DataSpeak

Clusters, maps, and hotspots: Small area analysis in maternal and child health

August 12, 2015

Today's Presenters

- **Russell S. Kirby, PhD, MS, FACE,** is the Distinguished University Professor and Marrell Endowed Chair, Department of Community and Family Health at the University of South Florida. Dr. Kirby will set the stage by reviewing basic principles of mapping and small area analysis, using examples from the literature, and describing the features of geographic information systems (GIS).

- **Michael Kramer, PhD, MMSc.,** is an Assistant Professor of Epidemiology at the Rollins School of Public Health at Emory University. Dr. Kramer will discuss the motivation for and several approaches to the production of statistically robust small area estimates of disease rates or other health relevant parameters. These methods are useful for mapping small area variation in disease occurrence or as inputs for subsequent spatial analysis.

- **Thomas J. Stopka, PhD, MHS**, is an Assistant Professor with the Department of Public Health and Community Medicine at the Tufts University School of Medicine. In his presentation, Dr. Stopka will describe the use of spatial epidemiological methods that can be employed to identify and characterize hotspot clusters of unmet needs related to nutrition supplementation and public health services.

Previous Events

DataSpeak Archives

2015 Series:

- Vitally Important: Improving the Timeliness of Vital Statistics to Advance MCH

2014 Series:

- Effects of the Built Environment on Maternal and Child Health

2013 Series:

- Measuring the Return on Investment in Maternal and Child Health Programs
- Findings from the 2011-2012 National Survey of Children's Health

DataSpeak:
The Case for Small Area Analysis in Maternal and Child Health

Russell S. Kirby, PhD, MS, FACE

Marrell Endowed Chair and Distinguished University Professor

Department of Community and Family Health

College of Public Health, University of South Florida

Overview

This presentation will provide:

1) Brief review of basic principles of mapping and its uses for studying spatial aspects of health phenomena

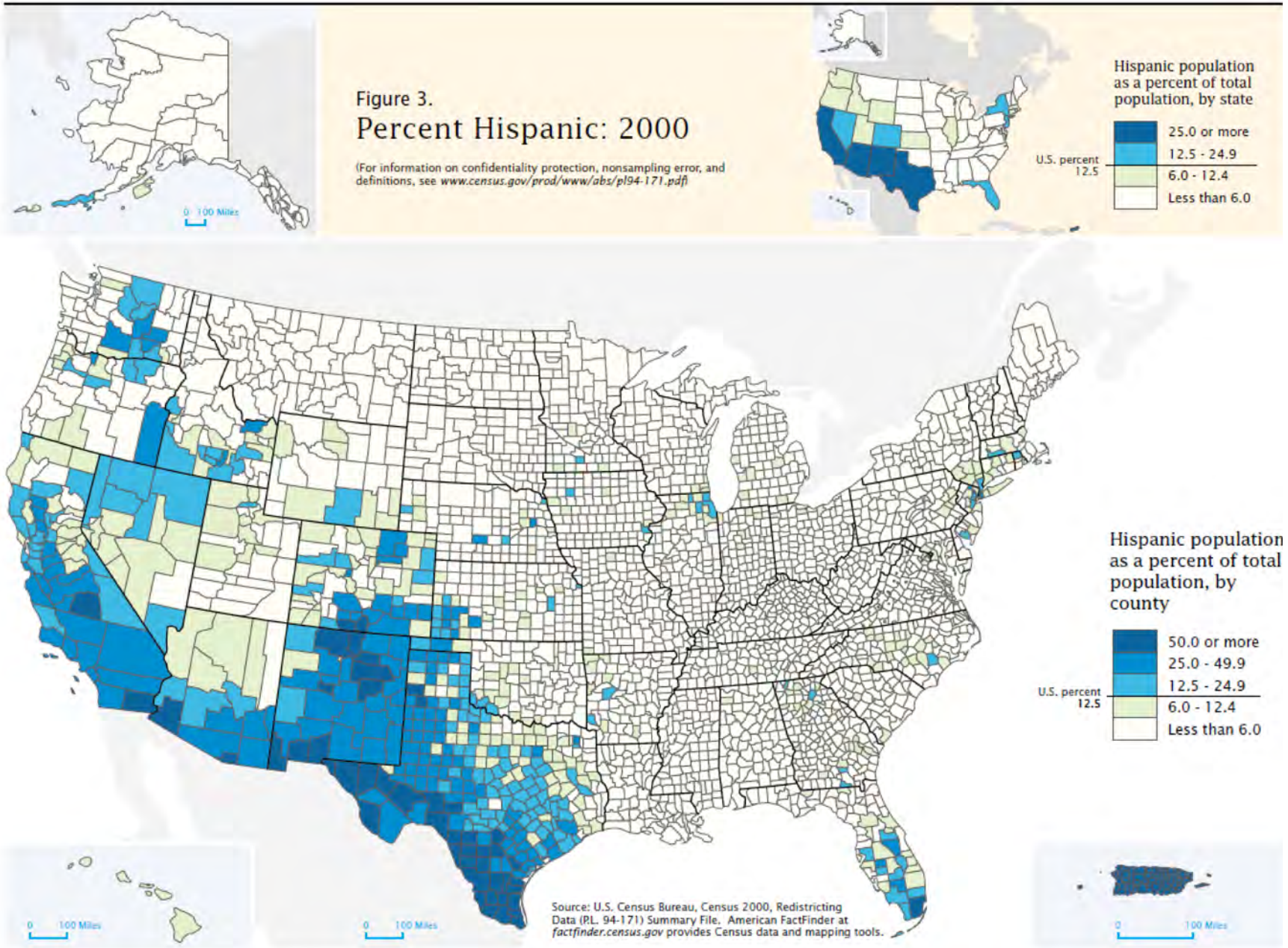
2) Introduction to small area analysis

Graphics for Visual Display of Data

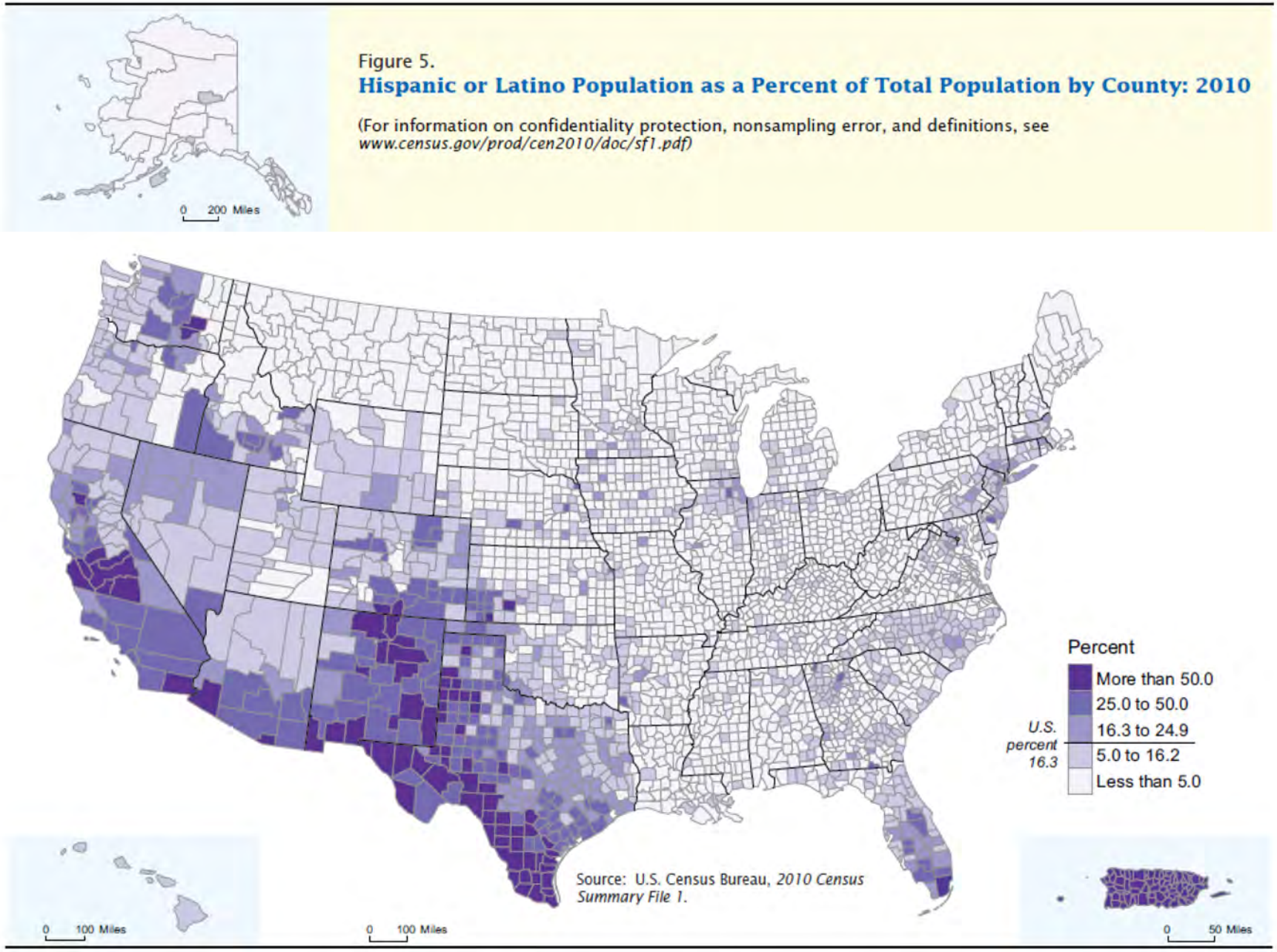
- Many of you many be familiar with the work of Edward Tufte *(The Visual Display of Quantitative Information, 1984).*
- More broadly, Bertin in his *The Semiology of Graphics (1967)* provided a framework for thinking about structures for displaying statistical data.
- More recently Wilkinson *(The Grammar of Graphics, 2nd edition 2005)* provides a language and syntax for creating virtually any kind of data graphic based on understanding the structure of the underlying data.

Pattern versus Process

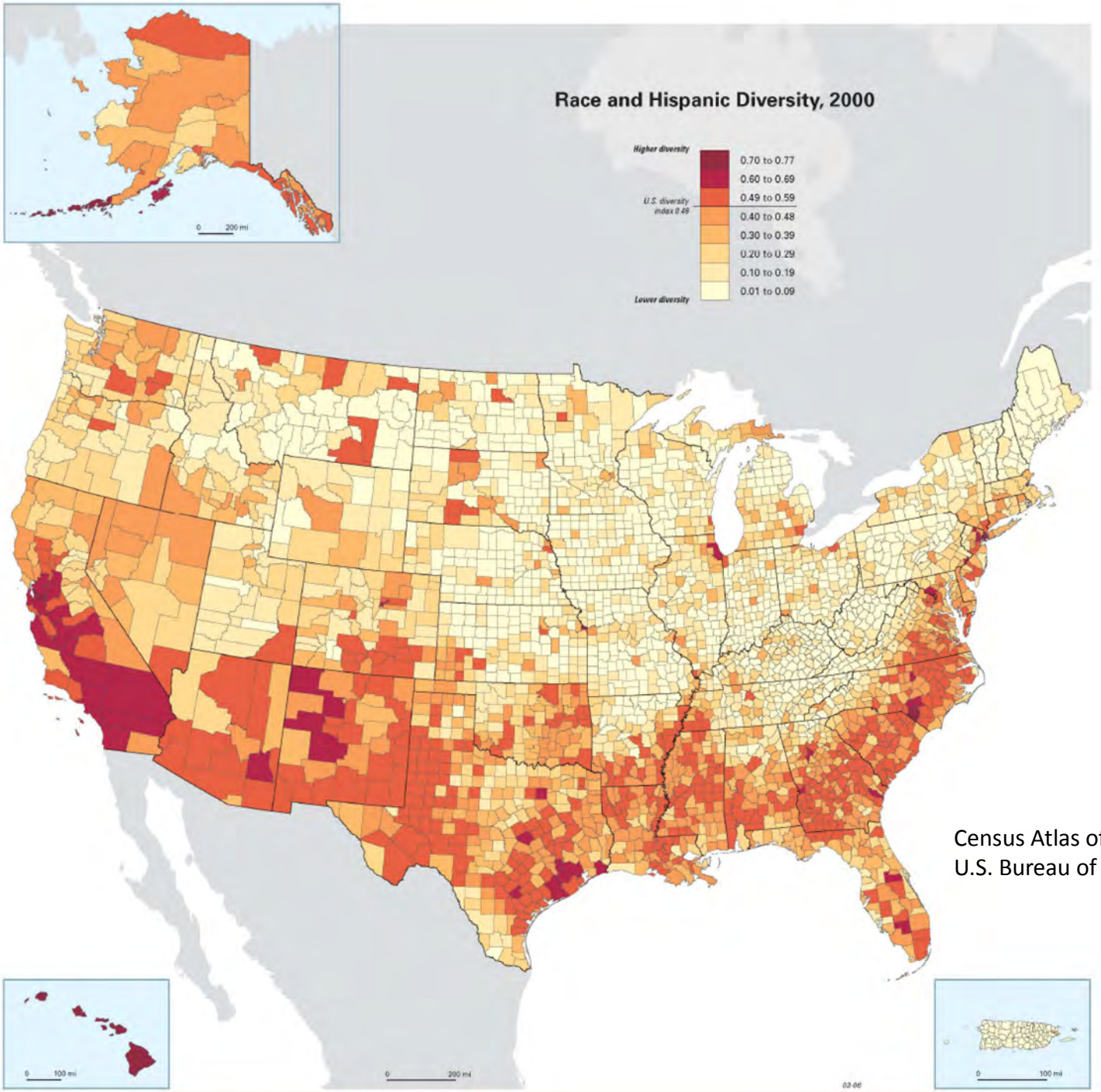
- An infinite number of maps can be created from any given set of data tagged with some form of geocode.
  - The term 'geocode' refers to an identifier that locates a record in a dataset by state, county, ZIP code, census tract, x,y coordinate pair.
- Most maps are presented to display a single variable, typically as a choropleth map.
- Frequently overlooked are opportunities to use maps as vehicles to understanding the processes that generated the spatial distributions shown.
- Let's illustrate some issues with a brief look at two series of maps displaying data concerning the distribution of the Hispanic population of the United States and on the crude birth rates of US states.



Census 2000 Brief: The Hispanic Population. (May 2001)



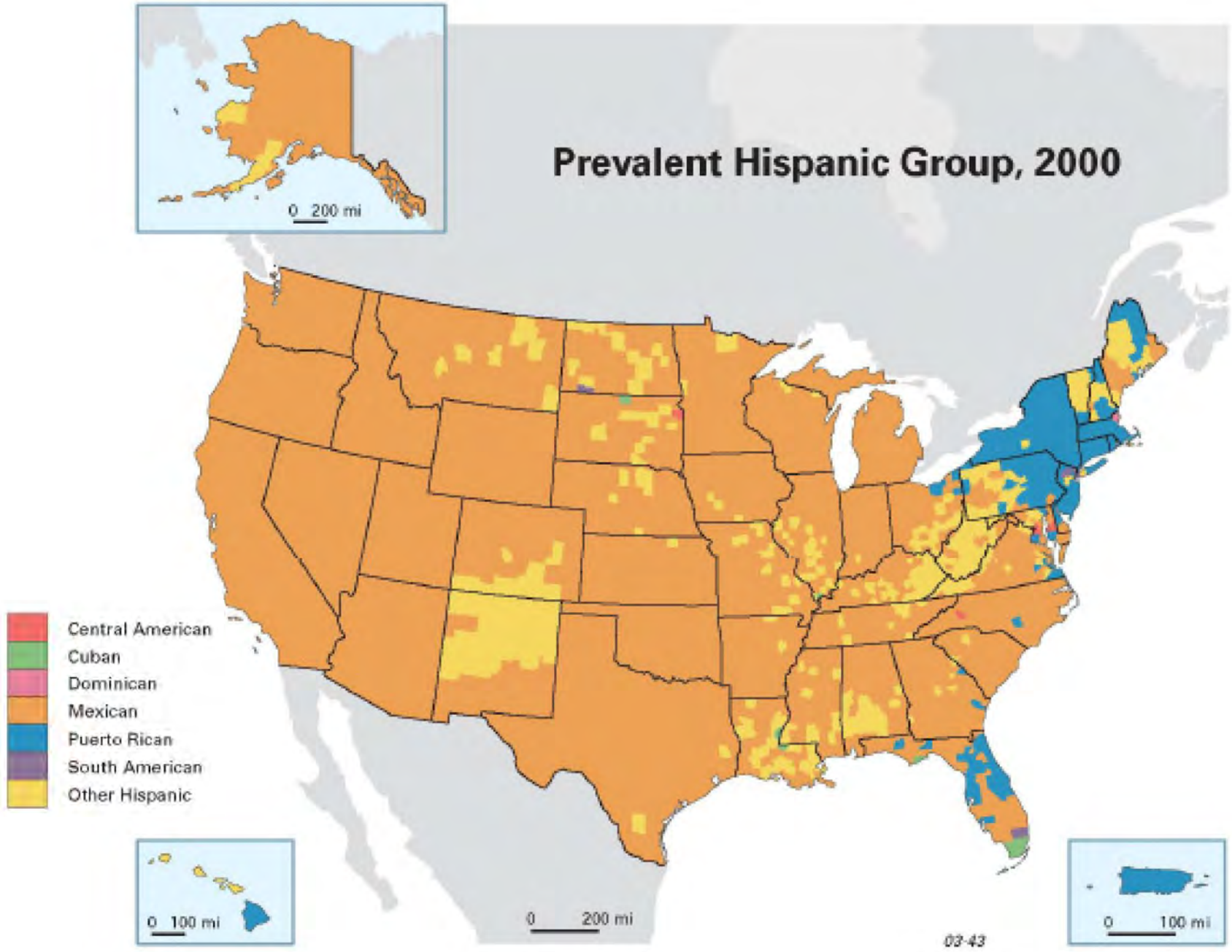
Census 2010 Brief: The Hispanic Population. (May 2011)

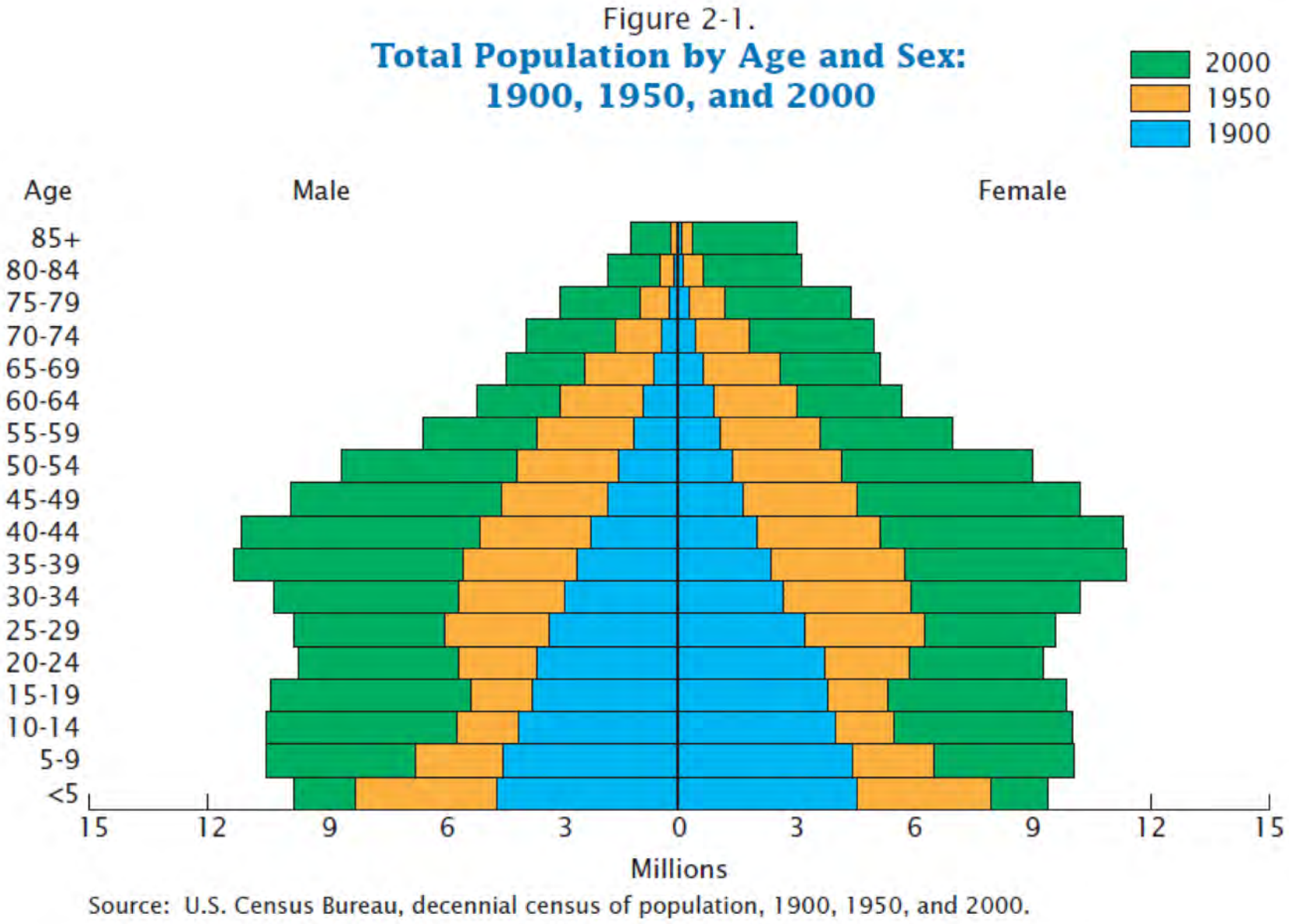


Census Atlas of the United States,
U.S. Bureau of the Census, 2007

Diversity Index

- The probability that two persons selected at random from a county population would be of different races, or that only one of the two would be Hispanic.


Prevalent Hispanic Group, 2000
0 200 mi
Central American
Cuban
Dominican
Mexican
Puerto Rican
South American
Other Hispanic
0 100 mi
0 200 mi
0 100 mi
03-43



Demographic Trends in the 20th Century. U.S. Bureau of the Census, 2002.

How to Lie with Maps

- Cartographer Mark Monmonier published a book entitled *How to Lie with Maps* in the early 1990s, based on his observations of map use in the media and elsewhere.
- This work built on the classic book by Huff, *How to Lie with Statistics*, published in the 1950s and still in print.
- Let's take a look at some of Monmonier's more recent observations . . .

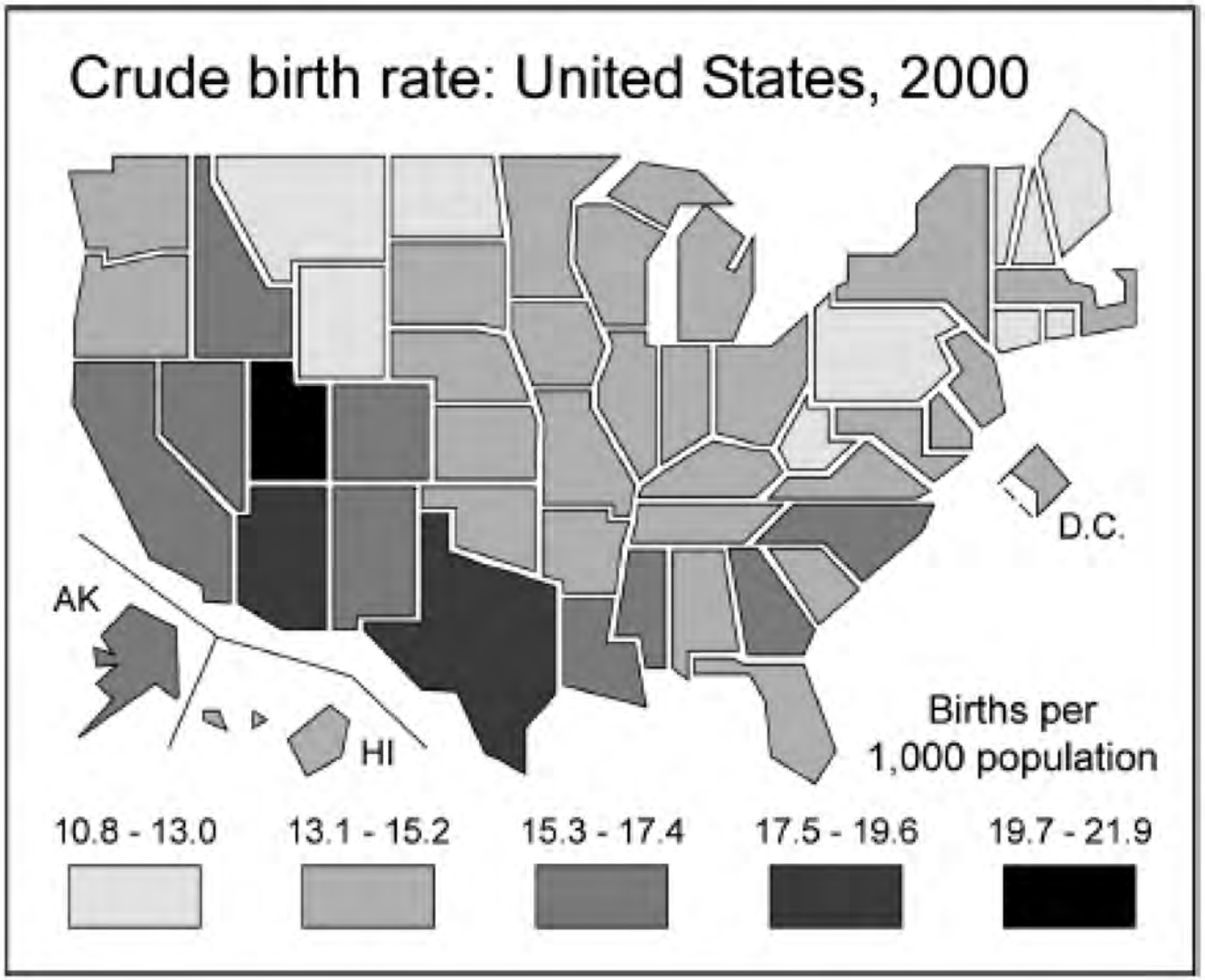


FIG. 2. *Crude birth rates, 2000, by state, based on equal-intervals cut-points and plotted on a visibility base map.*

Source: Monmonier M. *Statistical Science*, 20,3;2005

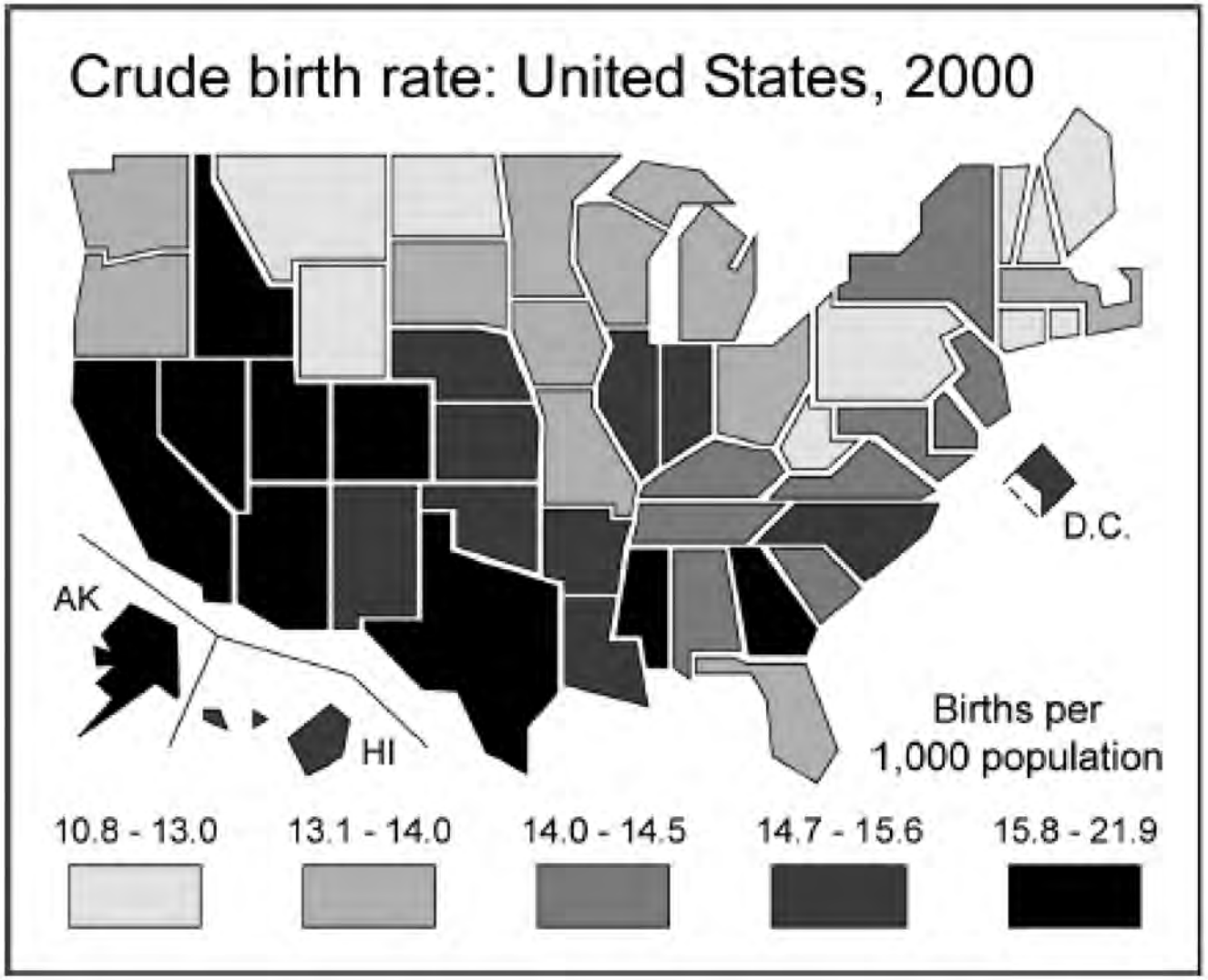


FIG. 3. *Crude birth rates, 2000, by state, based on quantile cut-points and plotted on a visibility base map.*

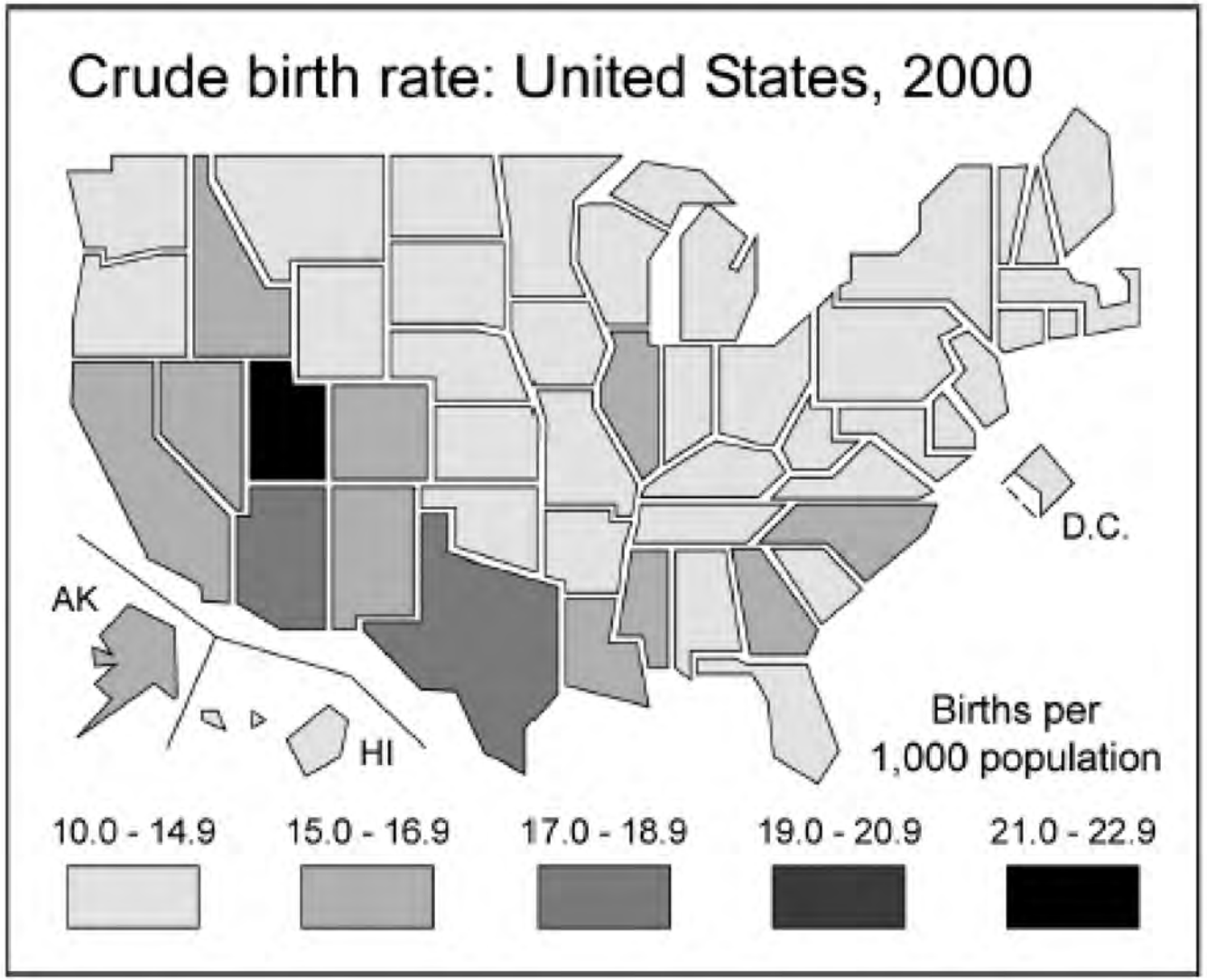


FIG. 5. *Crude birth rates, 2000, by state, categorized to suggest dangerously low rates overall.*

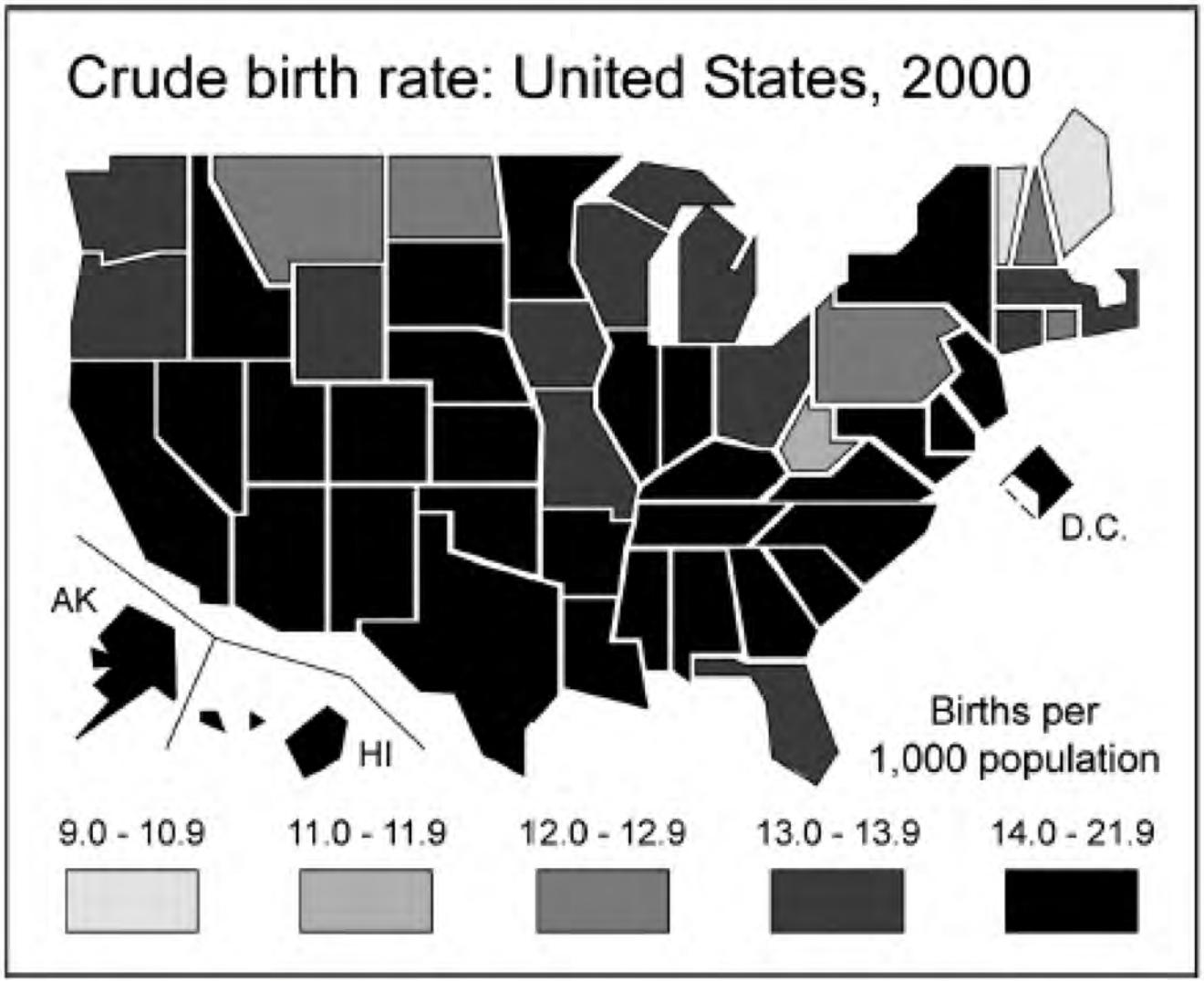


FIG. 6. *Crude birth rates, 2000, by state, categorized to suggest dangerously high rates overall.*

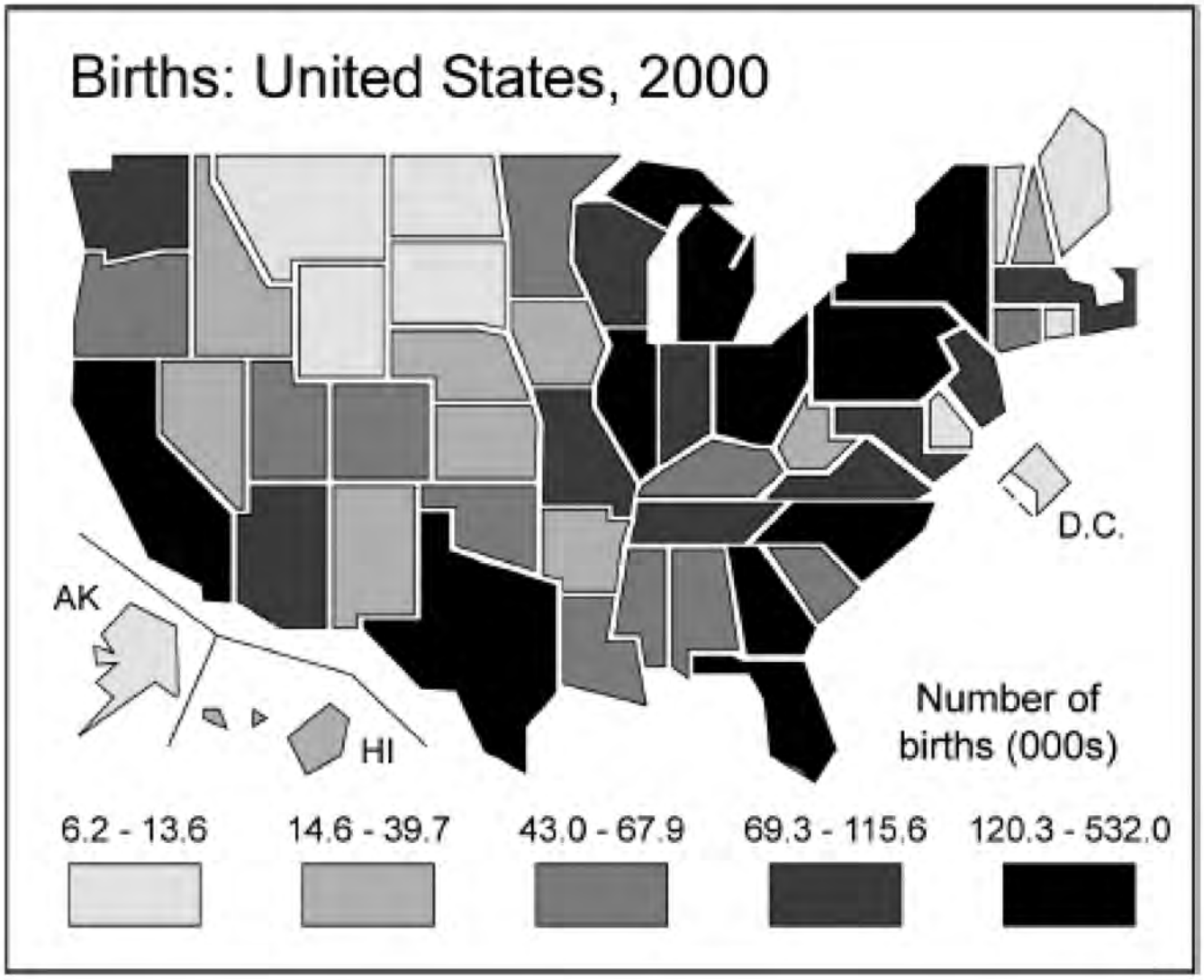


FIG. 7. *The darker-is-more-intense metaphor of choropleth maps offers a potentially misleading view of numbers of births.*

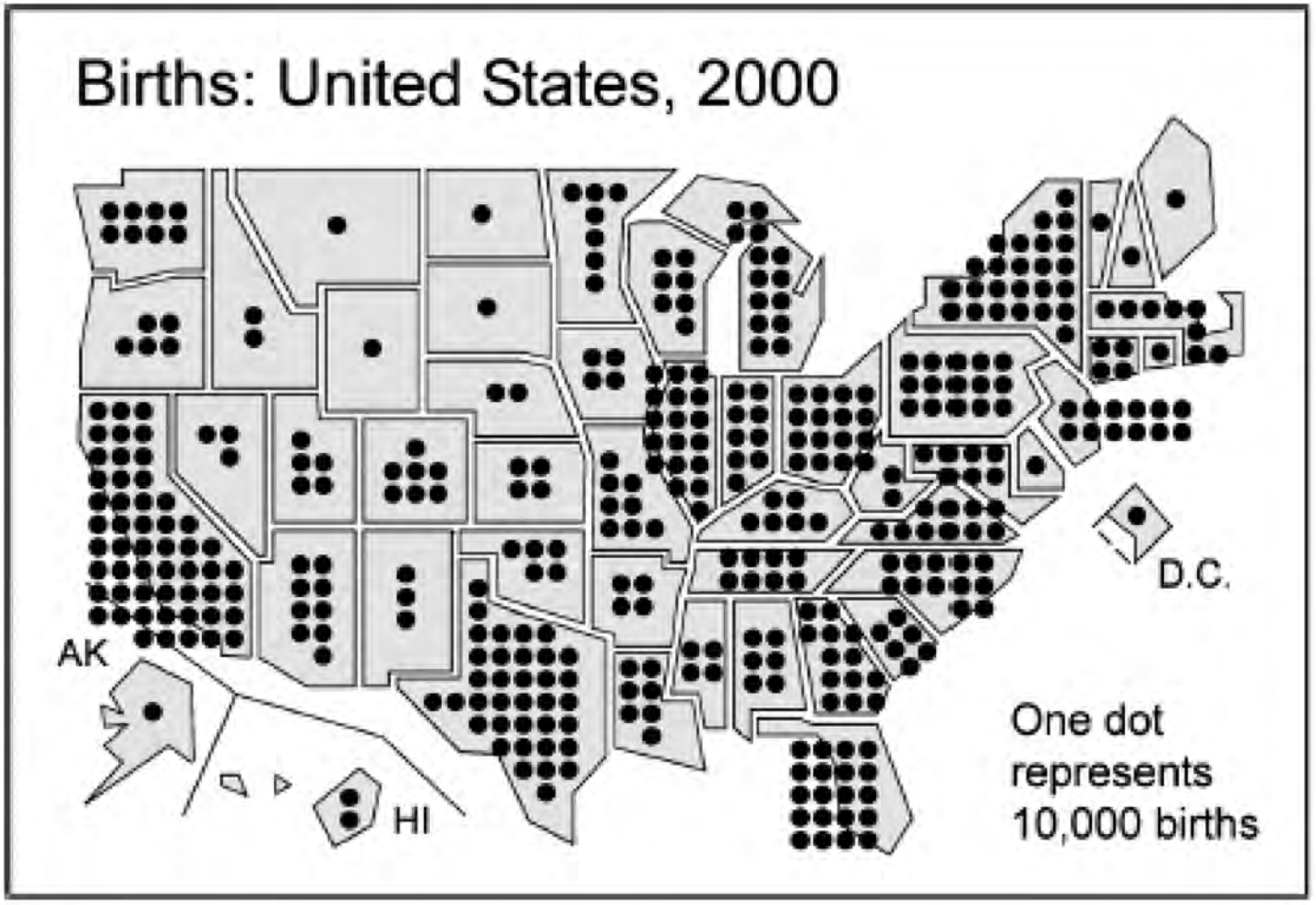


FIG. 8. *The bigger-means-more metaphor of this dot-array map affords a more appropriate treatment of the count data in Figure 7.*

Place Matters

- In descriptive epidemiology, we focus on the triad of person, place and time.
- Of these, the dimension of place is very frequently absent from the analysis, other than as a vehicle for data collection.
- Small area analysis is a set of methods to analyze phenomena at the local level, using data at county, ZIP Code, census tract, block group, block or latitude-longitude coordinate levels.

Does Place Matter?

- **Diseases and health conditions have been shown to:**
  - **Cluster in specific locations or regions**
  - **Have spatial gradients of incidence/prevalence**
  - **Vary across states or regions**
  - **Vary dramatically in incidence across nations**

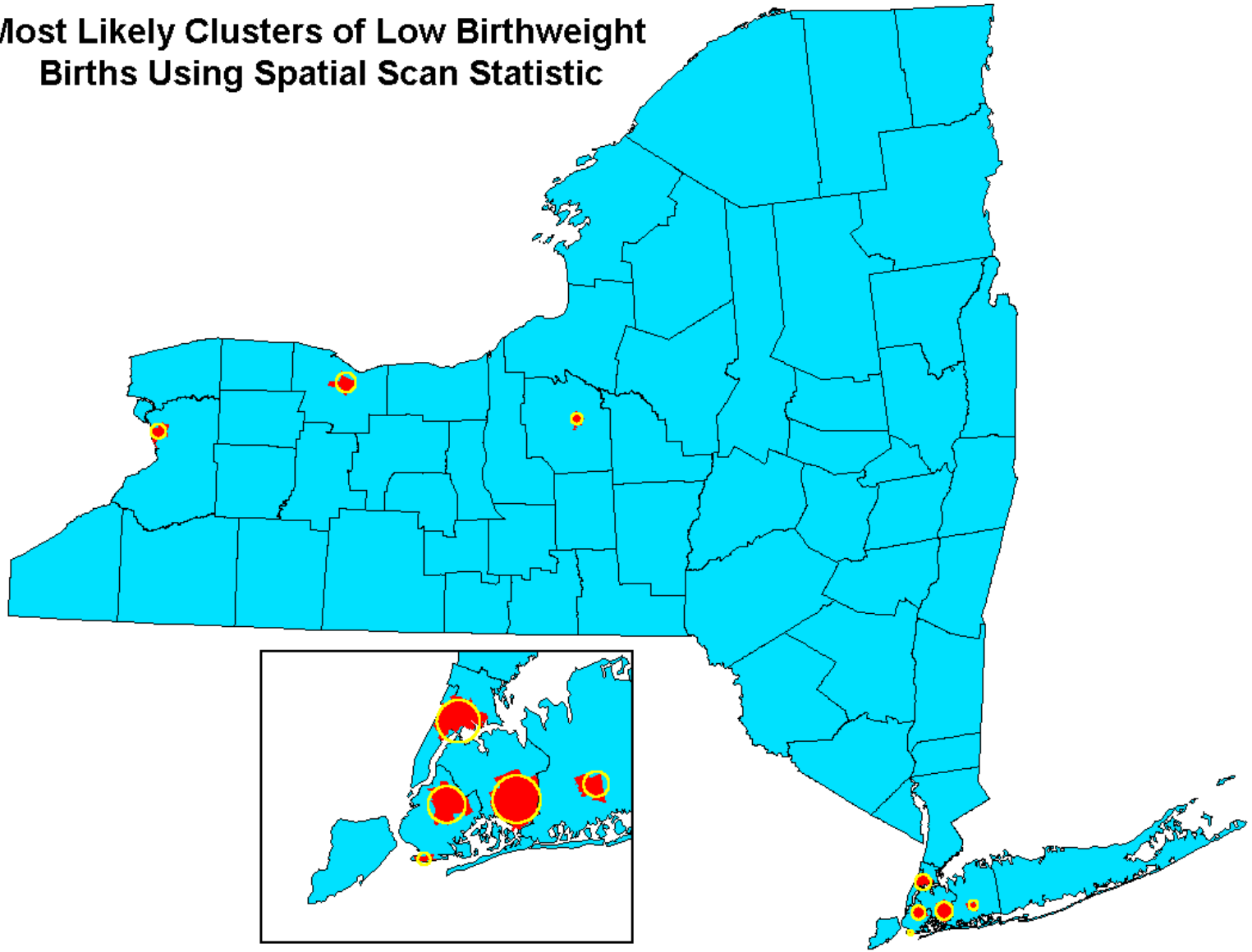
Does Place Matter? 2/3

- **Disease Clusters: Examples**
  - **Homicide, injuries**
  - **Environmental exposures**

    point source (toxic waste site, factory)

    non-point source (groundwater, ozone)
  - **Occupational, foodborne illness**

Most Likely Clusters of Low Birthweight Births Using Spatial Scan Statistic



p<0.05 Restictions; no cluster can contain more than 10% of births.

Does Place Matter?

- **Spatial Gradients: Examples**
  - **Multiple Sclerosis: tends to increase in prevalence through the mid-latitudes on both sides of the equator**
  - **Exposures to environmental contaminants tend to decrease with distance from a point source (very complex models may necessary to identify these patterns: i.e. Chernobyl and birth defects in NW Europe)**
- **Regional and National Variation**
  - **Why does the US have higher infant mortality rates than most western nations?**
  - **Why have C-Section rates tended to be higher in the southern US?**

Place Also Matters in More Complex Ways

- Immigrant women and birth outcomes
  - Why do these women tend to have better outcomes than members of their national origin or ethnic group born in the receiving country?
- SES and educational status
  - Why do persons of similar educational attainment living in poorer neighborhoods tend to have poorer health status than those living in neighborhoods with higher median income?

Summary

- An infinite number of maps can be created for any variable measured across small areas.
- Maps allow the user to better understand patterns in their data, but the ultimate goal should be to understand the processes that create the patterns observed.
- Small area analysis is a set of tools for understanding patterns in health data displayed spatially. The remainder of today's webinar will illustrate some applications of small area analysis.

Contact Information

- Russell S. Kirby, PhD, MS, FACE
  - Department of Community and Family Health
    College of Public Health, University of South Florida
  - Email: Russell Kirby
  - Telephone: 813-396-2347

Approaches to small area estimation in MCH

Michael R Kramer, PhD

Emory University

EMORY

ROLLINS SCHOOL OF PUBLIC HEALTH

Why do we need small area data?

- **Population heterogeneity**
  - A single average for an area or population may obscure important difference
- **Stakeholders want higher resolution data**
  - Resource allocation
  - Program planning and evaluation
  - Communications
- **As inputs for further spatial analysis**
  - Mapping, cluster analysis, spatial regression, sub-group analysis

Small area data: (some) problems and (some) solutions

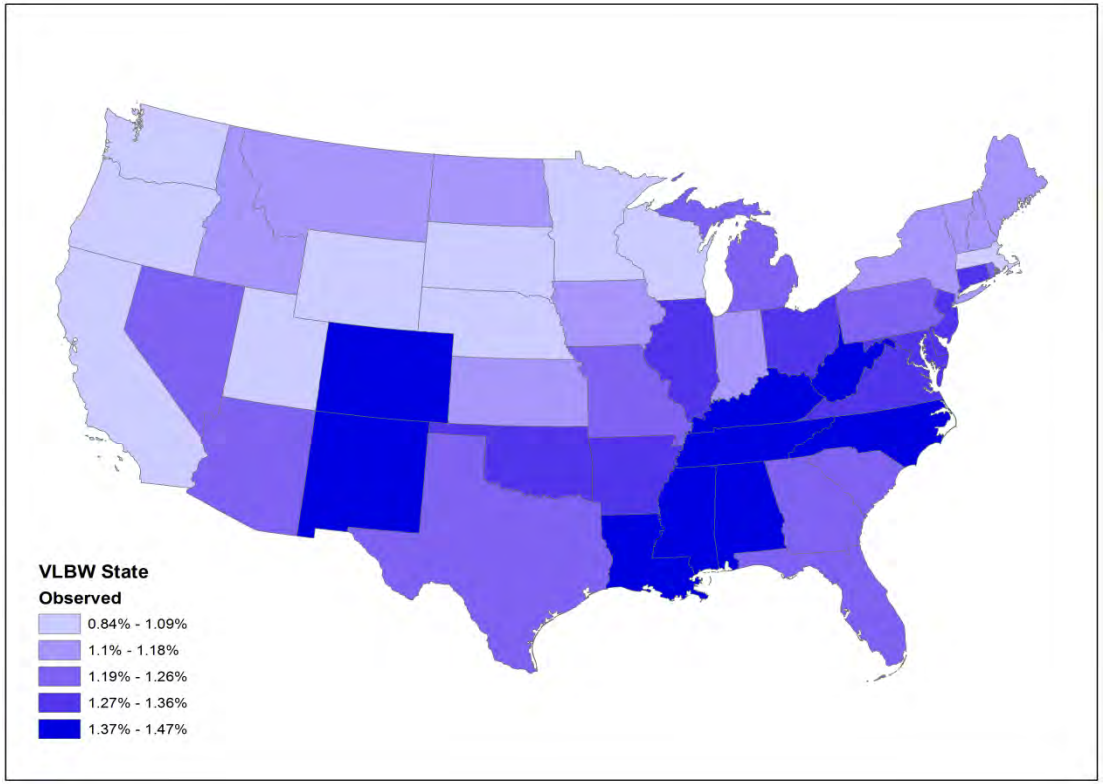
Problems

- Privacy and unintended disclosure of PII
- Statistical instability of rates with few events, small denominator

Solutions

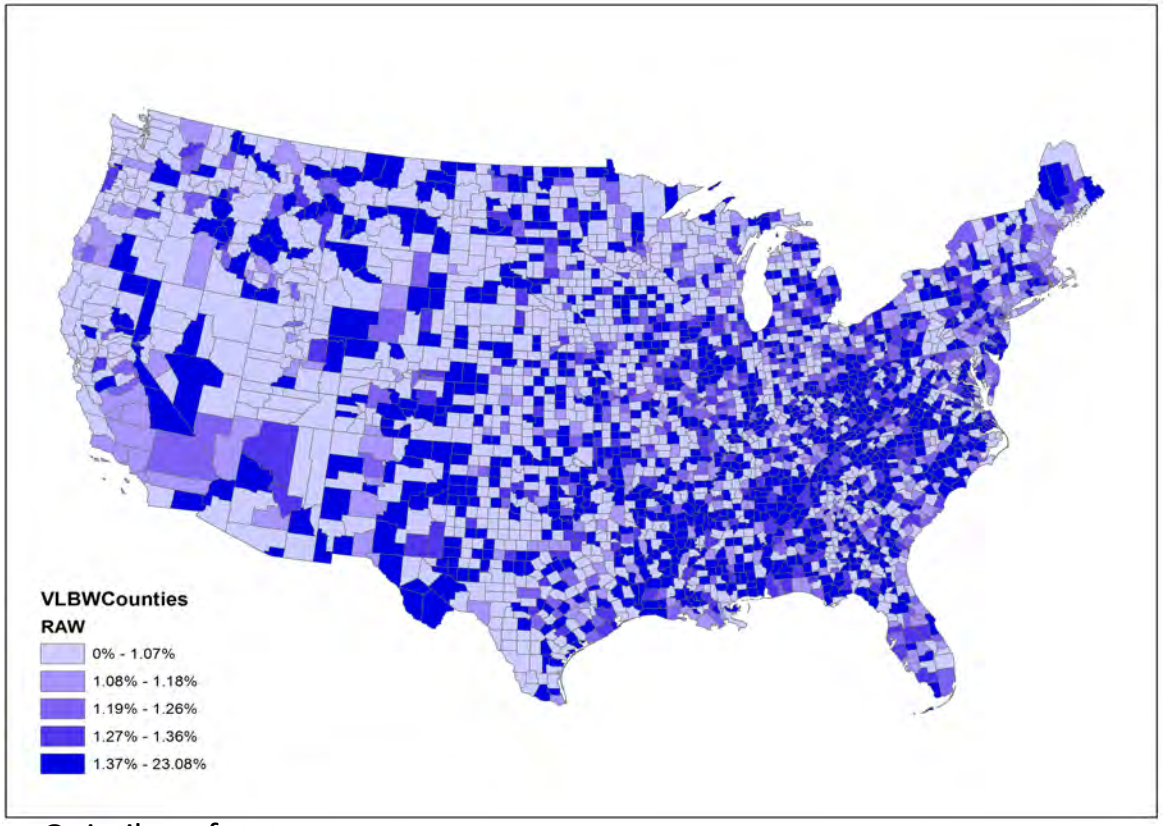
- Aggregation or pooling
- Suppression
- Geomasking
- Suppression (NCHS, CDC)
- Stratified sampling design (PRAMS, NSFG, BRFSS)
- Model based estimation**

Very Low birth weight (VLBW; <1500 grams), White women, 2005-7



Quintiles of states

VLBW – Counties, White women, 2005-7



Quintiles of states

Raw VLBW rates

- **Extreme values**
  - N=273 counties have VLBW prevalence of 0%
  - N=25 counties of VLBW > 4% (national mean is ~ 1.1%)
    - 6 have >100 births
    - Most have < 50 births and 1-4 VLBW

- **Uncertain rates**
  - Relative standard error: (S.E. / Estimate) * 100; some recommend caution with RSE > 30%
  - N=1753 counties of a relative standard error >30%

(Partial) solution: borrow auxiliary information

WHAT DO WE KNOW?

1. The numerator and denominator in [the small area]
2. The distribution of rates in other areas (the total study area, the neighboring areas)

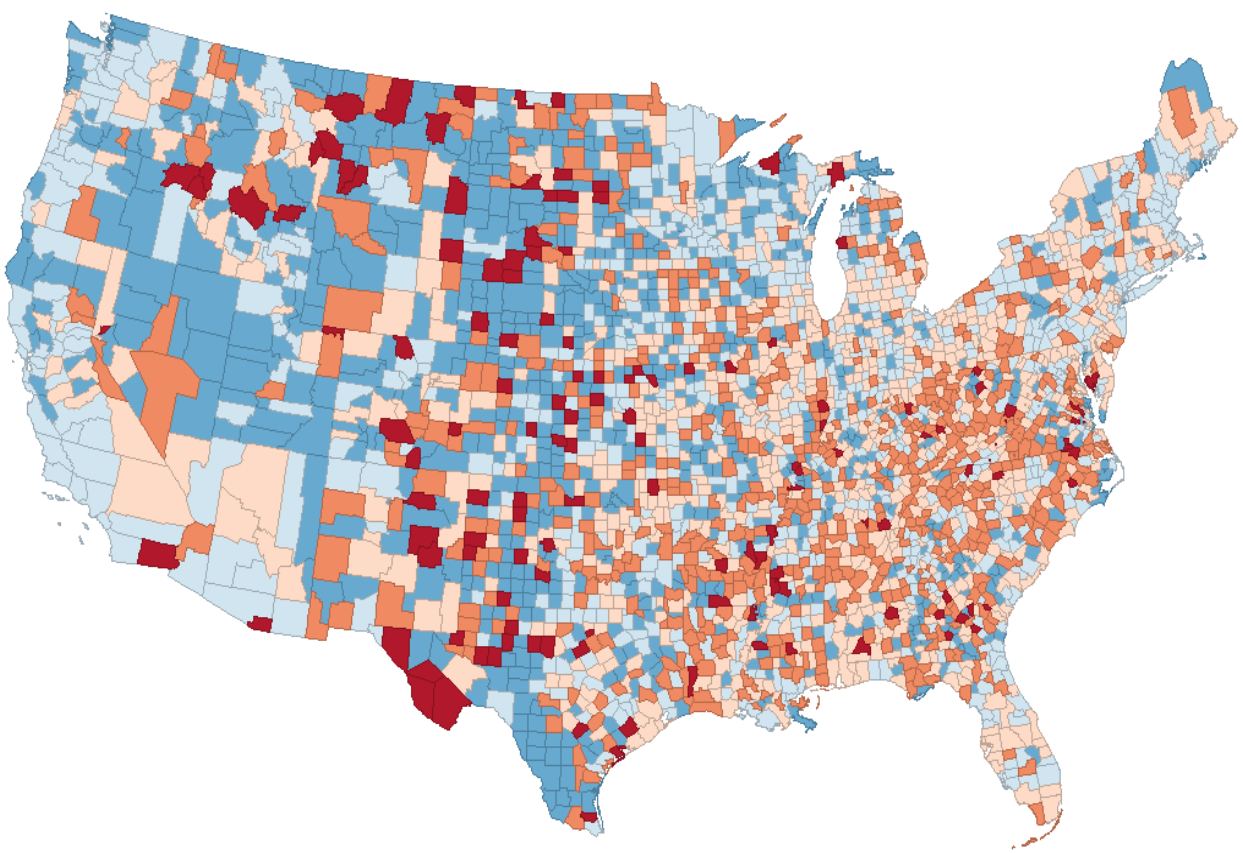
WHAT CAN WE DO?

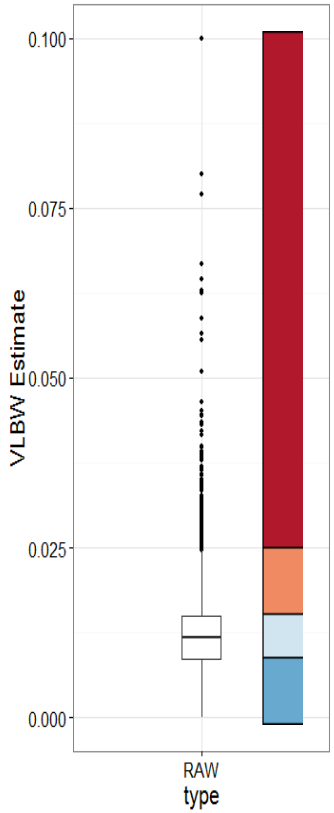
1. Inform our estimate of the local rate by information about the overall distribution of rates
2. Allow true outliers to be different but insure sparse data does not stray too far

Parameter shrinkage

1. **Aspatial Multilevel Regression**
   - Shrinks each area towards a regional mean (e.g. counties within states)
2. **Aspatial Empirical Bayes**
   - Shrinks each area towards a global mean
3. **Spatial Empirical Bayes**
   - Shrink towards a spatially local mean (e.g. the average rate in the neighboring counties)

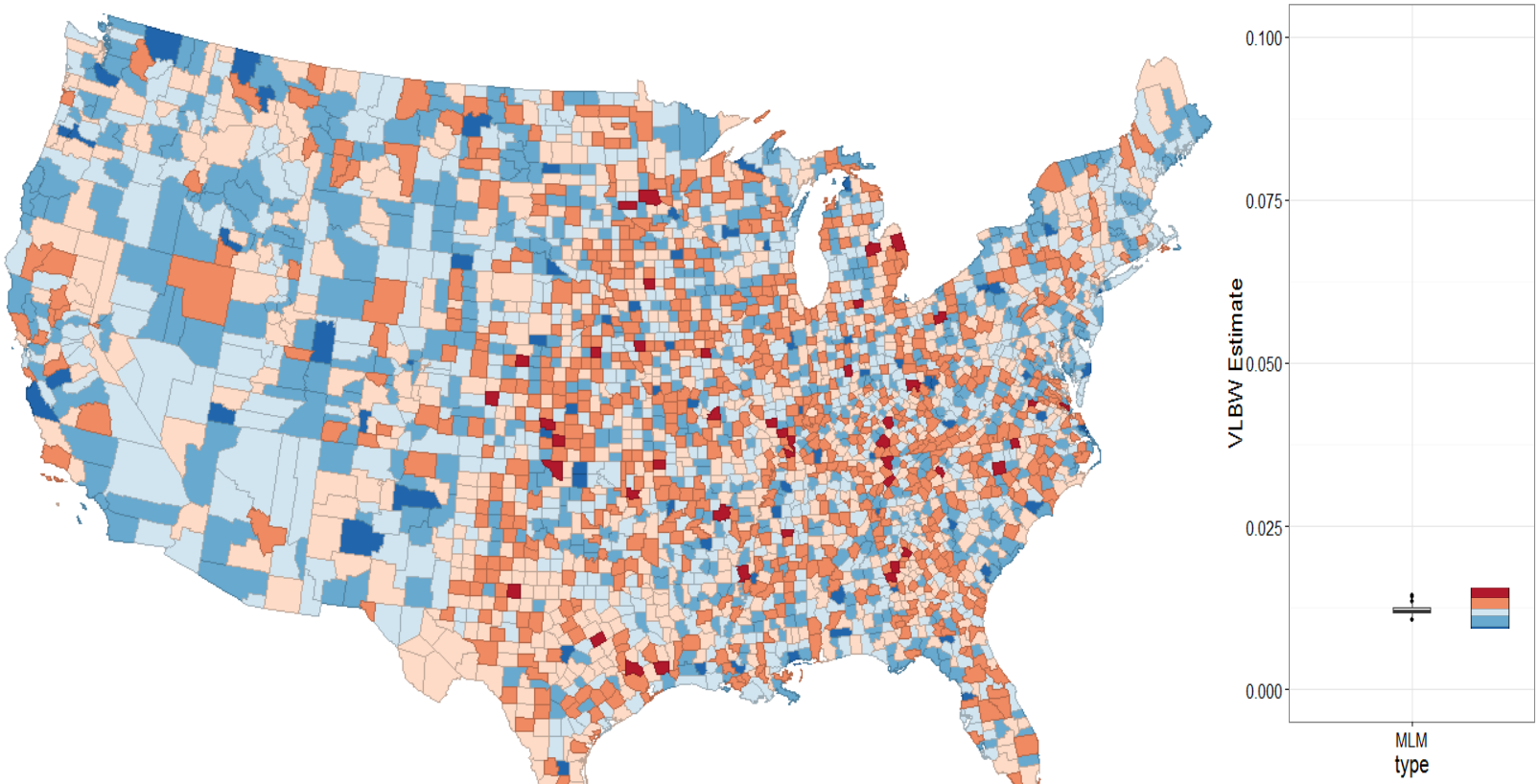
Raw VLBW Box Plot Map





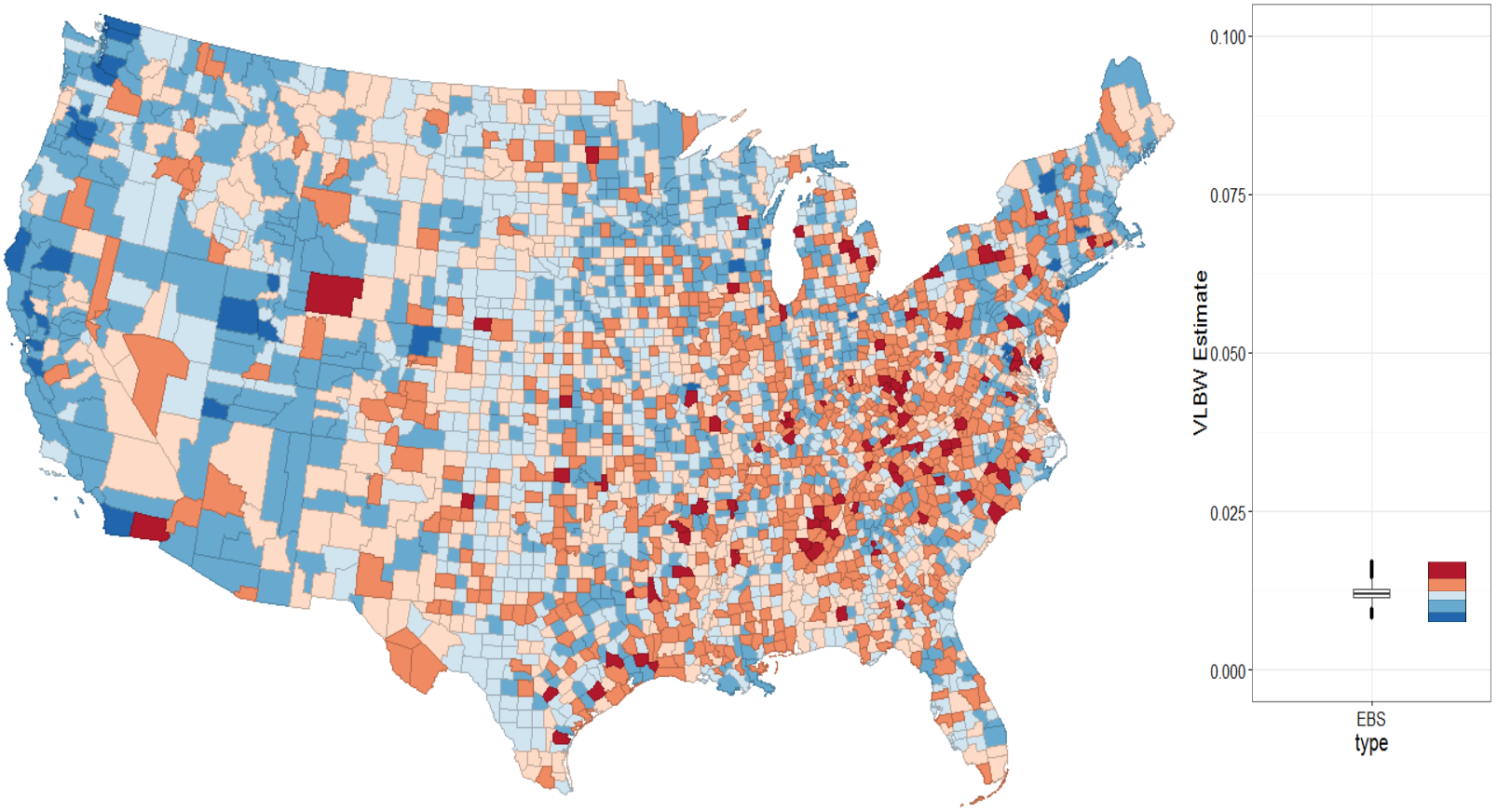
RSE > 30% in 1,736 Counties

Aspatial Multilevel Model Estimates



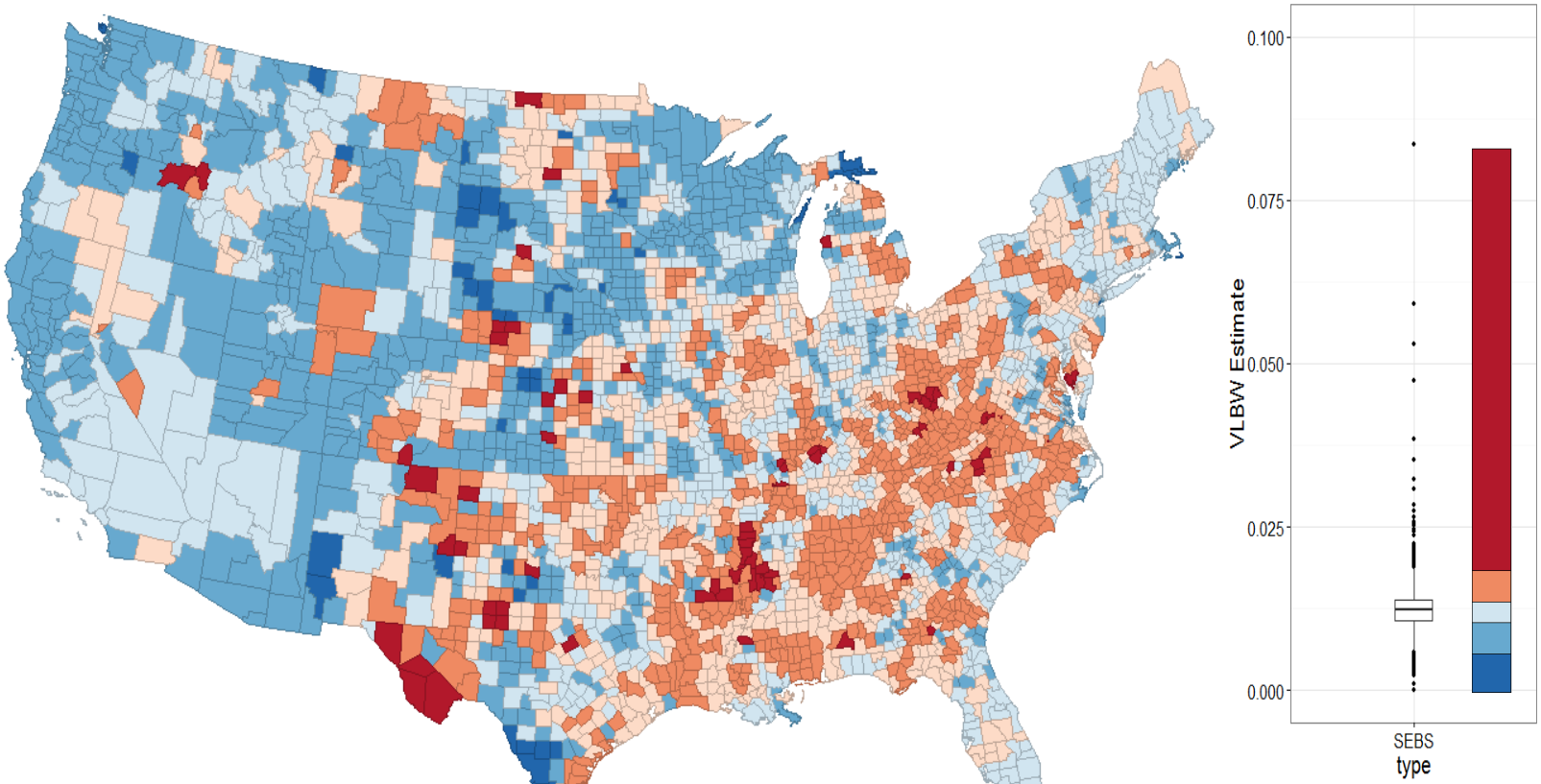
RSE > 30% in 26 Counties

Empirical Bayes estimates



RSE > 30% in 28 Counties

Spatial Empirical Bayes



RSE > 30% in 61 Counties

Comparing Methods

Goals

- Reduce unrealistic outliers
- Preserve true heterogeneity

Results

- MLM > EB reduced outliers and made rate distribution ‘normal’
- RSE > 30% in N=26 counties in EB vs N=1736 counties in RAW
- Spatial EB reduced outliers somewhat but maintained (and stabilized) rate variation in sparse counties

Other approaches to small area estimation

- Fully Bayesian disease mapping
- Kernel Density smoothing
- Spatial krieging
- Iterative weighted head-banging

Doing small area estimation

- Multilevel (aspatial) shrinkage
  - SAS (GLIMMIX) or R (lmer())

- Empirical Bayes aspatial and spatial shrinkage
  - GeoDa software
  - R

Conclusion

- **Goals of small area estimati**on:
  - Validly maximize the information in our data
  - Describe small area variations

- **Model-based small area estimation**
  - Borrowing statistical information to stabilizes local estimates
  - Balancing rate stabilization and realistic rate variation

A few relevant references…

- Rao, J. N. (2003). *Small area estimation*. John Wiley & Sons, Inc..
- Anselin, L., Lozano, N., & Koschinsky, J. (2006). Rate transformations and smoothing (PDF).
- Goodman, M. S. (2010). Comparison of small-area analysis techniques for estimating prevalence by race. *Preventing Chronic Disease*, *7*(2), A33.
- Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data* (Vol. 368). John Wiley & Sons.
- Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.

Small Area Analysis Using GIS and Spatial Epidemiology: Assessing WIC Unmet Needs

Thomas J. Stopka, PhD, MHS

Dept. of Public Health and Community Medicine
Tufts University School of Medicine

DATASPEAK
AUGUST 12, 2015

Tufts University School of Medicine

Acknowledgements

- California Department of Public Health, MCAH
  - Chris Krawczyk, PhD
  - Mike Curtis, PhD
- California WIC Program
  - Pat Gradziel, PhD, RD
  - Charlene Manning
- Mentor in GIS and Spatial Analysis:
  - Estella Geraghty, MD, MS, MPH/CPH

UCSF University of California San Francisco

Maternal, Child Adolescent Health Program, Center for Family Health, California Department of Public Health

California WIC Women, Infants & Children

California Department of Public Health CDPH

The University of California Davis

Tufts University School of Medicine

Use of Spatial Epidemiology and Hot Spot Analysis to Target Women Eligible for Prenatal Women, Infants, and Children Services

Thomas J. Stopka, PhD, MHS, Christopher Krawczyk, PhD, Pat Gradziel, PhD, RD, and Estella M. Geraghty, MD, MPH

Objectives. We used a geographic information system and cluster analyses to determine locations in need of enhanced Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) Program services.

Methods. We linked documented births in the 2010 California Birth Statistical Master File with the 2010 data from the WIC Integrated Statewide Information System. Analyses focused on the density of pregnant women who were eligible for but not receiving WIC services in California's 7049 census tracts. We used incremental spatial autocorrelation and hot spot analyses to identify clusters of WIC-eligible nonparticipants.

Results. We detected clusters of census tracts with higher-than-expected densities, compared with the state mean density of WIC-eligible nonparticipants, in 21 of 58 (36.2%) California counties ($P < .05$). In subsequent county-level analyses, we located neighborhood-level clusters of higher-than-expected densities of eligible nonparticipants in Sacramento, San Francisco, Fresno, and Los Angeles Counties ($P < .05$).

Conclusions. Hot spot analyses provided a rigorous and objective approach to determine the locations of statistically significant clusters of WIC-eligible nonparticipants. Results helped inform WIC program and funding decisions, including the opening of new WIC centers, and offered a novel approach for targeting public health services. (*Am J Public Health*. Published online ahead of print December 19, 2013: e1–e7. doi:10.2105/AJPH.2013.301769)

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), a federally funded nutrition and health program, was established as a pilot program in 1972.[1] WIC provides nutrition education, referrals, breastfeeding support, and nutritious supplemental food for low- to moderate-income families with nutritionally at-risk pregnant and breastfeeding women, infants, and children up to the 5th birthday.[1,2] With an annual federal budget of $6.2 billion, the WIC program serves nearly 9 million participants each month through 1900 local agencies in state public health departments, Indian tribal organizations, and US territories.[1,2] Additionally, the WIC Overseas Program provides services to Americans and dependents living overseas at US military bases. Nearly 30% of pregnant women participate in the WIC program nationally each year.[2] In 2010, the WIC program served 62.6% of all who were eligible. WIC served

Stopka et al. (2014). *AJPH*.
PMID: 24354821

Background: WIC

- Special Supplemental Nutrition Program for Women, Infants and Children (WIC)
  - Provides health education
  - Supplemental food vouchers
- Approximately 1 of 4 pregnant women and roughly 50% of all infants born in the U.S. participate in WIC
- More than half (51%) of pregnant women enroll in WIC during 1st trimester
- In California, WIC agencies provide services locally to nearly 1.5 million women, infants and children each month at >600 sites.

Background: Need for GIS and Spatial Analysis in Good Times and Bad...

- In 2012, the WIC Program faced a potential $832 million funding reduction
- Up to 500,000 low-income women and children would have been denied services
- Spatial analyses could help assess needs and unmet needs

Study Questions

- Where are statistically significant clusters of WIC eligible women* located California?
- Where do micro-level clusters of WIC eligible women exist within counties?

***WIC eligible women:** Received MediCal during pregnancy (i.e., WIC eligible) but did not receive WIC services

Data

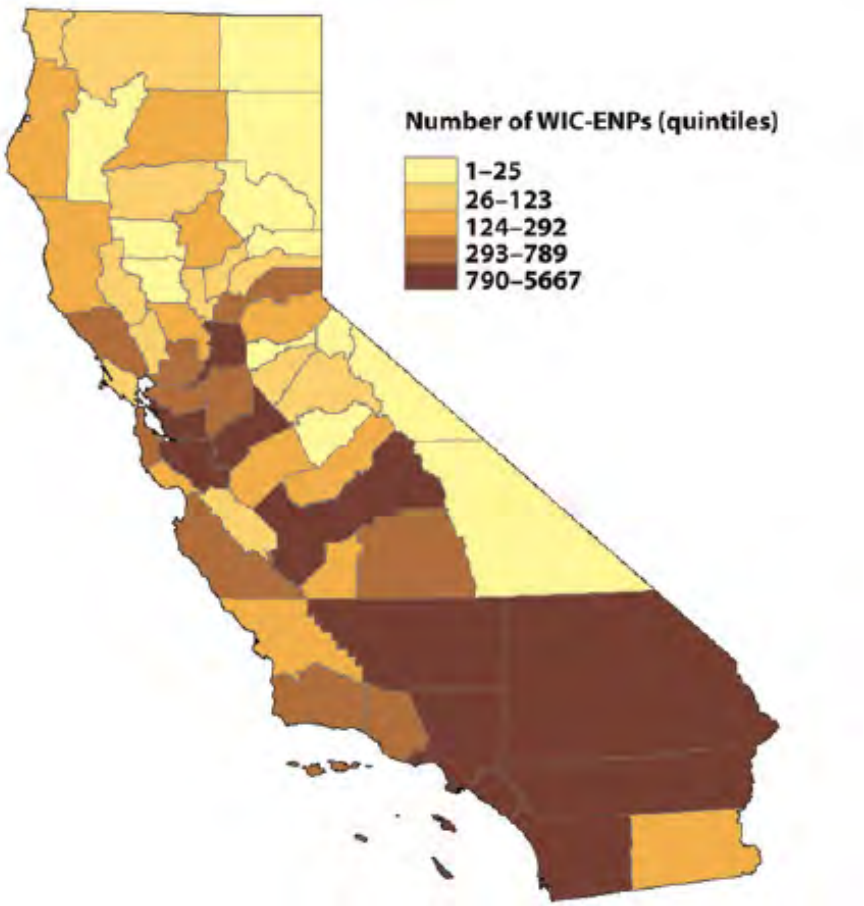
- Multi-step algorithm to merge 2 large datasets
  - Birth Statistical Master File for live births in CA in 2010 (n=501,907)
  - 2010 WIC-ISIS data file (n=279,288)
  - 2010 Not in WIC-ISIS (n=222,619)

- Outcome of interest:
  - WIC eligible women not receiving WIC Services (n=30,697)
  - Density of WIC eligibles/square mile on the census tract level

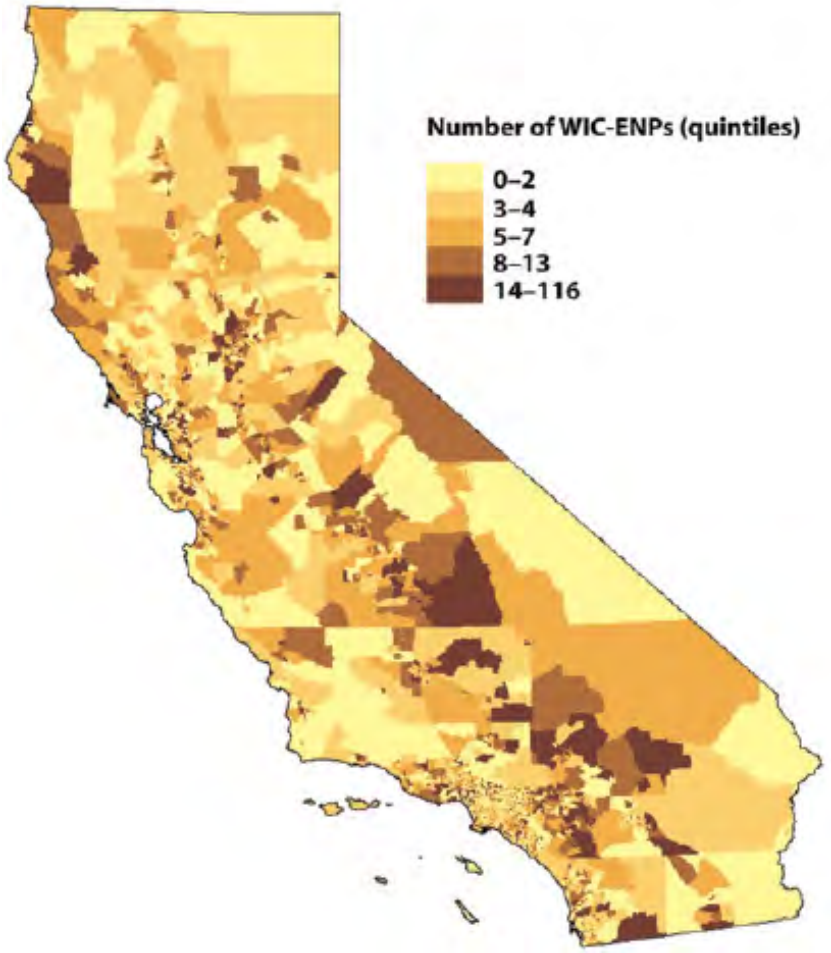
Methods

- Descriptive GIS Maps
  - Thematic choropleth maps
  - Dot density maps
- Spatial Epidemiology (many options)
  - Used 5-step geoprocessing approach
  - Conduct hotspot cluster analyses
- Software: ArcGIS 10.1

Number of WIC Eligible Women Not Receiving WIC Services: CA Counties, 2010



Number of WIC Eligible Women Not Receiving Services: CA Medical Service Study Areas, 2010



Dot Density of WIC Eligibles not receiving Services: CA Census Tracts, 2010

Questions:

- How do we know that these patterns are not due to chance alone?
- Where are statistically significant clusters of WIC eligible women located in California?
  - Hint: We need to conduct hot-spot cluster analyses to find out…

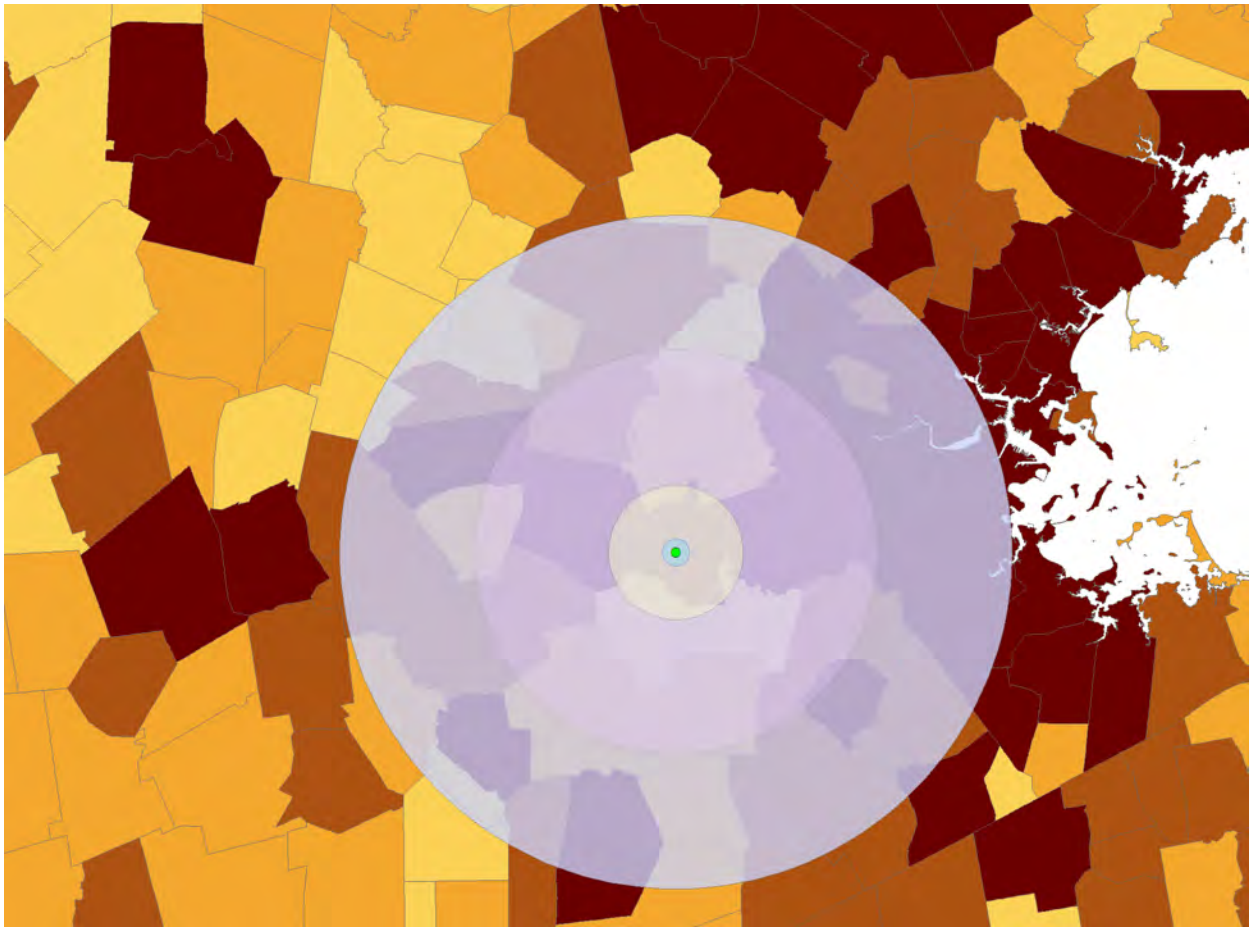
Spatial Epidemiology / Geostatistical Analysis

Methods: Hot-Spot Analysis

Getis-Ord Hot-Spot Analysis (Gi*)

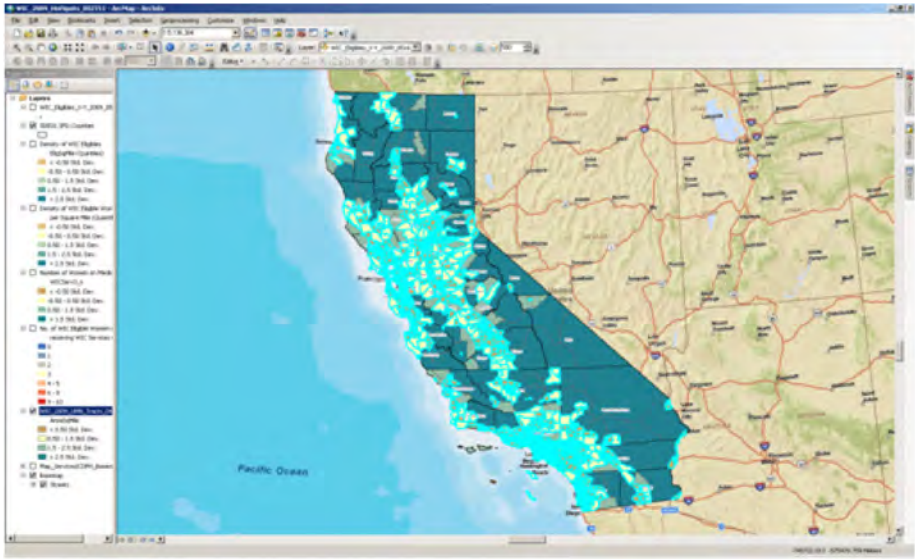
- Spatial analysis tool in ArcGIS
- Used to pinpoint locations of clusters
  - Looks at each feature within the context of neighboring features. A feature with a high value is a statistically significant hot spot if it is also surrounded by other features with high values.
  - The local mean for a feature and its neighbors is compared proportionally to the "global mean" of all features.
  - When the *observed* local mean is much different than the *expected* local mean, and that difference is too large to be the result of random chance, a statistically significant Z-score results.

Exemplar Sphere of Spatial Influence (1km, 5 km, 15km, 25km?)



Methods: Hot-Spot Geoprocessing Tasks

1) Calculate area for polygons and exclude areas that are > 1.5 SD above mean tract area to account for variation in polygon size

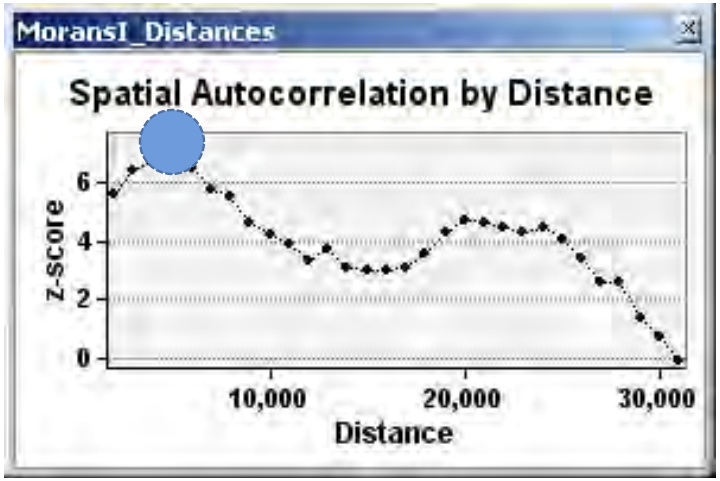


2) Find the appropriate spatial scale for selected tracts (i.e., distance from each tract to 2 nearest neighbors)
   - Starting Distance
   - Incremental Distance

Methods: Hot-Spot Geoprocessing Tasks

3) Conduct incremental spatial autocorrelation analysis (Moran's I)

- Determine multiple distances at which clustering peaks
- Find distance of first statistically significant peak (Z-score; p-value)



4) Generate a spatial weights matrix file to be included in analyses; account for large polygons

5) Conduct hot-spot analysis

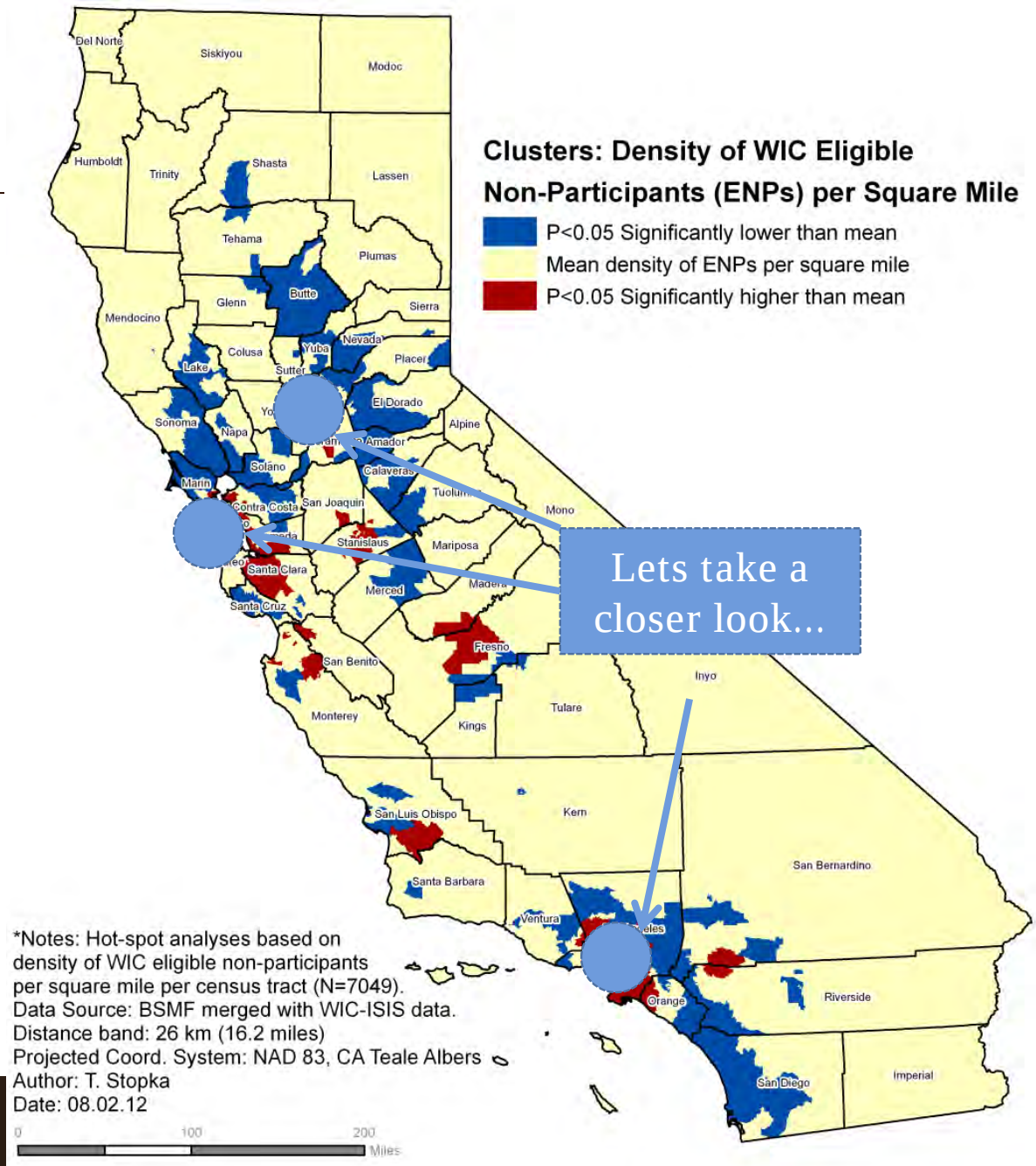
- Determine location of statistically significant clusters

Hot Spot Analysis Output

- Pinpointed statistically significant clusters
  - Across the **state** of CA
  - Within selected **counties**
- Results
  - P-values, Moran's *I*, and **Z-scores** (map layer)
    - Larger Z-score, more intense the clustering of high values (**a hot spot**)
    - Smaller Z-score, more intense the clustering of low values (**a cold spot**)

Results: Statewide Hot-Spot Analyses

Hot-Spot Clusters:

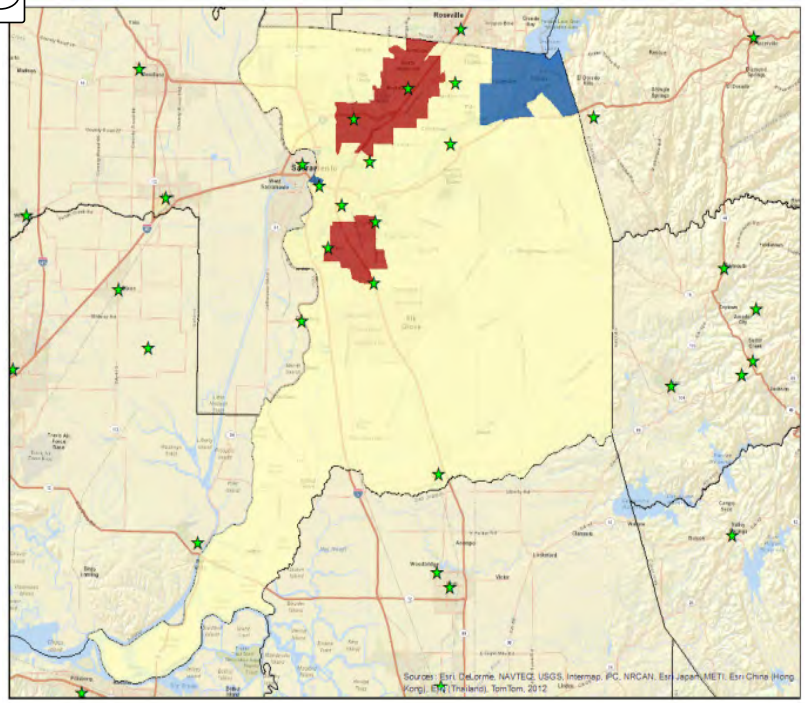


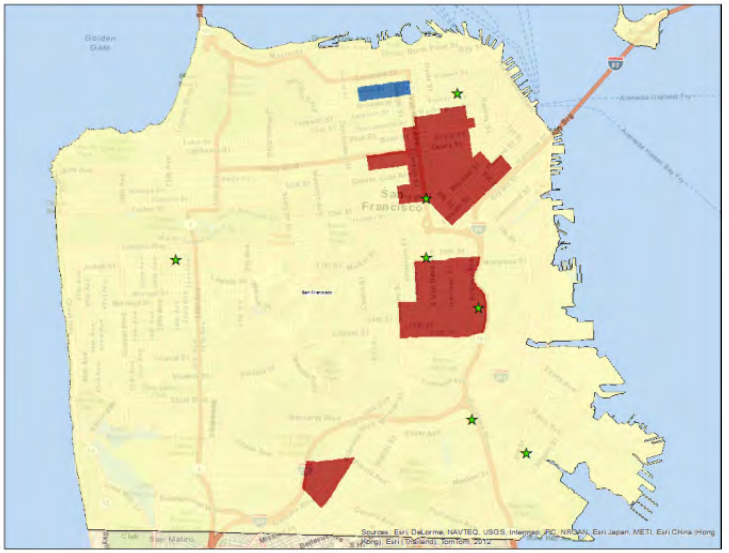
= Hot Spot (p<0.05)

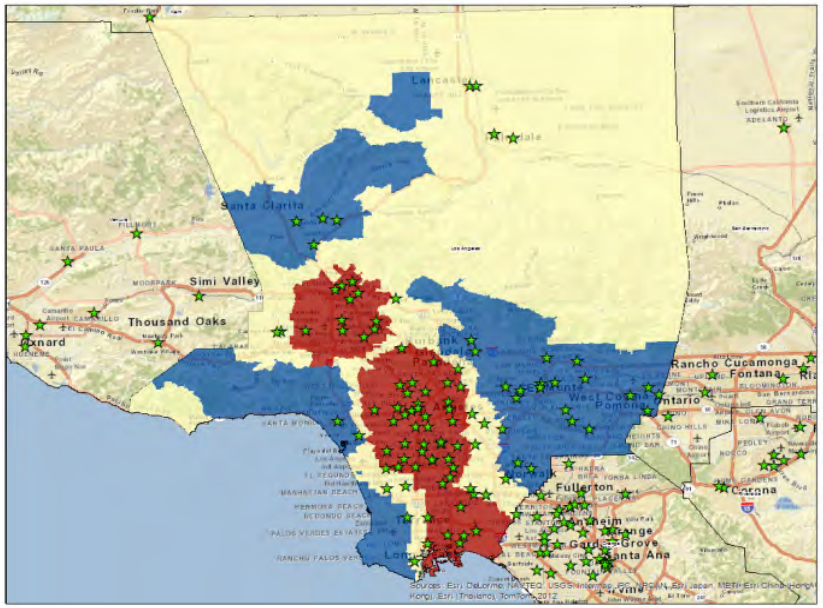
= Cold Spot (p<0.05)

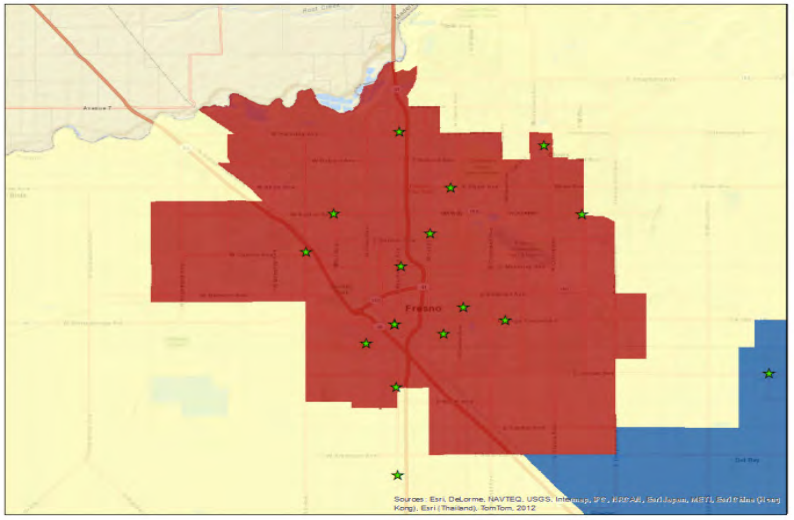
Stopka (2014). *AJPH*.
PMID: 24354821

Local Hot Spots

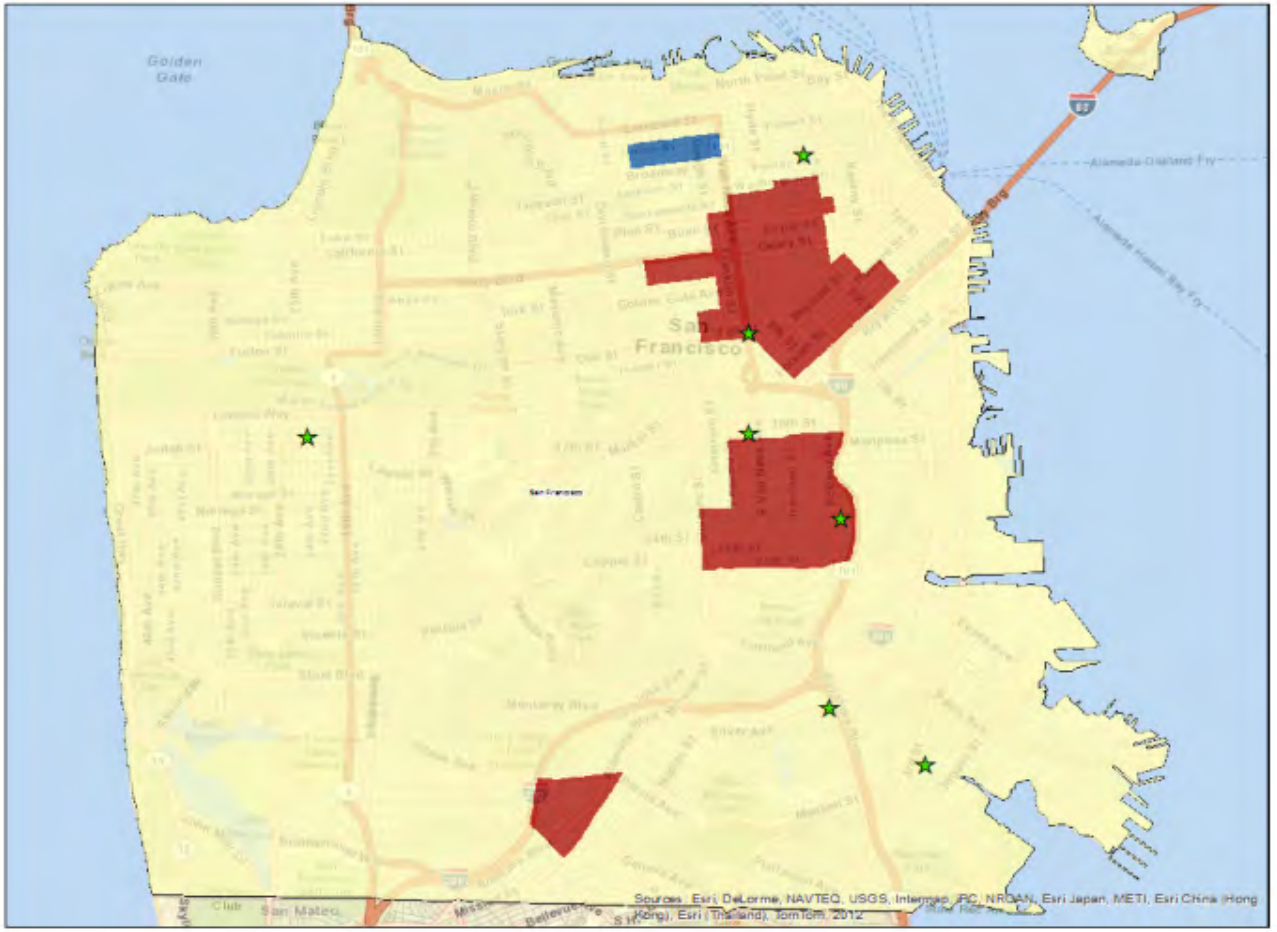




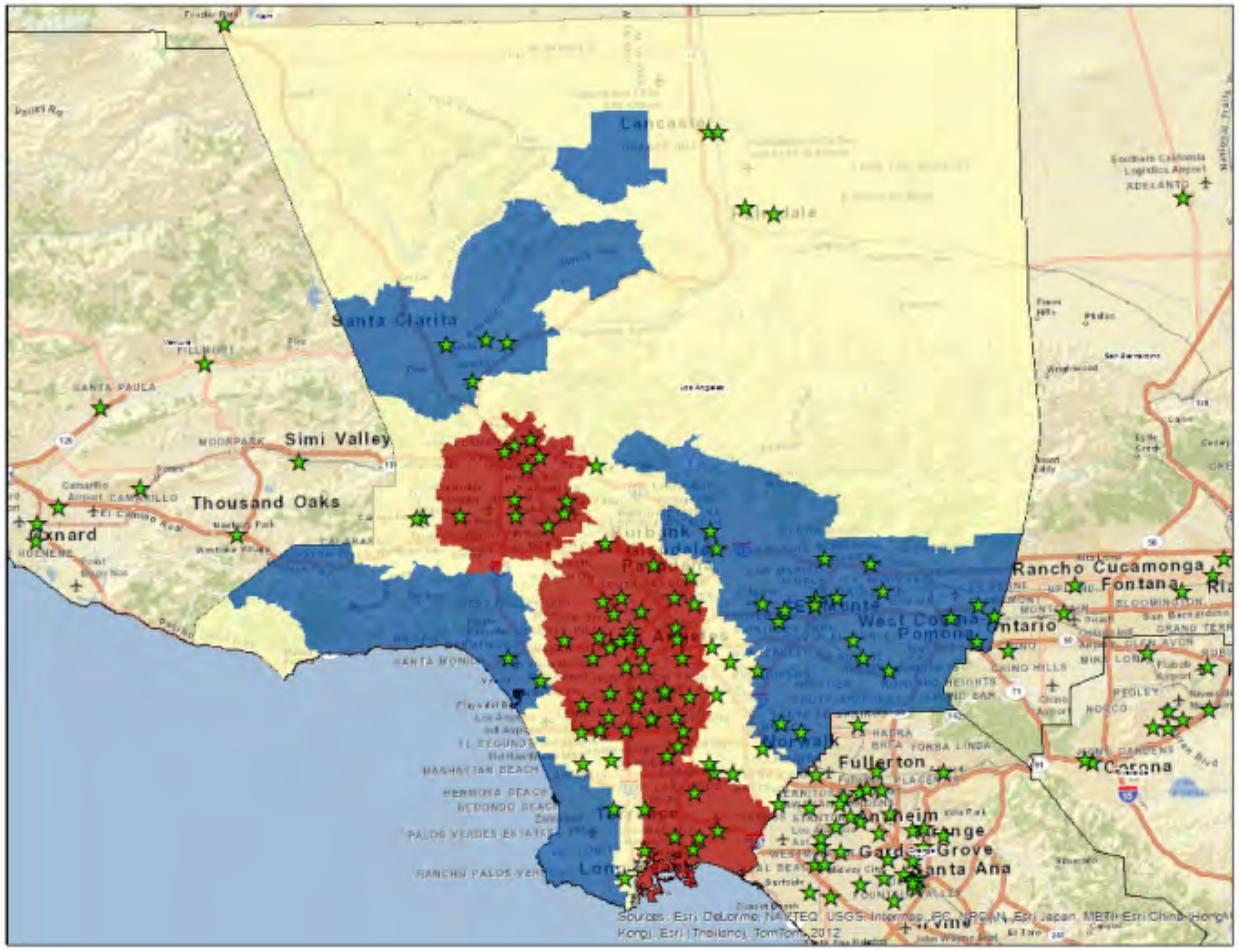




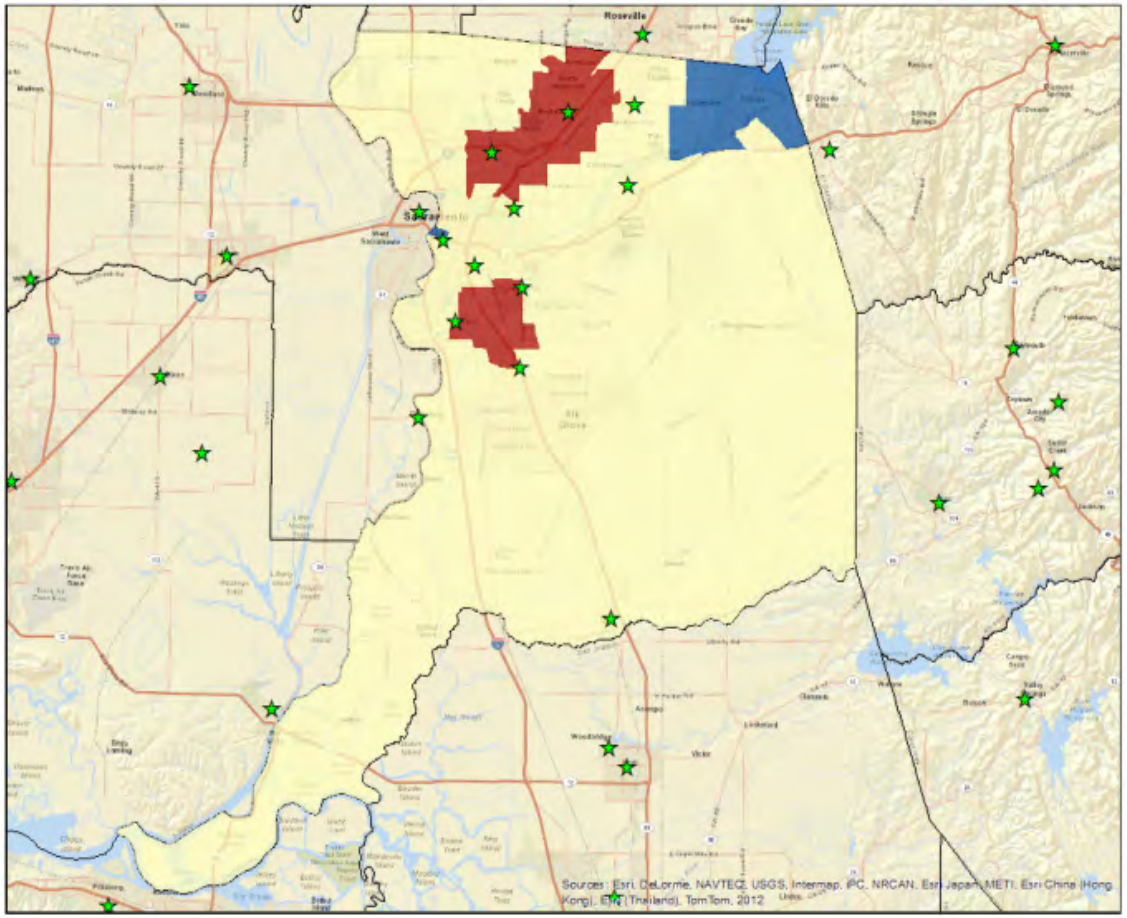
County-Specific Hot-Spot Analyses: San Francisco



County-Specific Hot-Spot Analyses: Los Angeles



County-Specific Hot-Spot Analyses: Sacramento



Grand Opening of New Clinic

MATERNAL, CHILD ADOLESCENT HEALTH
CA Dept of Public Health

CALIFORNIA WIC
WOMEN, INFANTS & CHILDREN

Tufts UNIVERSITY | School of Medicine

Discussion

- The 5 geoprocessing steps for hot-spot analyses provided a systematic, rigorous, and **objective approach**
- **State level hot-spot analyses** helped locate statistically significant clusters of WIC eligible women in key CA **counties**
- **County level hot-spot analyses** allowed us to locate clusters of highest WIC need on the local **neighborhood level**
- Findings helped **inform WIC** program and funding decisions on the state and local level

References

References

1. US Department of Agriculture. Food and Nutrition Service. Women, Infants and Children. Available at: http://www.fns.usda.gov/wic. Accessed September 14, 2011.

2. Oliveira VJ, Frazão E. *The WIC Program Background, Trends, and Economic Issues.* Economic Research Service report no 73. Washington, DC: US Department of Agriculture, Economic Research Service; 2009.

3. Martinez-Schiferl M, Zedlewski S, Giannarelii L. *National and State-Level Estimates of Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) Eligibles and Program Reach, 2010.* Washington, DC: US Department of Agriculture, Food and Nutrition Service, Office of Research and Analysis; 2013.

4. California WIC Association. WIC Program at a glance. Available at: http://www.calwic.org. Accessed September 22, 2011.

Thank you! Questions?

Thomas J. Stopka, PhD, MHS

Additional Questions

dataspeak@altarum.org

Thank You

Thank you for participating.

Complete feedback on today's program.

(the link will open in a new window)