CDC's National Environmental Public Health Tracking Program
A framework for evaluating denominator data sources

Samantha Lindsey, MPH – Contractor with Carter Consulting, Inc

Heather Strosnider, PhD, MPH – Science Team Lead

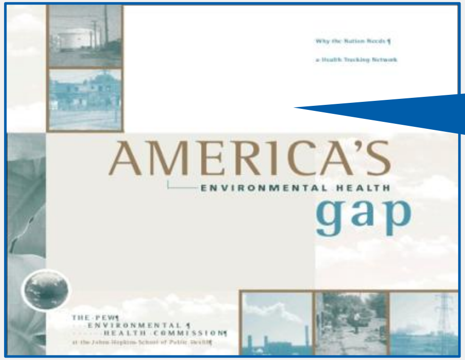
Today's Presentation

- Information on Environmental Public Health Tracking Network ("Tracking")
- Tracking Program's Sub-county data efforts
- Framework to evaluating potential differences in health outcome rates generated from multiple population sources

THE TRACKING PROGRAM

National Environmental Public Health Tracking Program

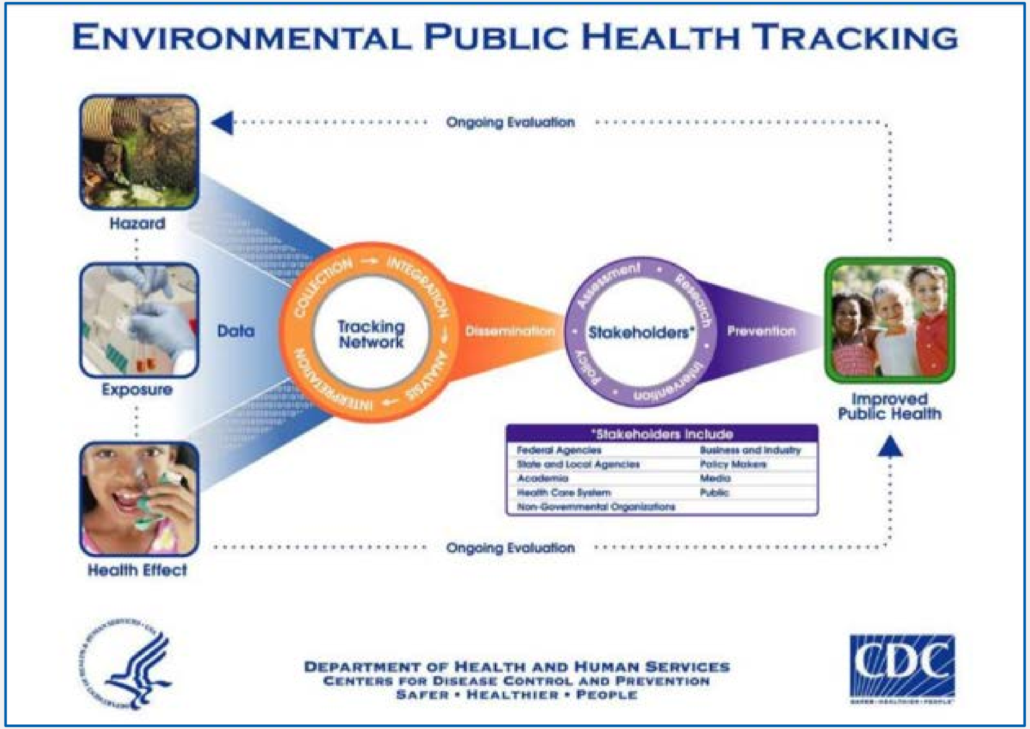
- Created in 2002 in response to Pew Commission report
- Identified gap in critical knowledge hindering efforts to reduce or eliminate diseases that may be prevented by managing environmental factors
- Recommended a “Nationwide Health Tracking Network for diseases and exposures”



“…create a federally supported Nationwide Health Tracking Network with the appropriate privacy protection, that informs consumers, communities, public health practitioners, researchers, and policymakers on chronic diseases and related environmental hazards and population exposures.”

National Environmental Public Health Tracking Program

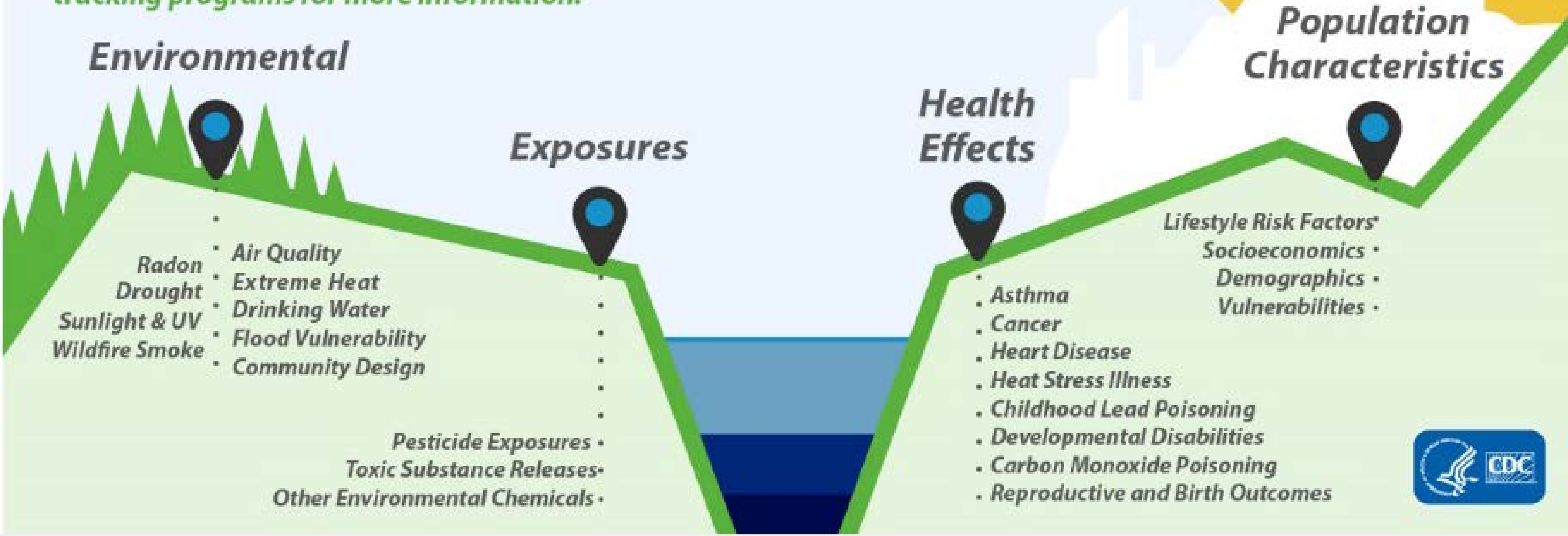
- **Vision:** Healthy informed communities
- **Mission:** To provide information from a nationwide network of integrated health and environmental data that drives actions to improve the health of communities



ENVIRONMENTAL PUBLIC HEALTH TRACKING

CONNECTS ENVIRONMENT & HEALTH INFORMATION

Check out CDC's data explorer and state and local tracking programs for more information.



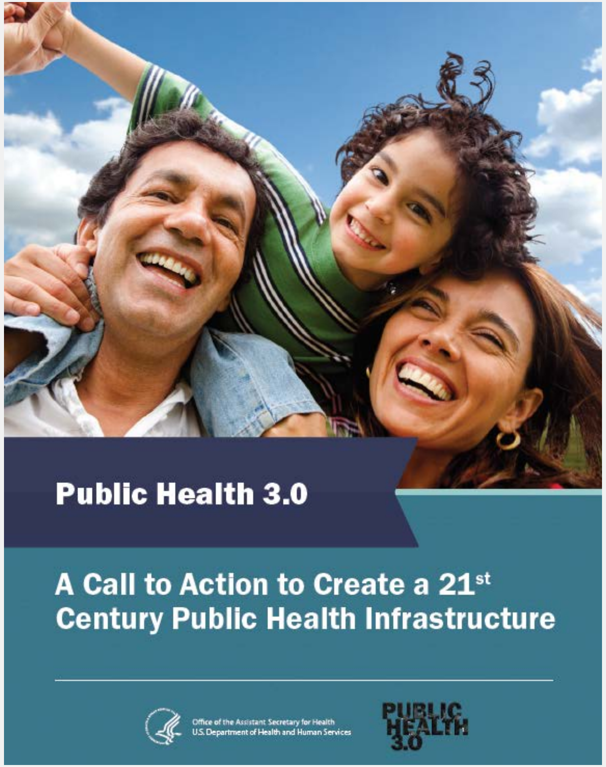
CDC

SUB-COUNTY DATA

Public Health 3.0: A Call to Action

- "A key need in PH3.0 is an understanding of how federal public health agencies can support local public health—especially with regard to data, metrics, and analytics tools."

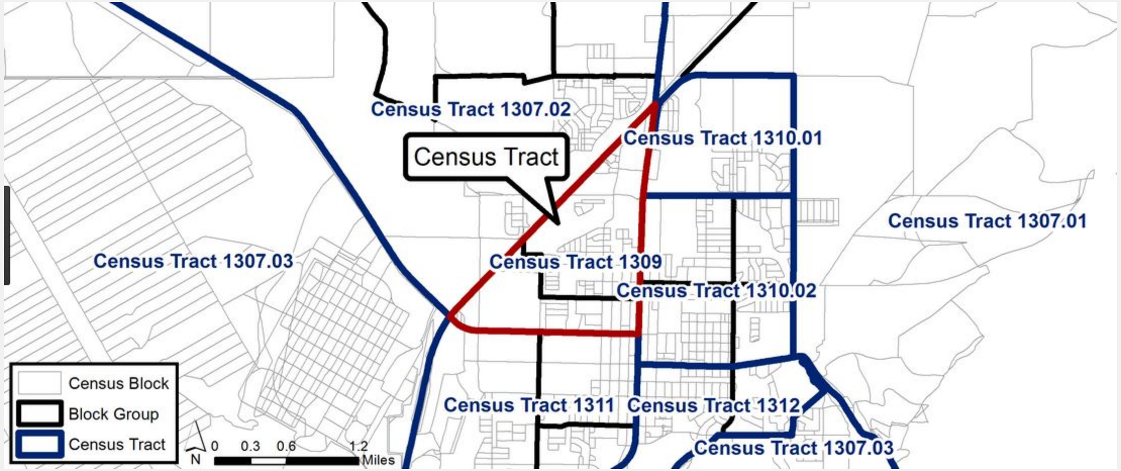
- "Today, a person's ZIP code is a stronger determinant of health than their genetic code."



Why use sub-county data?

- Small area data can:
  - Highlight local variation
  - Allow for a better understanding of EH processes and impacts
  - Improve surveillance
  - Target interventions

- Small area data can also:
  - Create data reliability issues
  - Confidentiality issues



FRAMEWORK

Question posed to Tracking:

Which denominator is best for calculating rates from any particular year?

Background

In small areas (census tracts, sub-county areas) small changes in the denominator (population count) may affect a rate in a much greater way than at a county- or state-level.

“Small numbers” problem

I propose that we ask

How are my rates different by population choice?

(a testable question—scientific method)

Instead of

Which denominator is best for calculating rates from any particular year?

Background—Denominator choice

Where can we find data for denominators?

- Population counts
  - Public sources
  - Private sources
  - Academic sources (not discussed here)

Public vs private sources

- Public (like Census): free, methods explained
- Private (like GeoLytics): need to pay for access, black box on methods—do they start with census numbers and improve them?

Only considered public sources for this project

Project

AIM:

Develop repeatable framework for comparing rates generated using different denominators

Methodology – Compare rates

- **Distribution**
  - Centiles
  - Visual checks
- **Geography to geography**
  - Percent difference
  - “Extreme” geographies
- **Applied public health perspective**
  - Choropleths
  - LISA analysis

Methodology

- **Distribution**
  - Centiles
  - Visual checks—QQ Plots and kernel density plots

“What is the shape of the distribution?” “Do the distributions look similar?”

Methodology

- **Geography to geography**
  - Percent difference
  - “Extreme” geographies

“Will the same rate be categorized differently by denominator?” “What is the magnitude in difference between one outcome over different denominators?”

Methodology

- **Applied public health perspective**
  - Choropleths
  - LISA analysis

“What are the conclusions we might draw from the data?” “What will happen when you analyze the data? Will you come to the same conclusions?”

About these data

- Using census tract data submitted from 4 Tracking states during Fall 2018
- CT, MO, NH, and RI
- Rate = health outcome/denominator *10,000
- Health outcome → number of visits to the emergency room for AMI during 2015 (by census tract)
- Denominators→ 5 year ACS ending in 2015 and the 2010 decennial census

About these data

- Using census tract data submitted from 4 Tracking states during Fall 2018
- CT, MO, NH, and RI
- Rate = health outcome/denominator *10,000
- Health outcome → number of visits to the emergency room for AMI during 2015 (by census tract)
- Denominators→ 5 year ACS ending in 2015 and the 2010 decennial census

* 2 outlier census tracts removed

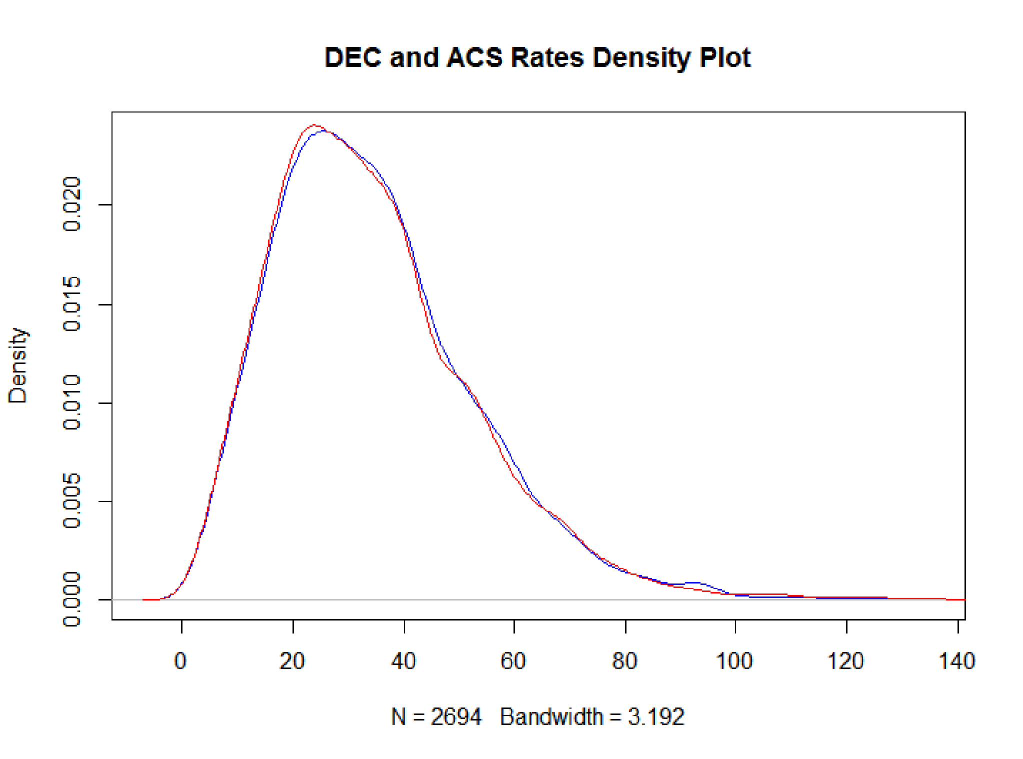
RESULTS

Distribution—Centiles

	0%	1%	5%	10%	25%	50%	75%	90%	95%	99%	100%
DEC	2.61	5.34	10.15	13.93	21.54	32.05	44.61	58.63	68.01	91.75	126.15
ACS	2.42	5.15	9.99	13.76	21.25	31.66	44.14	58.72	68.63	92.61	135.82

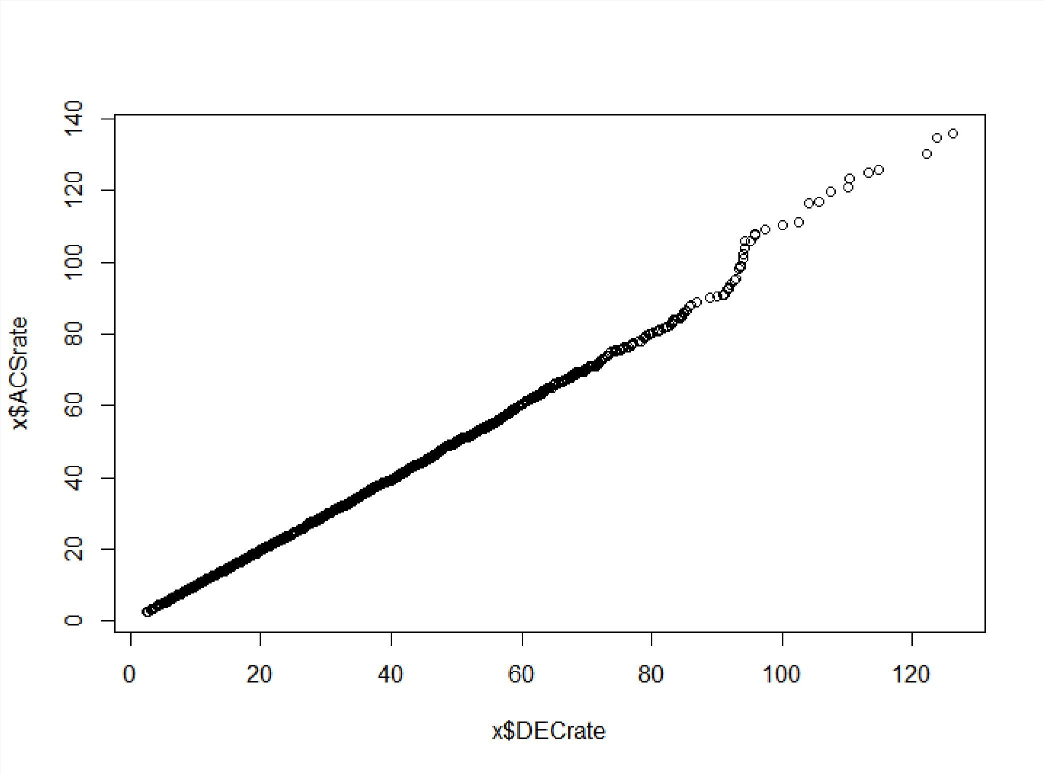
No statistical test for this—but nothing appears largely deviant except for the extreme values at highest centile

Distribution—Visual check



Distributions look similar—again, no statistical test

Distribution—Visual check



Similar until the largest rates (normal for QQ plots)

Geography to geography—Percent difference

	0%	1%	5%	10%	25%	50%	75%	90%	99%	100%
Percent Difference	0.00087	0.030358	0.176536	0.380088	1.042493	2.256529	4.083187	6.22614	12.32264	29.7782

90% of the rates for the census tracts have rates that are about 6% different or less

This is a proxy for percent difference between denominators

Geography to geography— “extreme” geographies

How the Decennial survey would categorize the census tracts

How the ACS would categorize the census tracts

	Very low	Low	Average	High	Very High
Very low	125	9	0	0	0
Low	9	109	18	0	0
Average	0	17	2105	26	5
High	0	0	30	85	20
Very high	0	0	1	24	110

Defining grouping…
Very low = 0-5th, low = 6-10th, average 11-89th, high = 90-94th, very high = 95-100th

Want to look at **misclassifications** that are more than one category different

Geography to geography— "extreme" geographies

How the Decennial survey would categorize the census tracts

How the ACS would categorize the census tracts

	Very low	Low	Average	High	Very High
Very low	125	9	0	0	0
Low	9	109	18	0	0
Average	0	17	2105	26	5
High	0	0	30	85	20
Very high	0	0	1	24	110

Defining grouping…

Very low = 0-5th, low = 6-10th, average 11-89th, high = 90-94th, very high = 95-100th

Most concerned by **misclassifications** that are more than one category different

Applied PH perspective

Applied public health perspective

- **Choropleths**
- LISA analysis

Choropleths made with the same bins might look different if geographies have large differences in value based on the denominator used to in the calculation of the rate

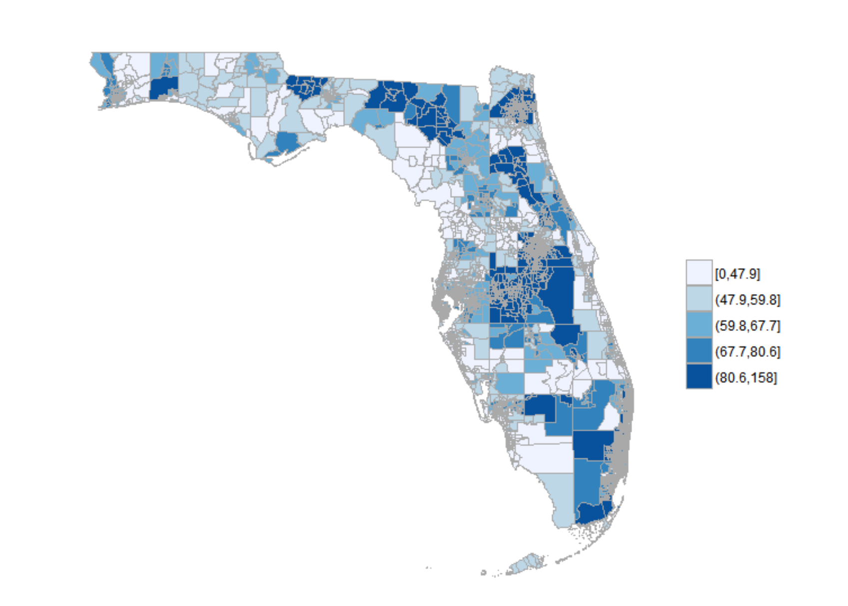
Applied PH perspective

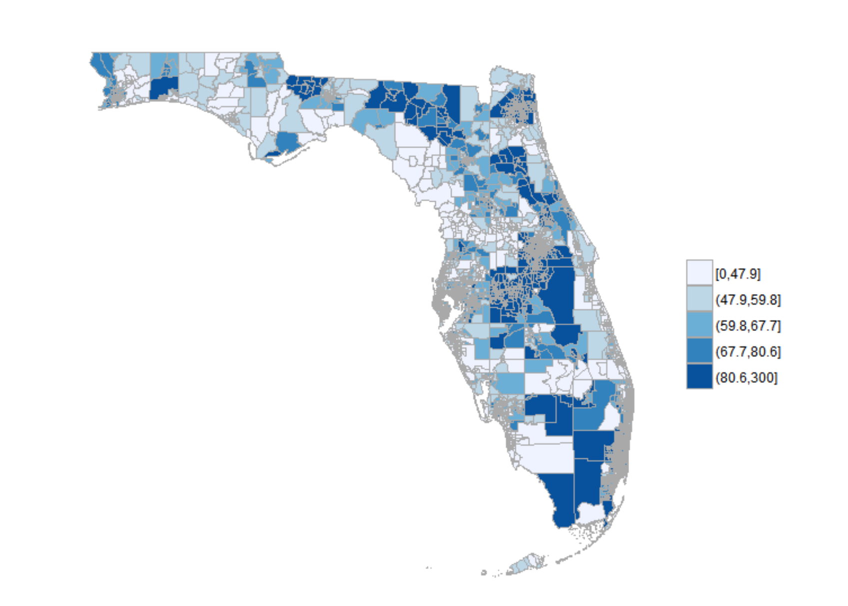
Applied public health perspective

- Choropleths
- **LISA analysis**

Commonly used in spatial analysis—defines clusters

Applied PH perspective—choropleths Florida





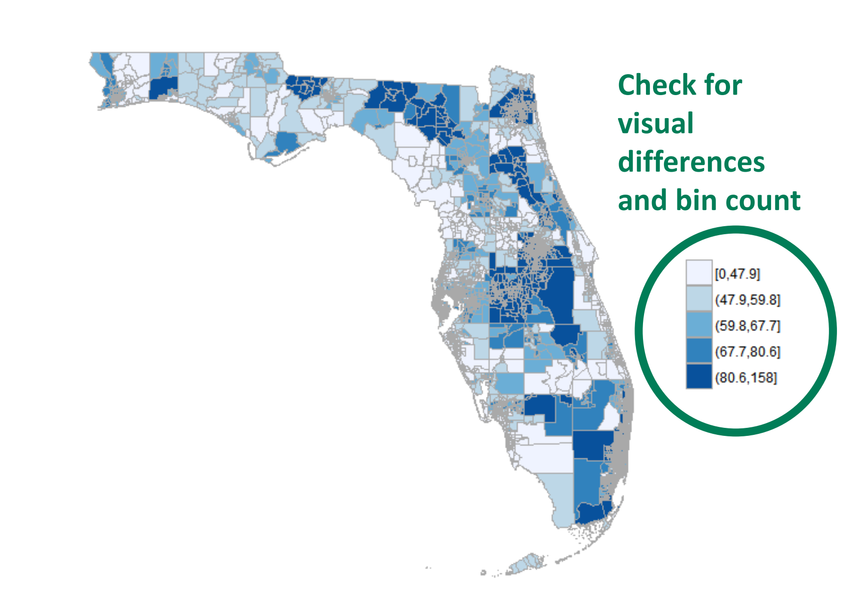
69% of census tracts were classified the same bins by both denominators

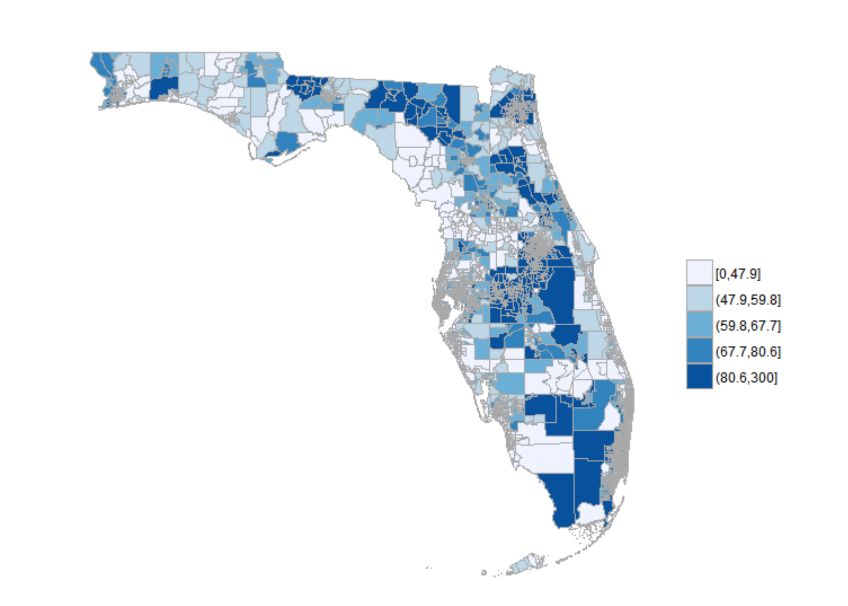
0.5% of census tracts “changed” by 2 or more bins

Caution: Increasing or decreasing the number of bins will change the rate of agreement. The researcher should test multiple bin categories for a thoroughly demonstration of the potential changes

*expected census tract counts

Applied PH perspective—choropleths Florida



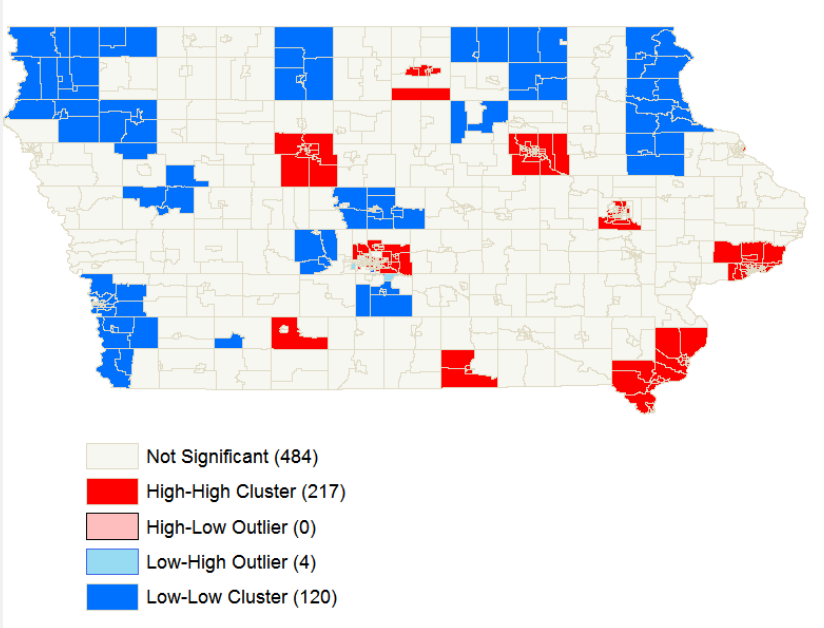


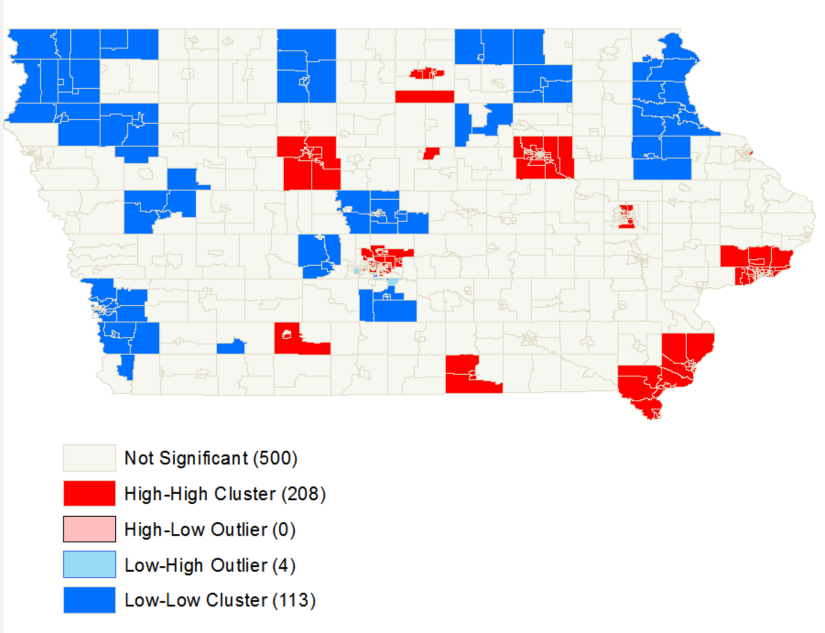
69% of census tracts were classified the same bins by both denominators

0.5% of census tracts "changed" by 2 or more bins

Caution: Increasing or decreasing the number of bins will change the rate of agreement. The researcher should test multiple bin categories for a thoroughly demonstration of the potential changes

Applied PH perspective—LISA analysis





The changes are small with only 32 of 825 census tracts changing clusters.

*SAT scan

Recap

- **Distribution**
  - Centiles
  - Visual checks
- **Geography to geography**
  - Percent difference
  - “Extreme” geographies
- **Applied public health perspective**
  - Choropleths
  - LISA analysis

Recap

Agreement: Looking at how rates and classifications agree by denominator selection

Usefulness: Can be repeated across years and different geography levels

Subjective: Identifying a "meaningful difference"

Investigation: This analysis framework is a starting point, further investigation may be required for geographies that change/do not agree consistently through denominator choice

Population count vs rate: our experience is that doing the analysis with rates provided better information than raw population counts, but a number of pieces of this analysis can be run with population

Limitations

- Crude rates—not age adjusted
- Exploratory only; does not explain differences
- Publically available datasets which may “use” one another
- Other metrics exist (agreement statistics)

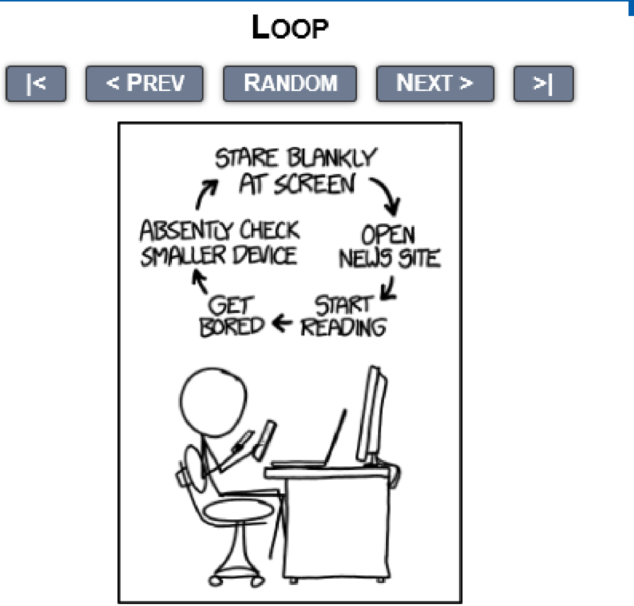
Future Analyses

- Age-adjusted rates
- Alternate population source
- Other health outcomes
- Testing sensitivity

Acknowledgements

- Angie Werner
- Lance Waller
- Pilot participants—Jessie Sagona (NH), Lillian Morris (WA), Jen Shriber

Thank you!



https://xkcd.com/51411/

For more information, contact NCEH
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov