

Introduction

The geographic aggregation tool (GAT) is a free tool that runs in R to create aggregated areas. GAT was developed by New York State (NYS) Environmental Public Health Tracking (EPHT) to create meaningful sub-county areas to display environmental health indicators. This could help identify patterns of disease that may be masked at county level while allowing stable rates and protecting confidentiality.

GAT's process

GAT runs a series of steps in 4 sections:

1. Request user inputs, including:
 - a. Values to aggregate (population, cases, deaths)
 - b. Aggregation method to use (geographic or population weighted centroid, similar ratio)
 - c. Special conditions or exclusion criteria
 - d. Paths to read and save files
2. Run aggregation
3. Create maps for evaluation
4. Output shapefiles and documentation

Aggregation types

The geographic centroid method aggregates to the area with the closest geographic centroid. The similar ratio method selects areas to aggregate based on a ratio of two numeric values, for example percent of population living below poverty and rate of a disease. Areas are aggregated to their neighbors based on which neighbor has the most similar ratio of values.

Sources:

Stamm & Babcock. gatpkg: Geographic Aggregation Tool (GAT). <https://github.com/ajstamm/gatpkg>

US Census Bureau. TIGER/Line Shapefiles. <https://www2.census.gov/geo/pvs/tiger2010st/>

US Census Bureau. American Community Survey. <https://www.census.gov/programs-surveys/acs>

Funding: CDC EPHT grant #NUE1EH001341-03-01

Comparing aggregation methods

The maps at right show aggregated areas using each method in Rensselaer and Albany, NY. In the geographic centroid method, tracts joined to tracts with similar poverty levels, based on geographic principles where adjacent areas tend to be more similar, as well as very different poverty levels. In Figure 1, borders of areas with mixed poverty levels 2-3 quantiles apart are highlighted. In similar ratio aggregation using median income and percent minority population, tracts joined mostly to tracts with similar poverty levels. In Figure 2, borders of areas with similar poverty levels are highlighted, in contrast to the same areas in Figure 1.

Aggregation model	R2
geographic centroid	0.8253
similar ratio	0.8973

Table 1: Aggregation model results, Albany & Rensselaer Counties

In preliminary analyses, a higher proportion of the variance in percent poverty was explained by the similar ratio model than the geographic centroid model (see Table 1). This suggests the similar ratio model results in more homogeneous areas than the geographic centroid model.

Conclusion

Users can customize GAT's aggregation process to create small areas that meet their project needs. When aggregated on variables relevant to health equity, GAT can create areas useful for exploring social determinants of health.

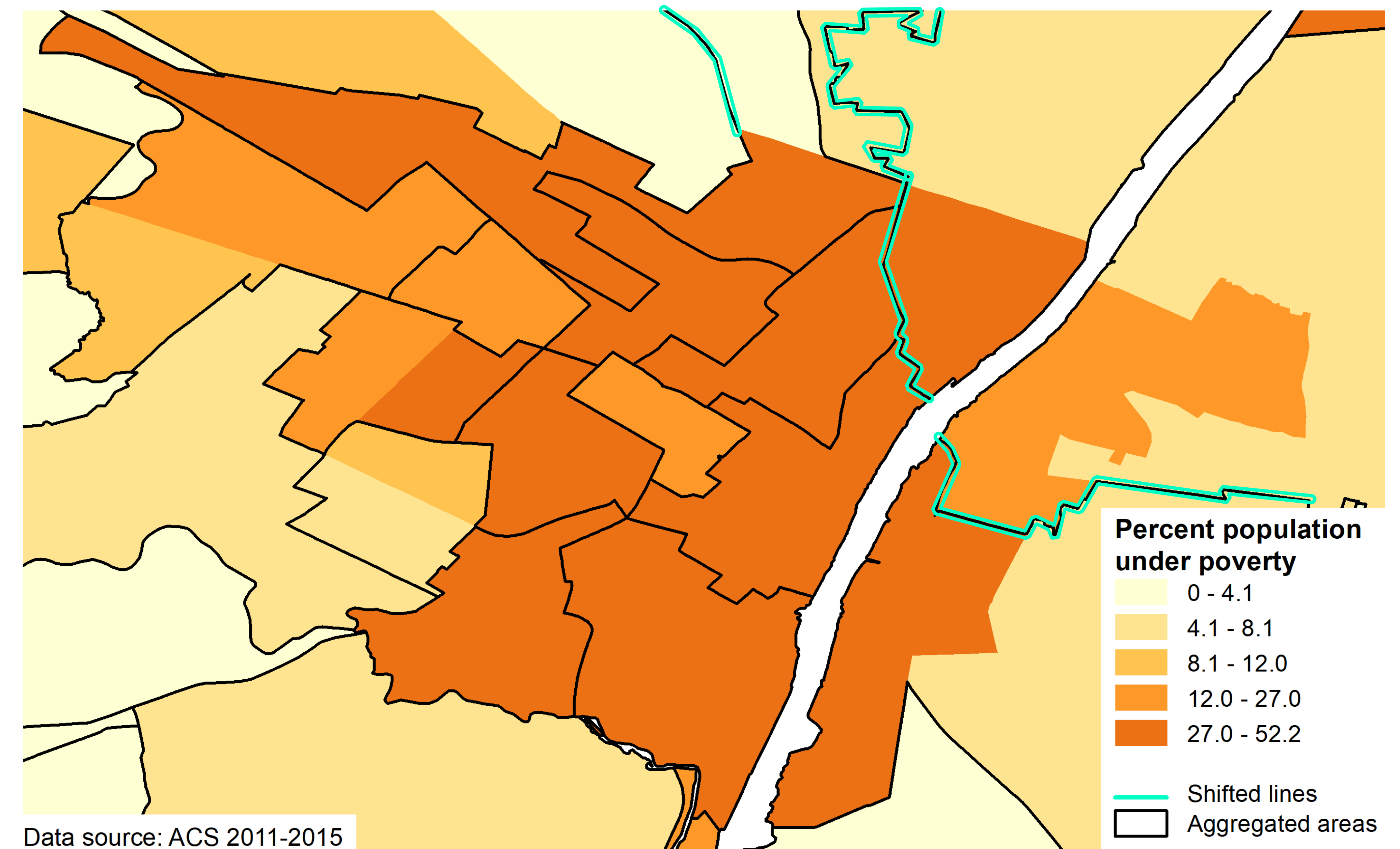


Figure 1: Geographic centroid aggregation, Albany & Rensselaer cities, NY

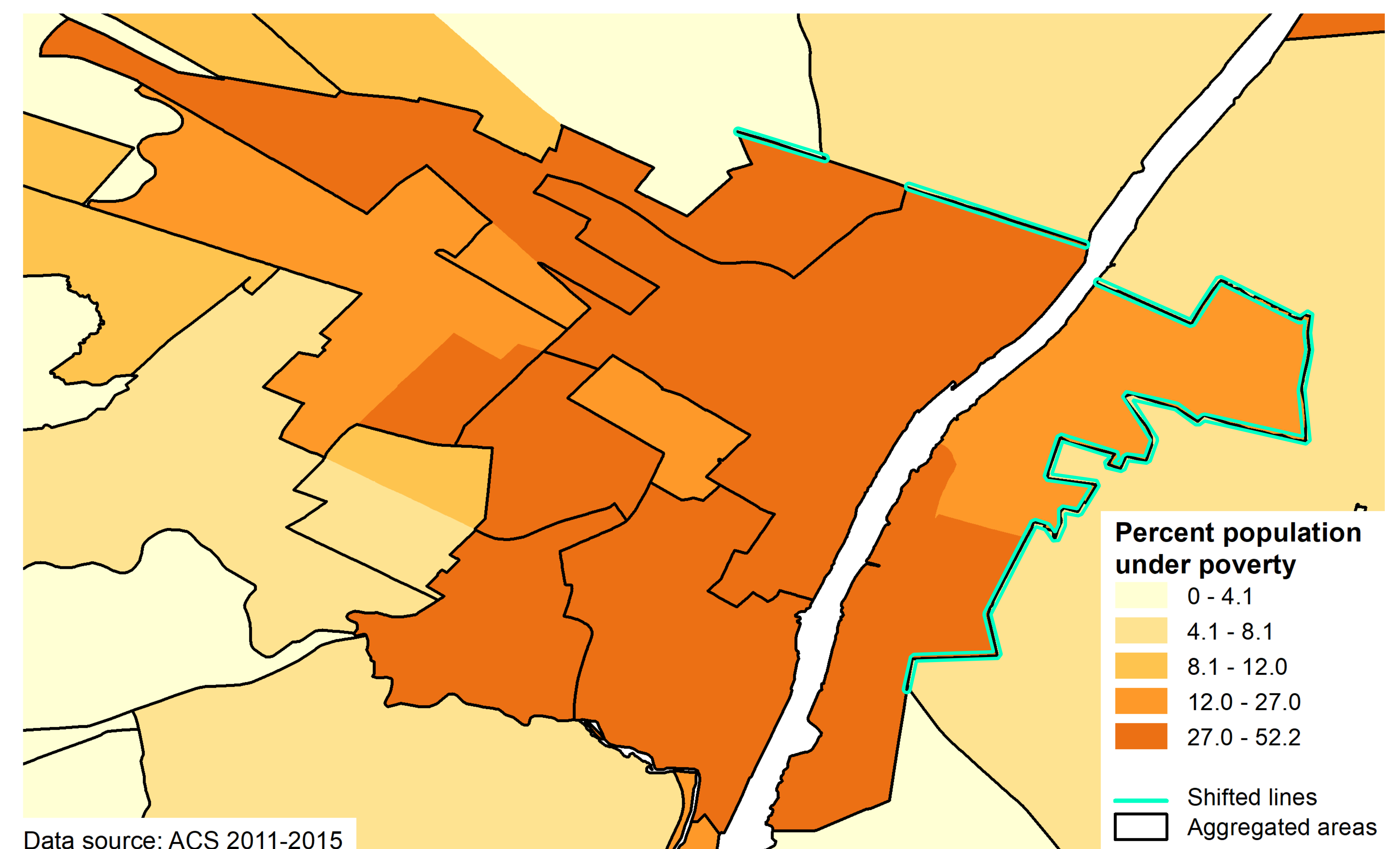


Figure 2: Similar ratio aggregation, Albany & Rensselaer cities, NY (percent minority residents divided by median income)