Sidestepping the box: Designing a supplemental poverty indicator for school neighborhoods

Doug Geverdt
National Center for Education Statistics

Laura Nixon
U.S. Census Bureau

2018 Annual meeting of the Federal Committee on Statistical Methodology
March 9, 2018 Washington, DC

Poverty Data

- Poverty impacts a variety of educational outcomes
- Poverty data needed for programs -1974 ESEA (Title I)
- Problem: Level of importance < > available indicators
- Common sources for educational programs:
  - School district Title I estimates (Census Bureau)
  - Title I school indicator (NCES-CCD)
  - National School Lunch Program (USDA/FNS; NCES-CCD)

NSLP program data

- Standard measure for school poverty (and SES proxy)
- Lunch data has many benefits for research/administration:
  - Near universal participation by school systems
  - Stable program infrastructure; regularly updated and accessible
  - Cost-effective (for education purposes)
- Lunch data also has limitations
- Community Eligibility Provision (CEP) and 100% eligibility
- We need multiple measures of poverty in/around schools

Neighborhood proxies in educational research

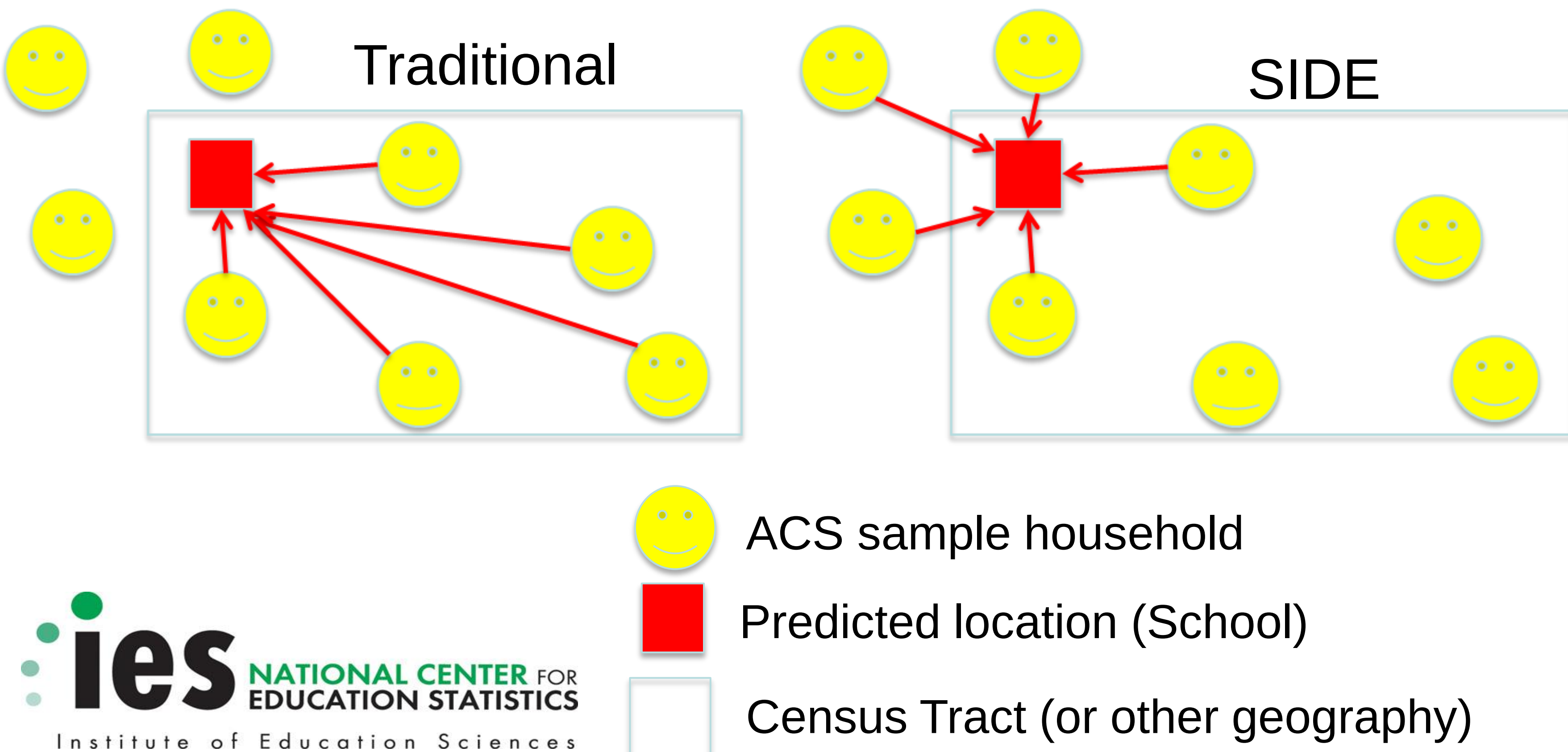
- How to define? How to operationalize?
- Source for all: American Community Survey (ACS)
- Census block groups, tracts, ZCTAs (ZIP)
- School districts
- Conclusions:
  - Proxies selected based on whether data are available
  - Proxies may provide a poor match (lack of construct validity)
  - Proxies pose difficulties for detecting neighborhood effects

Proposed solution: What if we…

- Re-frame problem to focus on point locations instead of polygons
- Use spatial statistics with ACS to create interpolated poverty surface
- Define neighborhoods as a set of neighbors nearest to schools
- Join school point locations to poverty surface to create estimates
- Result = poverty estimate for school neighborhoods
- Spatially interpolated demographic estimates (SIDE)

Linking outside the box

Estimates for non-sampled locations based on nearest neighbors instead of a default geographic container

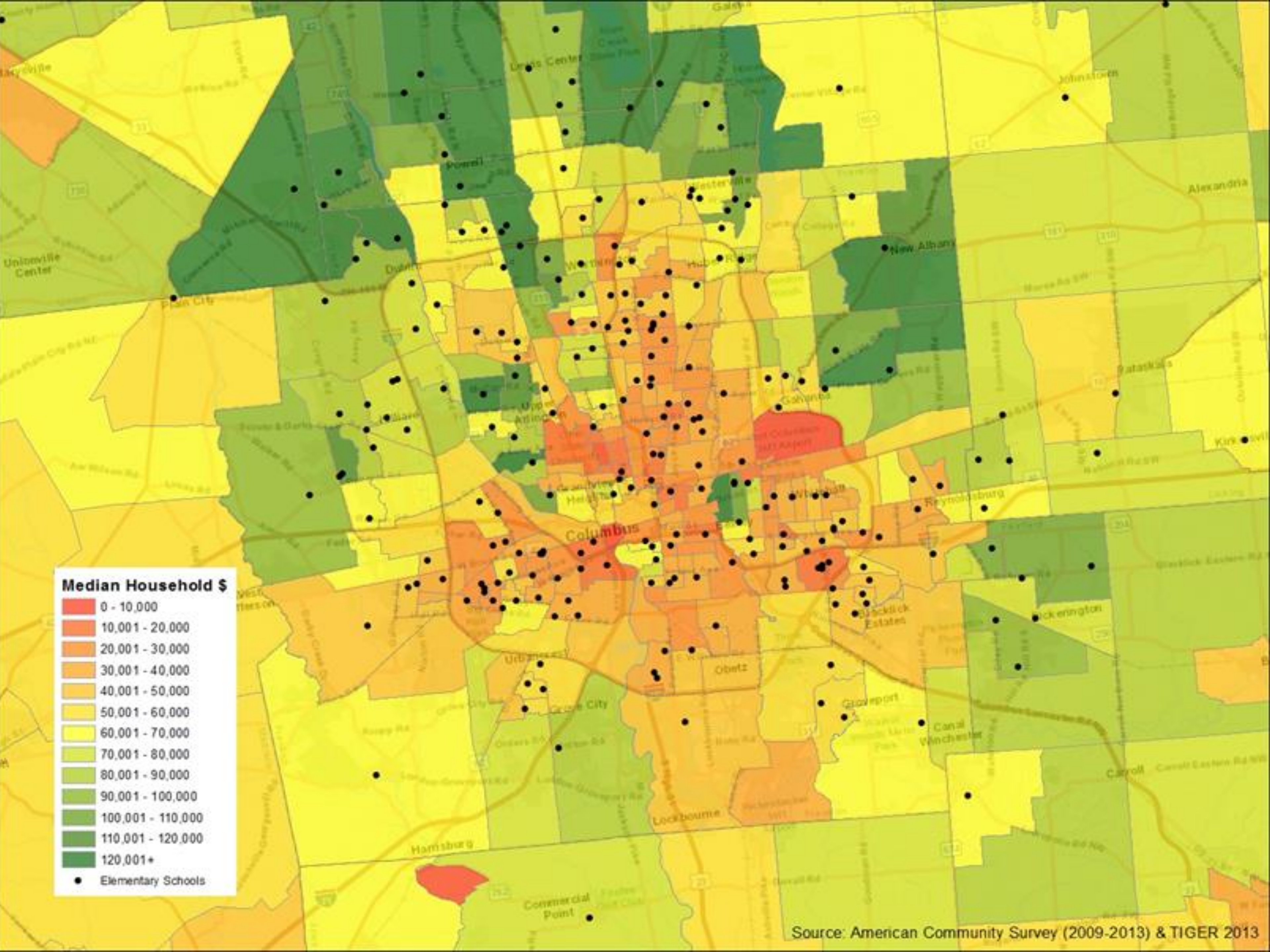


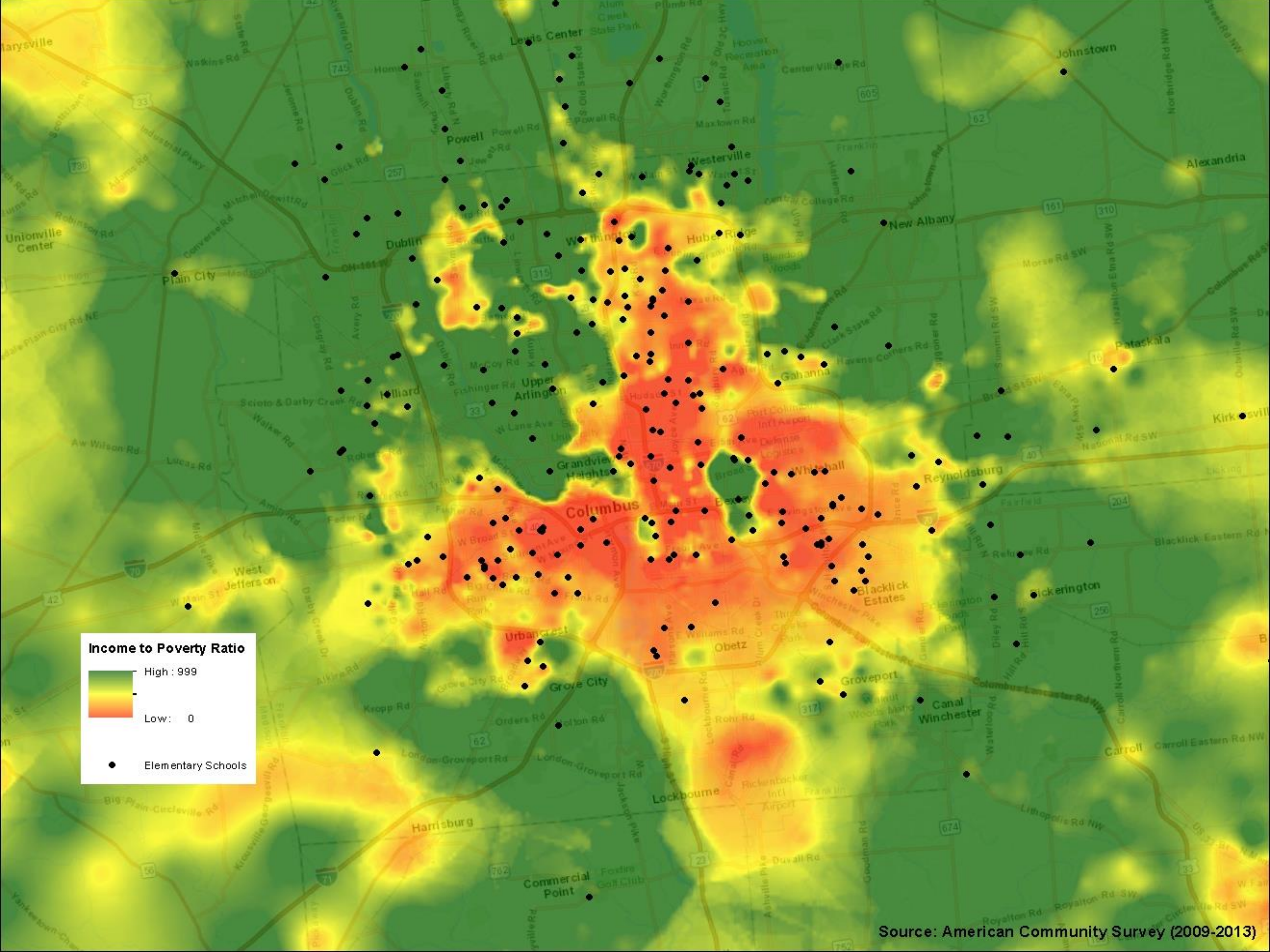
Case Study Prototype

- Elementary schools in Ohio (n=1,793)
- ACS households from 5-year sample (2009-2013)
- Census block centroids for household locations (TIGER 2013)
- Model the income-to-poverty ratio (IPR)
- Empirical Bayesian kriging for estimation (ESRI)
- Neighborhoods based on 25 nearest neighbors
- School locations from NCES (2012-2013)

Kriging

- Least squares interpolator that uses the weighted sum of values from measured locations to predict values in unmeasured locations. The closer the location, the greater the weight. (Cressie, 1993)
- Based on Tobler's first law of geography
- Kriging assumes data are spatially autocorrelated
- Kriging uses the data twice:
  - Modeled with a semivariogram to identify how dfferences vary by distance.
  - Applied to the data to predict values for unsampled locations
- Results in a continuous prediction surface across study region


Median Household $
0 - 10,000
10,001 - 20,000
20,001 - 30,000
30,001 - 40,000
40,001 - 50,000
50,001 - 60,000
60,001 - 70,000
70,001 - 80,000
80,001 - 90,000
90,001 - 100,000
100,001 - 110,000
110,001 - 120,000
120,001+
Elementary Schools
Columbus
Powell
Dublin
Westerville
New Albany
Gahanna
Reynoldsburg
Pickerington
Groveport
Canal Winchester
Obetz
Grove City
Urbancrest
Lockbourne
Harrisburg
Commercial Point
Plain City
Johnstown
Alexandria
Pataskala
Carroll
Hilliard
Upper Arlington
Worthington
Whitehall
Bexley
Blacklick Estates
Unionville Center
Source: American Community Survey (2009-2013) & TIGER 2013

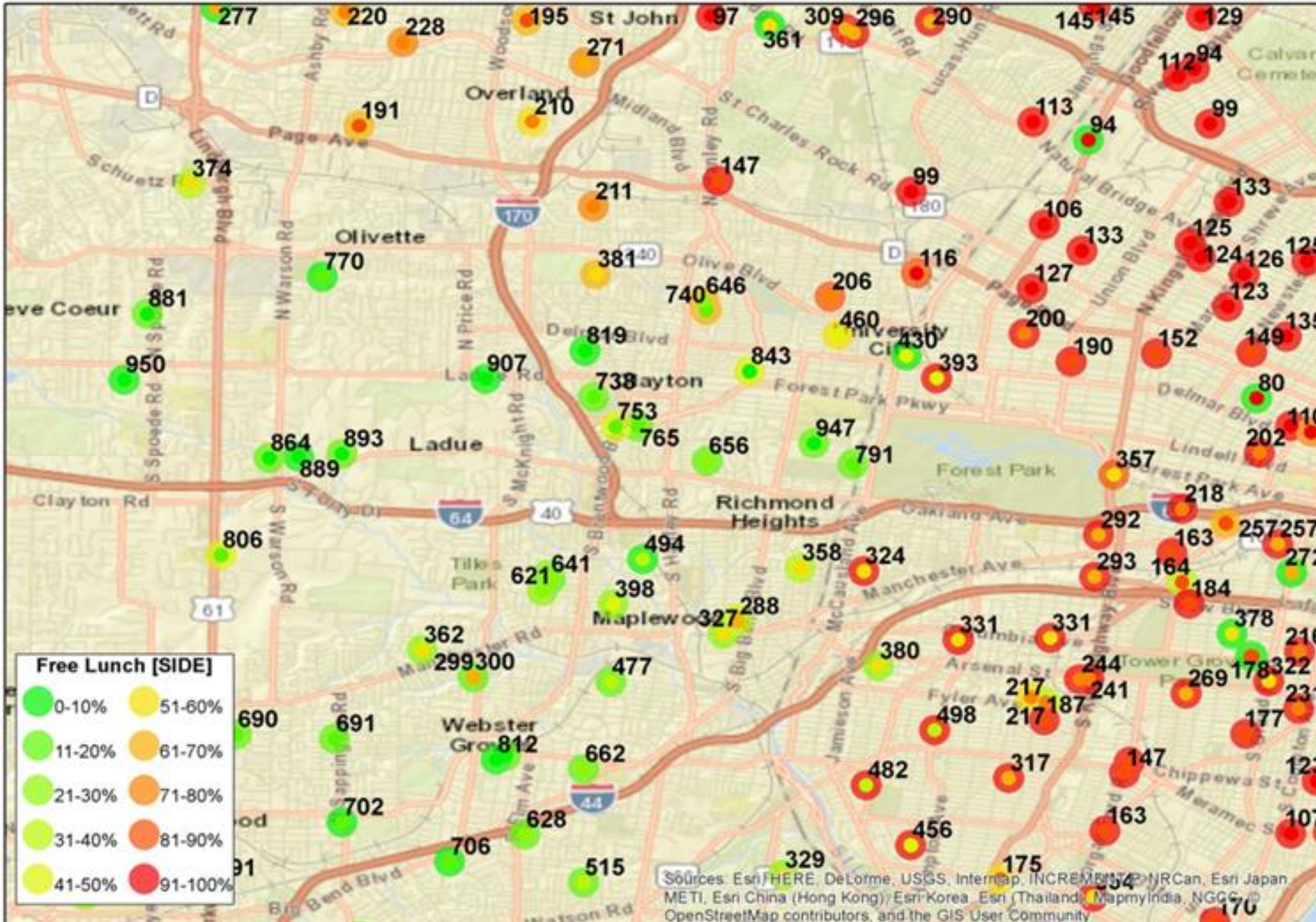

Income to Poverty Ratio
High : 999
Low : 0
Elementary Schools
Source: American Community Survey (2009-2013)
Columbus
Lewis Center
Powell
Westerville
Dublin
Worthington
Huber Ridge
New Albany
Johnstown
Alexandria
Unionville Center
Plain City
Hilliard
Upper Arlington
Grandview Heights
Gahanna
Whitehall
Bexley
Reynoldsburg
Pataskala
Kirkersville
West Jefferson
Blacklick Estates
Pickerington
Urbancrest
Grove City
Obetz
Groveport
Canal Winchester
Lockbourne
Carroll
Harrisburg
Commercial Point

Results

1. Estimates for household type by race followed the same pattern as general poverty estimates. No surprises.
2. Physical extent of SIDE neighborhoods tended to be smaller than census tracts, and estimates had a consistent sample size.
3. SIDE statistical prediction surface provided a more nuanced representation of school location vs. tract-level estimates.
4. Correlated with school lunch data (r = -.46)
5. Production was feasible (our server survived!)
6. SIDE estimates satisfied Census disclosure requirements

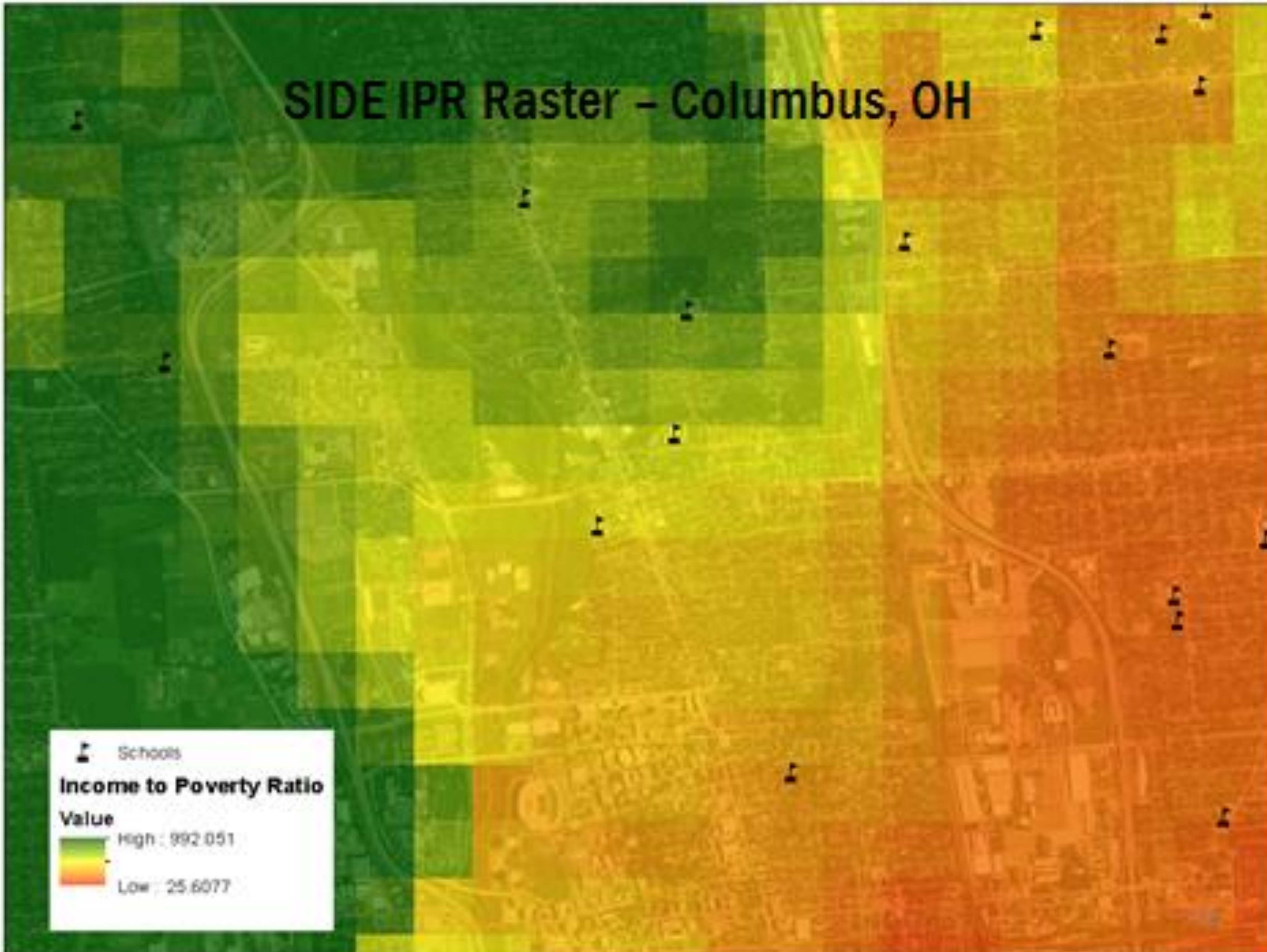
Potential benefits of SIDE IPR

- Leverage existing data sources (thus cost-effective)
- Provide a new covariate to help control for SES
- Create new poverty option for small areas (address-level)
- Avoid disclosure concern for small areas through modeling
- Avoid impact of boundary changes
- Identify mismatch between school and student neighborhoods


Free Lunch [SIDE]
0-10%
11-20%
21-30%
31-40%
41-50%
51-60%
61-70%
71-80%
81-90%
91-100%
St John
Overland
Olivette
Ladue
Clayton
Richmond Heights
Forest Park
Maplewood
Webster Groves
Tilles Park
Page Ave
Olive Blvd
Delmar Blvd
Forest Park Pkwy
Lindell Blvd
Manchester Ave
Clayton Rd
Big Bend Blvd
Watson Rd
Sources: Esri, HERE, DeLorme, USGS, Intermap, INCREMENT P, NRCan, Esri Japan, METI, Esri China (Hong Kong), Esri Korea, Esri (Thailand), MapmyIndia, NGCC, © OpenStreetMap contributors, and the GIS User Community

Potential benefits of SIDE IPR

- Leverage existing data sources (thus cost-effective)
- Provide a new covariate to help control for SES
- Create new poverty option for small areas (address-level)
- Avoid disclosure concern for small areas through modeling
- Avoid impact of boundary changes
- Identify mismatch between school and student neighborhoods
- Offer potential public SIDE IPR layer to create IPR estimates for other addresses or locations


SIDE IPR Raster – Columbus, OH
Schools
Income to Poverty Ratio
Value
High : 992.051
Low : 25.6077

Next steps

- R&D report – rationale and details of case study
- Experimental file of IPR estimates for *all* public schools
- Explore additional covariates
- Continue working on SIDE raster data product
- Explore potential internal applications for NCES programs.
- Gather feedback

Questions?

Doug Geverdt
National Center for Education Statistics
douglas.geverdt@ed.gov

Laura Nixon
U.S. Census Bureau

2018 Annual meeting of the Federal Committee on Statistical Methodology
March 9, 2018 Washington, DC