Spatial Analysis of Health Risk

Gerard Rushton
Department of Geography and College of Public Health
The University of Iowa

Specialist Meeting on Spatial Analysis of Health Risk Perception
Center for Spatially Integrated Social Science
University of California, Santa Barbara
October 10, 2003

Common Approaches

- A variety of methods for searching for statistically significant spatial clusters after adjusting rates for known risk factors.
- Data are generally “area-based” but sometimes are “point-based.”
- Methods are generally different for the two types of data.

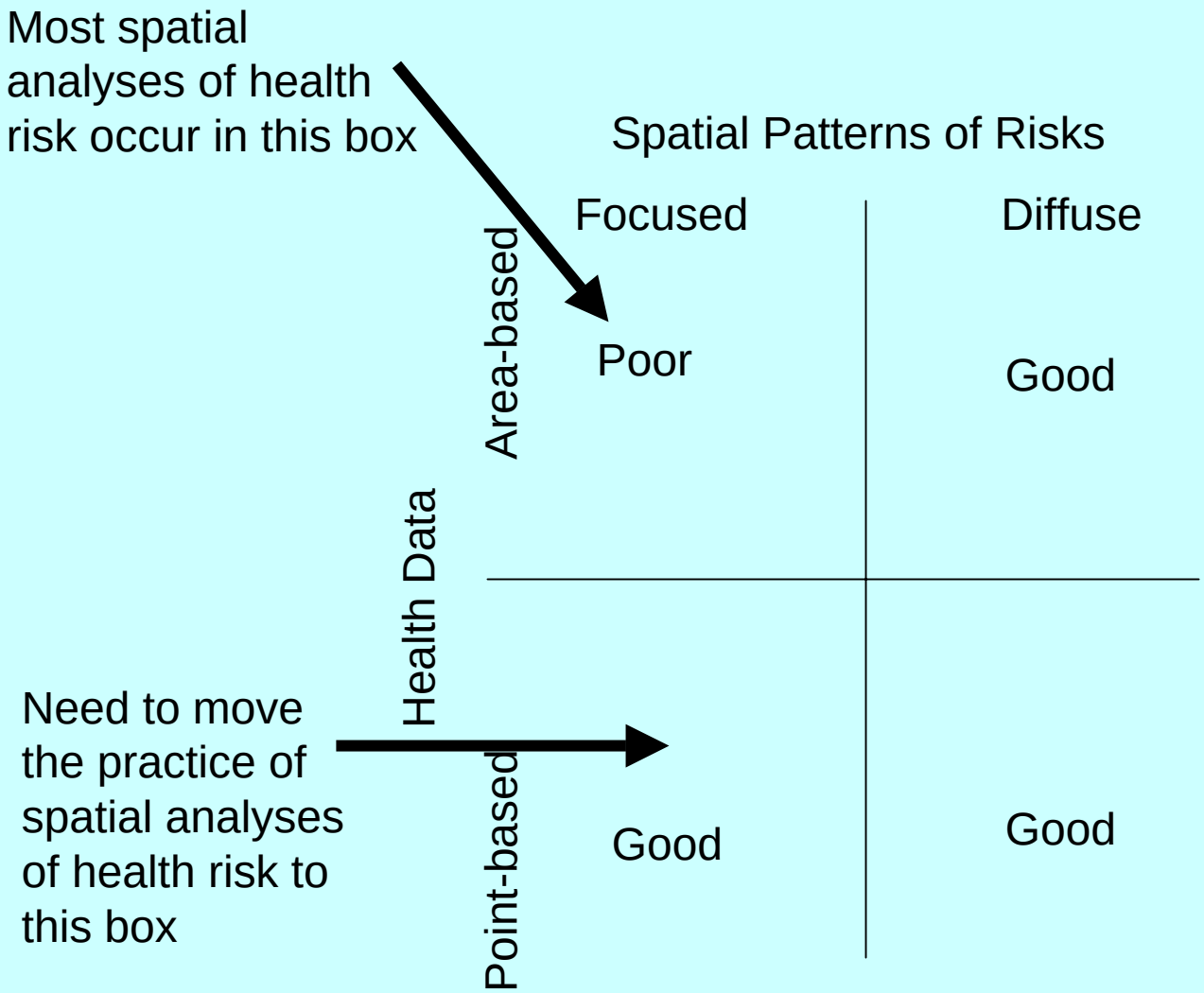
Common problems in these approaches: Measurement Errors

- exposures of individuals or groups
- health outcomes (population based registries?)
- locations of people or exposures—geocoding errors
- spatial relationships between people and exposures
- co-variates
- in protecting the privacy of individuals and institutions

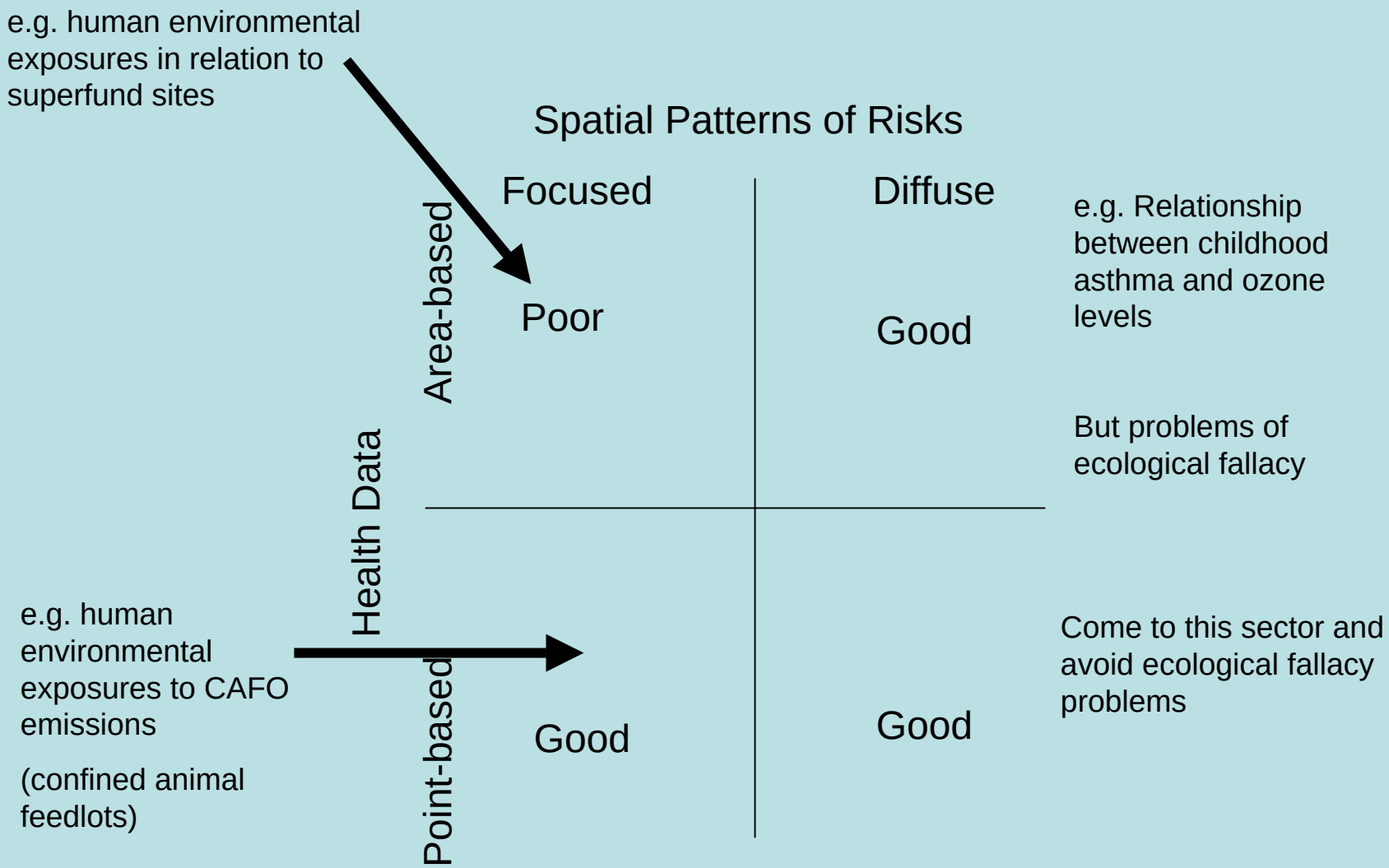
Common problems: accounting for co-variates that condition effects of health risks

- adjusting for age and sex effects
- Adjusting for social or economic effects
- Adjusting for differential screening effects

Ability to Measure Health Risk by Spatial Analysis

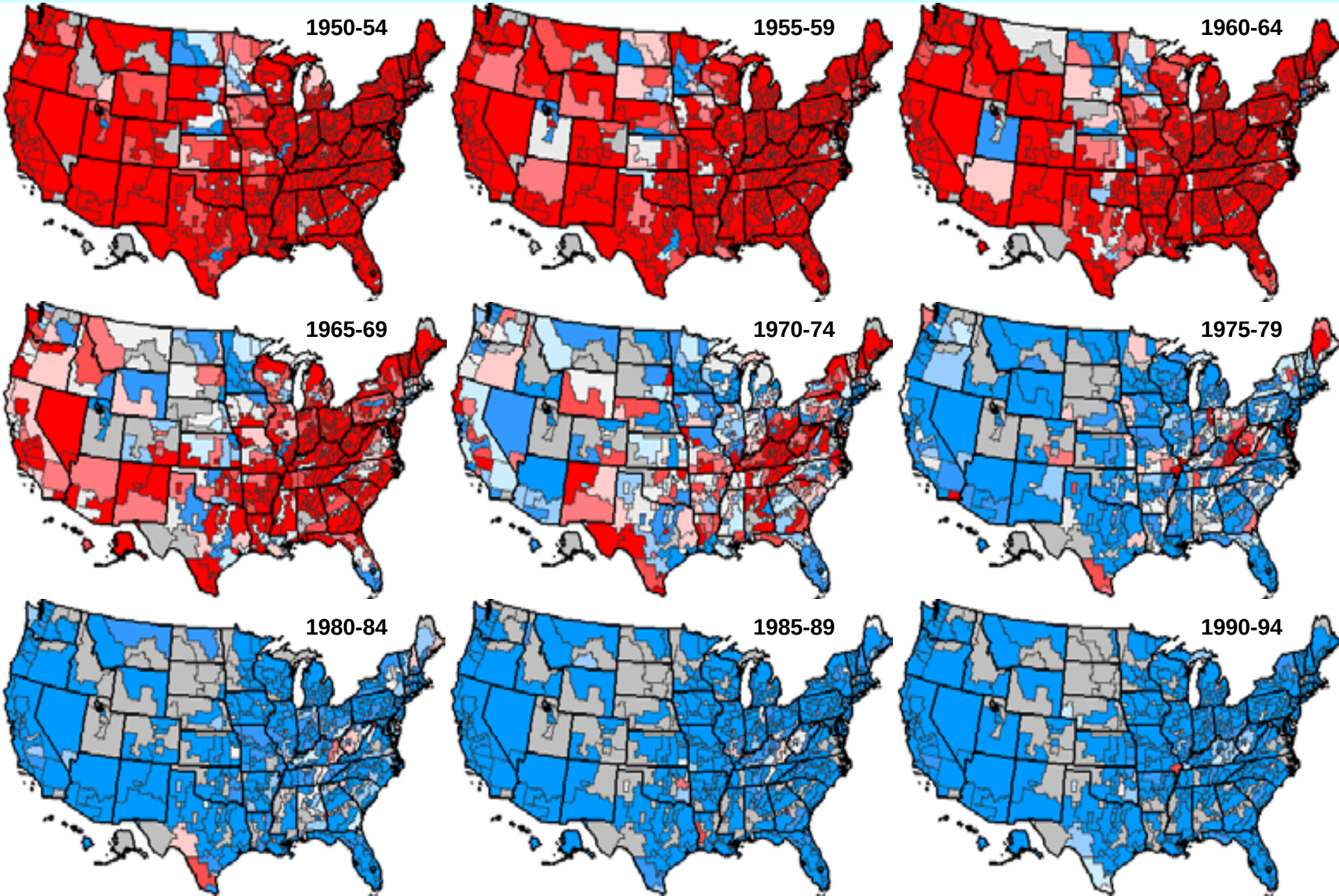


Ability to Measure Health Risk by Spatial Analysis

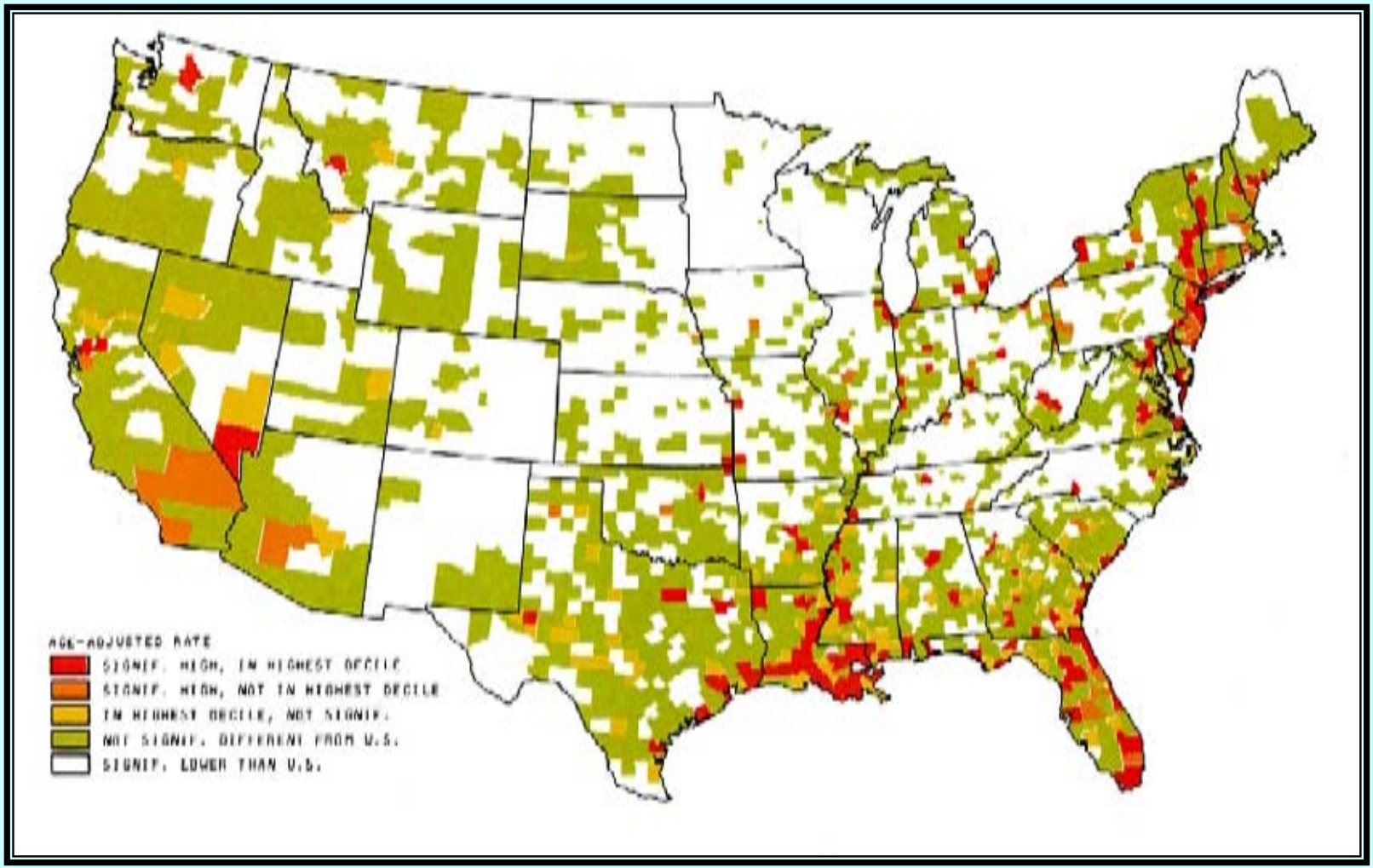


Cervix uteri cancer, white females, 1950-94, state economic area
Time trends

Source: www.nci.nih.gov/atlasplus

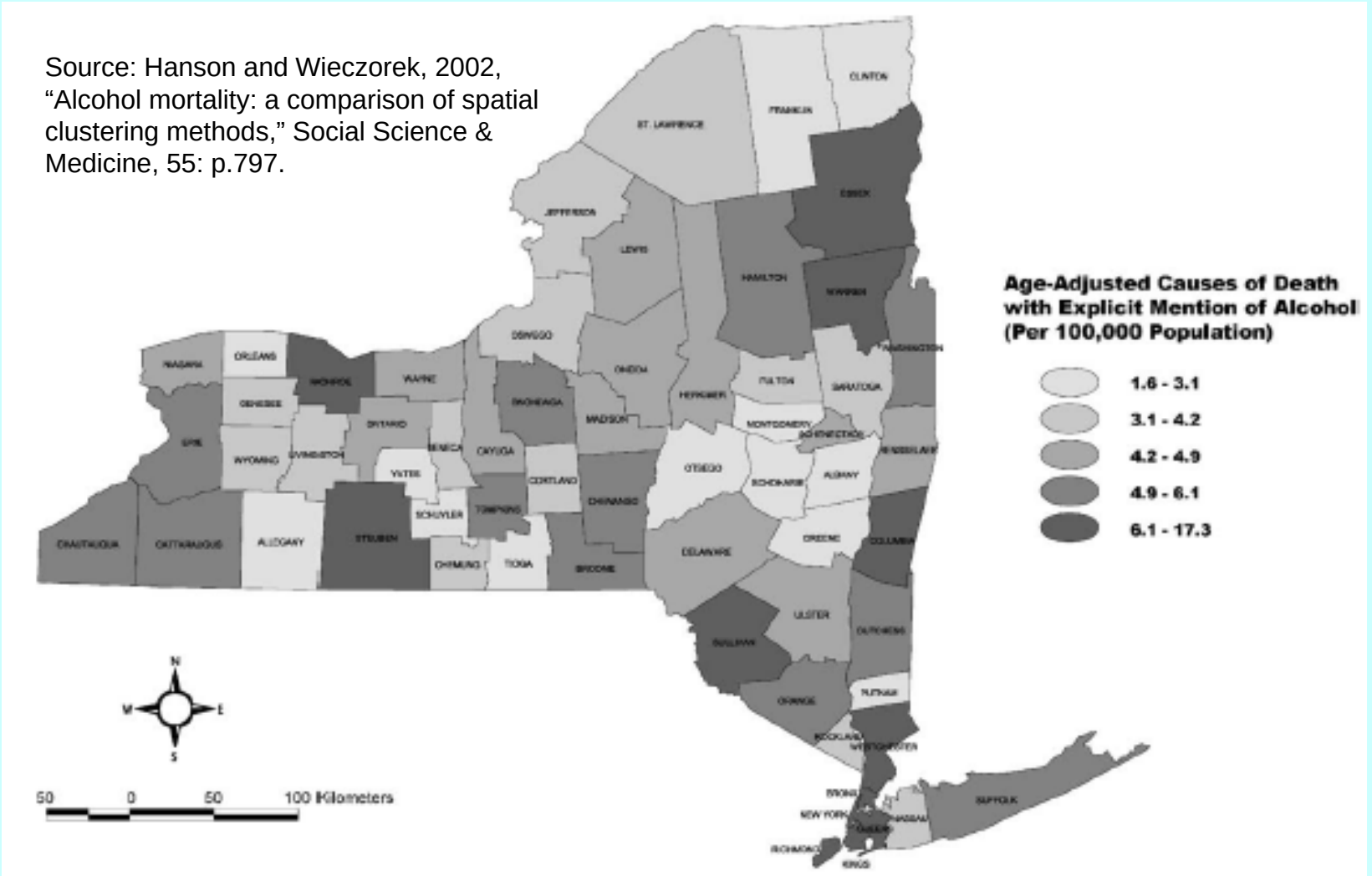


Lung Cancer Mortality Rates, 1950-69, White Males

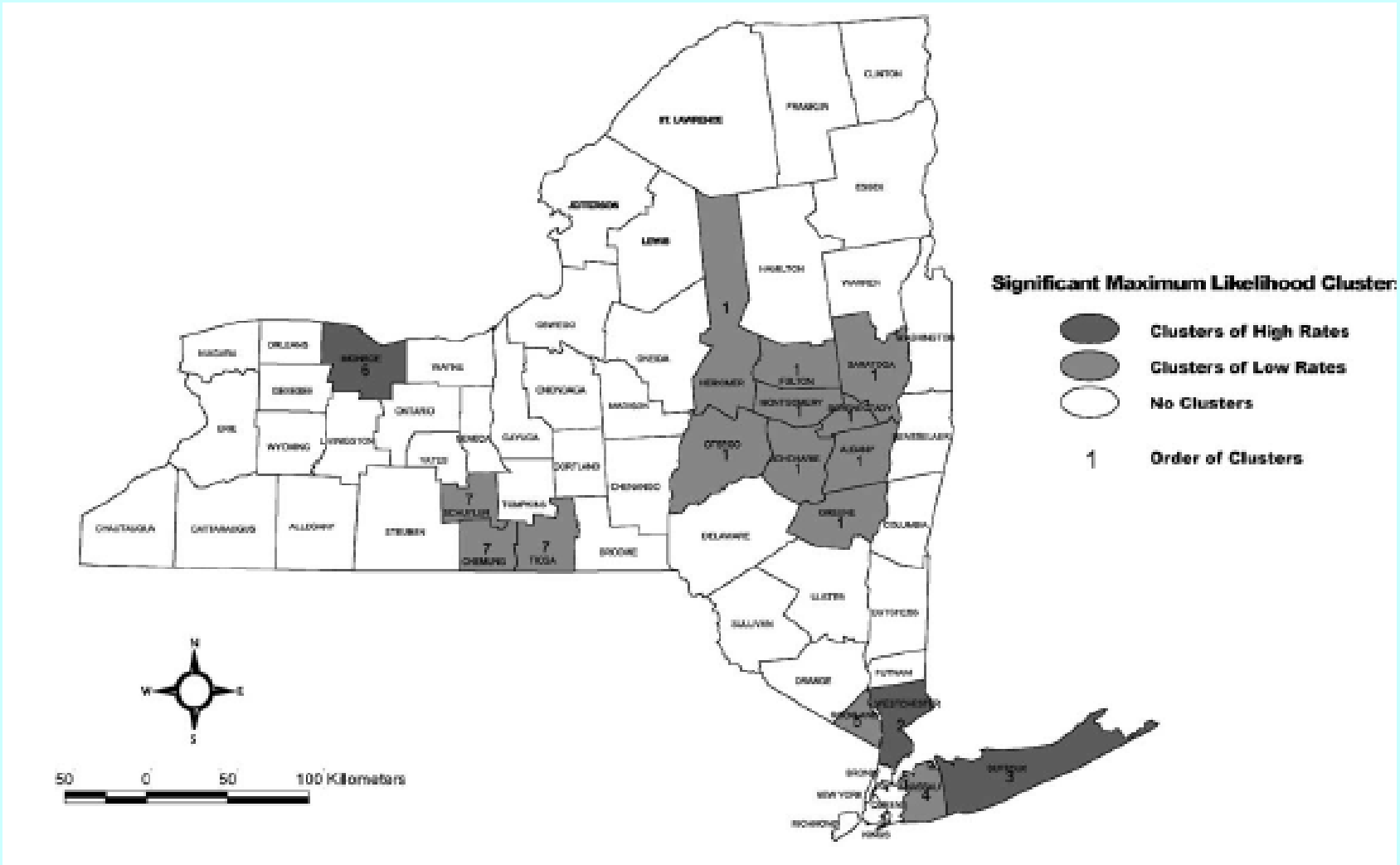


Source: Mason et al., ***Atlas of Cancer Mortality for U.S. Counties,*** **1950-1969. NIH, 1975.**

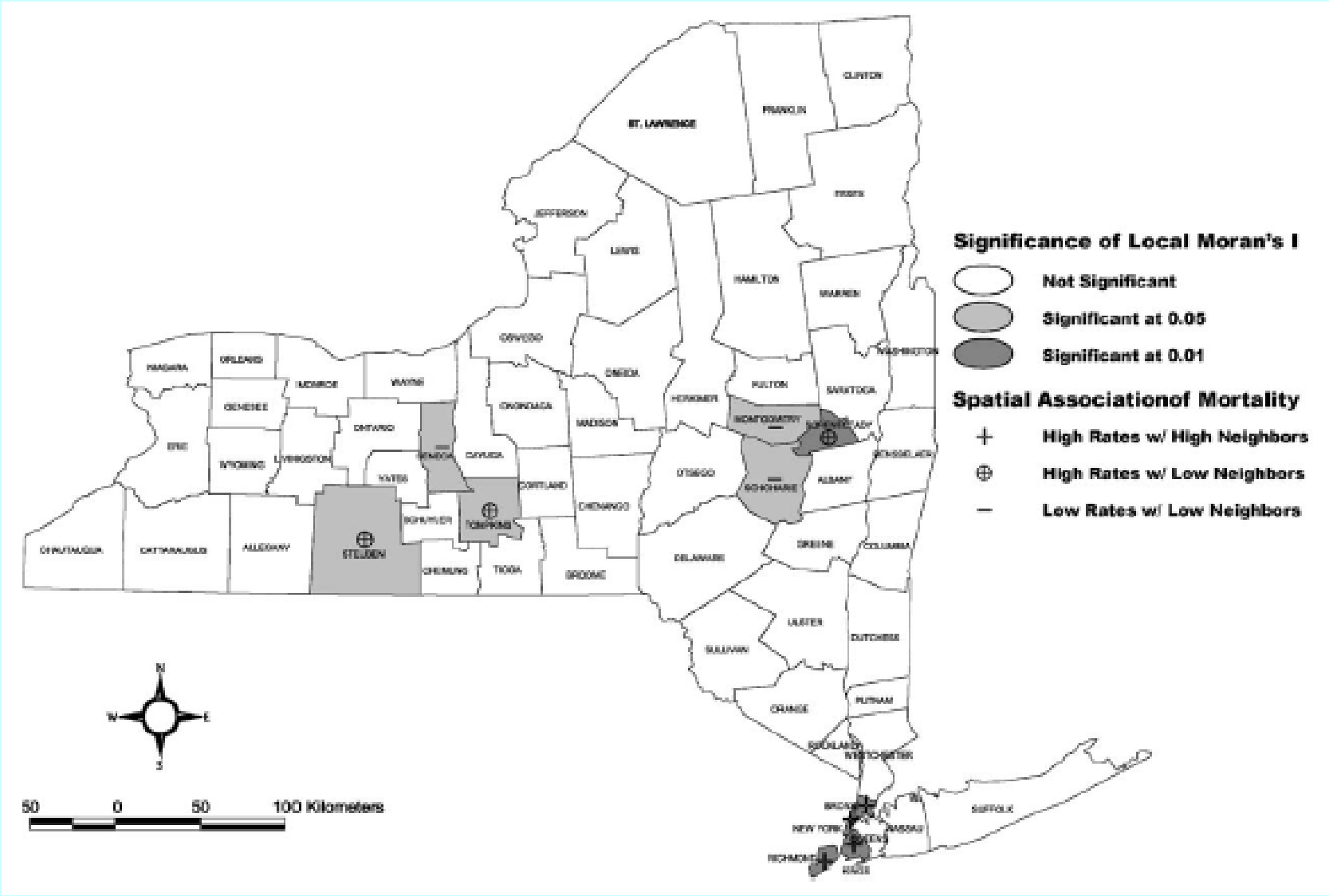
Map of direct age-adjusted cause of death with explicit mention of alcohol (per 100,000 population).



SATScan (Kulldorff, 1997) Significant spatial scan clusters of alcohol-explicit mortality. Source: Hanson and Wieczorek, 2002, SS&M, p.799.



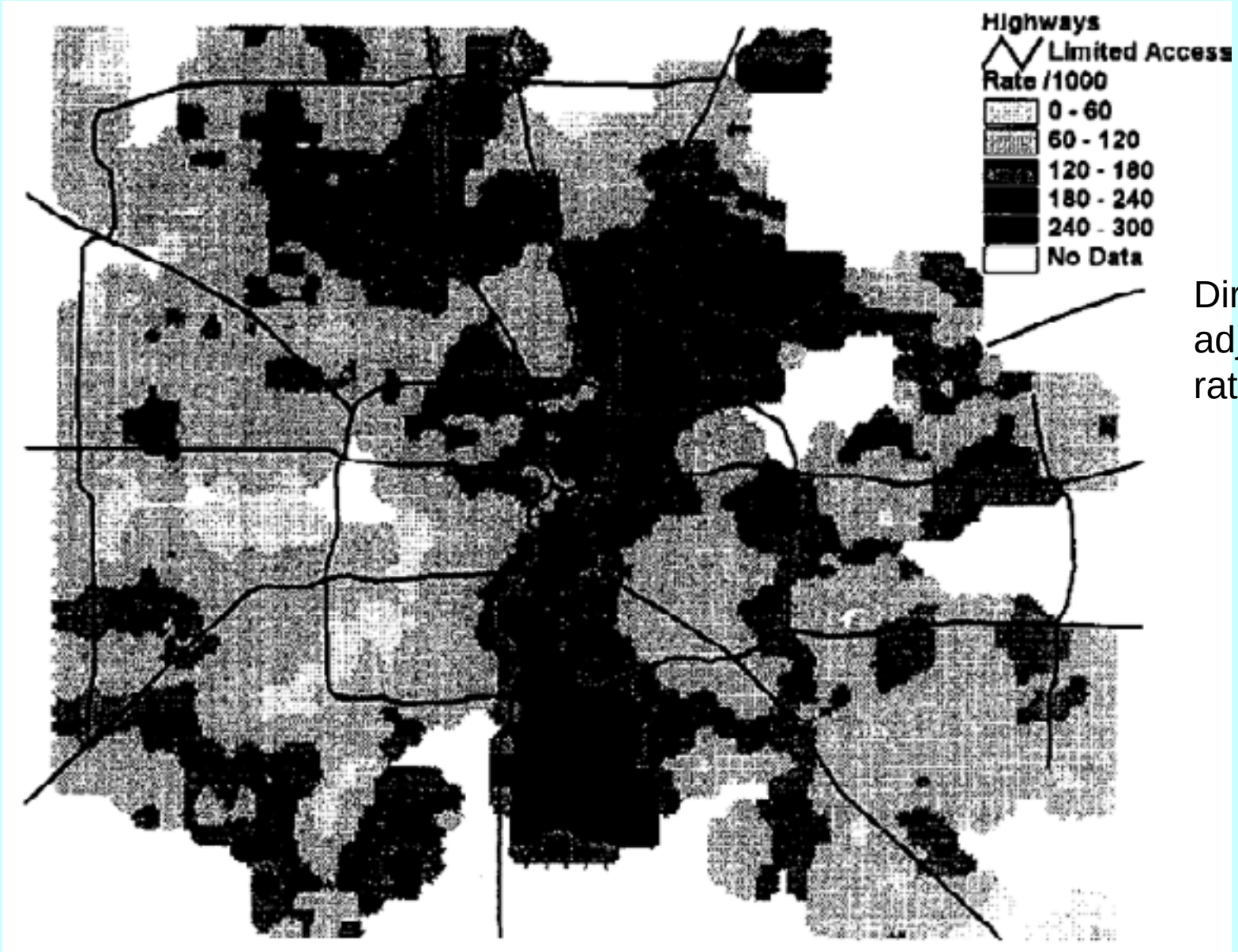
LISA (Anselin 1995, G.A.) Areas of significance of spatial association of alcohol-explicit mortality. Hanson and Wieczorek, 2002, SS&M, p.798



Comparison of Three Methods of Identifying Areas of Risk

Source: B. Sue Bell.2002. Spatial Analysis of Disease – Applications. Ch. 8 in C. Beam, ed. Biostatistical Applications in Cancer Research. Boston: Kluwer Academic Publishers, 2002, pp. 151-182.

“The research question was “Does the birth outcome of intrauterine growth retardation (IUGR) co-vary ecologically with characteristics at the neighborhood level?” IUGR (Usher and McLean 1969; Frisbie et al. 1997) is an outcome measure to categorize births that are small for gestational age. The project team analyzed geocoded birth records for Harris County (Houston), Texas for 1991.” p. 160.



Direct adjusted rates

Fig. 1. Map based on spatial filtering using a half-mile grid and a half-mile filter

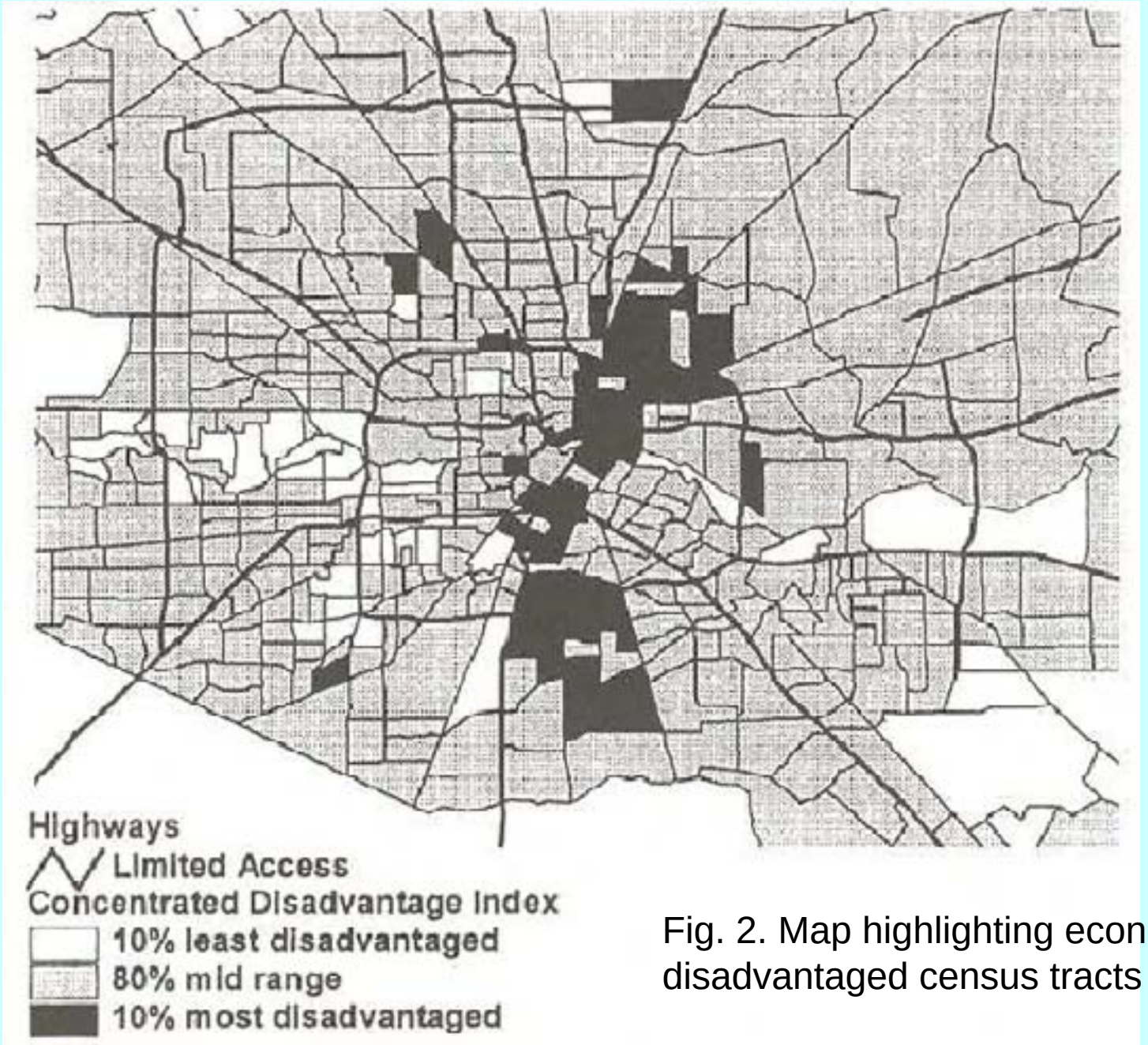


Fig. 2. Map highlighting economically disadvantaged census tracts

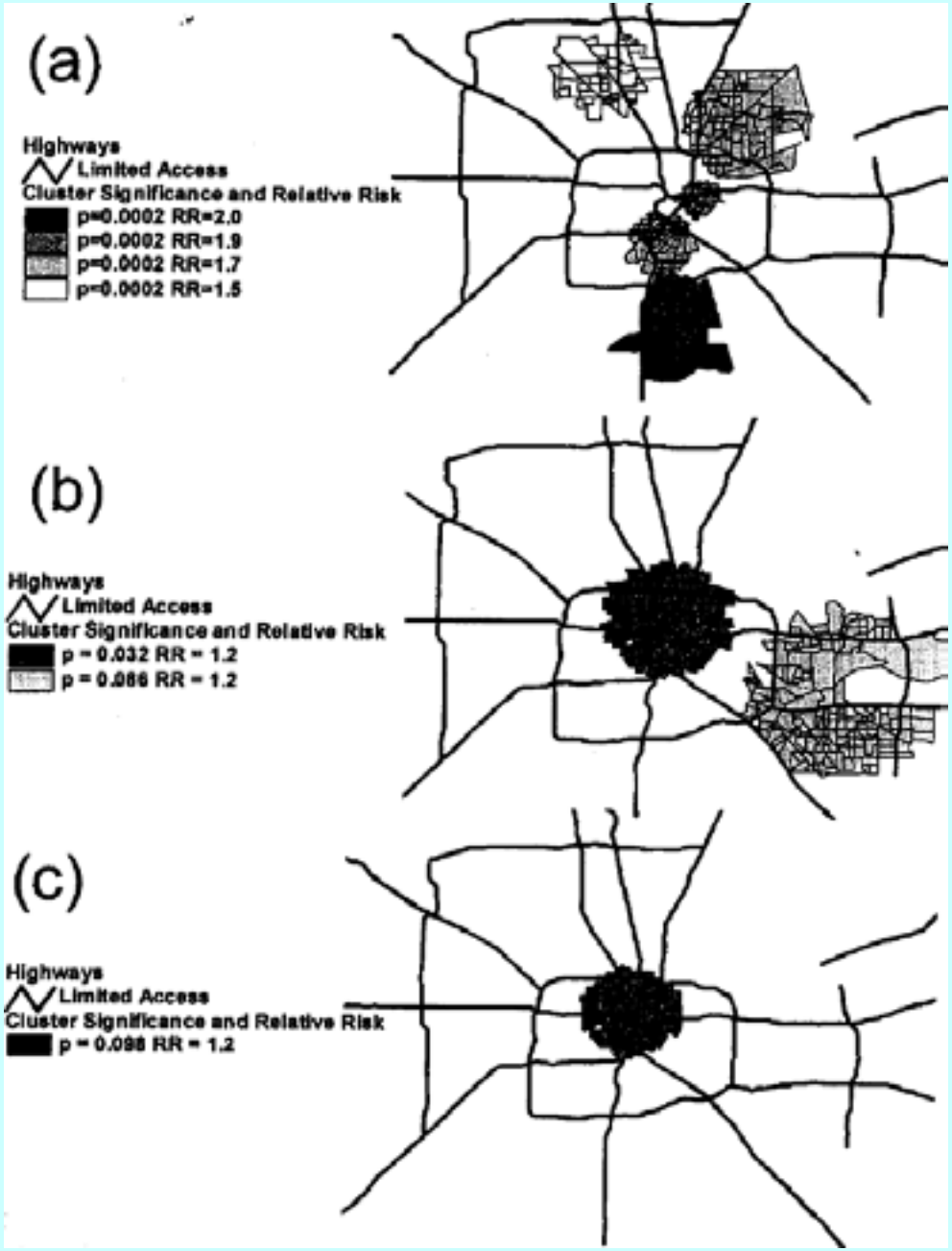


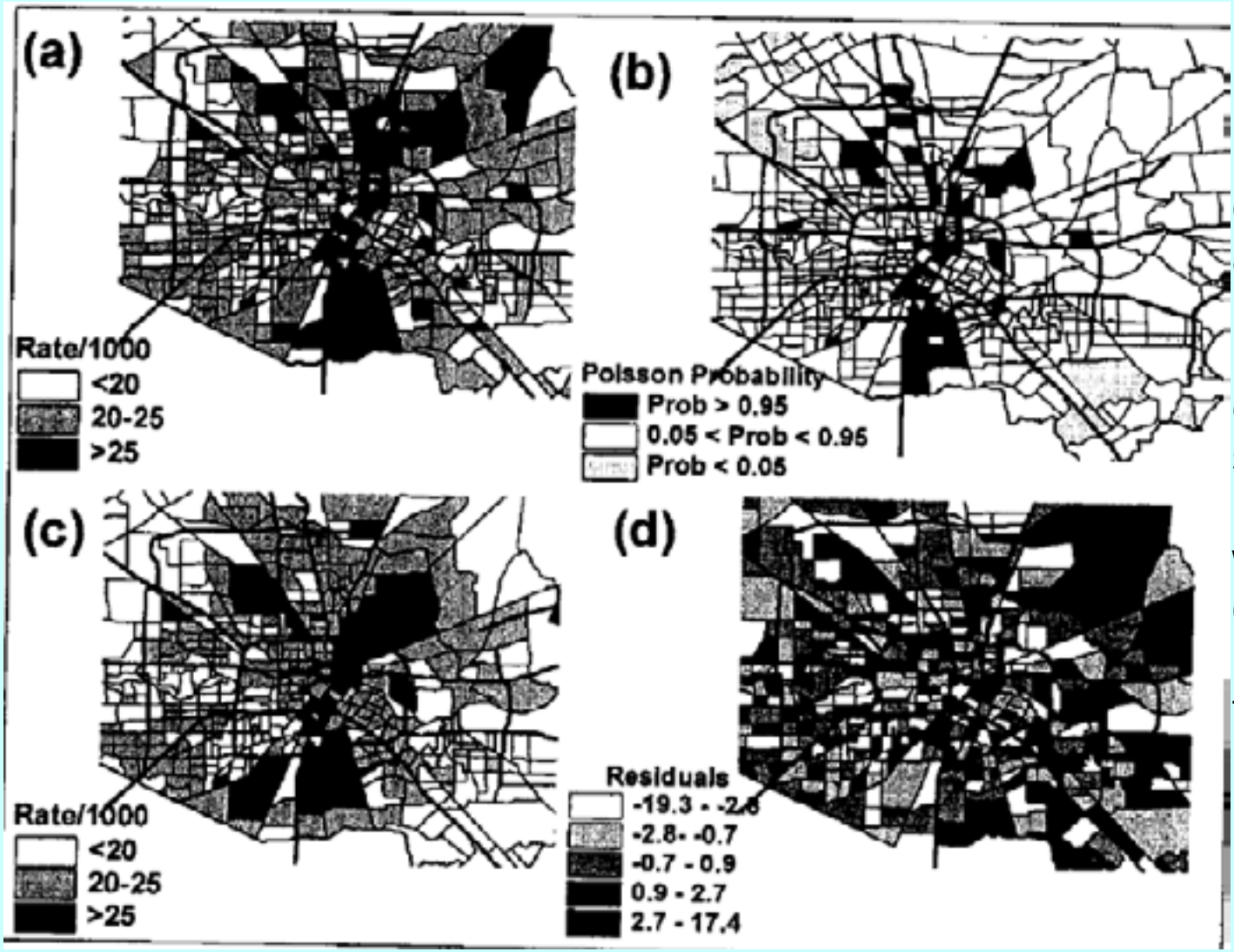
Fig. 3. Comparison of clustering results from SaTScan with

(a) no adjustment for confounding variables,

(b) adjustment for race/ethnicity of the mother, and

(c) adjustment for race/ethnicity and age of mother. (Note: first cluster in each group with the darkest shading is the primary cluster).

Source: Bell, 2002, p. 168



(c). The CAR model is a conditional autoregression model. .."It incorporates the concept of spatial neighbors where the definition of neighbor is left to the analyst." (Bell 2002, p.173.

Fig. 5. Maps of rates, probabilities and residuals (a) rate before spatial modeling (b) map of statistically high rates and statistically low rates assuming rates have a Poisson distribution with constant mean, © map of predicted rates fit by a CAR spatial model, (d) map of residuals after fit by a CAR spatial model. (Bell, 2002, p.177).

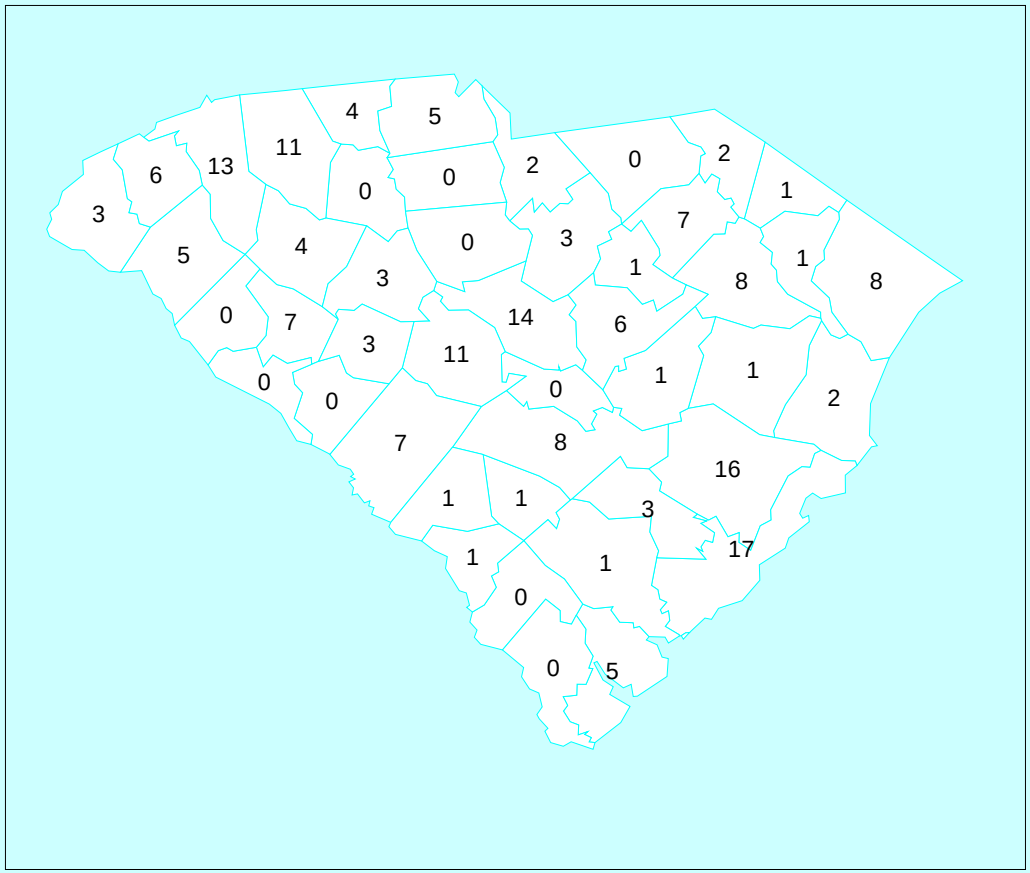
Introduction to Bayesian Mapping Methods

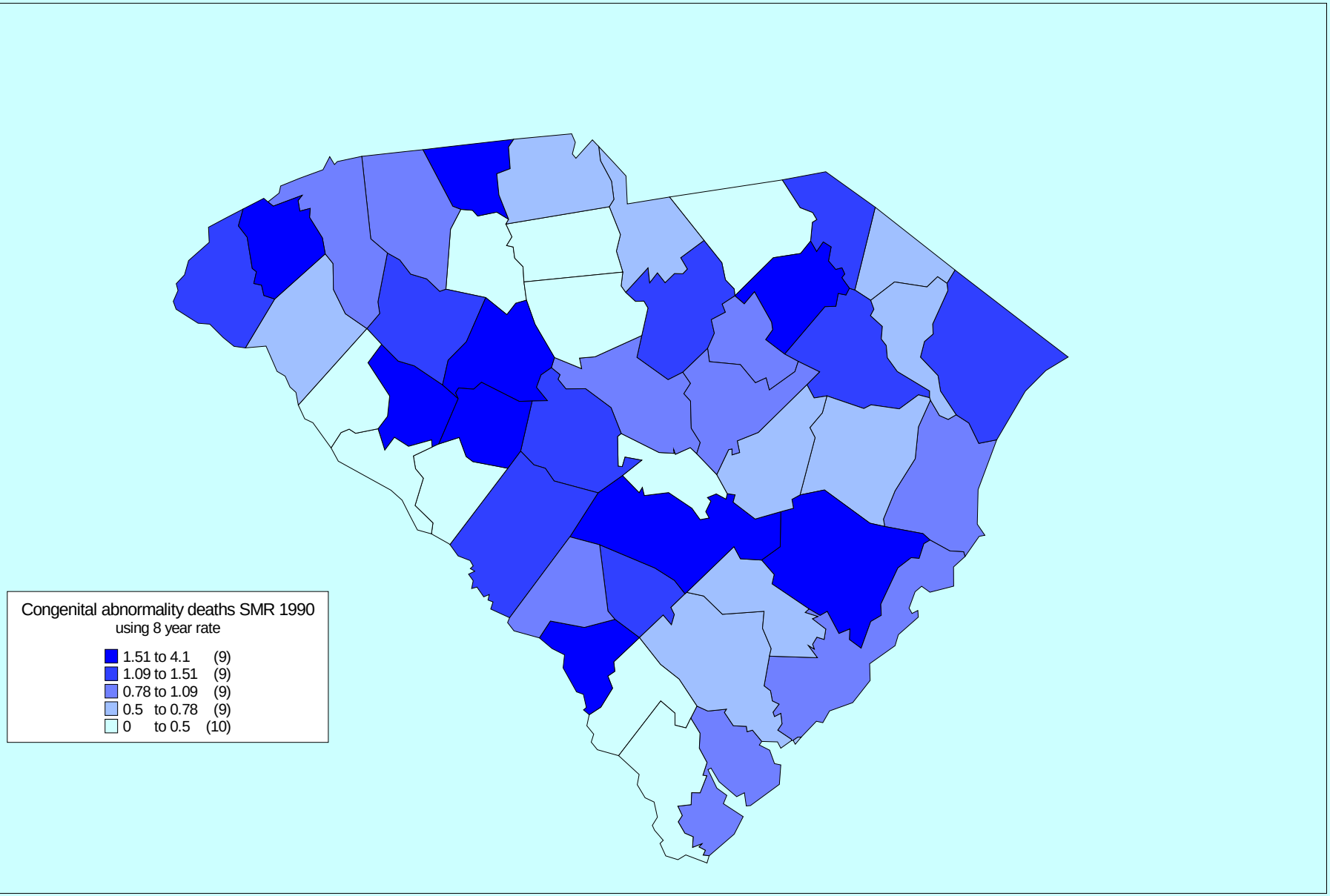
Andrew B. Lawson

Arnold School of Public Health

University of South Carolina

- South Carolina congenital abnormality deaths 1990




Congenital abnormality deaths SMR 1990
using 8 year rate
1.51 to 4.1 (9)
1.09 to 1.51 (9)
0.78 to 1.09 (9)
0.5 to 0.78 (9)
0 to 0.5 (10)

Some notation

- For each region on the map:
- y_i is the count of disease in the i^{th} region
- e_i is the expected count in the i^{th} region
- θ_i is the relative risk in the i^{th} region

- The SMR is just $smr_i = y_i / e_i$
- This is just an estimate of θ_i

SMR problems

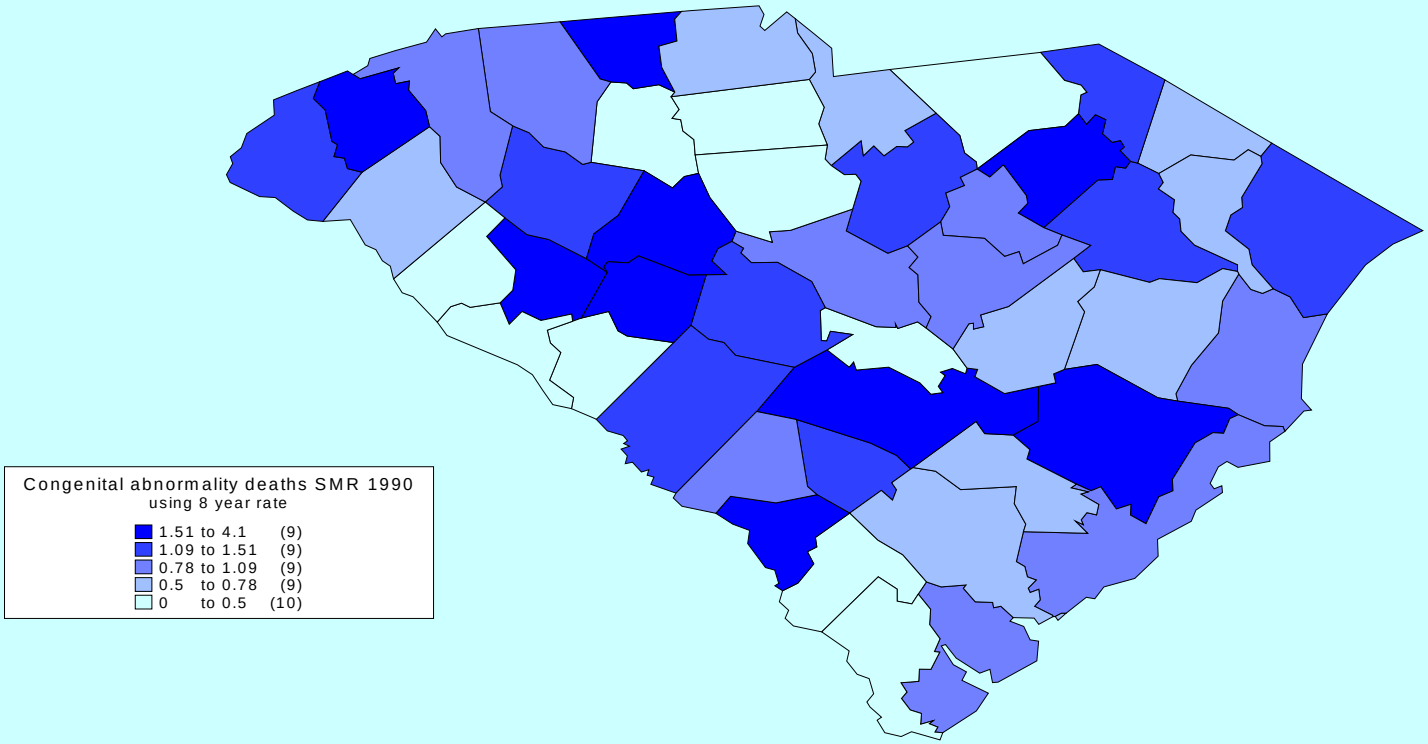
- Notoriously unstable
- Small expected count can lead to large SMRs
- Zero counts aren’t differentiated
- The SMR is ***just the data!***

Bayesian Modeling

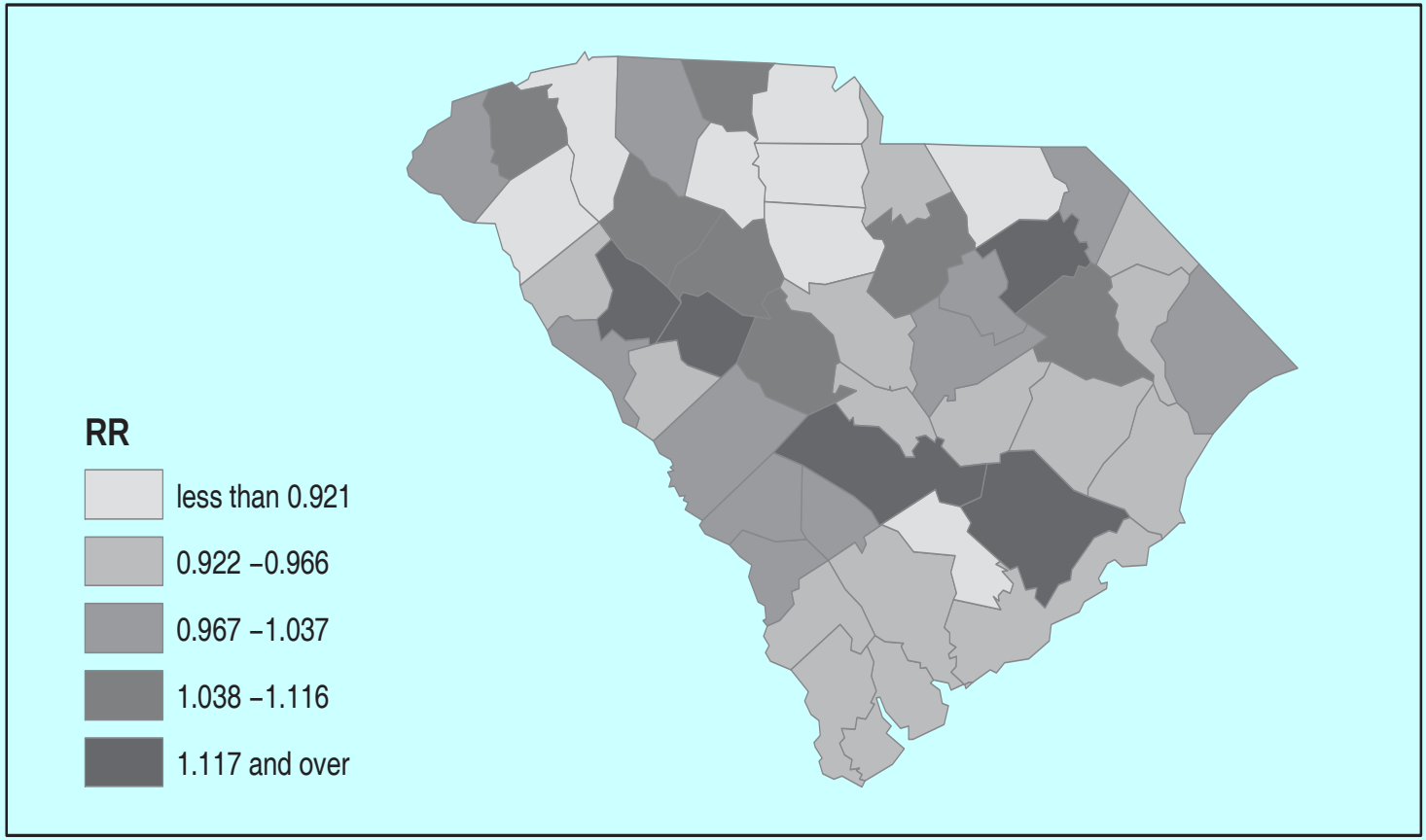
Some statistical ideas:

- Likelihood…….we usually assume that counts of disease have a Poisson distribution so that y_i has a Poisson distribution with expected value $e_i \theta_i$
- We usually write this as $y_i \sim \text{Pois}(e_i \theta_i)$ for short
- The counts have a Poisson likelihood

SMR for congenital anomalies



Gamma Poisson model: WinBUGS

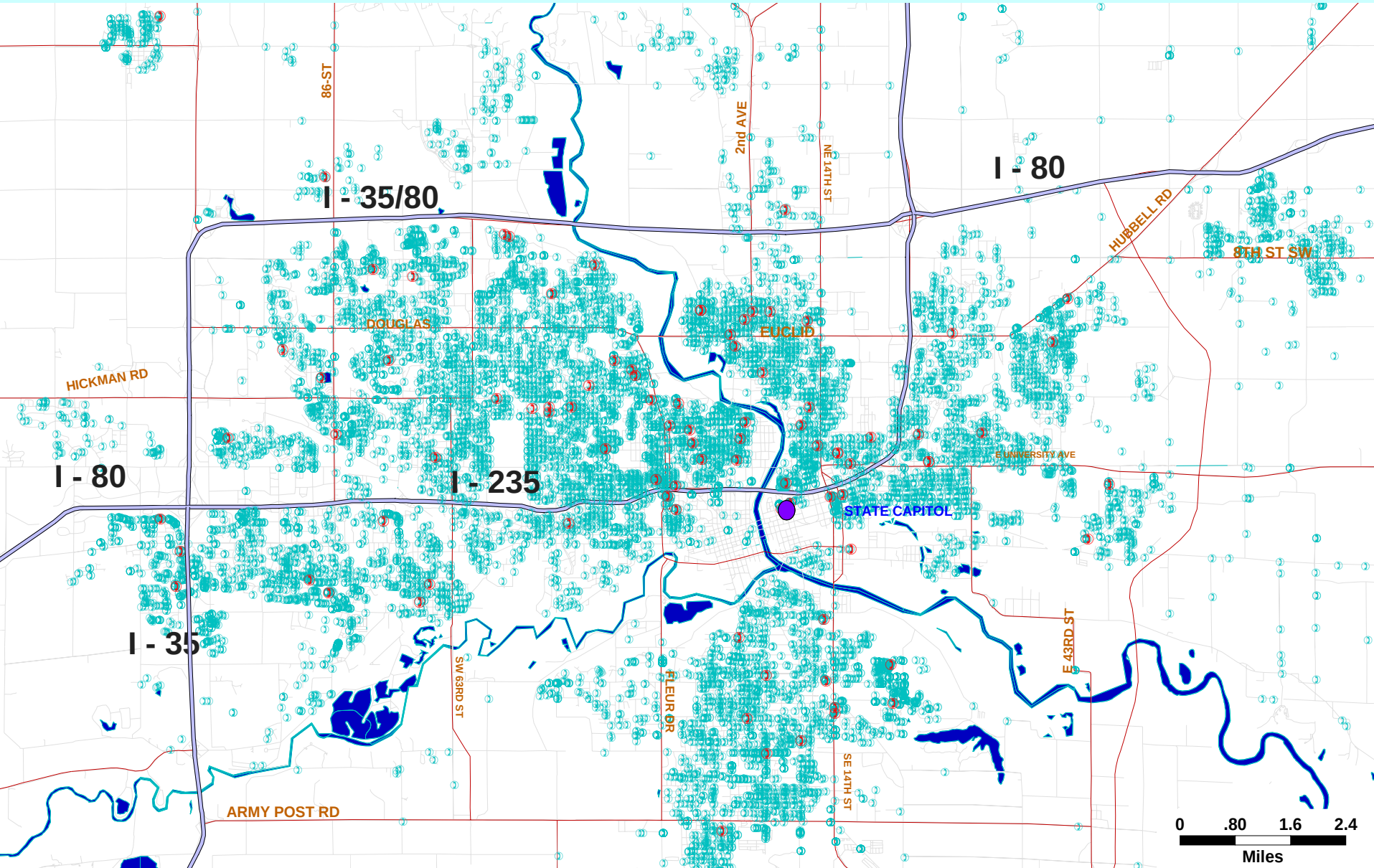


Conclusions (Lawson 2003)

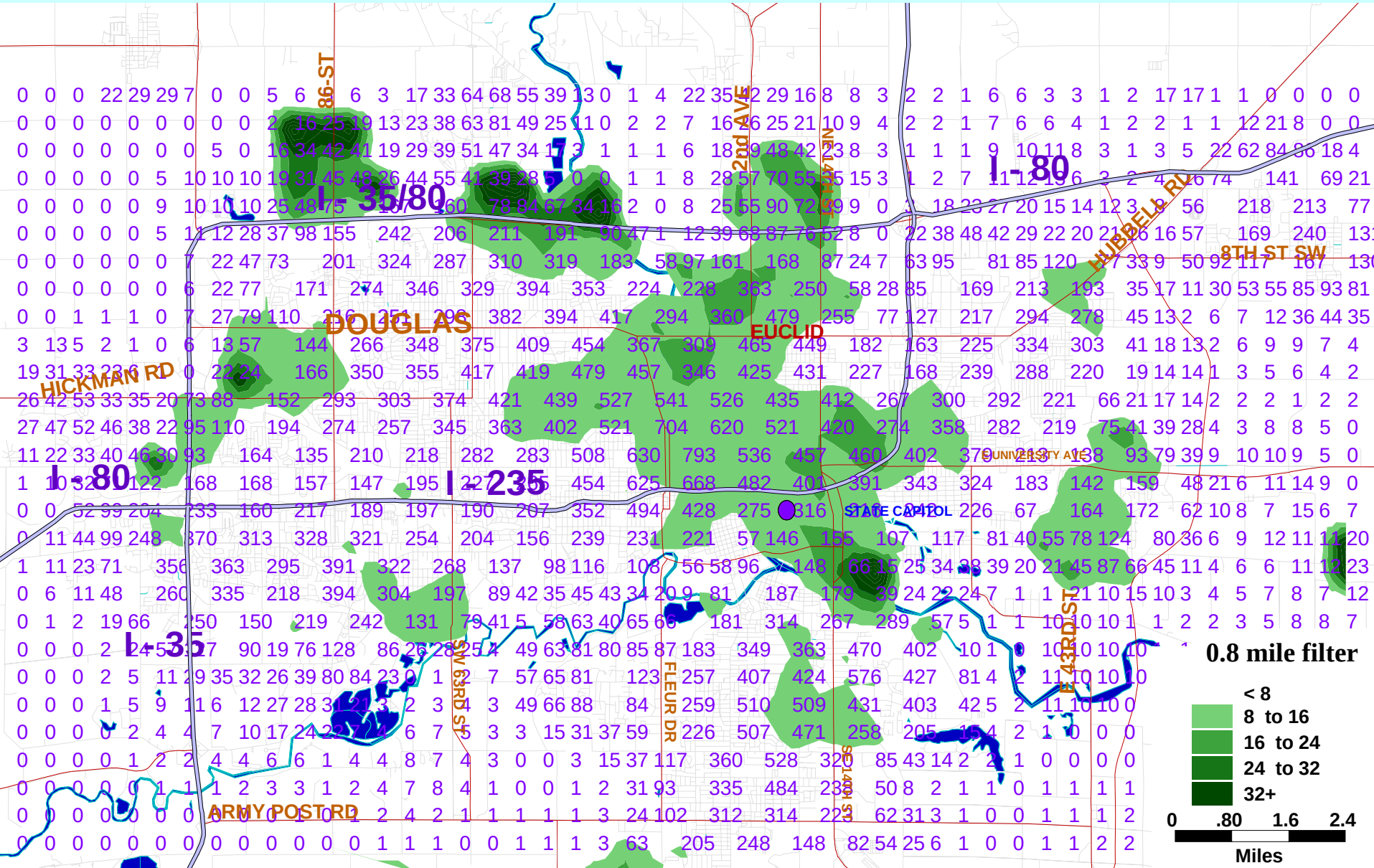
- WinBUGS provides a *free* and relatively easy-to-use tool for disease mapping with small area count data
- Allows state-of-the-art approach to relative risk and ecological regression
- Available from:

  www.mrc-bsu.cam.ac.uk/bugs

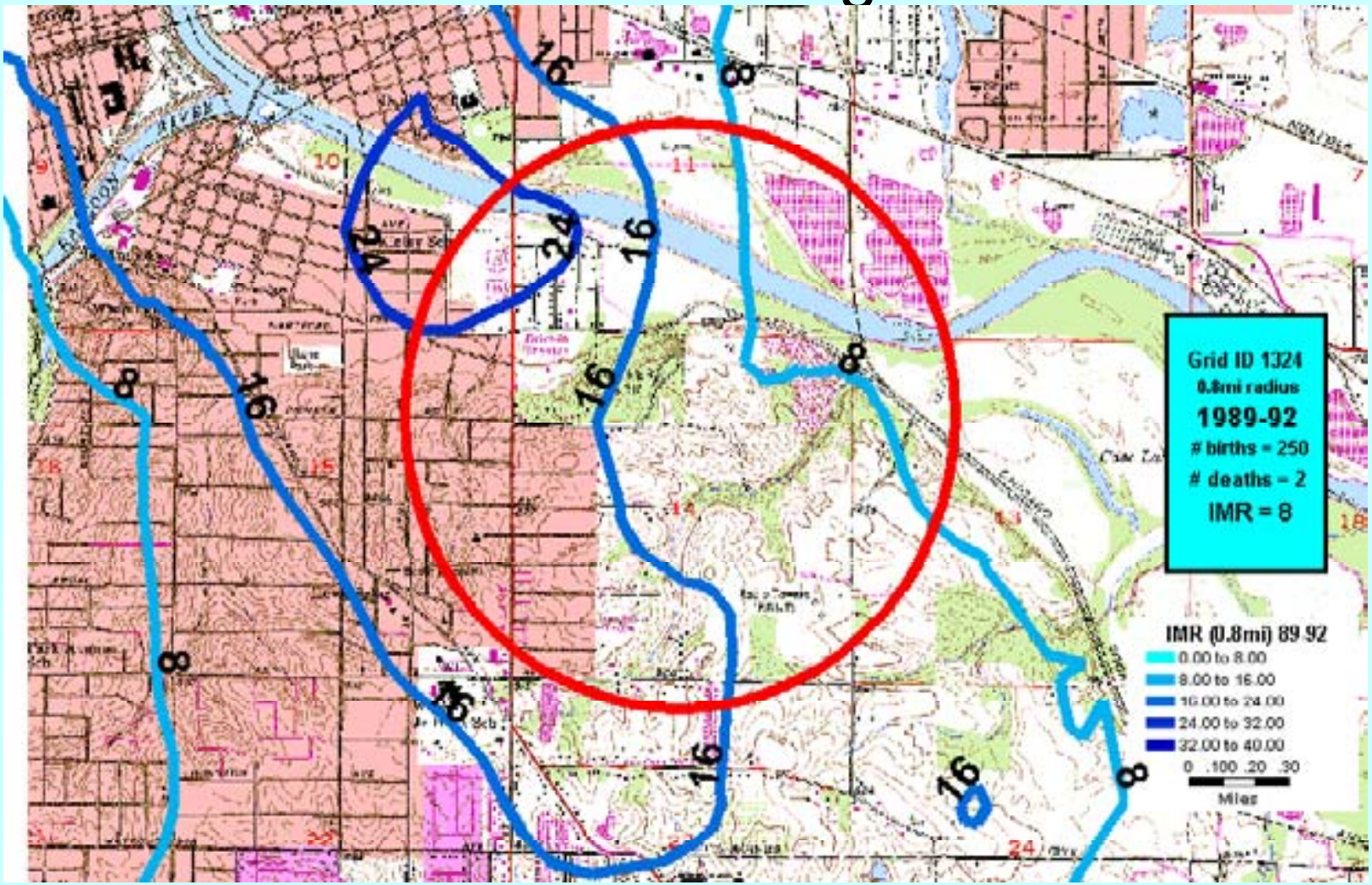
Location of Infant Births and Deaths: 1996-98



Des Moines, Iowa: Births and Infant Mortality Rates: All Races, 1996-98



Part of Des Moines, Iowa: infant mortality rates 1989-92 draped on USGS 1:24,000 TIF Image



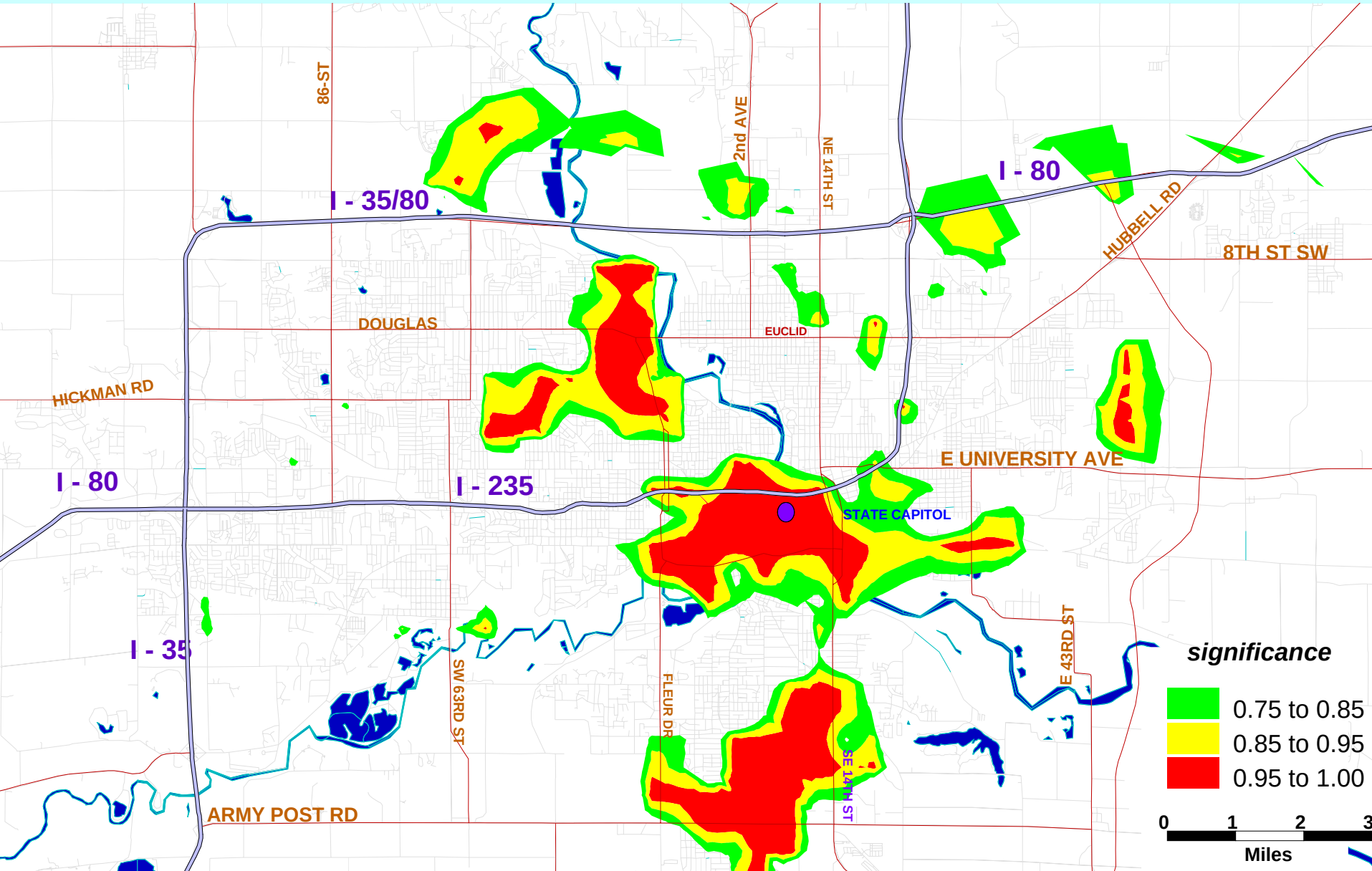
Not all areas of the map have equally reliable *rates*

- Some local rates are based on a large amount of information; other local rates are based on little information.
- Need to compute the local reliability of the observed rate
- Generate a map of the expected disease pattern according to the null hypothesis that each person at risk has an equal and identical chance of having the disease

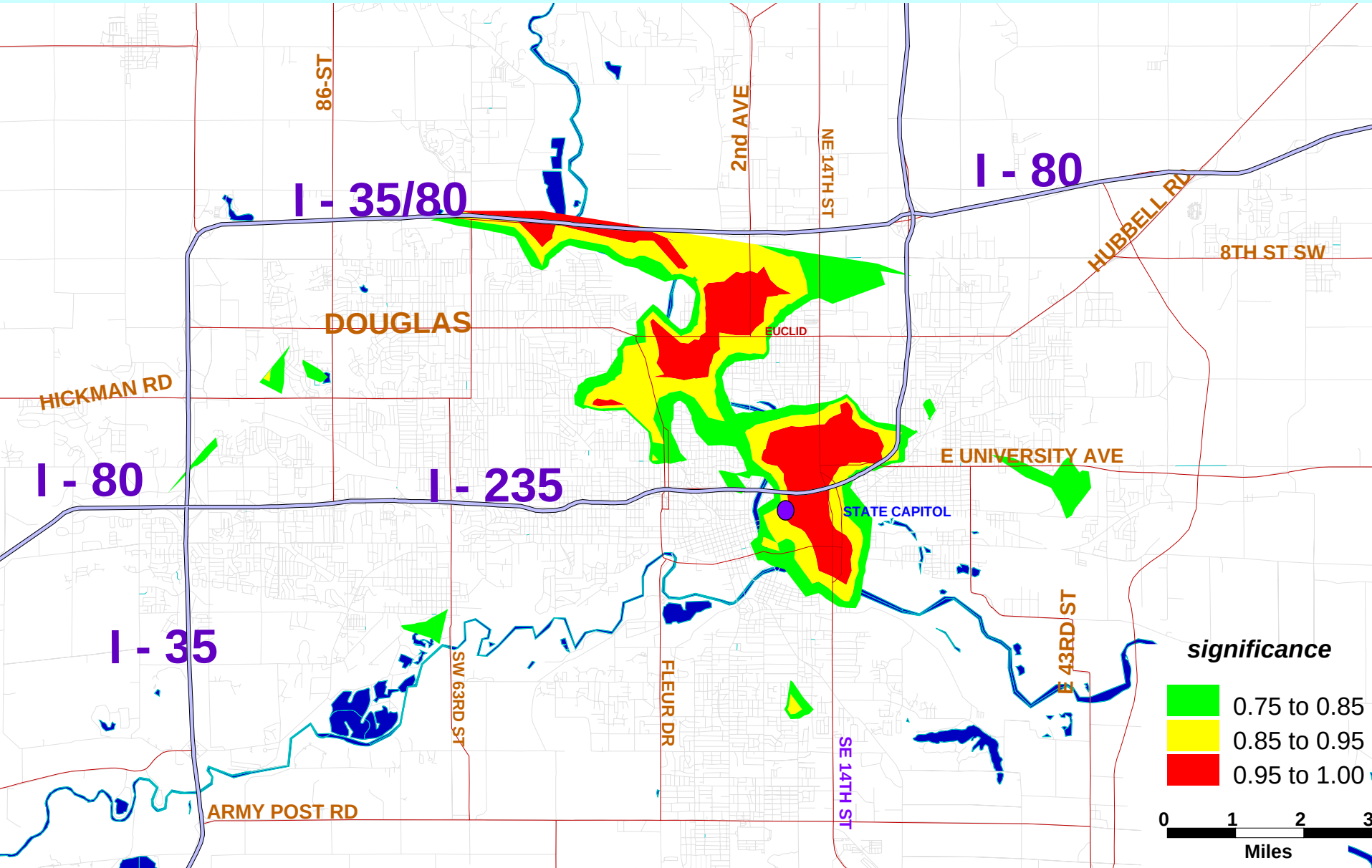
Significance of Rates

- let each birth in 1989-91 have an identical probability equal to the county-wide infant mortality rate of 9.5 of becoming an infant death
- generate 1,000 geographical distributions of simulated infant deaths
- for each of these maps compute the infant mortality rate at each grid location
- compute the proportion of the simulated rates at each grid location that are smaller than the observed infant mortality rate
- define regions with significantly high infant mortality rates where a high proportion of the simulates rates are smaller than the observed rate

Significance of Infant Mortality Rates: 1993-95

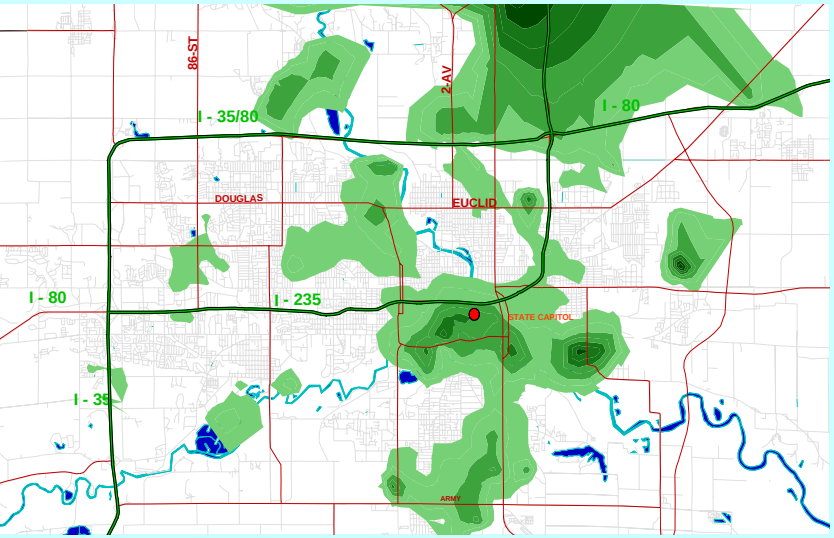


Significance of Infant Mortality Rates: 1996-98

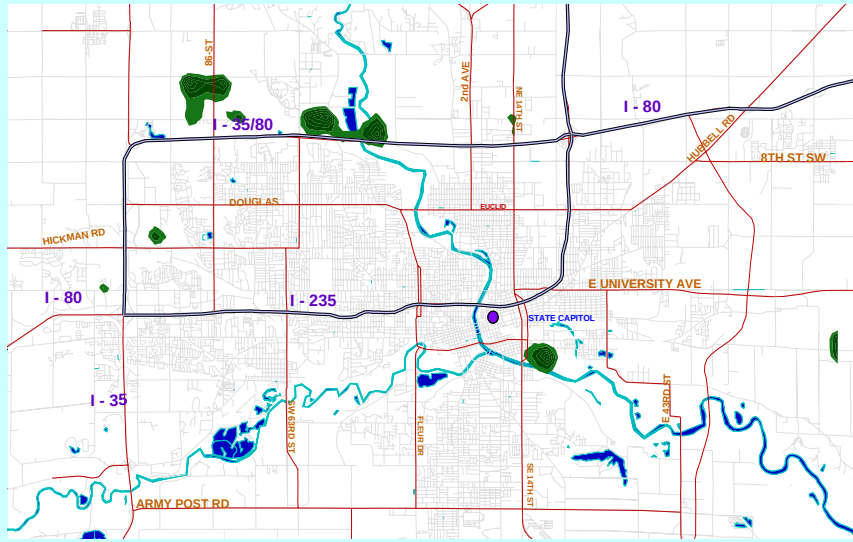


Change in Infant Mortality 1993-1998

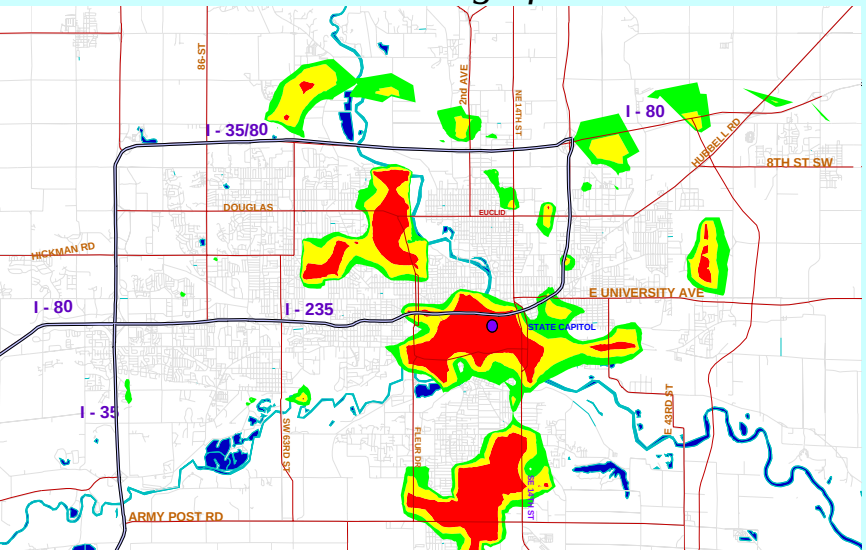
1993-95 IMR



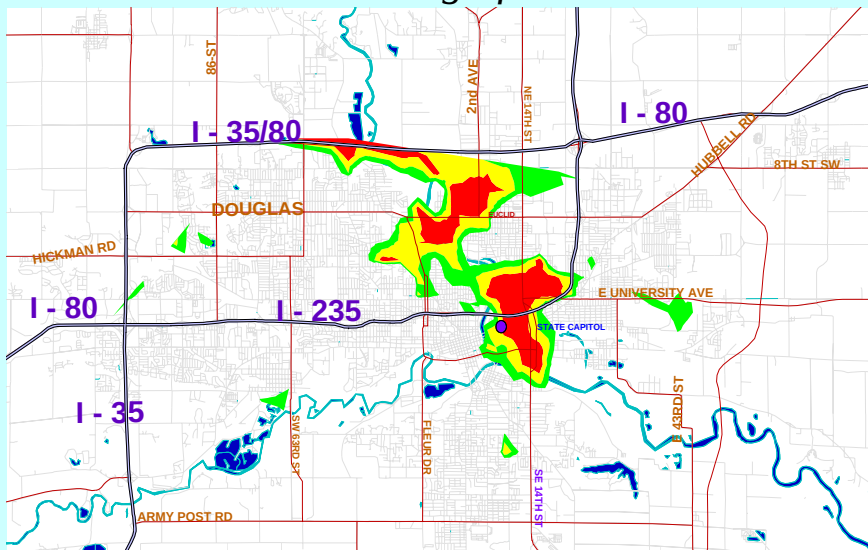
1996-98 IMR



1993-95 Significance



1996-98 Significance

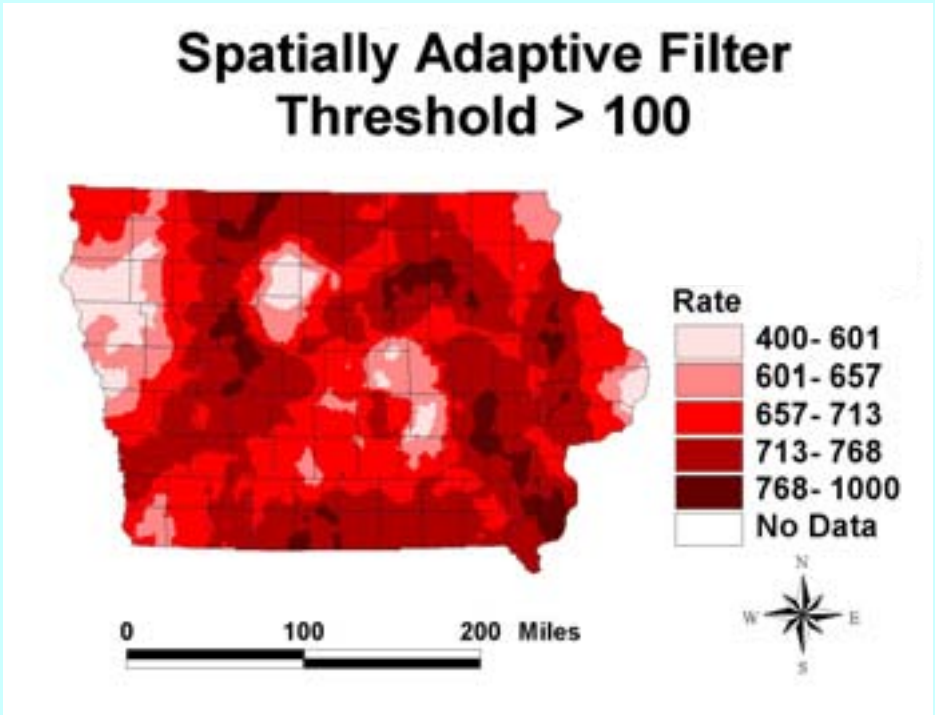


Examples

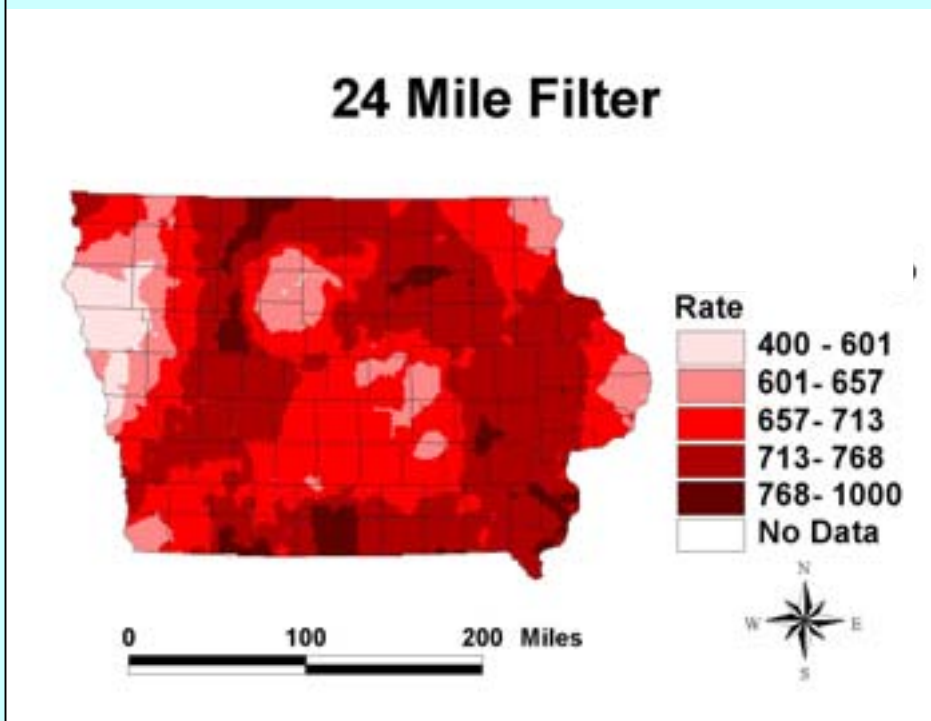
- Most searches for spatial clusters of cancer use geographic units such as health service areas, state economic areas or counties.
- But suppose areas of high risk for cancer are much more localized?

Spatially Adaptive Filters

Rates of Late-Stage Colorectal Cancer at Time of First Diagnosis



Using Spatially Adaptive Filters

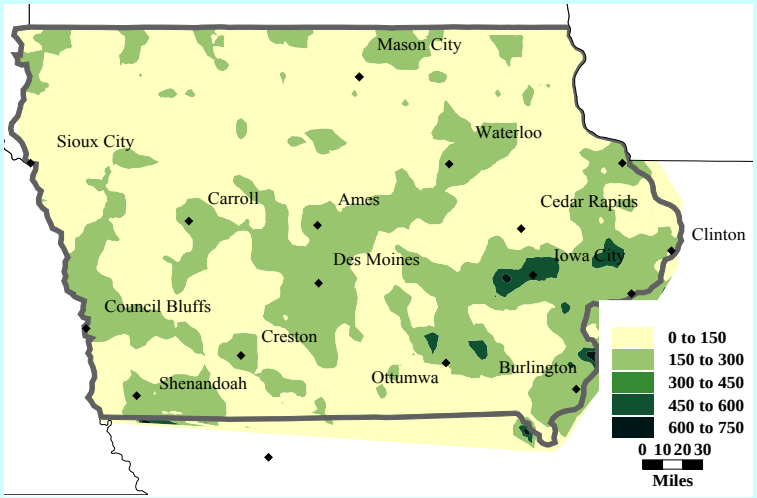


Using Fixed Filter Size of 24 Miles

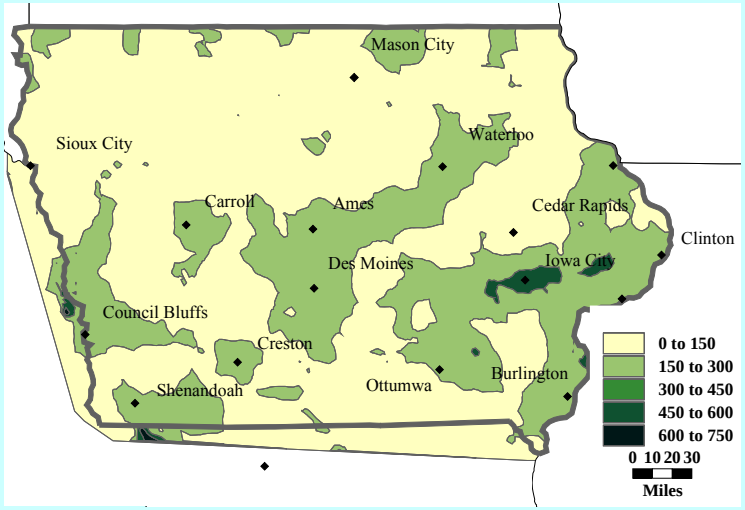
Source: Chetan Tiwari

Lumpectomy With Radiation Rates

10 Mile Filter

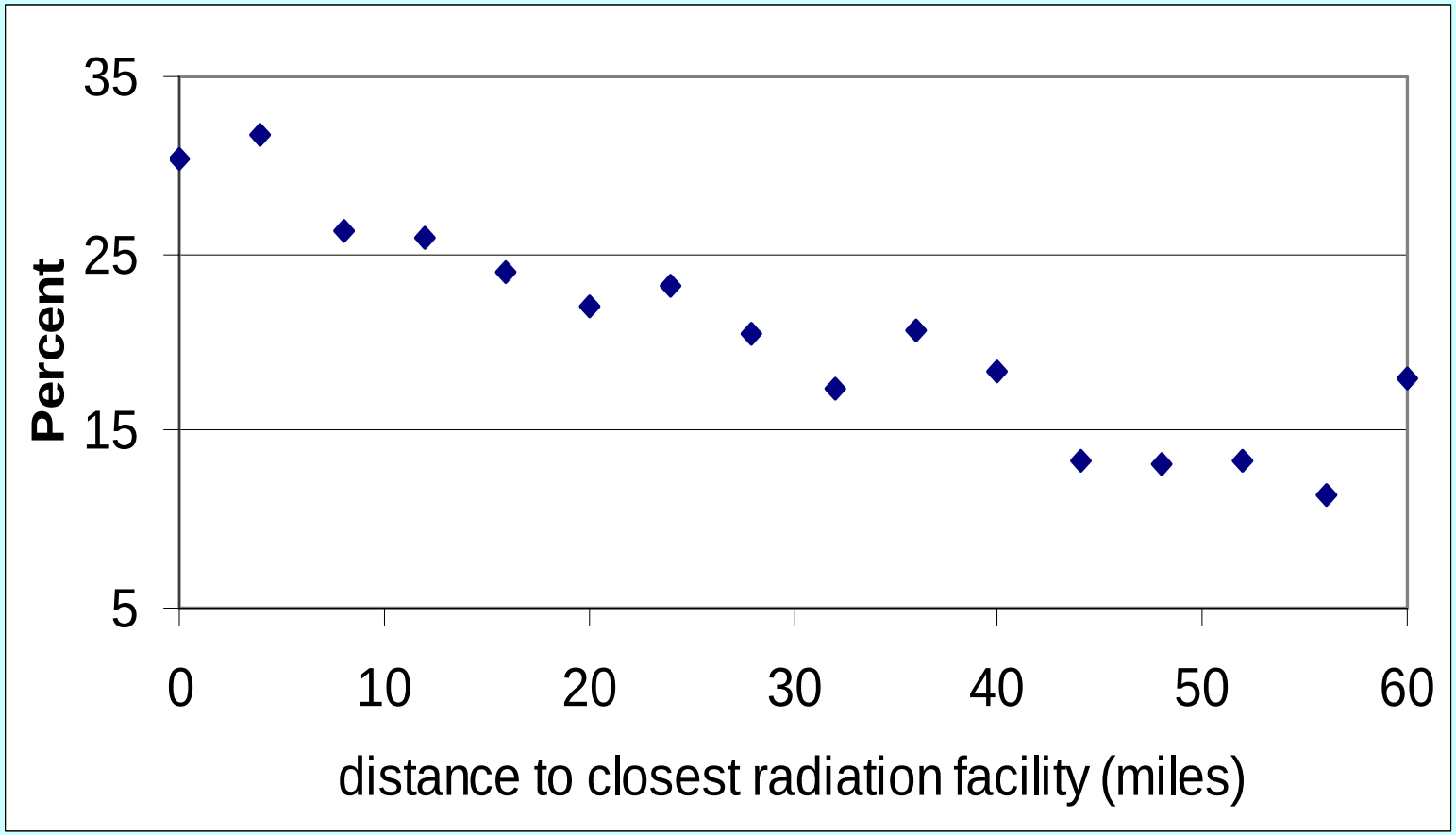


12.5 Mile Filter



15 Mile Filter

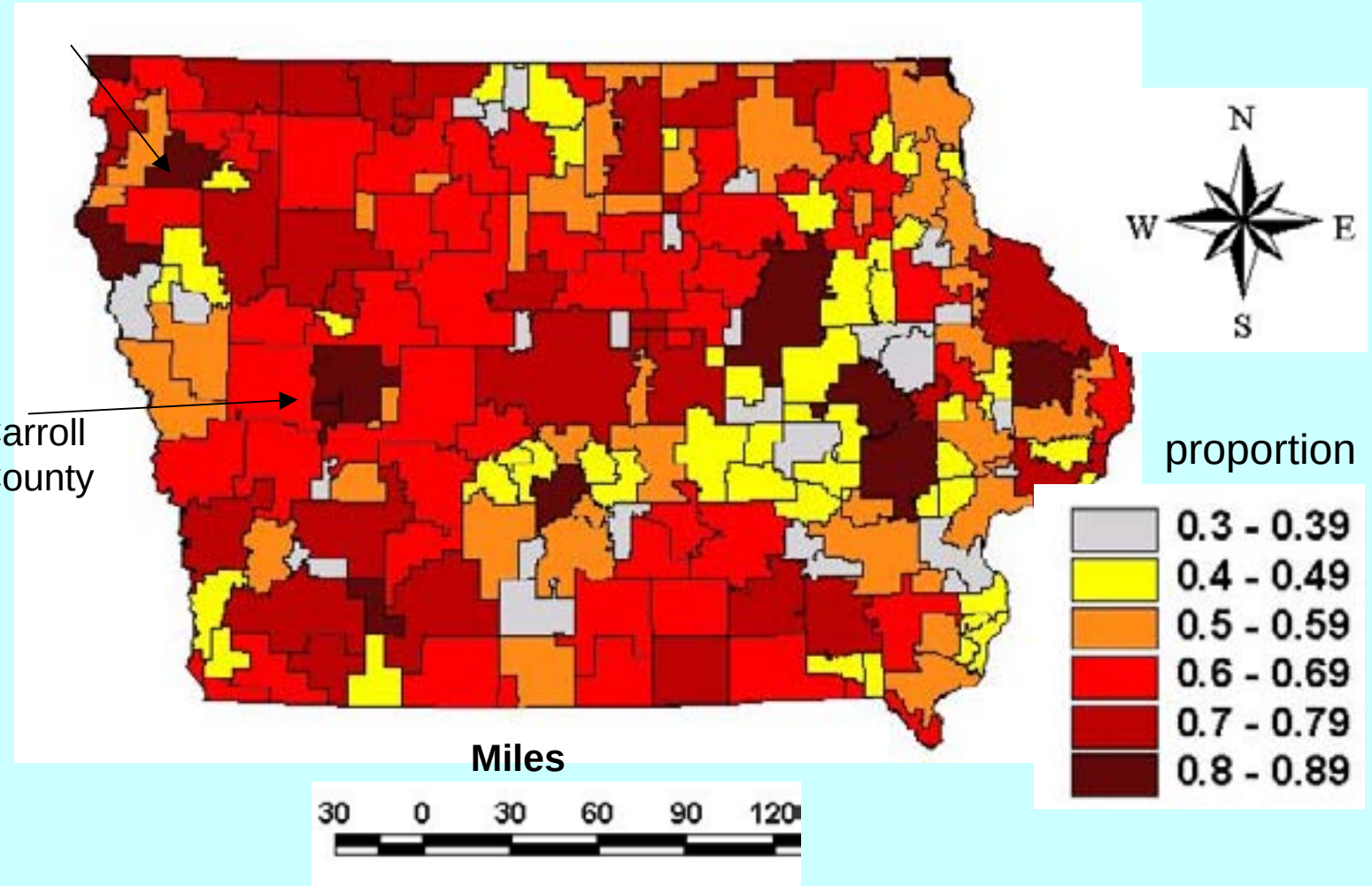
Percent Selecting Lumpectomy/Radiation Treatment



Problem with geocoding practices

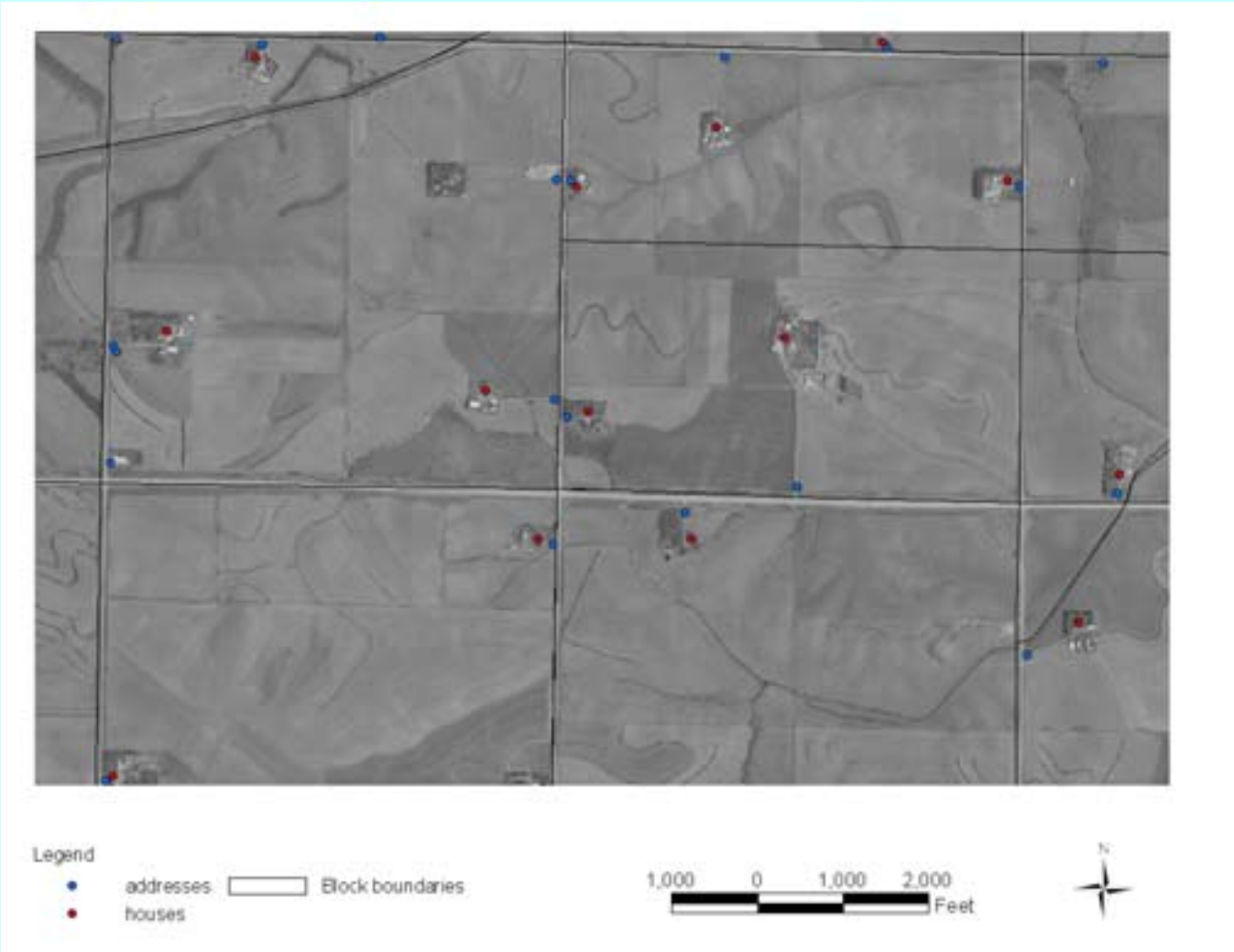
- Geocoding is not taken too seriously in most environment exposure and health applications
- The principle of ‘fitness for use’ must be understood and used.

Iowa: Primary Care Service Areas—proportion of primary care patient visits made inside the local service area

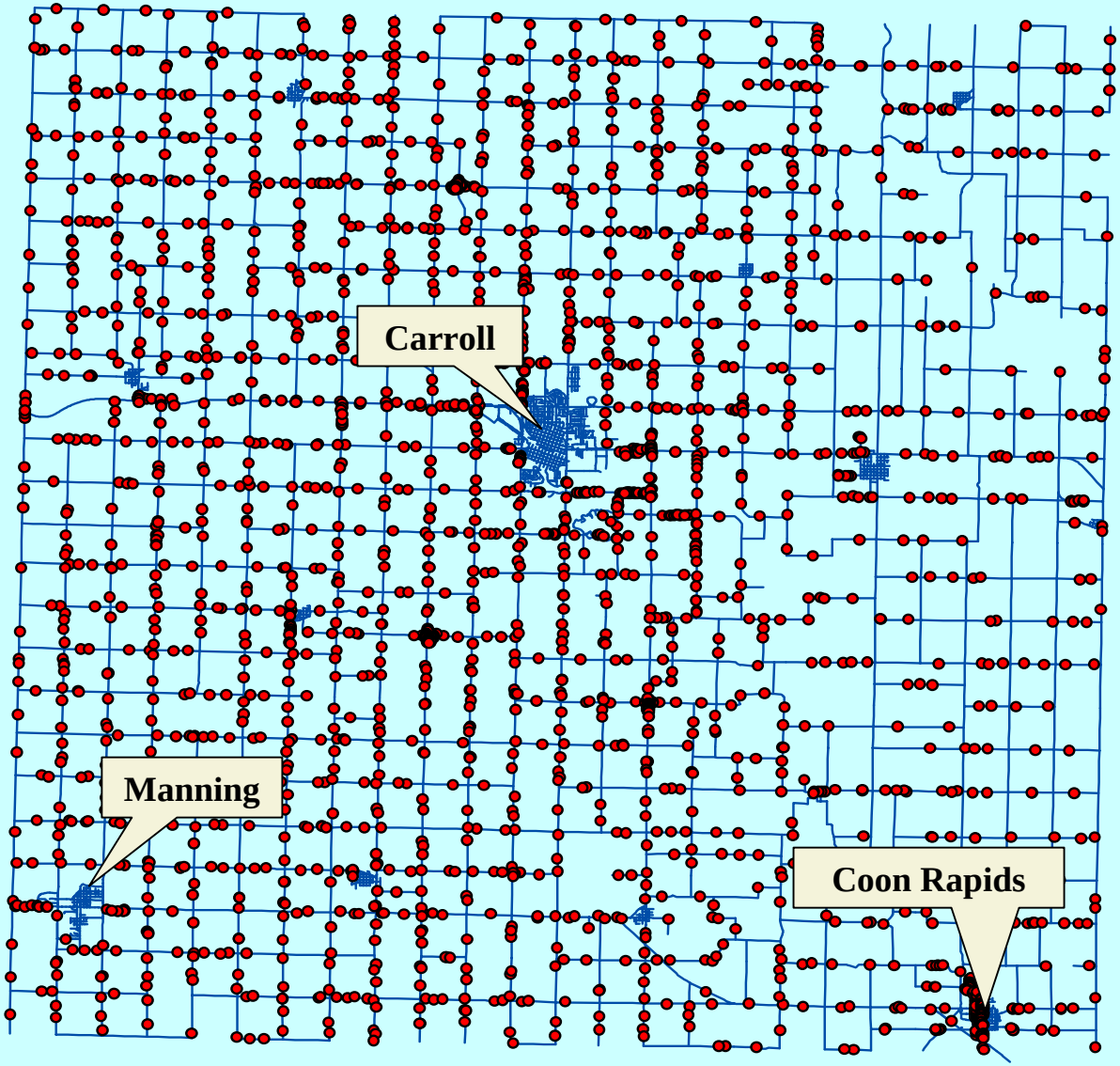


Source: data files from pcsa.hrsa.gov See Goodman et al. 2003

In this map, there are 14 houses and 20 addresses. Each house has a corresponding address point located where the entrance road to the house leaves the public road. There are two possible reasons for the other six addresses that do not have a corresponding house. The first is that the address is not a residential address, the second is that the house of the resident address can not be identified by this interpretation of the orthophoto map.



Address point map containing the locations of separately addressable units in Carroll County



This map(07/21/03) shows the current extent of the address point file. This file contains the locations of 100% of the addresses in the E-W direction and 2/3rds of the addresses in the N-S direction.

Legend

- Address point
- Roads

Long Island Breast Cancer Study and the GIS-H (Health)

NATIONAL CANCER INSTITUTE

Division of Cancer Control & Population Sciences

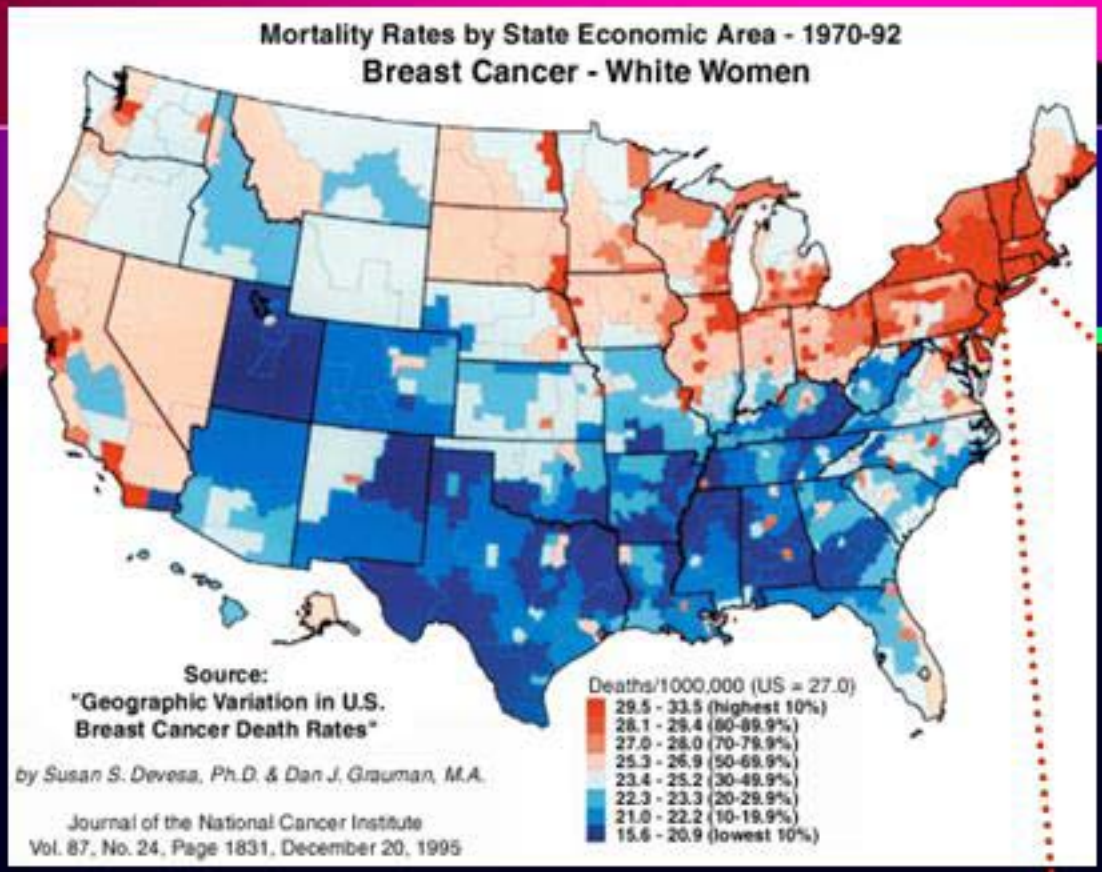
Edward J. Trapido, Sc.D.

Associate Director

Epidemiology and Genetics Research Program, DCCPS/NCI

COMPREHENSIVE APPROACHES TO CANCER CONTROL

September, 2003 Atlanta, GA



NATIONAL CANCER INSTITUTE

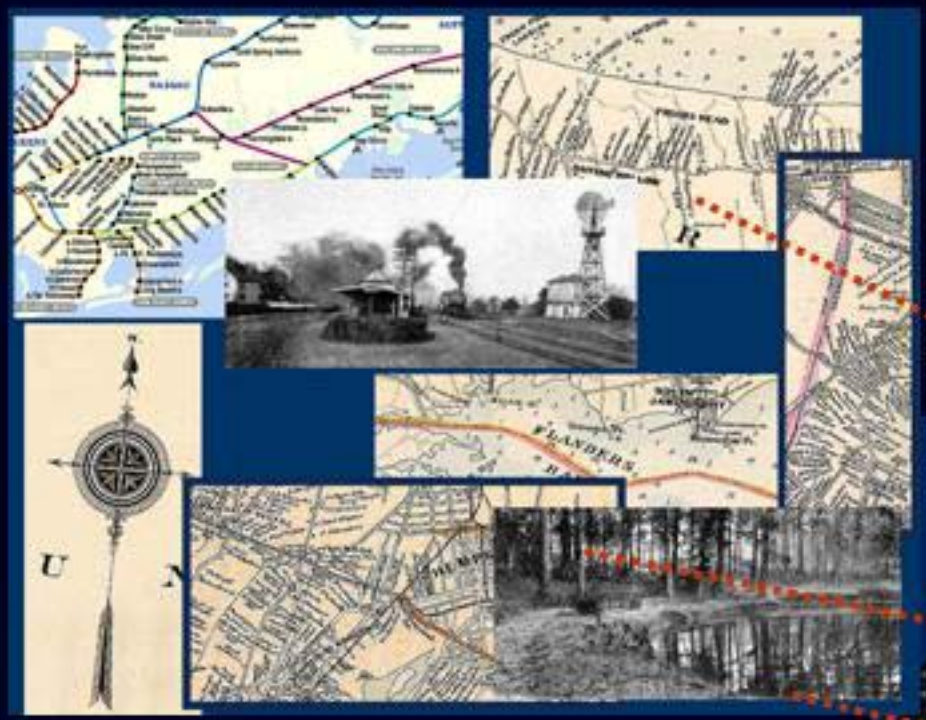
Division of Cancer Control & Population Sciences

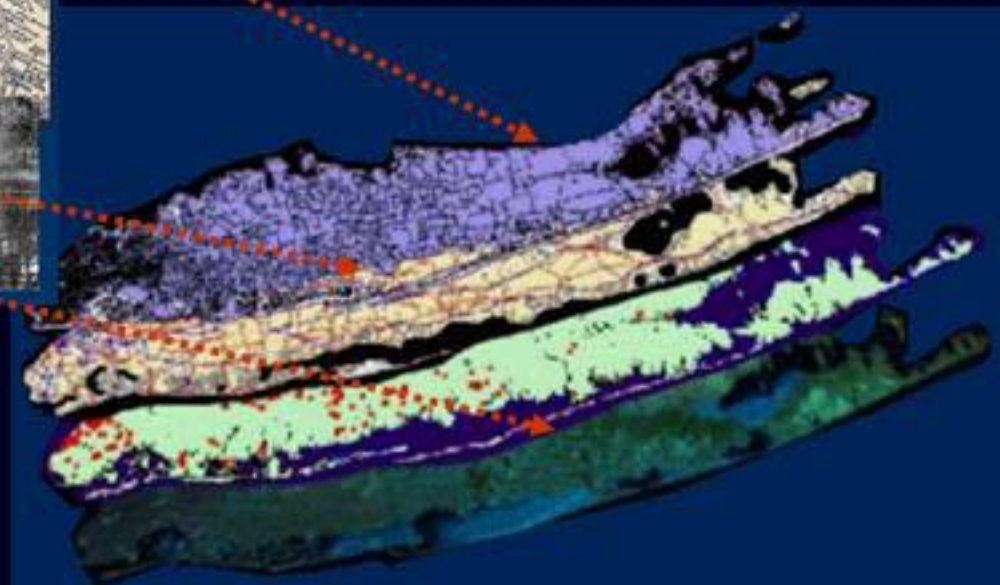
The northeastern United States has had high rates of breast cancer.

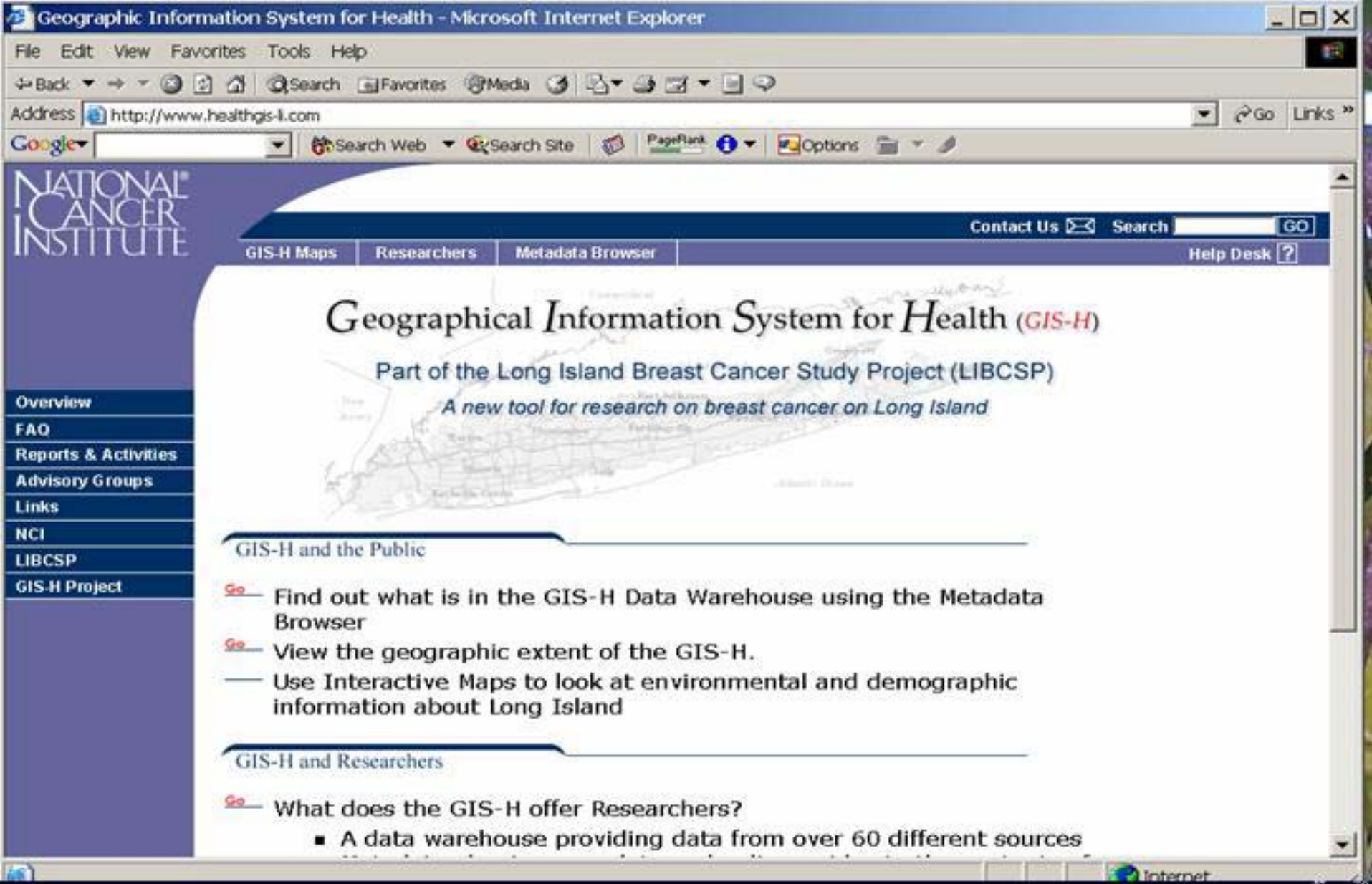
The Long Island Breast Cancer Study Project (LIBCSP) focuses on Long Island (Nassau and Suffolk counties) in New York.

A Tool for Studying Environment & Breast Cancer

NATIONAL CANCER INSTITUTE
Division of Cancer Control & Population Sciences





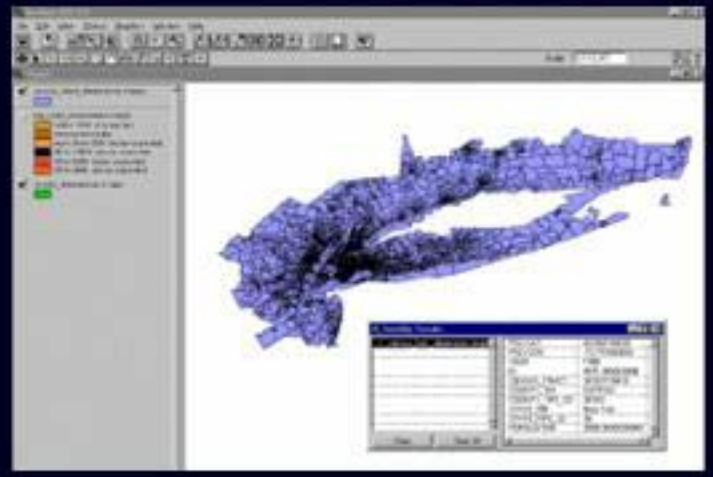


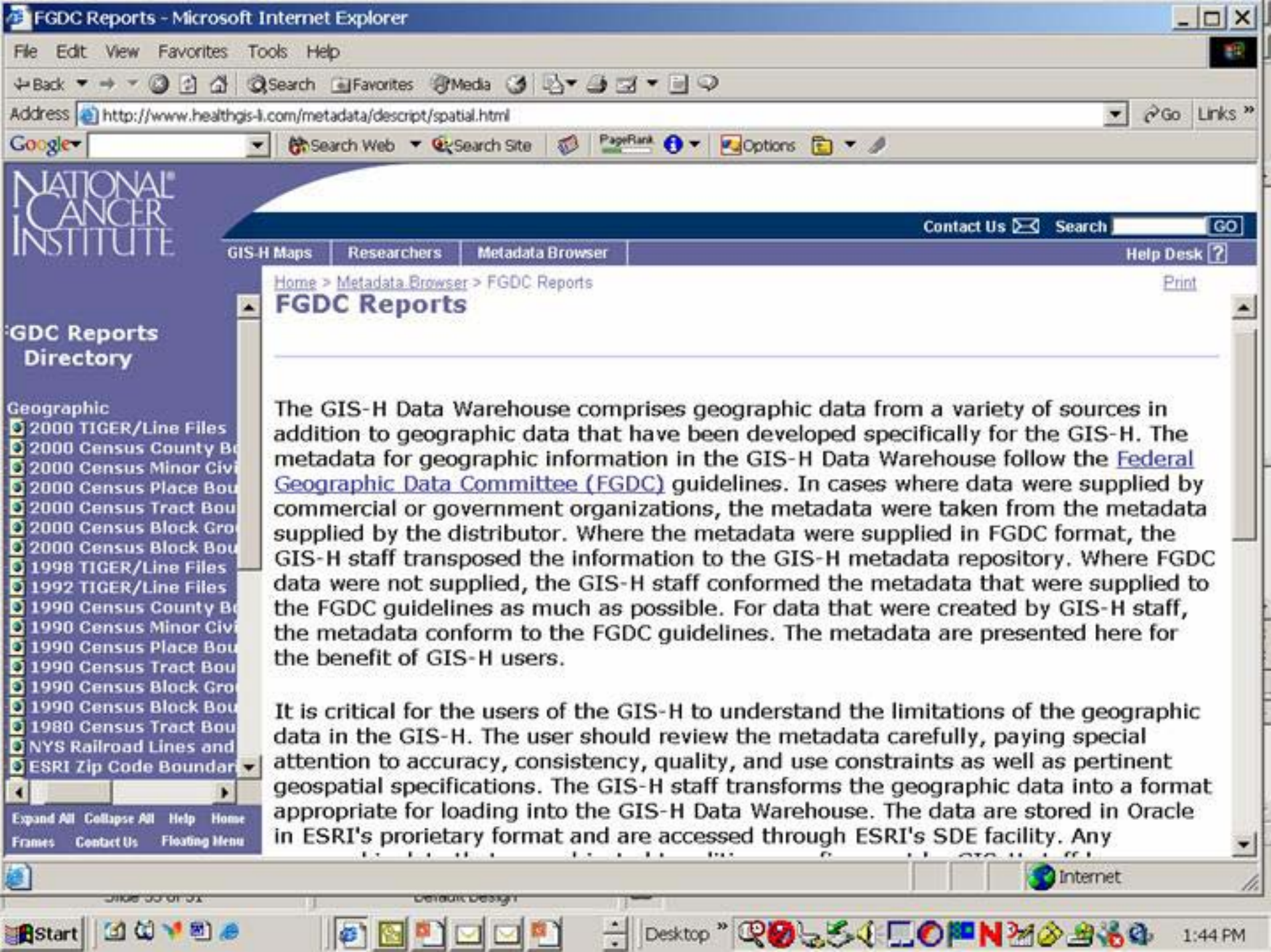
www.healthgis-li.com

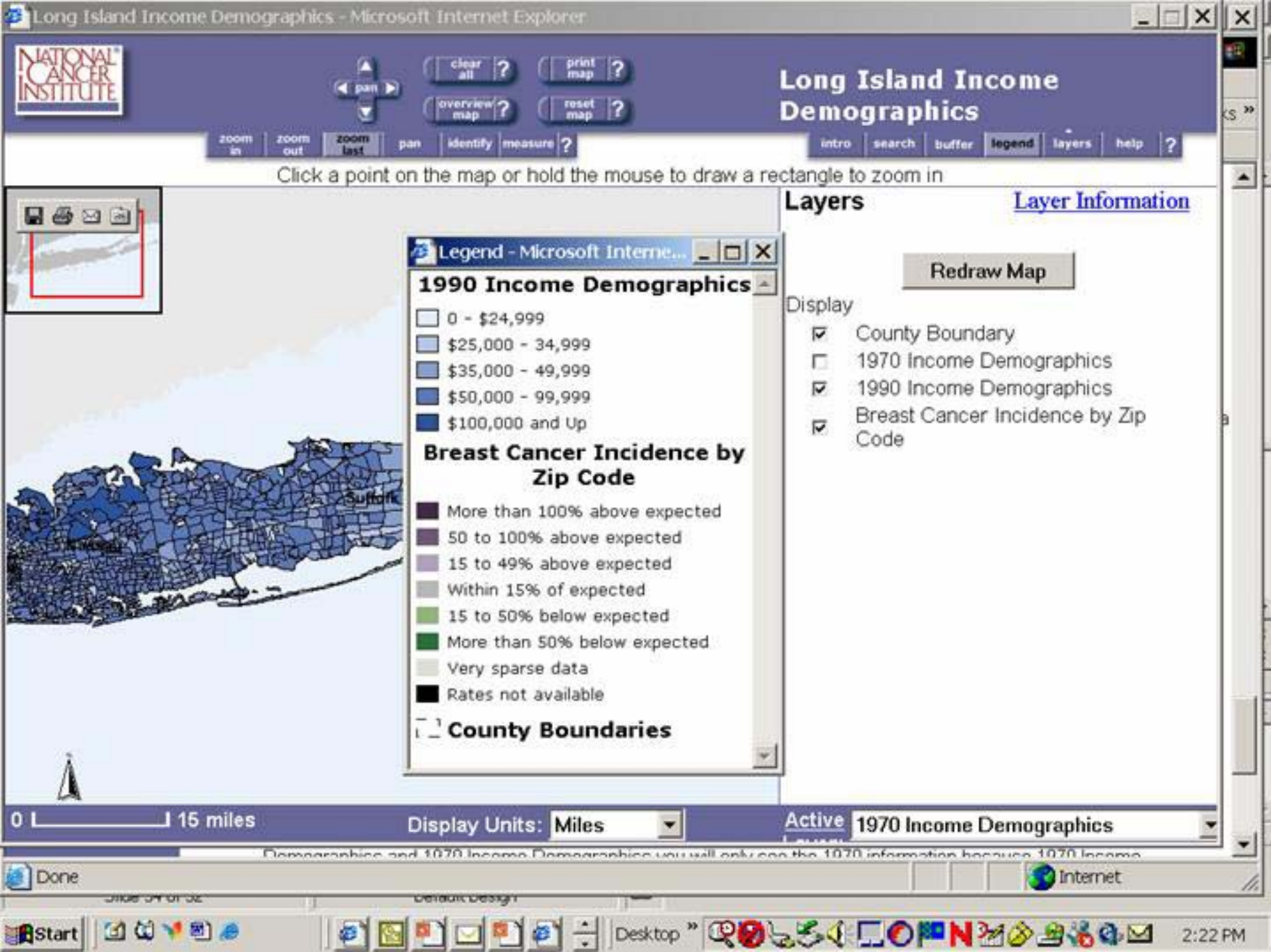
Data Included in the GIS-H

NATIONAL CANCER INSTITUTE
Division of Cancer Control & Population Sciences

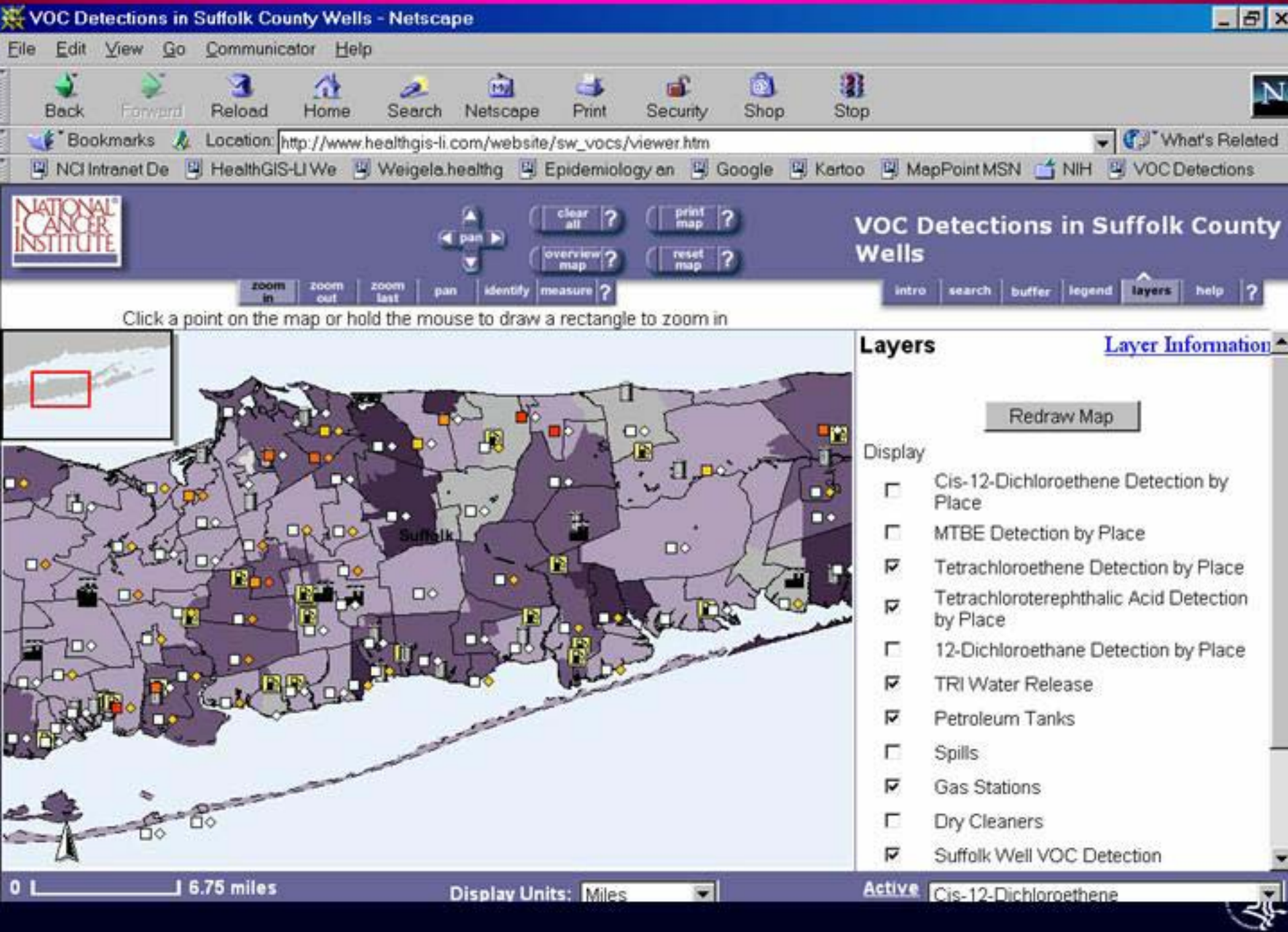
- Geospatial
- Demographic and Behavioral
- Health
- Environmental



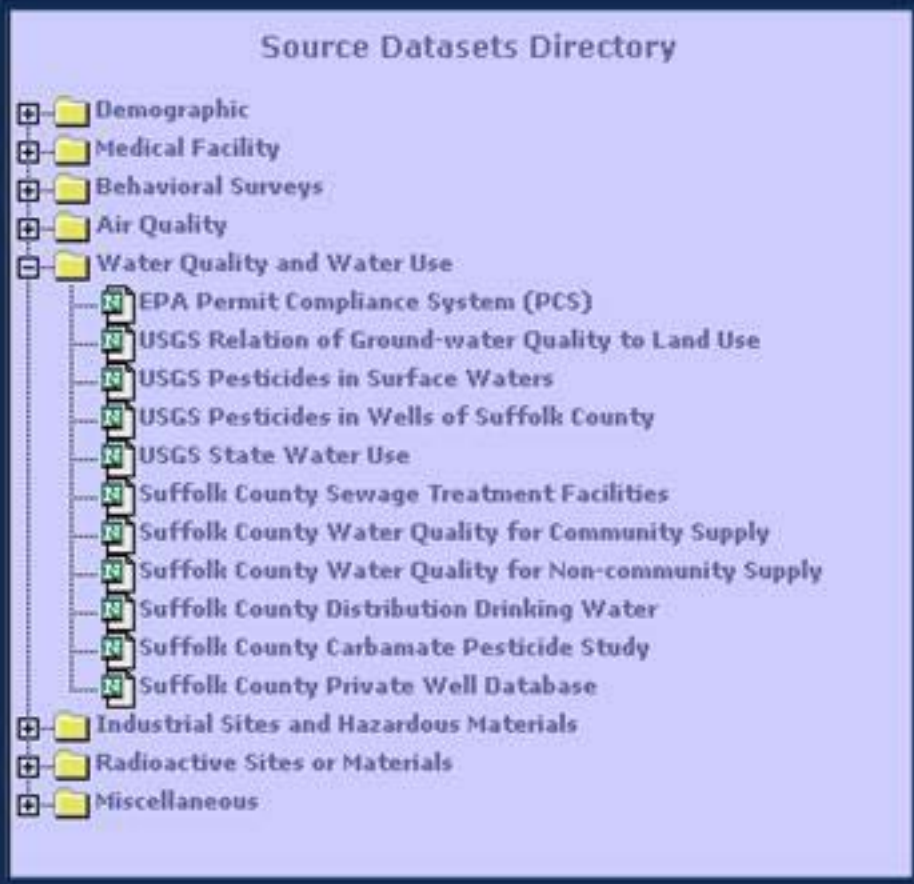

FGDC Reports - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.healthgis-li.com/metadata/descript/spatial.html
Go Links
Google Search Web Search Site PageRank Options
NATIONAL CANCER INSTITUTE
Contact Us Search GO
GIS-H Maps Researchers Metadata Browser Help Desk
Home > Metadata Browser > FGDC Reports
Print
FGDC Reports
GDC Reports Directory
Geographic
2000 TIGER/Line Files
2000 Census County Bo
2000 Census Minor Civi
2000 Census Place Bou
2000 Census Tract Bou
2000 Census Block Gro
2000 Census Block Bou
1998 TIGER/Line Files
1992 TIGER/Line Files
1990 Census County Bo
1990 Census Minor Civi
1990 Census Place Bou
1990 Census Tract Bou
1990 Census Block Gro
1990 Census Block Bou
1980 Census Tract Bou
NYS Railroad Lines and
ESRI Zip Code Boundar
Expand All Collapse All Help Home
Frames Contact Us Floating Menu
The GIS-H Data Warehouse comprises geographic data from a variety of sources in addition to geographic data that have been developed specifically for the GIS-H. The metadata for geographic information in the GIS-H Data Warehouse follow the Federal Geographic Data Committee (FGDC) guidelines. In cases where data were supplied by commercial or government organizations, the metadata were taken from the metadata supplied by the distributor. Where the metadata were supplied in FGDC format, the GIS-H staff transposed the information to the GIS-H metadata repository. Where FGDC data were not supplied, the GIS-H staff conformed the metadata that were supplied to the FGDC guidelines as much as possible. For data that were created by GIS-H staff, the metadata conform to the FGDC guidelines. The metadata are presented here for the benefit of GIS-H users.
It is critical for the users of the GIS-H to understand the limitations of the geographic data in the GIS-H. The user should review the metadata carefully, paying special attention to accuracy, consistency, quality, and use constraints as well as pertinent geospatial specifications. The GIS-H staff transforms the geographic data into a format appropriate for loading into the GIS-H Data Warehouse. The data are stored in Oracle in ESRI's prorietary format and are accessed through ESRI's SDE facility. Any
Internet
Start
Desktop
1:44 PM


Long Island Income Demographics - Microsoft Internet Explorer
NATIONAL CANCER INSTITUTE
pan
clear all
print map
overview map
reset map
Long Island Income Demographics
zoom in
zoom out
zoom last
pan
identify
measure
intro
search
buffer
legend
layers
help
Click a point on the map or hold the mouse to draw a rectangle to zoom in
Legend - Microsoft Interne...
1990 Income Demographics
0 - $24,999
$25,000 - 34,999
$35,000 - 49,999
$50,000 - 99,999
$100,000 and Up
Breast Cancer Incidence by Zip Code
More than 100% above expected
50 to 100% above expected
15 to 49% above expected
Within 15% of expected
15 to 50% below expected
More than 50% below expected
Very sparse data
Rates not available
County Boundaries
Suffolk
Layers
Layer Information
Redraw Map
Display
County Boundary
1970 Income Demographics
1990 Income Demographics
Breast Cancer Incidence by Zip Code
0
15 miles
Display Units:
Miles
Active
1970 Income Demographics
Done
Internet
Start
Desktop
2:22 PM


Long Island Health Facilities - Microsoft Internet Explorer
NATIONAL CANCER INSTITUTE
pan
clear all
print map
overview map
reset map
Long Island Health Facilities
Find an address
zoom in
zoom out
zoom last
pan
identify
measure
intro
search
buffer
legend
layers
help
Click a point on the map or hold the mouse to draw a rectangle to zoom in
Suffolk
Layers
Layer Information
Redraw Map
Display
Health Facilities
Mammography Clinics
Zip Code Boundaries
County Boundary
Breast Cancer Incidence by Zip Code
0
7 miles
Display Units:
Miles
Active
Health Facilities
Other ...
385 Items, 227 Unread


VOC Detections in Suffolk County Wells - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://www.healthgis-li.com/website/sw_vocs/viewer.htm
What's Related
NCI Intranet De HealthGIS-LI We Weigela.healthg Epidemiology en Google Kartoo MapPoint MSN NIH VOC Detections
NATIONAL CANCER INSTITUTE
pan
clear all
print map
overview map
reset map
VOC Detections in Suffolk County Wells
zoom in zoom out zoom last pan identify measure
intro search buffer legend layers help
Click a point on the map or hold the mouse to draw a rectangle to zoom in
Suffolk
Layers
Layer Information
Redraw Map
Display
Cis-12-Dichloroethene Detection by Place
MTBE Detection by Place
Tetrachloroethene Detection by Place
Tetrachloroterephthalic Acid Detection by Place
12-Dichloroethane Detection by Place
TRI Water Release
Petroleum Tanks
Spills
Gas Stations
Dry Cleaners
Suffolk Well VOC Detection
0
6.75 miles
Display Units: Miles
Active
Cis-12-Dichloroethene

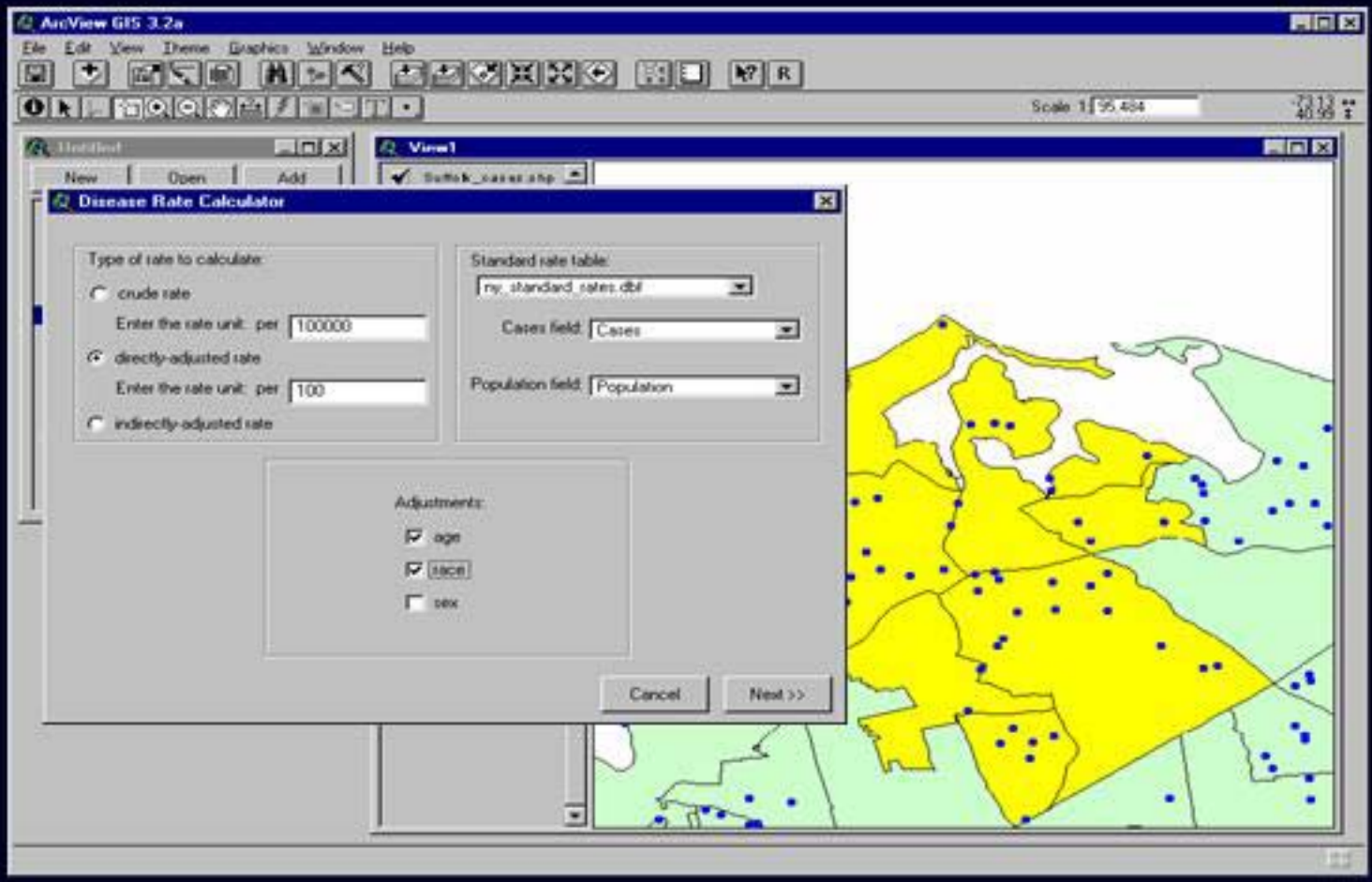
Metadata Browser



Disease Rate Calculator

Division of Cancer Control & Population Sciences

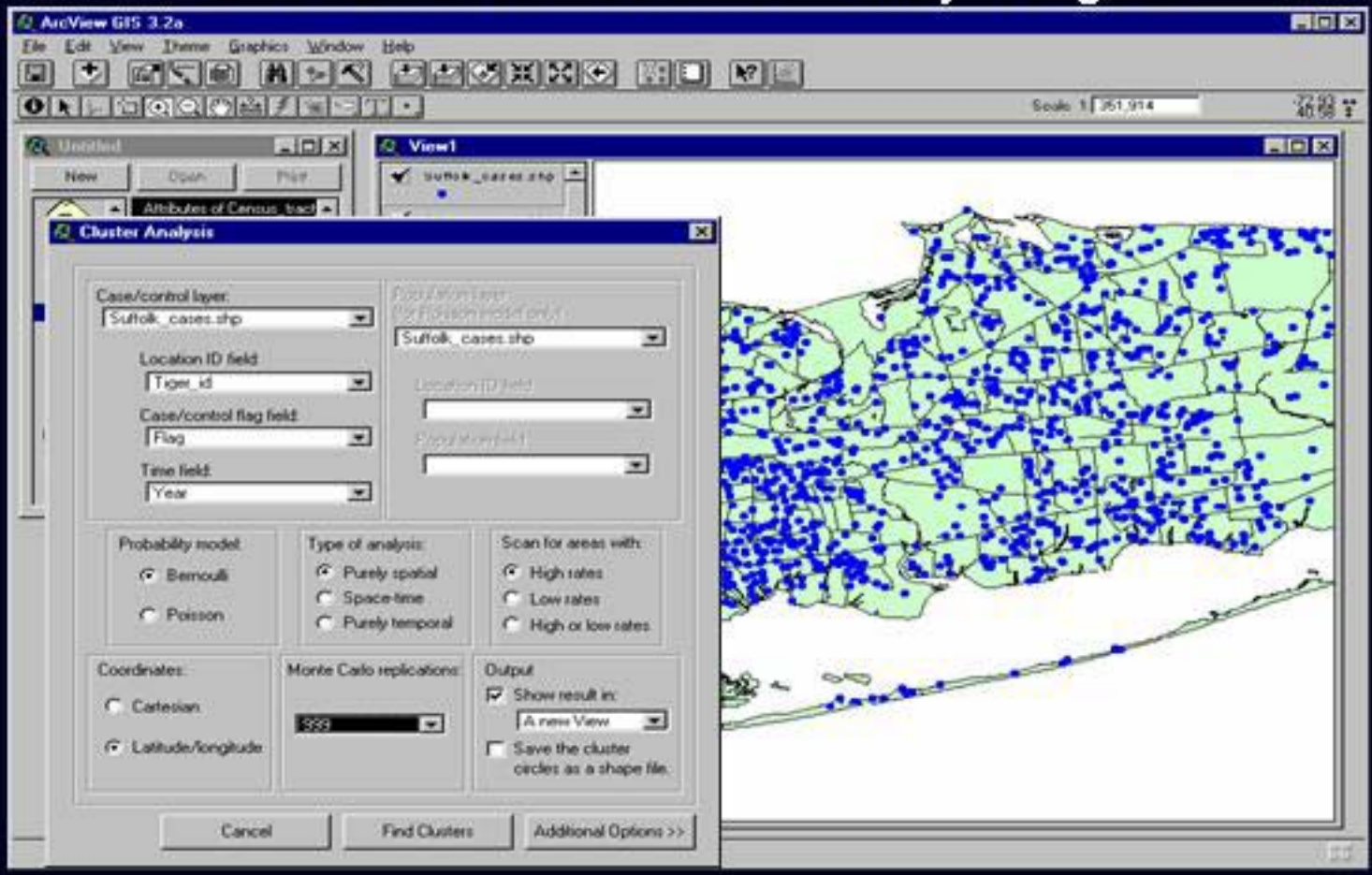
Calculating directly-adjusted rate for selected census tracts.



Cluster Analysis

Division of Cancer Control & Population Sciences

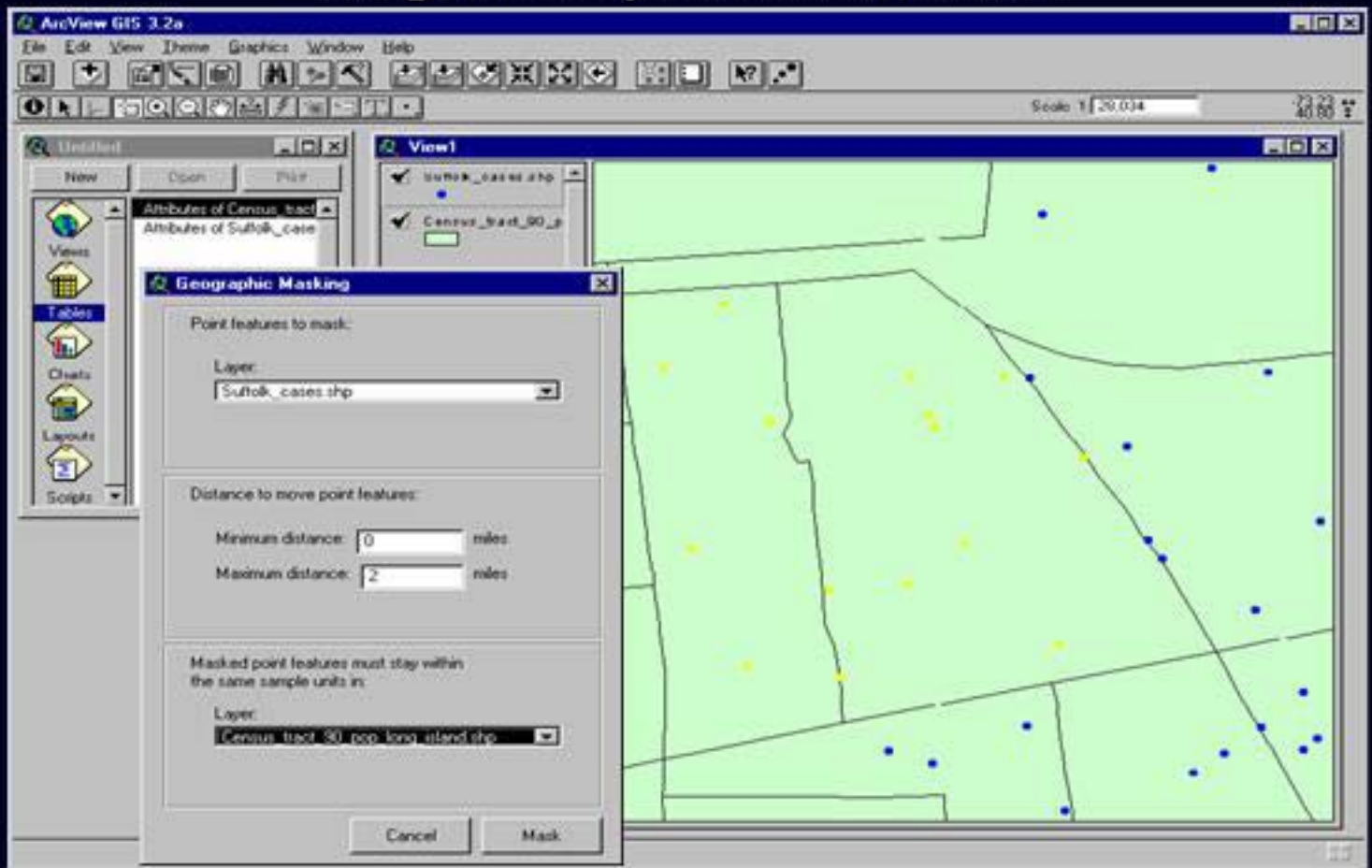
Checking for clusters of sample cases
uses SatScan software as cluster analysis engine.

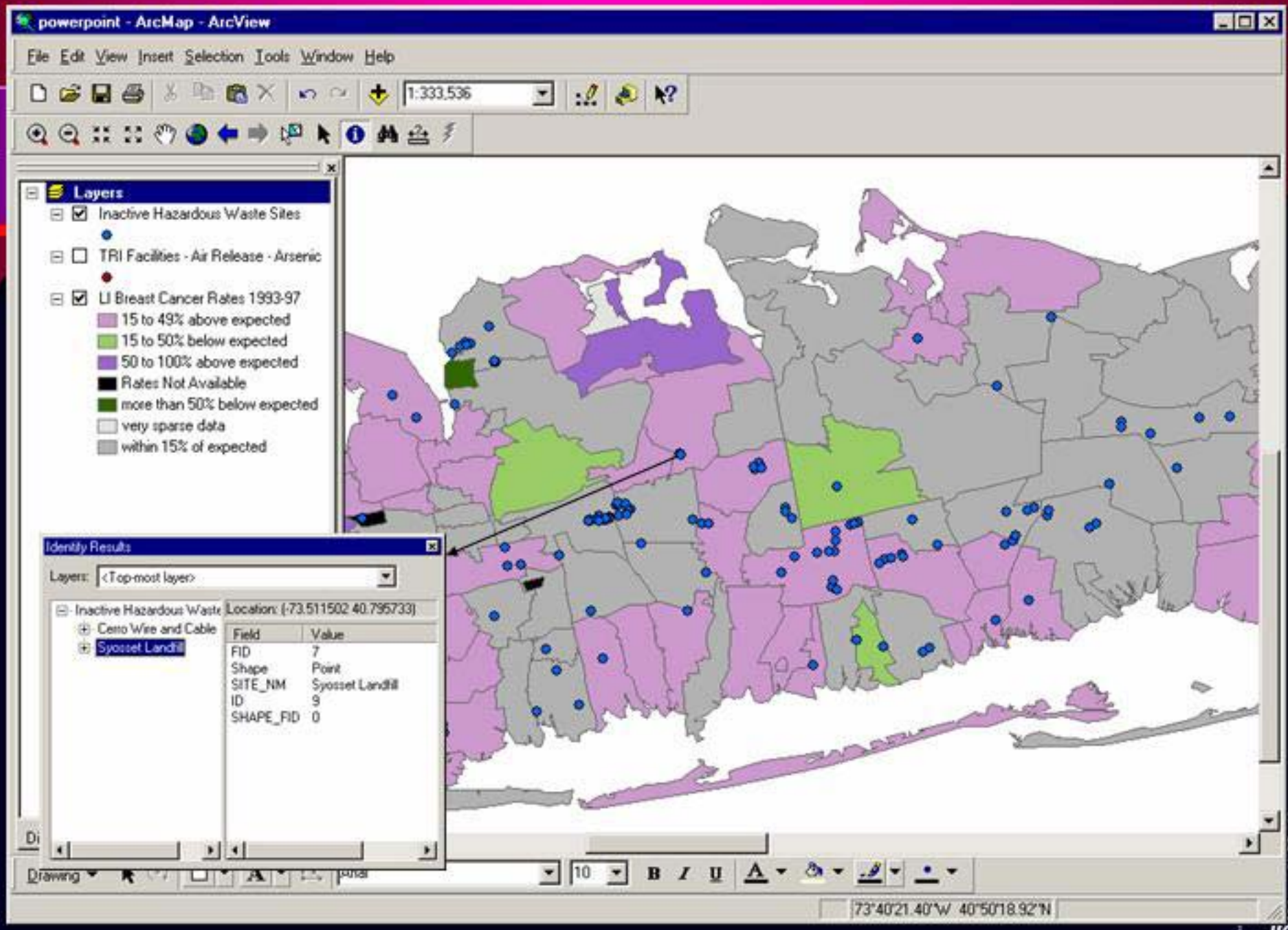


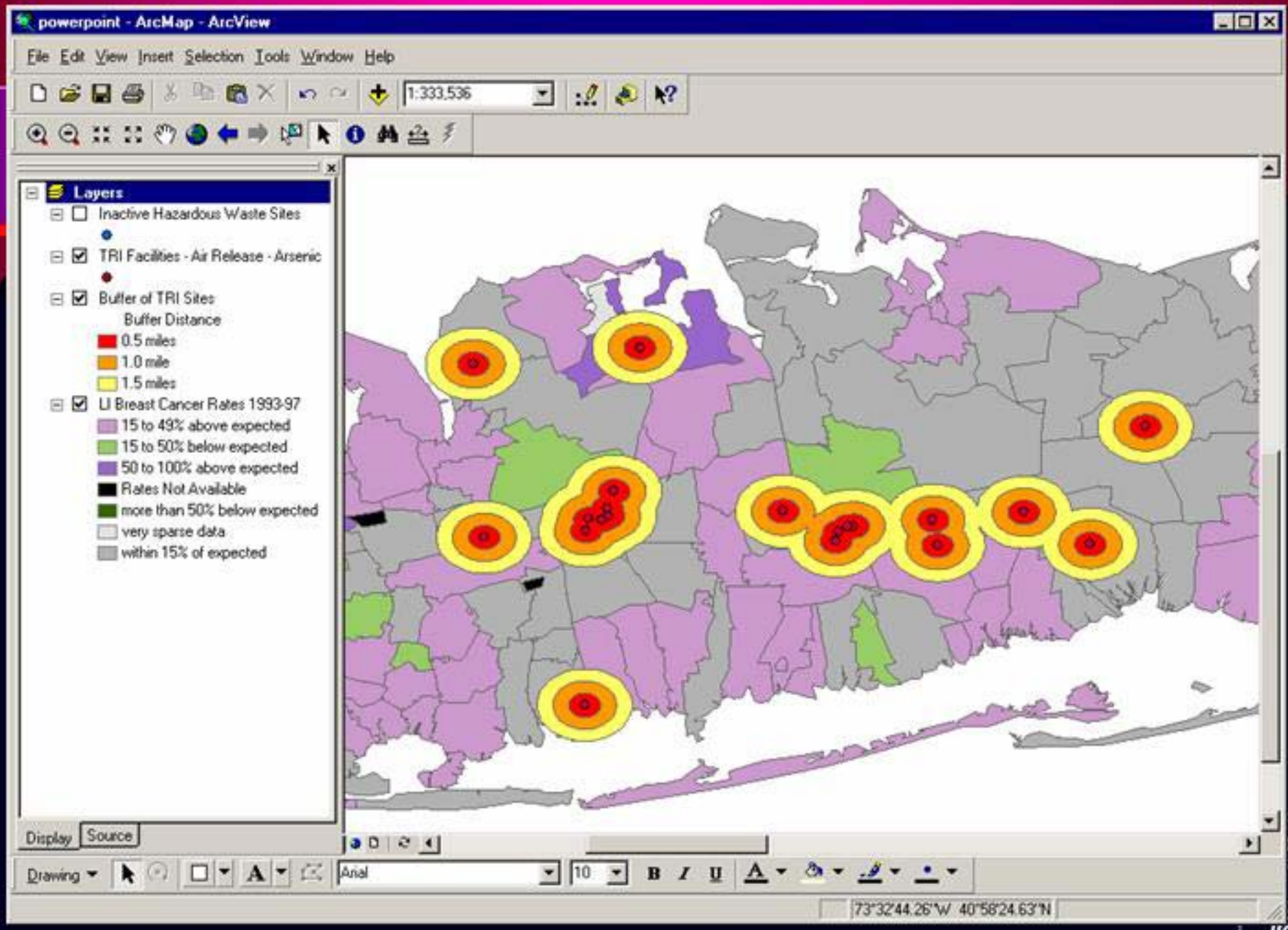
Geographic Masking

National Cancer Institute
Division of Cancer Control & Population Sciences

Masking selected (in yellow) sample cases using random perturbation method.




powerpoint - ArcMap - ArcView
File Edit View Insert Selection Tools Window Help
1:333,536
Layers
Inactive Hazardous Waste Sites
TRI Facilities - Air Release - Arsenic
LI Breast Cancer Rates 1993-97
15 to 49% above expected
15 to 50% below expected
50 to 100% above expected
Rates Not Available
more than 50% below expected
very sparse data
within 15% of expected
Identity Results
Layers: <Top-most layer>
Inactive Hazardous Waste
Location: (-73.511502 40.795733)
Cerro Wire and Cable
Syosset Landfill
Field Value
FID 7
Shape Point
SITE_NM Syosset Landfill
ID 9
SHAPE_FID 0
Drawing
10
73°40'21.40"W 40°50'18.92"N


powerpoint - ArcMap - ArcView
File Edit View Insert Selection Tools Window Help
1:333,536
Layers
Inactive Hazardous Waste Sites
TRI Facilities - Air Release - Arsenic
Buffer of TRI Sites
Buffer Distance
0.5 miles
1.0 mile
1.5 miles
LI Breast Cancer Rates 1993-97
15 to 49% above expected
15 to 50% below expected
50 to 100% above expected
Rates Not Available
more than 50% below expected
very sparse data
within 15% of expected
Display
Source
Drawing
Arial
10
73°32'44.26"W 40°58'24.63"N

NATIONAL CANCER INSTITUTE

Division of Cancer Control & Population Sciences

What is Availability of GIS-H?

- Available now to researchers with approved projects
- Public mapping features available soon

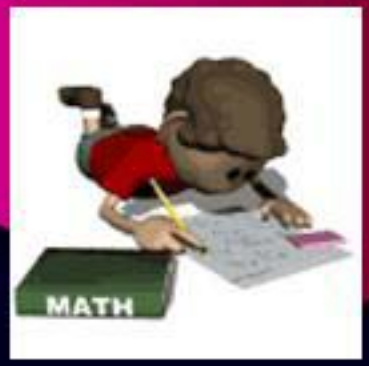
National Cancer Institute

Division of Cancer Control & Population Sciences

In Summary, the GIS-H is…

- Comprehensive, integrated data warehouse (> 80 datasets)
- Flexible and expandable
- Can integrate external datasets
- Sophisticated researcher's toolbox
- Community input and access
- Systematic attempt to include high quality data, comprehensive metadata
- A prototype and resource for future studies

Apply!

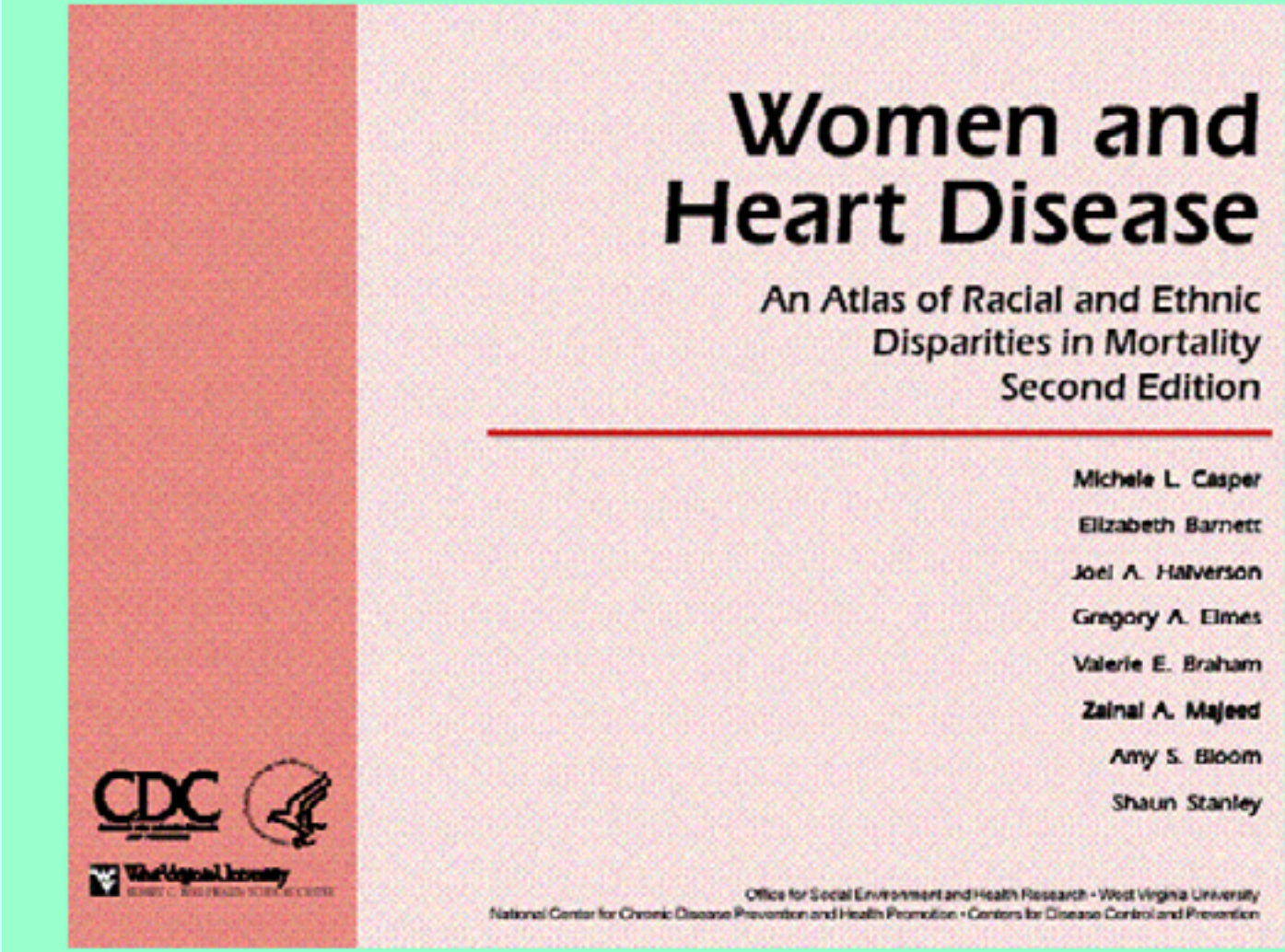


- Access to researcher site is limited to investigators with approved protocols
- For additional information, visit GIS-H website

www.healthgis-li.com

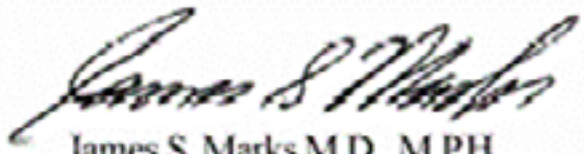
Inquiries may be directed to:

- Burdette (Bud) Erickson, M.Sc.
- 301.435.4913
- berikso@mail.nih.gov


Women and Heart Disease
An Atlas of Racial and Ethnic Disparities in Mortality
Second Edition
Michele L. Casper
Elizabeth Barnett
Joel A. Halverson
Gregory A. Elmes
Valerie E. Braham
Zainal A. Majeed
Amy S. Bloom
Shaun Stanley
CDC
Office for Social Environment and Health Research • West Virginia University
National Center for Chronic Disease Prevention and Health Promotion • Centers for Disease Control and Prevention

"*Men and Heart Disease* indicates where those programs are most needed and can have the greatest benefit.

It is my hope that *Men and Heart Disease: An Atlas of Racial and Ethnic Disparities in Mortality* will be used to guide the distribution of funds and resources to those communities of men experiencing excess mortality from heart disease and will promote the development of culturally sensitive prevention strategies."



James S. Marks M.D., M.P.H.
Director, National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention

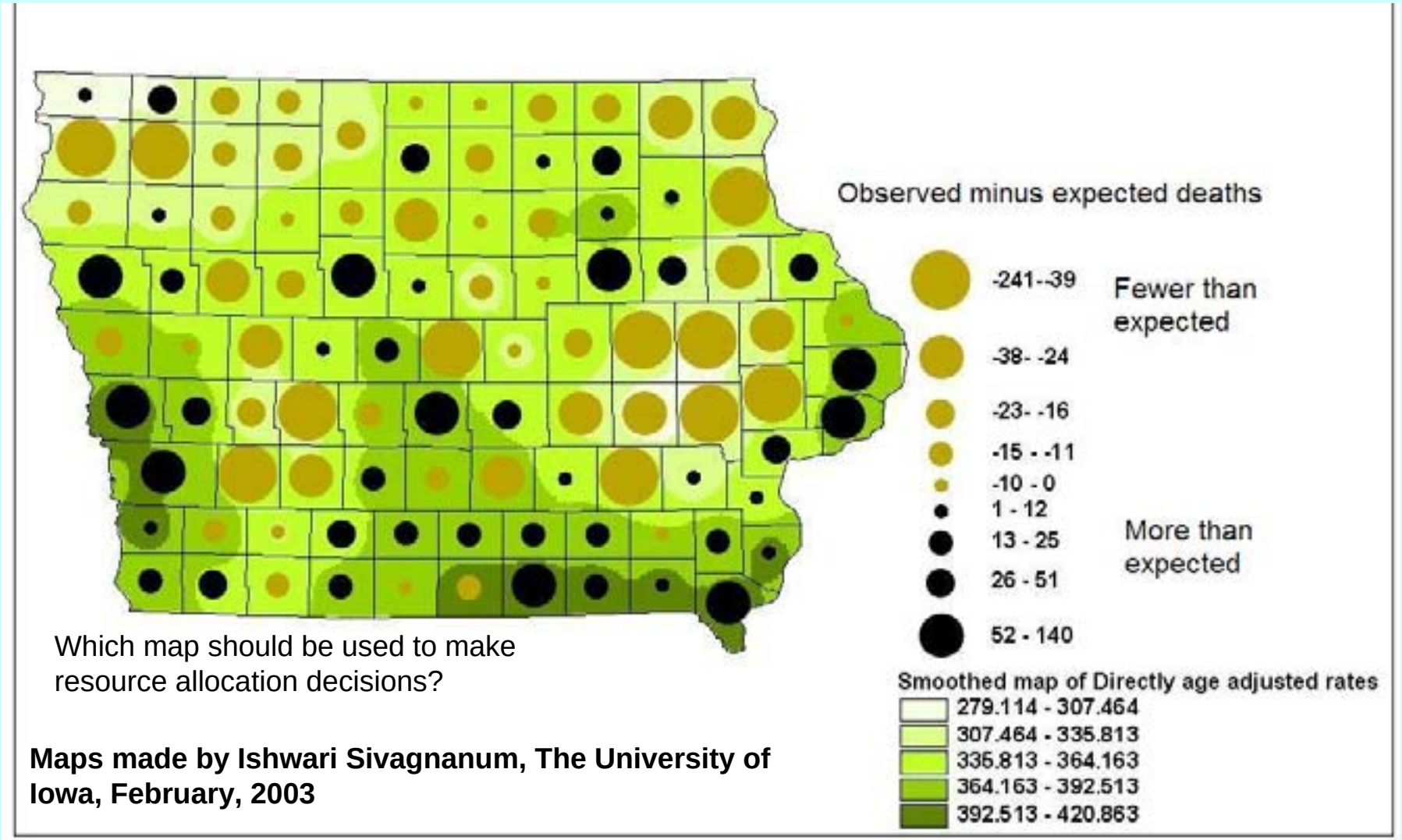
In *Direct* Adjustment:

- Locally measured disease rates are applied to a standard population
- "The formula calls for computing the weighted average of the age-specific death rates in a given area, using as weights the age distribution of the standard population." M&M Demography p. 419
- This answers the question: given the age-sex specific disease rates of this area, how many more (or less) incidences would occur in this area if it had the demographic characteristics of the standard population.

In *indirect* adjustment

- In indirect adjustment, national or regional rates are applied to the demographics of each small area.
- The ratio of actual/expected incidences is the standardized disease incidence rate of the area.
- This answers the question, "given the demographics of this area, how many incidences of the disease would we expect to find?"

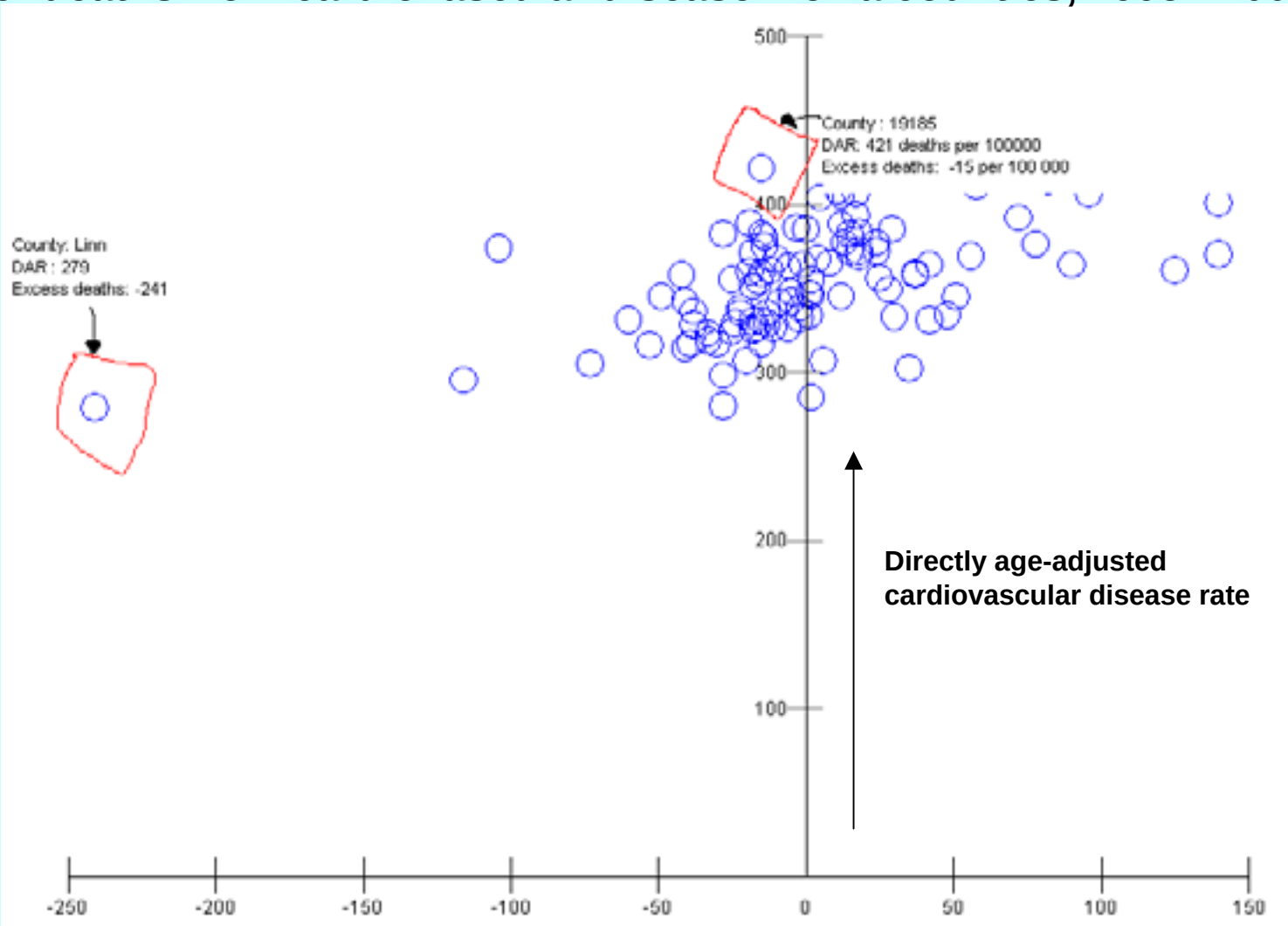
The Green Map is the Cardiovascular Death Rate in Iowa, 1995 - 2000.
(Directly Adjusted Rates--smoothed)
Circles show the difference between the observed and the expected number of deaths (from Indirectly Age Adjusted Rates). Black is excess deaths—greater than expected; Brown is fewer than expected number of deaths.



Which map should be used to make resource allocation decisions?

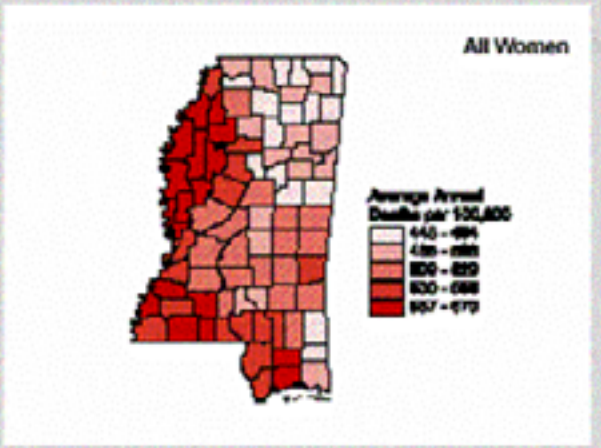
Maps made by Ishwari Sivagnanum, The University of Iowa, February, 2003

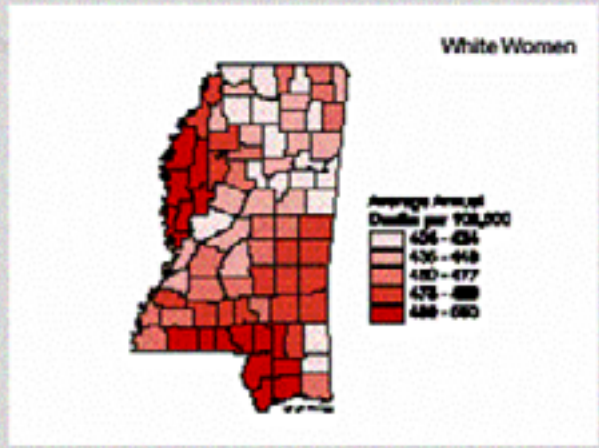
Relationship Between Direct Age-Adjusted Cardiovascular Disease Rates and the difference between observed and expected numbers of deaths from cardiovascular disease. Iowa counties, 1995 - 2000



Mississippi

Smoothed County Heart Disease Death Rates, 1991-1995

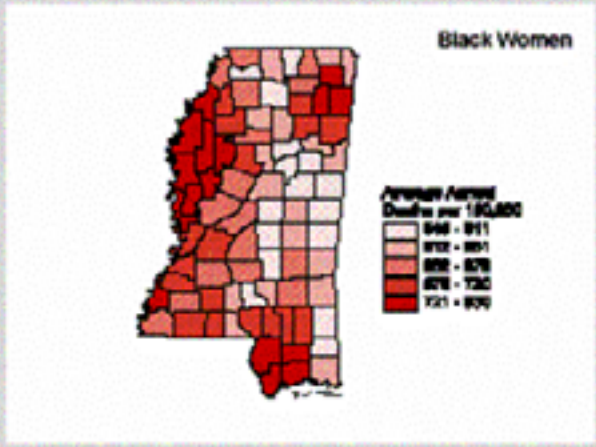




State Profile — Mississippi

Race or Ethnicity	State Population 1995	State Heart Disease Death Rate, 1991-1995*
All Women	690,720	531
American Indian and Alaska Native Women	1,720	337
Asian and Pacific Islander Women	3,434	Insufficient Data
Black Women	206,720	686
Hispanic Women	3,680	Insufficient Data
White Women	460,836	472

* Average annual age-adjusted rate (deaths per 100,000) for women ages 35 years and older. Data for Hispanics are also included within each of the four categories of race.



Indirectly adjusted rates can be mapped for smaller spatial units than directly adjusted rates

- the indirect rate involves multiplying national or statewide rates by local age-sex demographics;
- direct rates involve multiplying local age-sex rates by national or statewide demographics.
- the indirect rate involves multiplying stable disease rates by stable population totals;
- the direct rate involves multiplying unstable disease rates by stable population totals.

Alternative Data Models for GIS and Health

It is often necessary to work simultaneously with all three data models