Geography and Usability of the American Community Survey

Seth Spielman
Assistant Professor of Geography
University of Colorado
Goals

1. To convince you that the margins of error from the American Community Survey (ACS) are too high for many policy relevant variables.

2. Argue that this is a geographic problem that can be fixed with a new class of statistical geographies that are data-driven and built to yield estimates that meet specific usability thresholds.

3. Demonstrate that it is possible to efficiently create an ACS specific 1, 3, and 5 year geography that would cover the US AND provide usable data.
Since the 1930’s US cities have been divided, like a jigsaw puzzle, into statistical geographies such as census tracts and block groups.

- These are geographic entities that exist solely for the purpose of statistical tabulation.
- **Criteria** for these geographies are **stability** and **population thresholds**
The American Community Survey

• The American Community Survey (ACS) is the primary national source for demographic and economic data about neighborhoods.

• The ACS produces over 1400 tables for 74,000+ tracts and 200,000+ block groups each year.

• Data is widely used, but ...
Los Angeles Tracts: ACS 2007-2011
Number of Children Below Poverty Line

<table>
<thead>
<tr>
<th>CENSUS TRACT</th>
<th>KIDS =&lt;5 IN POVERTY</th>
<th>MARGIN OF ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census Tract 2048.10, Los Angeles County, California</td>
<td>92</td>
<td>142.2</td>
</tr>
<tr>
<td>Census Tract 2060.10, Los Angeles County, California</td>
<td>99</td>
<td>115.0</td>
</tr>
<tr>
<td>Census Tract 2122.03, Los Angeles County, California</td>
<td>61</td>
<td>174.1</td>
</tr>
<tr>
<td>Census Tract 1913.01, Los Angeles County, California</td>
<td>25</td>
<td>170.1</td>
</tr>
<tr>
<td>Census Tract 1098, Los Angeles County, California</td>
<td>55</td>
<td>141.3</td>
</tr>
<tr>
<td>Census Tract 1204, Los Angeles County, California</td>
<td>95</td>
<td>151.3</td>
</tr>
<tr>
<td>Census Tract 1976, Los Angeles County, California</td>
<td>12</td>
<td>165.6</td>
</tr>
<tr>
<td>Census Tract 1349.03, Los Angeles County, California</td>
<td>35</td>
<td>172.9</td>
</tr>
<tr>
<td>Census Tract 2060.32, Los Angeles County, California</td>
<td>291</td>
<td>162.9</td>
</tr>
<tr>
<td>Census Tract 2124.20, Los Angeles County, California</td>
<td>133</td>
<td>125.8</td>
</tr>
<tr>
<td>Census Tract 1233.01, Los Angeles County, California</td>
<td>87</td>
<td>146.6</td>
</tr>
<tr>
<td>Census Tract 1414, Los Angeles County, California</td>
<td>100</td>
<td>169.5</td>
</tr>
<tr>
<td>Census Tract 1276.04, Los Angeles County, California</td>
<td>54</td>
<td>109.7</td>
</tr>
<tr>
<td>Census Tract 1236.02, Los Angeles County, California</td>
<td>111</td>
<td>159.7</td>
</tr>
</tbody>
</table>
US Census Tracts: ACS 2009-2013
Number Of Children Under 6 In Poverty

Graph truncated at the 90th percentile to facilitate plotting

Margin of Error = 100% of Estimate

Margin of Error = 50% of Estimate

Margin of Error = 10% of Estimate

45% of Tracts MOE greater than Estimate

49% of Tracts MOE 50-100% of Estimate

Margin of Error (90%)
US Census Tracts: ACS 2009-2013
Number People In Poverty

Estimate

Margin of Error = 100% of Estimate
Margin of Error = 50% of Estimate
Margin of Error = 10% of Estimate
US Census Tracts: ACS 2009-2013
Median Household Income

Margin of Error (90%)

Margin of Error = 100% of Estimate
Margin of Error = 50% of Estimate
Margin of Error = 10% of Estimate
Social and Geographic Patterns in Data Quality

Median household Income Estimates, 2011 ACS.
Metro-Scale Variation in Estimate Quality

2011 ACS: Census Tracts.
Some Variables Are Better Than Others…
Temporal Dynamics

Has the number of poor kids in Boulder, CO changed?

Low 142 in 2012
High 743 in 2011

ACS 1 years estimates 2005-2013
A Geographic Problem?
Estimate Quality in the ACS

- The quality of ACS estimates is managed through temporal and geographic aggregation using geographies that existed prior to the advent of the “rolling” ACS.

  - These geographies defined using criteria that are **exogenous** to the survey; geographies independent from the data they summarize.

- This doesn’t work well...

  - Few places get 1 and 3 year estimates.

  - Margins of error on 1, 3 and 5 year estimates are often very high!
## Managing Estimate Quality in the ACS

<table>
<thead>
<tr>
<th>Geographic Areas</th>
<th>1-year estimates areas w/ 65,000 +</th>
<th>3-year estimates areas w/ 20,000 +</th>
<th>5-year estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>count</td>
<td>percent</td>
<td>count</td>
</tr>
<tr>
<td>County</td>
<td>825</td>
<td>26%</td>
<td>1,909</td>
</tr>
<tr>
<td>Census Tract</td>
<td>0</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>Block Group</td>
<td>0</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>Place</td>
<td>568</td>
<td>2%</td>
<td>2,157</td>
</tr>
<tr>
<td>MSA/M-MSA</td>
<td>530</td>
<td>55%</td>
<td>934</td>
</tr>
</tbody>
</table>
Geographies don’t fit the survey

• Margins of error are high.
• Few places receive 1 and 3 year estimates.
• Statistical geographies could be redesigned to yield better, more usable data.
New Criteria For ACS Statistical Geographies
## Criteria for Census Tracts


<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
<th>Active Since</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable boundaries</td>
<td>Boundaries remain constant over time</td>
<td>1934</td>
</tr>
<tr>
<td>Population thresholds</td>
<td>Min/max population</td>
<td>1934</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>Pop/housing stock should be approx. homogenous within tract</td>
<td>1934-1990</td>
</tr>
<tr>
<td>Shape</td>
<td>“The length should not be more than 3 times the width”</td>
<td>1937 p.6</td>
</tr>
<tr>
<td>Usability/Data Quality</td>
<td>Zones designed to yield efficient estimates.</td>
<td>1934 p.6, 1937 p.6</td>
</tr>
</tbody>
</table>
New Criteria for Statistical Geographies

- The traditional approach is to manually draw statistical geographies then employ sophisticated small area estimation procedures using those areas.

- It's possible to think about **usability criteria for statistical geographies**.
  
  - Early tract manuals emphasized usability and estimate quality.
  
  - Given some criteria we can identify “optimal” geographies.
  
  - Maximize geographic detail for some specific usability threshold.
  
  - For example, a set of geographies for which estimates have a margin of error that is no more than 20% of the estimate and the population is no less than 4000 people.
  
- This is computationally intensive but “tractable.”
Tracts

ACS Geography

Percent of Pop w/ Bachelor's Deg.

(2,25] [25,35] [35,50] [50,63] [63,96]
Areas that are reddish have a CV over 12%
## Criteria for ACS Areas

<table>
<thead>
<tr>
<th>ACS Areas</th>
<th>Current Tract Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability Thresholds</td>
<td>Population Threshold</td>
</tr>
<tr>
<td>Community engagement in survey response (response rate/mode drives geographic resolution)</td>
<td>Community engagement in design</td>
</tr>
<tr>
<td>Created from lower level small areas</td>
<td>Created from lower level small areas, nested within counties</td>
</tr>
</tbody>
</table>
Conclusions

• Tried to show that there are problems with ACS data.

• Presented the idea of **Usability-Based ACS Geographies**.
  
  • These new geographies don’t alter the status-quo and can be efficiently created.

  • Provide usable 1, 3, and 5 year estimates to the entire US.

  • We have created 5yr geographies for a bunch of cities. **Help with evaluation and dissemination.**
Uses of ACS Data

Survey of planners (n=169)

Many of these tasks are difficult with current data.
Conclusions

• There is not a single set of “optimal” geographies. A set of units drawn to optimize housing estimates may look different from one drawn for transportation estimates (different phenomena, different organization in space).

• A compromise geography?

• Requires trading geographic resolution for attribute resolution.
Acknowledgments

• David Folch, Florida State Geography has been central to much of work I presented here.

• The NCRN initiative has been a fantastic opportunity.

• Todd Gardner, Mike Ratcliffe and others at the Census Bureau who have taken the time to discuss ideas and data.

• Amy Griffin (University of New South Wales) and Jason Jurevitch (Portland State).

