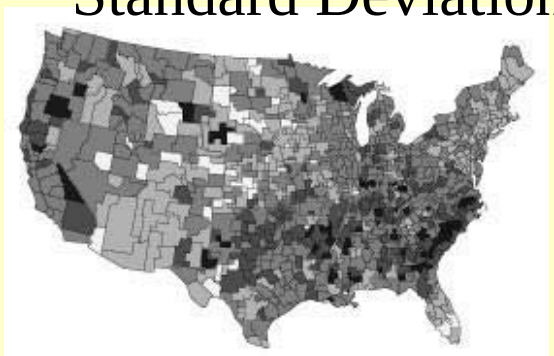
Quintile



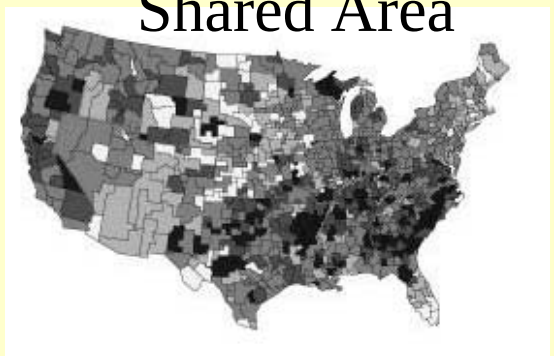
Minimum Boundary



Standard Deviation



Shared Area



Source: Brewer & Pickle, Annals of the AAG, 2002

Recommendations for rate map design

- Design for particular audience and purpose & TEST PROPOSED DESIGNS FOR THESE
- Quantile-categorized choropleth map works well
- Use standard legend design
- Colors should be chosen for visually impaired and consistent with conventions
- Identify unreliable rates, don’t blank out
- Accept that multiple maps are often needed
  - to address different questions,
  - to focus attention on different scales,
  - to compare modeled (smoothed) to observed...

Extensions of Map-based Research at NCI

- Extension of map research to computer-based maps, web-based data dissemination
- Development of new graphical tools for data exploration and communication
- Usability of interactive systems by the public
- Examples
  - **Visualization tools: Smoothing, Cluster identification**
  - Linking maps & graphs: Linked micromap plots, Exploratory Spatio-Temporal Analysis Tool (ESTAT)
  - Communication over the web
    - Cancer atlas
    - Long Island Breast Cancer GIS
    - State Cancer Profiles

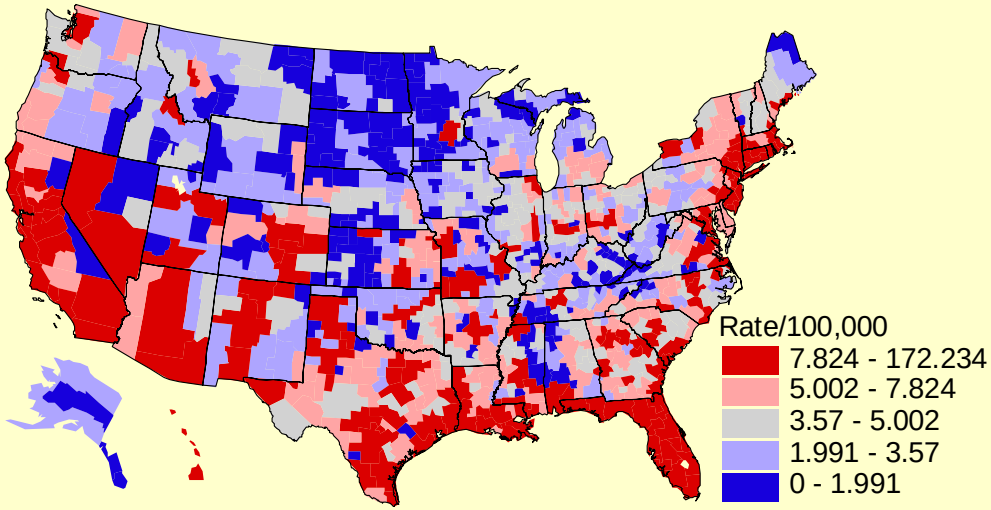
Map smoothing methods

- 2D Smoothing is a method of removing some variability in a quantitative map
- Maps of cancer rates for small areas can be difficult to interpret because of background “noise”
- Previous methods ignored population differences
- Methods now can incorporate weights
  - very stable rates are smoothed less
  - more unstable rates (due to small populations) are smoothed more

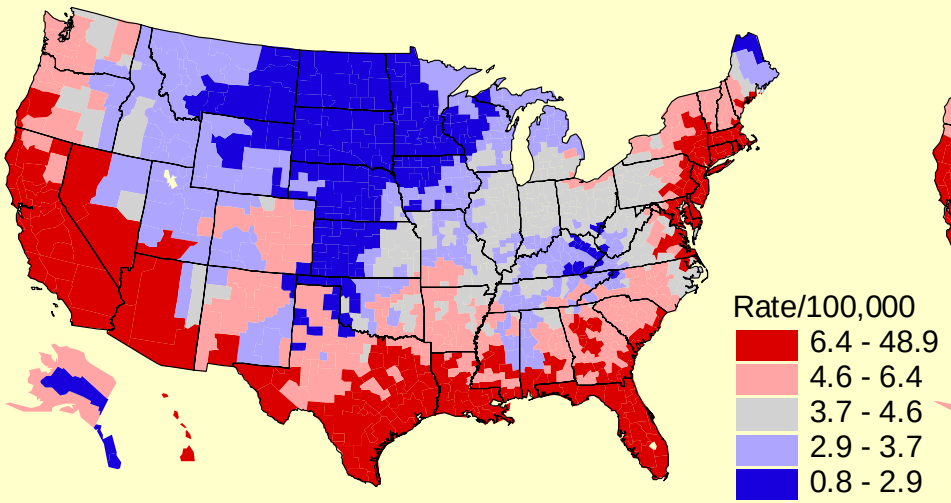
Source: Mungiole, Pickle, Simonson, *Statistics in Medicine*, 1999

HIV mortality rates, 1988-92

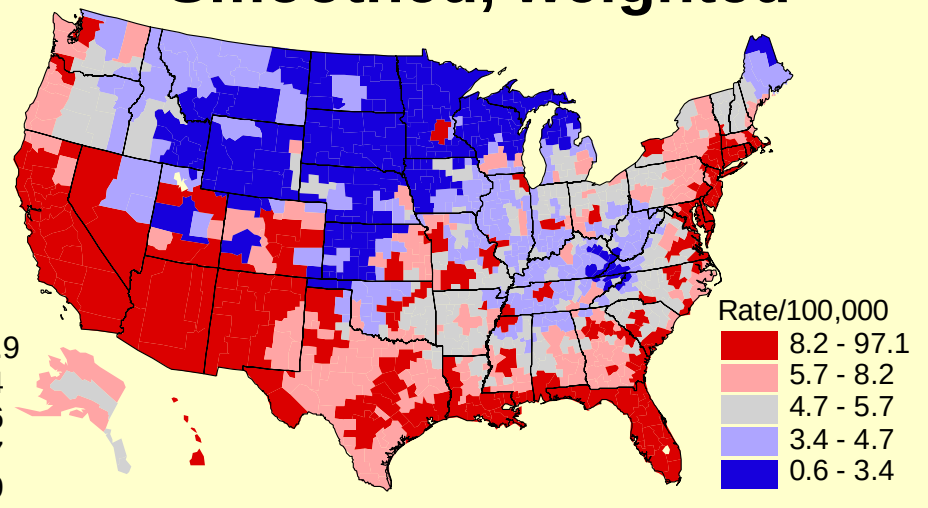
Original data:



Smoothed, unweighted



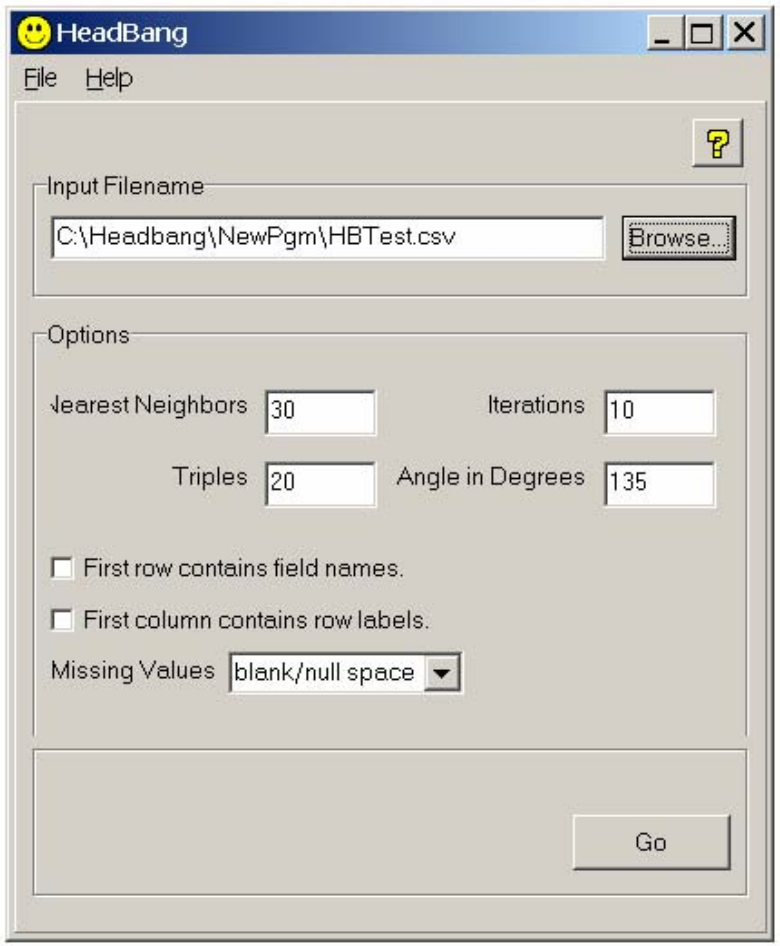
Smoothed, weighted



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

Headbang software available from http://srab.cancer.gov/headbang/

GUI Interface



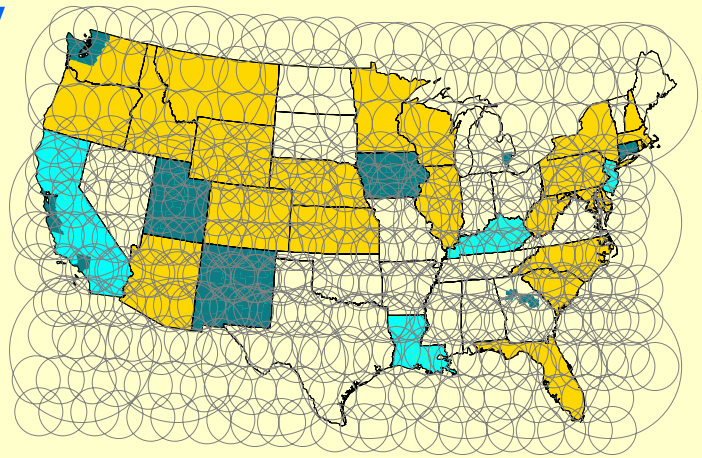
S+ call to C+ program

```
ResultNew<-dos(paste("headbang.exe",
nn,ntrip,niter,thetastar),rbind(X,Y,rate,wgt))
```

Developers: Katherine Hansen Simonson and IMS, Inc. staff

Cluster Identification

- SaTScan, a space and time scan statistic, was developed at NCI by Martin Kulldorff (see srab.cancer.gov and www.satscan.org)
- Tests null hypothesis that disease risk is the same all over the map
- Creates a set of circles (new version includes ellipses) centered on each geographic unit

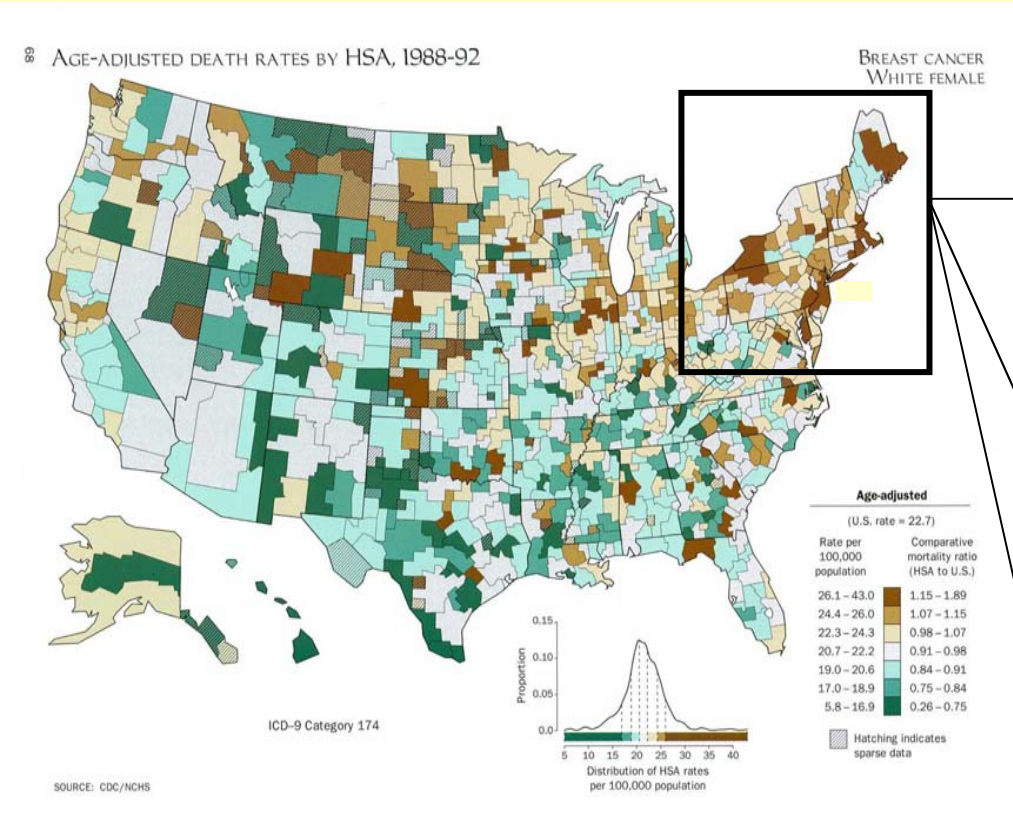


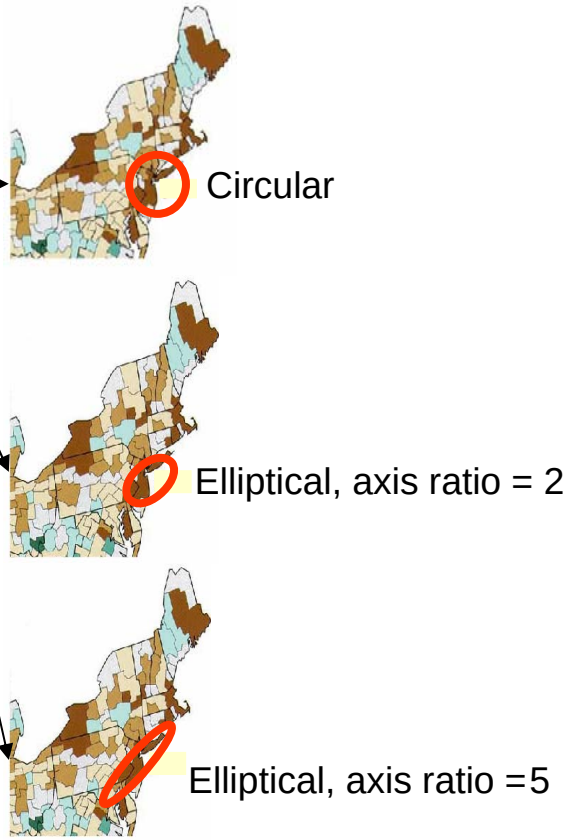
- Generates random replicas of the data under Ho, compares most likely clusters in real & random data sets to identify most likely cluster & its significance level

Example: Breast cancer clusters

Breast cancer mortality rates

Most likely cluster

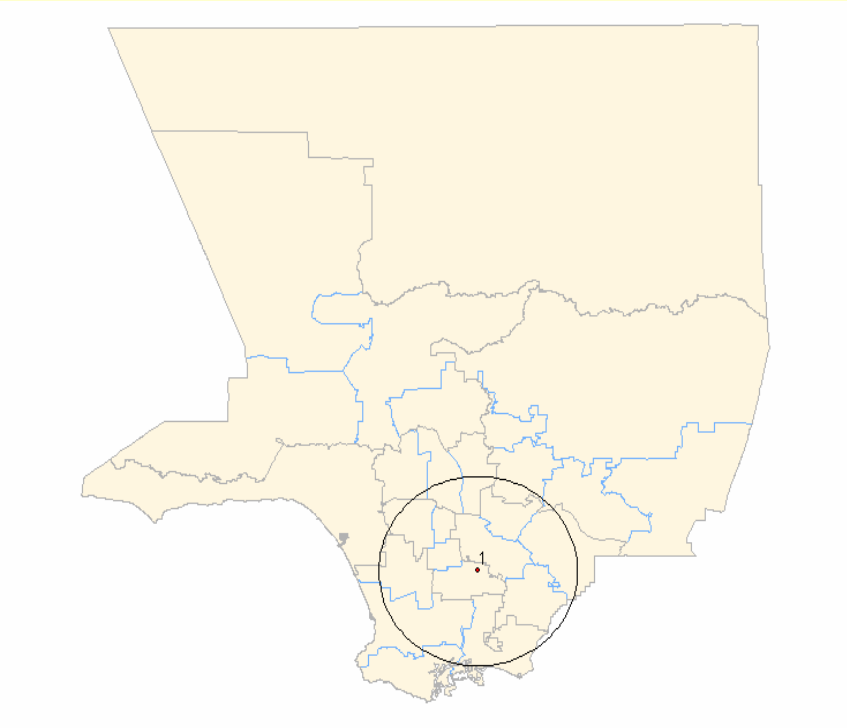


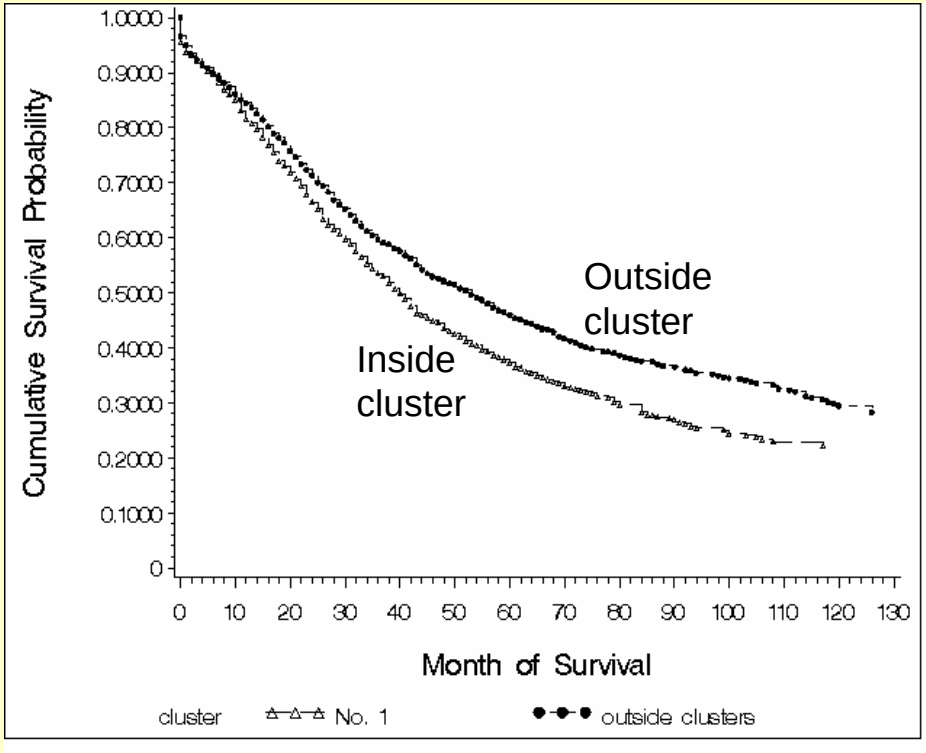


Pickle et al., Atlas of United States Mortality, NCHS, 1996

Spatial clustering of survival for stage III colorectal cancer in Los Angeles, among male cases diagnosed 1988-2002

Significantly short survival, p=0.01, radius=17km





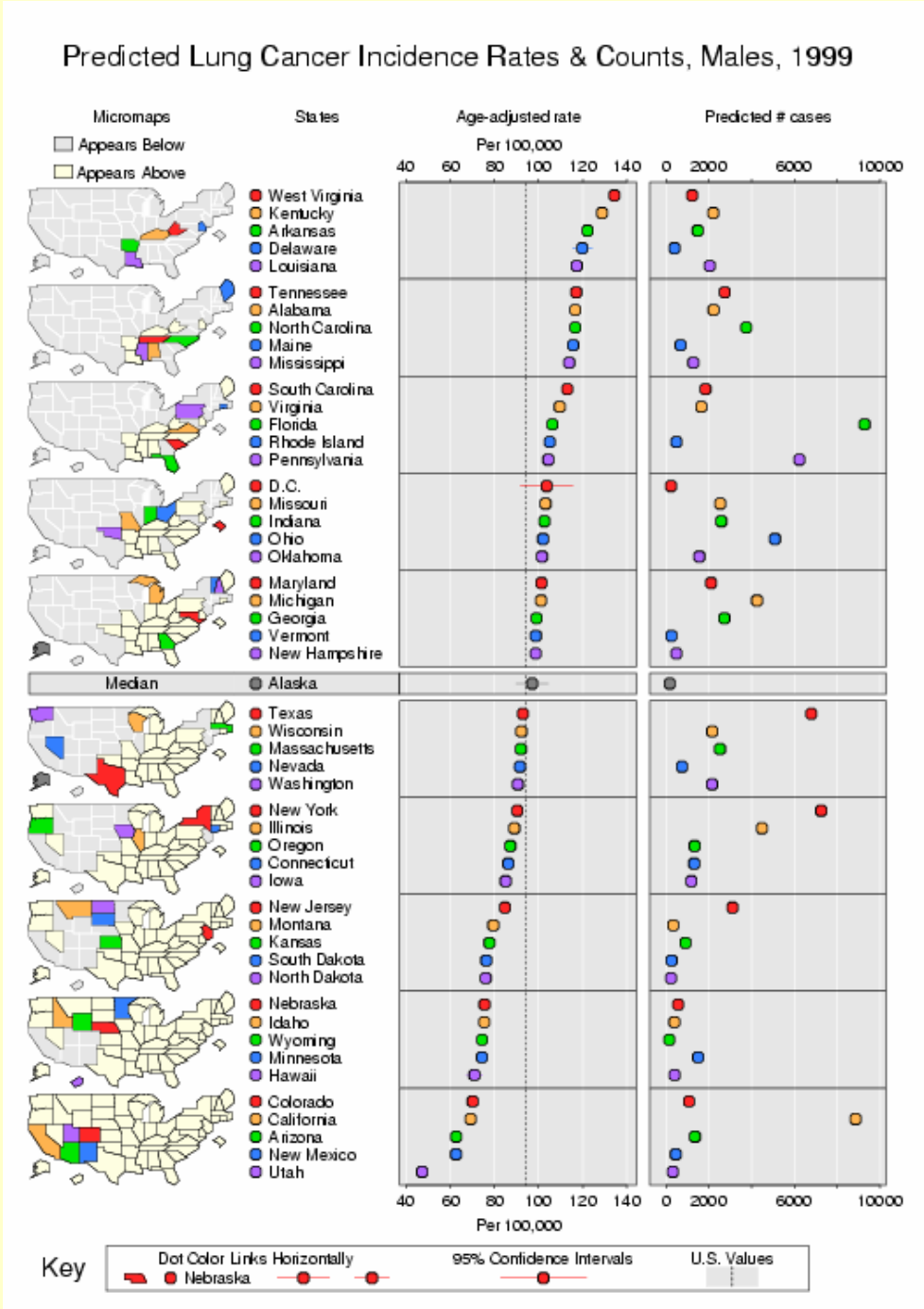
Source: Huang et al. (NCI), manuscript in preparation

Extensions of Map-based Research at NCI

- Extension of map research to computer-based maps, web-based data dissemination
- Development of new graphical tools for data exploration and communication
- Usability of interactive systems by the public
- Examples
  - Visualization tools: Smoothing, Cluster identification
  - **Linking maps & graphs**: Linked micromap plots, Exploratory Spatio-Temporal Analysis Tool (ESTAT)
  - Communication over the web
    - Cancer atlas
    - Long Island Breast Cancer GIS
    - State Cancer Profiles

Linked Micromap Plot

Linking geographic patterns with statistical detail



Source: Carr, Wallin & Carr, Stat in Med 2000

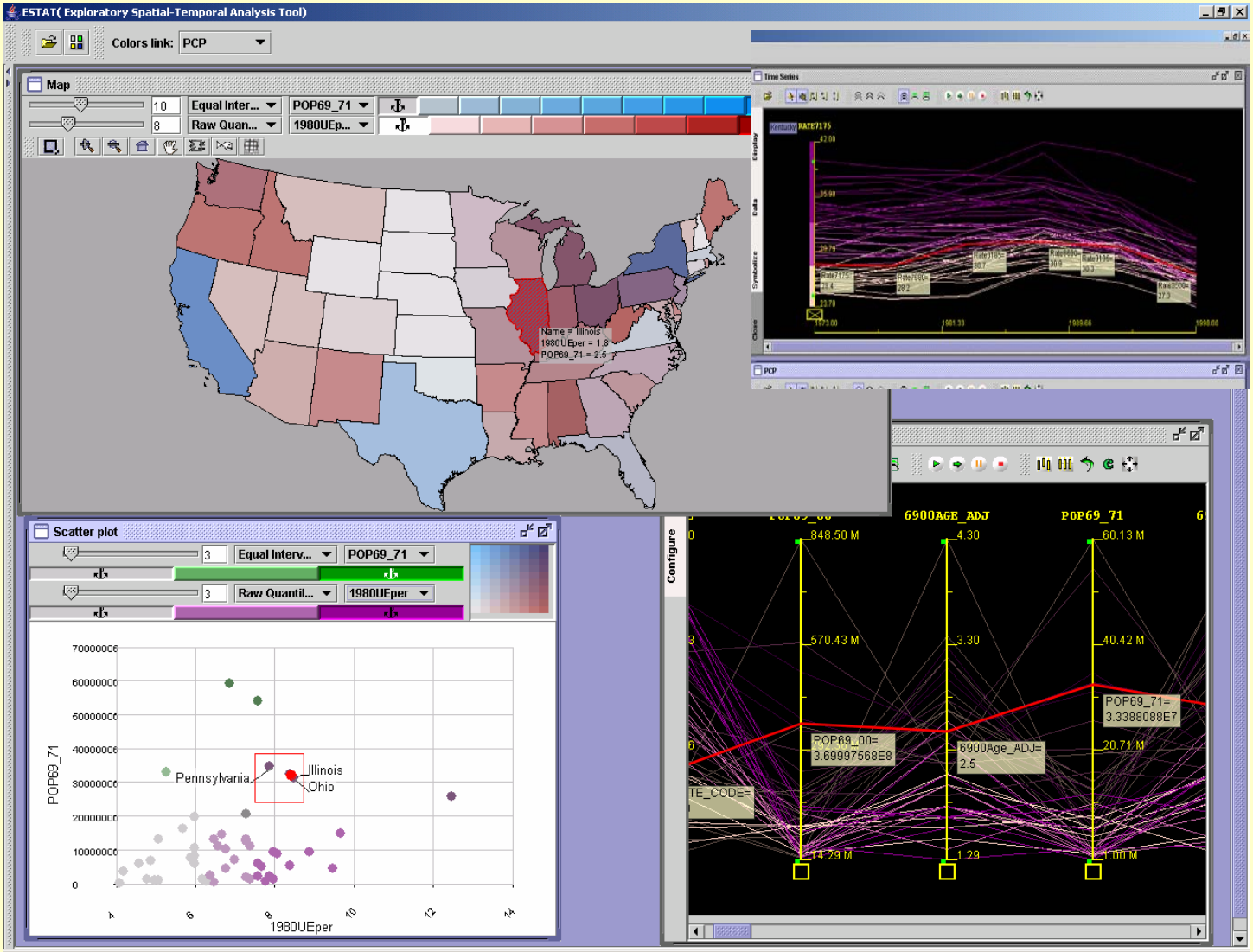
Exploratory Spatio-Temporal Analysis Tool (ESTAT)

Map

Rate Time Series Plot

Scatter plot

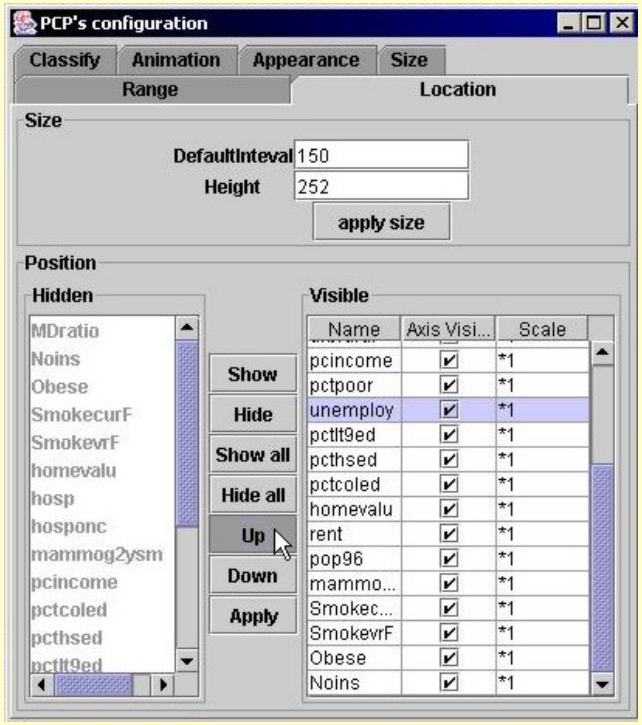
Covariate PCP plot



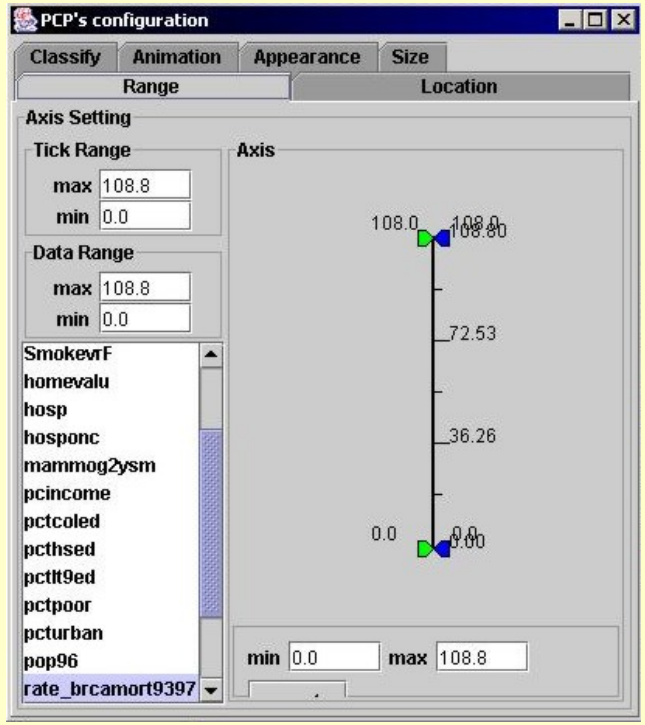
Developed by Alan MacEachren & GeoVista staff, Penn State University

Examples of ESTAT's extensive user controls

Select and order PCP variables



Restrict range of PCP axis



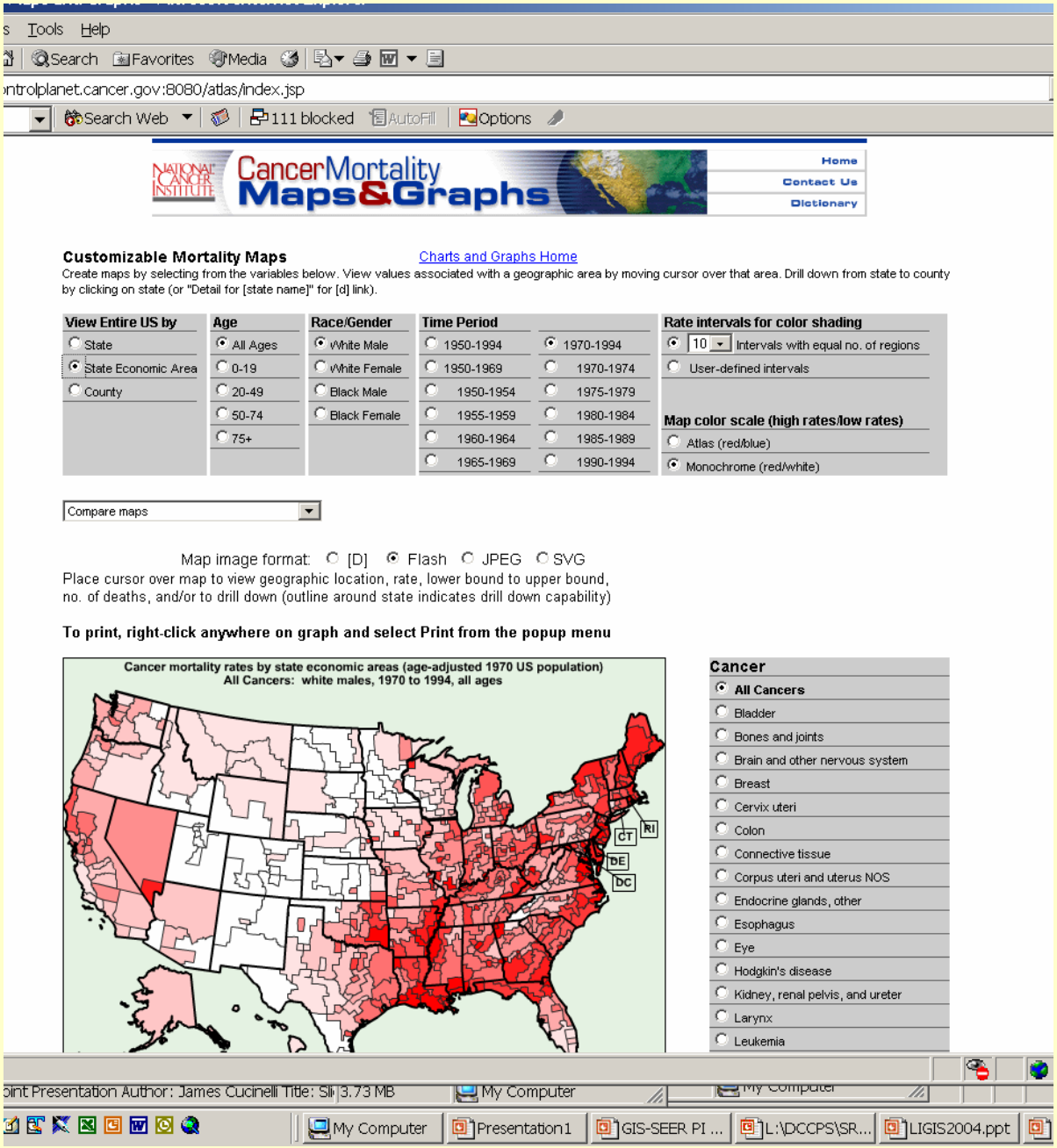
Also, dynamic linking & brushing, color specification, simple summary statistics, etc.

Extensions of Map-based Research at NCI

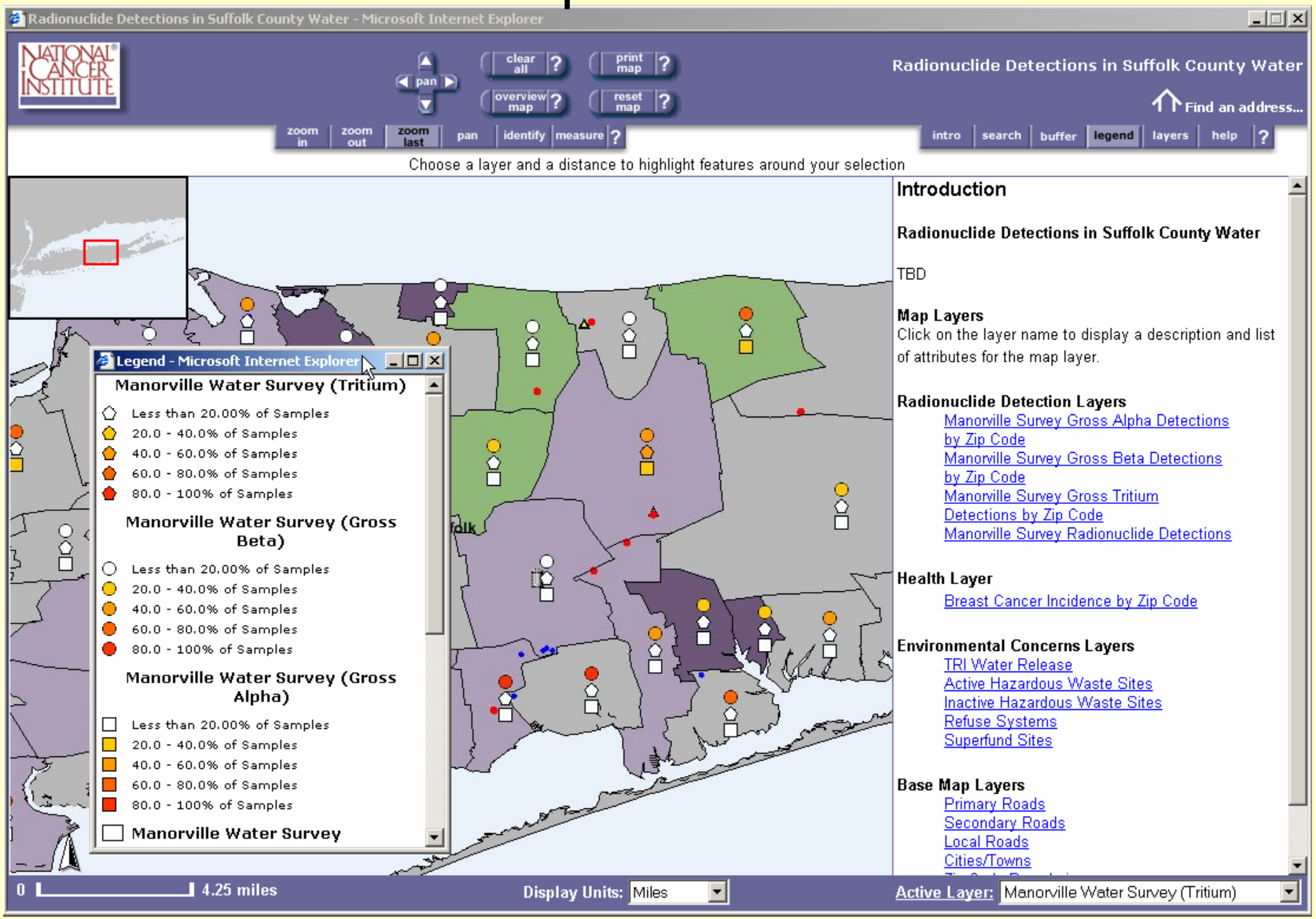
- Extension of map research to computer-based maps, web-based data dissemination
- Development of new graphical tools for data exploration and communication
- Usability of interactive systems by the public
- Examples
  - Visualization tools: Smoothing, Cluster identification
  - Linking maps & graphs: Linked micromap plots, Exploratory Spatio-Temporal Analysis Tool (ESTAT)
  - Communication over the web
    - **Cancer atlas**
    - **Long Island Breast Cancer GIS**
    - **State Cancer Profiles**

Web-based interactive cancer mortality maps

www.cancer.gov/atlasplus



Long Island Breast Cancer Study Project GIS: A more complex web-based tool



http://www.healthgis-li.com/

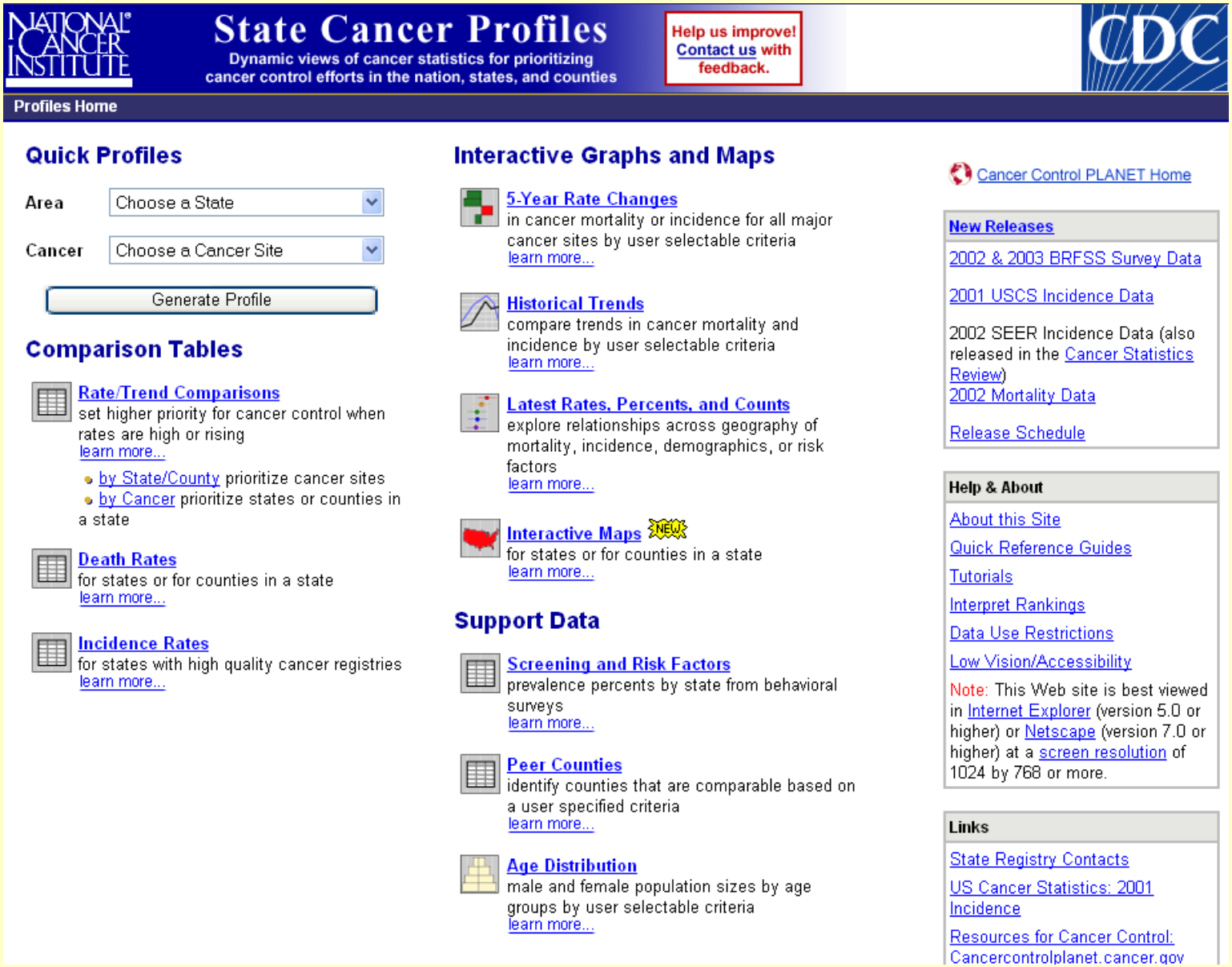
Web-based Communication of Cancer Statistics: State Cancer Profiles Web Site

- Goal: provide a system to characterize the cancer burden in a standardized manner to:
  - Motivate action
  - Integrate surveillance into cancer control planning
  - Characterize areas & demographic groups
  - Expose health disparities
- Target audiences:
  - Health planners
  - Policy makers
  - Cancer information providers

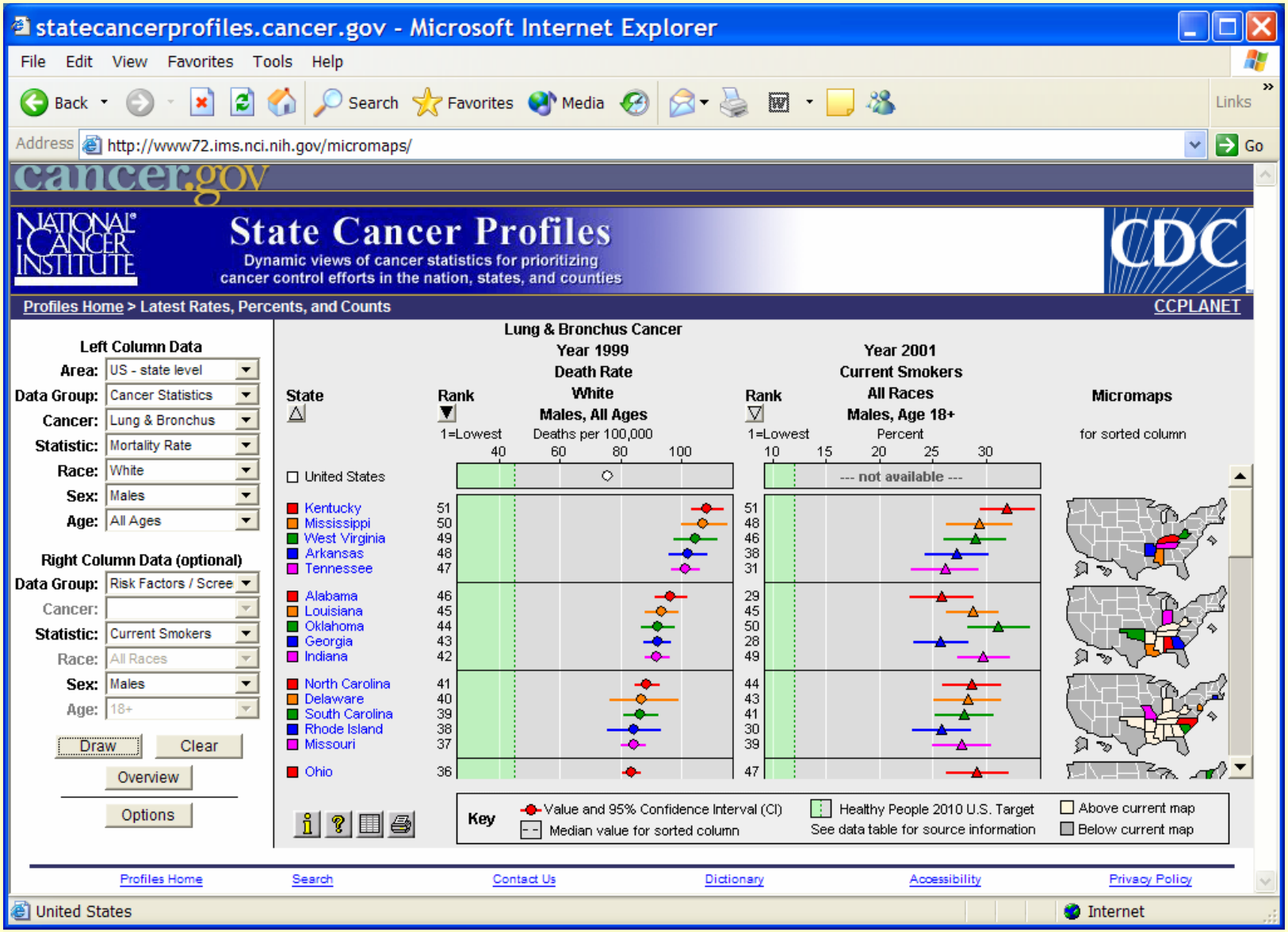
Extensive usability testing

- Tested at several professional meetings that members of target audience attended as well as in NCI Usability Lab
- Focus groups + hands-on testing conducted by a specialist in usability tests of web pages
- Tested on and/or approved by federal, state and local health department staff; cancer control professionals; policy makers
- Many iterations of prototypes
- Released to state health departments a week early so that they could verify their own data before general release

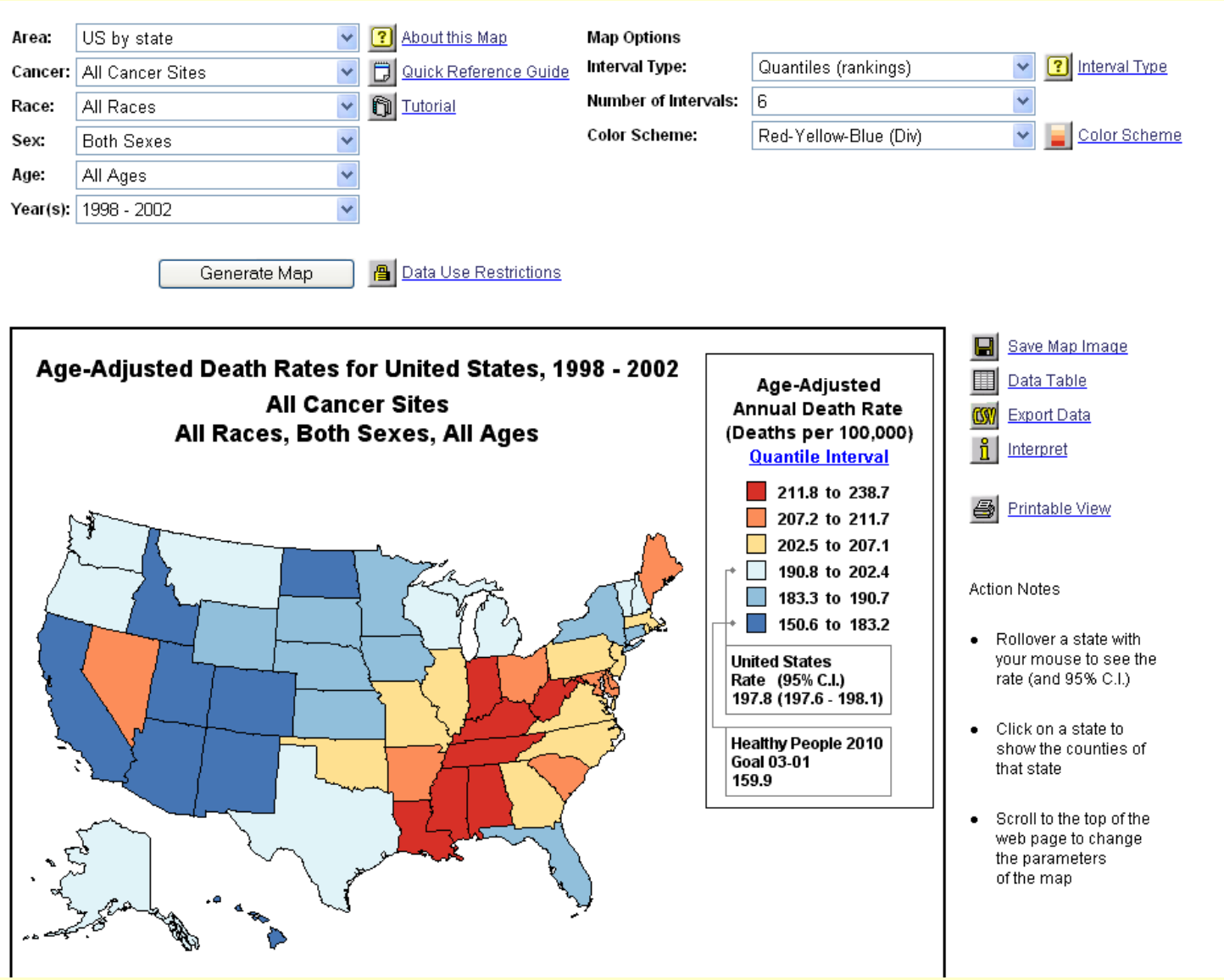
URL: statecancerprofiles.cancer.gov



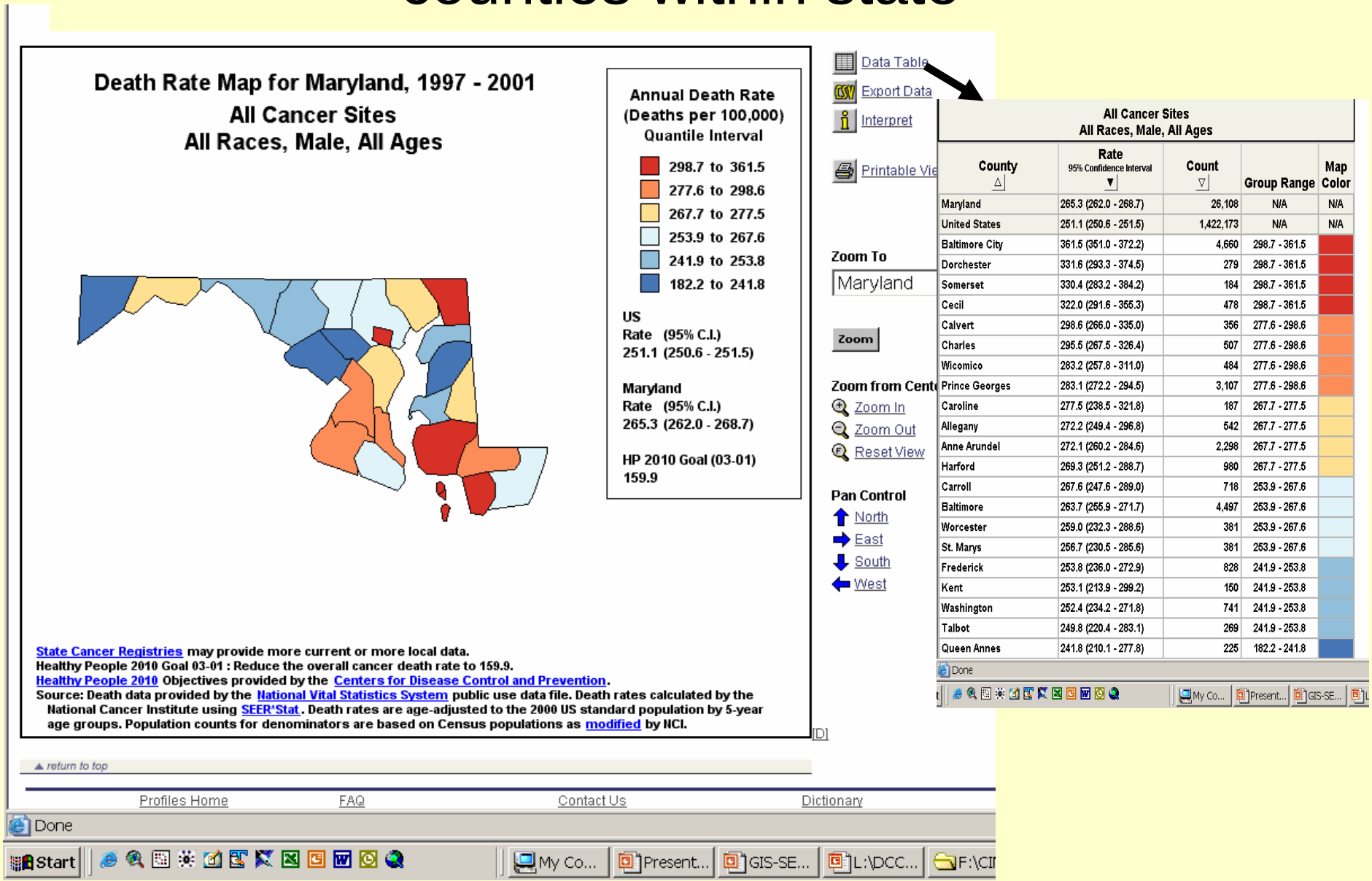
Includes linked micromap plots...



New interactive map feature



Basic mapping functions for states or counties within state



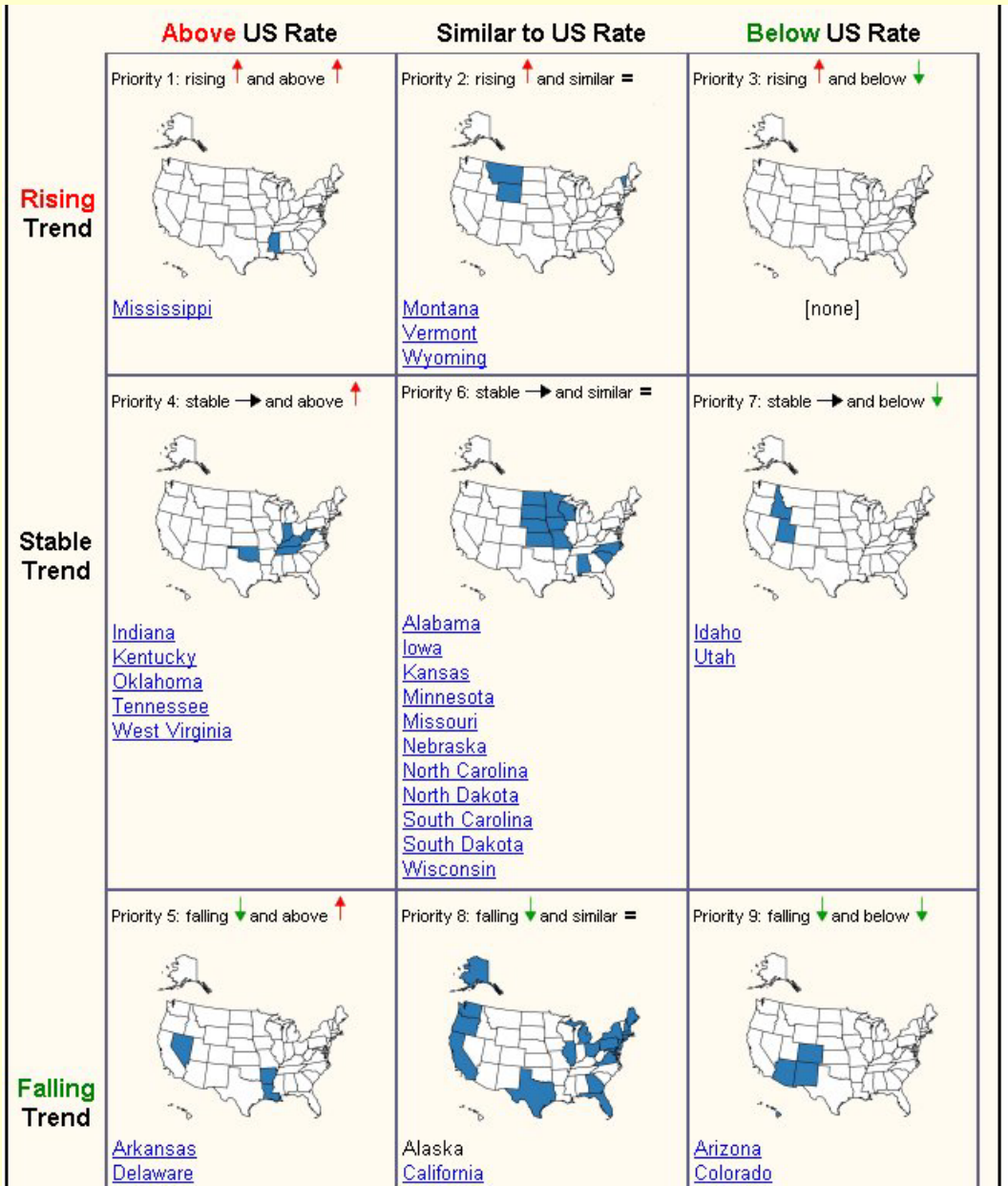
Where is there a problem?
Color & position classify rates & trends

Death Rate/Trend Comparison by Cancer, death years through 2001
US States versus US

Lung & Bronchus
All Races, Both Sexes

	Above US Rate	Similar to US Rate	Below US Rate
Rising Trend	Priority 1: rising ↑ and above ↑ Mississippi	Priority 2: rising ↑ and similar = Montana Vermont Wyoming	Priority 3: rising ↑ and below ↓ [none]
Stable Trend	Priority 4: stable → and above ↑ Indiana Kentucky Oklahoma Tennessee West Virginia	Priority 6: stable → and similar = Alabama Iowa Kansas Minnesota Missouri Nebraska North Carolina North Dakota South Carolina South Dakota Wisconsin	Priority 7: stable → and below ↓ Idaho Utah
Falling Trend	Priority 5: falling ↓ and above ↑ Arkansas Delaware Louisiana Nevada	Priority 8: falling ↓ and similar = Alaska California Connecticut District of Columbia Florida Georgia Illinois Maine Maryland Massachusetts Michigan New Hampshire	Priority 9: falling ↓ and below ↓ Arizona Colorado Hawaii New Mexico

Experimental Rate/Trend Table with Maps



Cf. Conditioned Choropleth Maps: Carr, Wallin, & Carr, *Statistics in Medicine*, 2000

Identification of Peer Counties in State Cancer Profiles system

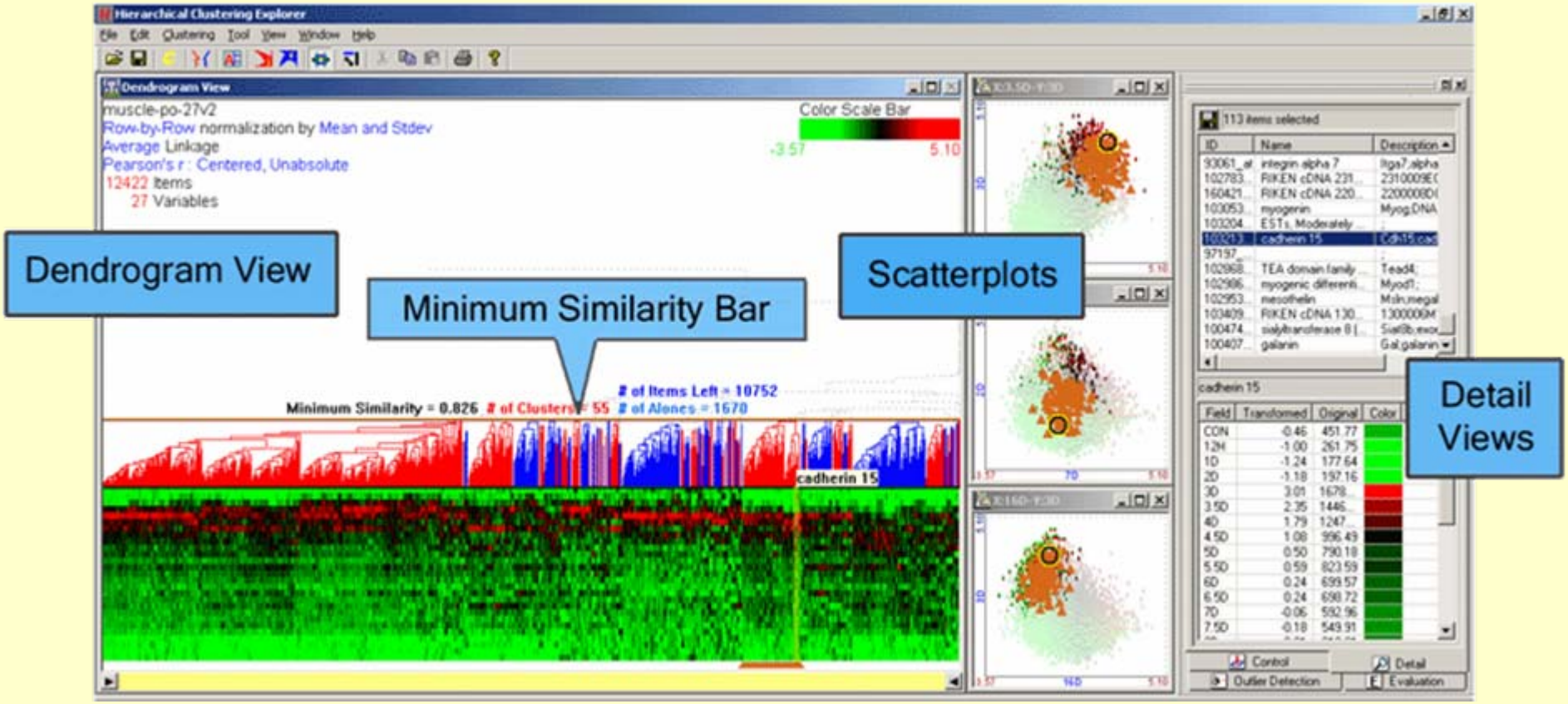
- A common question: One county in my state has unusually high cancer rates compared to the rest of the state, but we know that county is different from the others, e.g., in terms of income, education, etc.

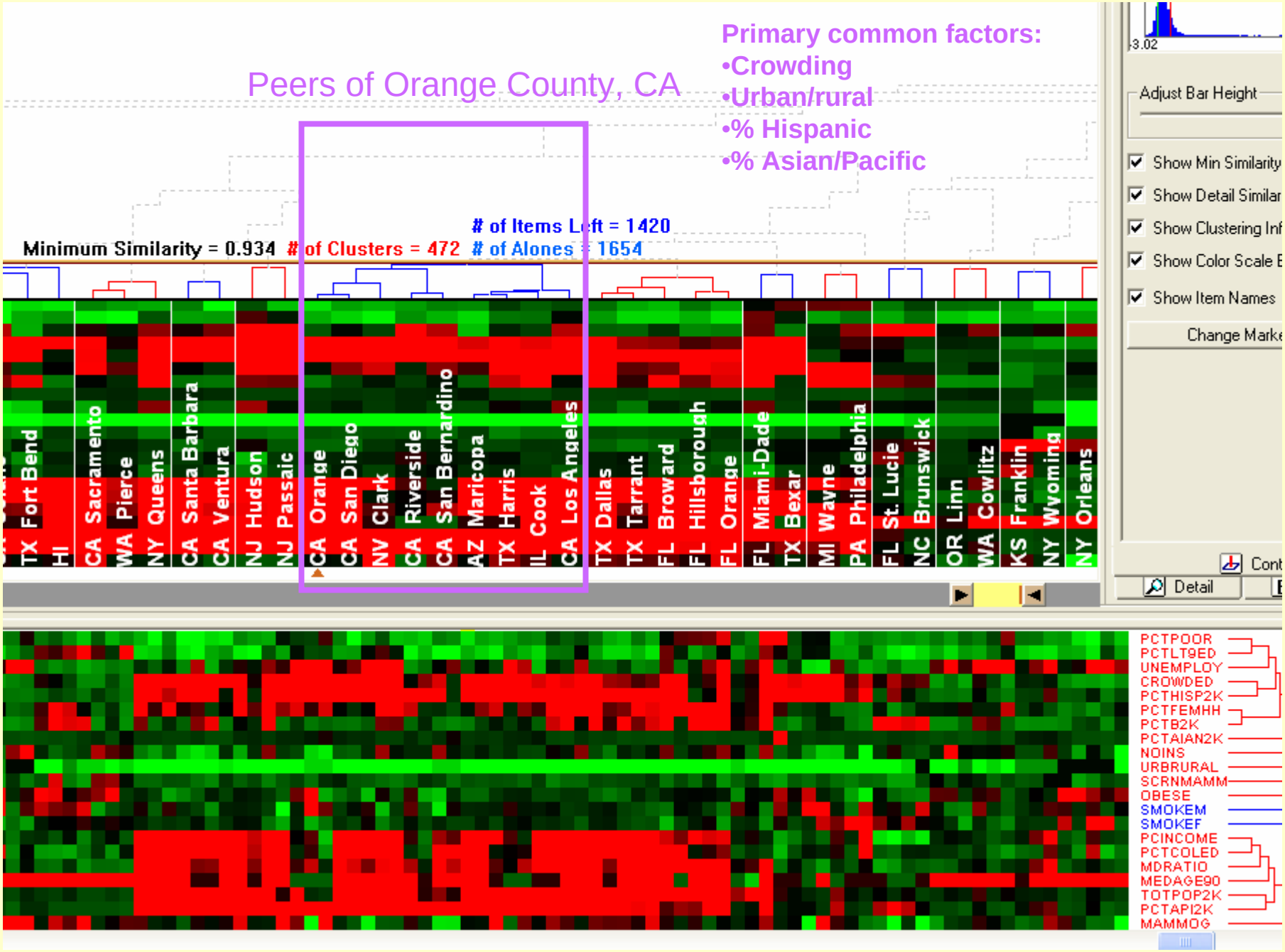
  How do the rates in this county compare with others in the US with a similar sociodemographic profile?

- How to identify “peer counties” for this comparison?

Hierarchical Clustering Explorer

for Interactive Exploration of Multidimensional Data




Peers of Orange County, CA
Primary common factors:
•Crowding
•Urban/rural
•% Hispanic
•% Asian/Pacific
# of Items Left = 1420
Minimum Similarity = 0.934 # of Clusters = 472 # of Alones = 1654
TX Fort Bend
HI
CA Sacramento
WA Pierce
NY Queens
CA Santa Barbara
CA Ventura
NJ Hudson
NJ Passaic
CA Orange
CA San Diego
NV Clark
CA Riverside
CA San Bernardino
AZ Maricopa
TX Harris
IL Cook
CA Los Angeles
TX Dallas
TX Tarrant
FL Broward
FL Hillsborough
FL Orange
FL Miami-Dade
TX Bexar
MI Wayne
PA Philadelphia
FL St. Lucie
NC Brunswick
OR Linn
WA Cowlitz
KS Franklin
NY Wyoming
NY Orleans
3.02
Adjust Bar Height
Show Min Similarity
Show Detail Similar
Show Clustering Inf
Show Color Scale B
Show Item Names
Change Marke
Cont
Detail
PCTPOOR
PCTLT9ED
UNEMPLOY
CROWDED
PCTHISP2K
PCTFEMHH
PCTB2K
PCTAIAN2K
NOINS
URBRURAL
SCRNMAMM
OBESE
SMOKEM
SMOKEF
PCINCOME
PCTCOLED
MDRATIO
MEDAGE90
TOTPOP2K
PCTAPI2K
MAMMOG

Collaborators

- **NCI was a partner in an NSF Digital Government Initiative grant to develop better visualization tools (web site: diggov.org)**

  Dan Carr, George Mason University

  Alan MacEachren, Penn State University

  David Scott, Rice University
- **NCI geographic information systems grant & contracts to develop ESTAT**

  Alan MacEachren, Penn State University
- **NCI sabbatical**

  Dan Carr, George Mason University
- **Web sites for more info:**
  - **gis.cancer.gov (for a poster on NCI research in GIS)**
  - **srab.cancer.gov (for headbang, SaTScan, etc)**
  - **statecancerprofiles.cancer.gov**
- **Email: PICKLEL@MAIL.NIH.GOV**