

Selected Slides from
New Spatially Referenced Microdata

Steve Ruggles
Minnesota Population Center

TXRDC Segregation Workshop
September 16, 2016
Texas A&M University

Limitations of segregation analysis using aggregate data

- Indices are constrained by census geography
 - Tracts, block groups, other available units are not necessarily the relevant kinds of units
 - PSHP example: block fronts, not grid cells
- Tract and block boundaries are not consistent over time
 - It you want to measure change, you must interpolate

GeoLytics Neighborhood Change Database

Provides estimates of tract characteristics in 1970, 1980, 1990, and 2000 using Census 2010 tract boundaries. Liabilities for academic research:

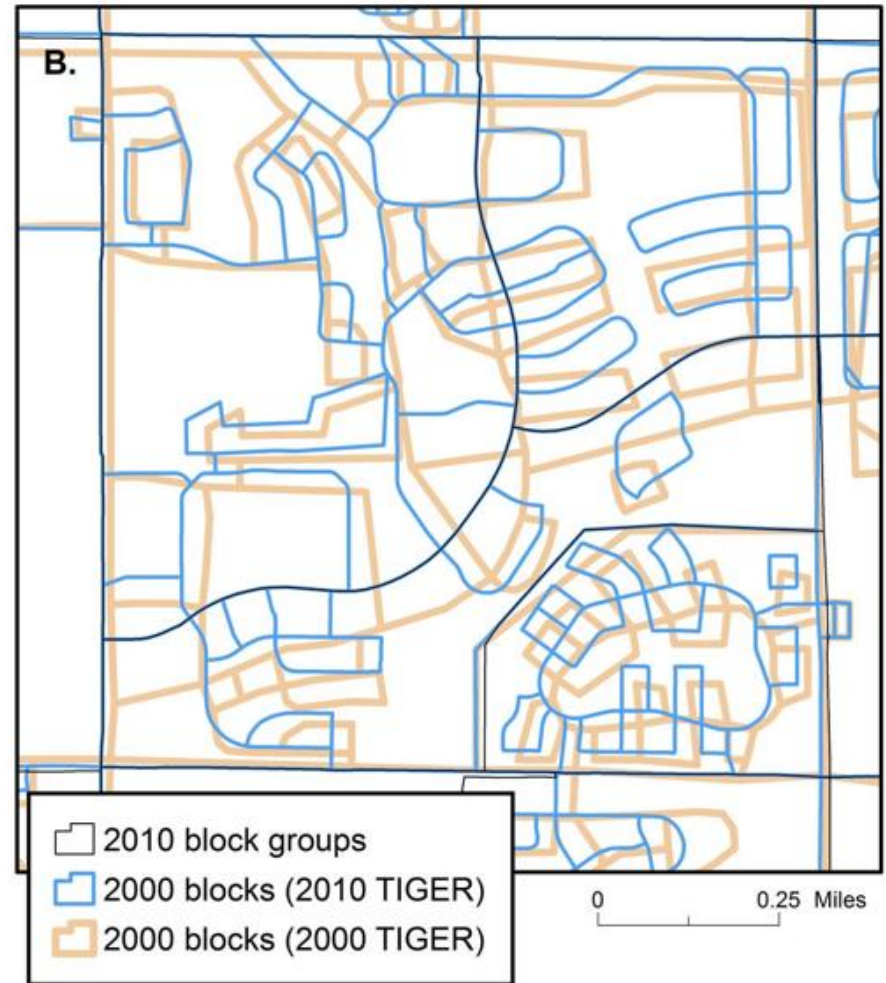
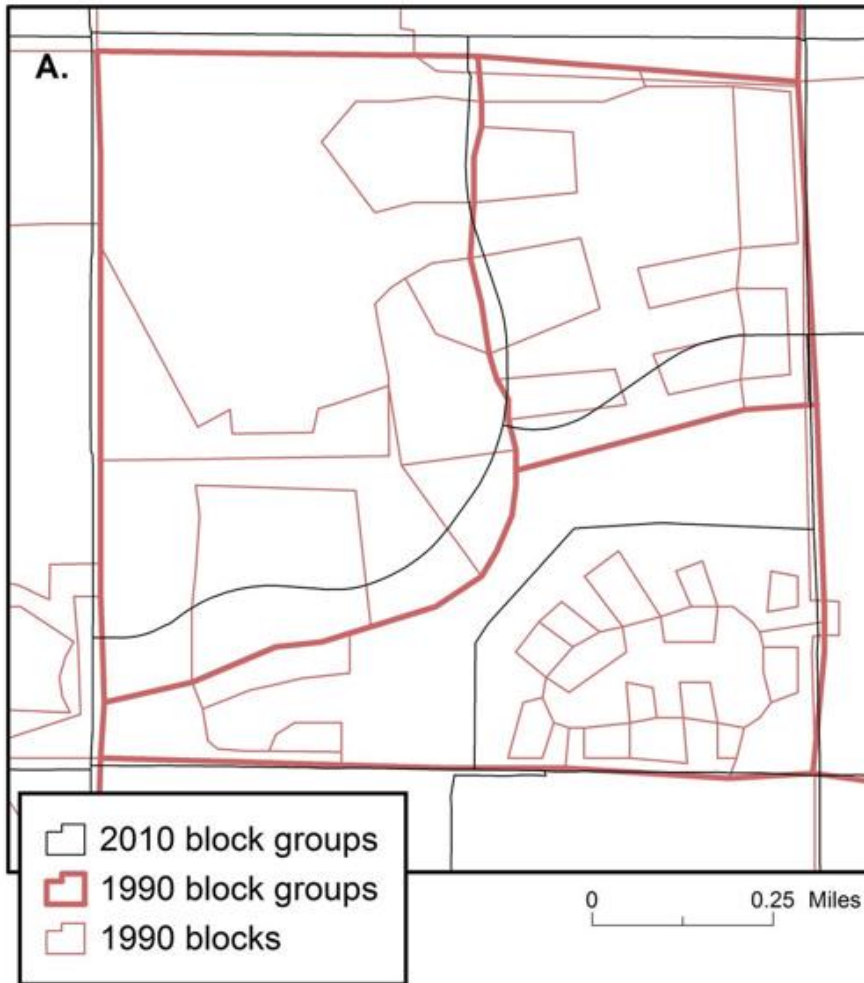
1. GeoLytics uses total population counts of blocks or block groups to reallocate all variables to Census 2000 boundaries.
2. GeoLytics assumes that the spatial distribution of population within tracts in 1970 and 1980 was identical to the 1990 distribution.

Geolytics

- GeoLytics estimates become increasingly blurry the further back they go. We would expect inter-tract differences to be systematically understated in 1980 and even more so in 1970.
- GeoLytics-based estimates of long-run trends in residential segregation probably understate segregation in the two earliest censuses.

New Alternatives

- US2010 Longitudinal Tract Data Base (LTDB) provides interpolated 1970, 1980, 1990, and 2000 estimates for 2010 tracts (Logan et al).
- NHGIS geographically standardized tables. These are new, limited coverage so far. NHGIS reaggregates data from the smallest source units for which the data are available using all variables. NHGIS also supplies lower and upper bounds based on the spatial relationship between the source units and standard units.



Examples of boundary misalignment in Broomfield County, CO. (A) The boundaries of 1990 blocks and block groups, as defined in 2000 TIGER/Line files, are imprecise and align poorly with 2010 boundaries from the improved 2010 TIGER. (B) The 2000 blocks offer a bridge between the two TIGER vintages; the 2000 TIGER version aligns well with 1990 blocks, and the 2010 TIGER version aligns well with 2010 area definitions.

Advantages of Microdata

Potential to overcome boundary changes

- Consistent units can be identified by aggregating lower geographic levels
- Many sources have the potential for full geocoding (long-lat), which enables novel measures of segregation
- In historical data, neighbors can be inferred.

Geographic sub-county source variables in historical microdata (1850-1940)

- ED (1850-1940)
- MCD (1850-1940)
- Ward (1850-1940, selected cities)
- Address (1880-1940, urban areas)
- Tract (1910-1940, selected cities)
- Block (1920-1940)

Consistent Geography in IPUMS

- SEA (1850-1950)
- CONSPUMA (1980-2000, ACS 2000-2012)
- CPUMA0010 (2000-2010, ACS 2013-present)
- 1940: Approximate longitude Latitude
- Logan: Geocodes for major cities, 1880.
1940 (and 1930) forthcoming
- Nagel-Buttenfield: imputed tract (1980-present)
- PUMA Density (1960-present)

Geographic source variables in internal Census Bureau microdata (1960-Present)

- All datasets have blocks, tracts
- No boundaries for blocks available before 1990
- 1990 has addresses
- 2000, 2010 are PIKed and linked to MAF, providing longitude-latitude

Census Bureau Microdata (1940-Present)

Proposed consistent geographies:

Table 1	
Integrated Census Geography	
Consistent County	1940-2020
Super-tract 1	1940-2020
Super-tract 2	1960-2020
1980 Tracts	1980-2020
1990 Tracts	1990-2020
2000 Tracts	1990-2020
2010 Tracts	1990-2020
1990 Block Groups	1990-2020
2000 Block Groups	1990-2020
2010 Block Groups	1990-2020