



**Department
of Health**

Geographic Aggregation Tool (GAT): A method for handling small numbers when calculating disease rates

**Abigail Stamm, Center for Environmental Health,
NYS Department of Health
August 2020**

Outline

- Need for subcounty data
- GAT
 - What it does
 - How it works
 - Application examples

Why display subcounty data?

Need: High risk areas

Issues:

- Smoothing/masking (county)
- Small numbers (tract, town)

Solution: Aggregation

GAT's objective

Aggregate small areas to:

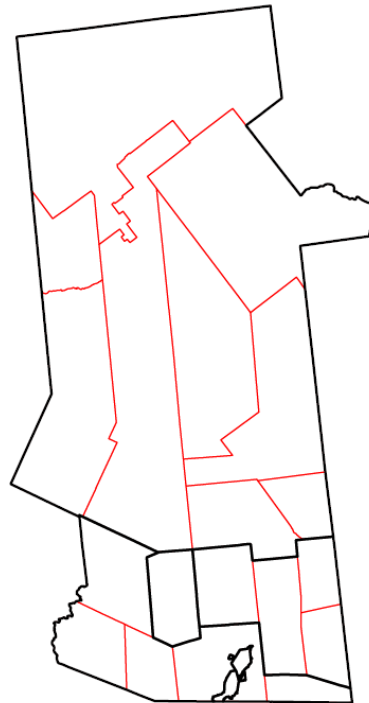
1. Meet minimum counts
2. Standardize process

GAT's process

1. Request user inputs
2. Run aggregation
3. Output shapefiles and documentation

Map comparing original and aggregated areas

- Original areas
- Aggregated areas



Merge type: closest population-weighted centroid
Merged variable: 6,000 to 15,000 TOTAL_POP



Department
of Health

User inputs

- Shapefile
- Minimum and maximum values
- Boundaries
- Exclusions
- Aggregation method

Step 11: Review settings

1. File to aggregate: C:/Users/AStamm/Documents/R/win-library/3.6/gatpkg/extdata/hftown
2. Identifying variable: ID
3. Boundary variable: COUNTY required
4. Minimum and maximum values:
6,000 to 15,000 TOTAL_POP
Areas excluded (value over maximum): 1 of 21
5. Exclusion criteria: MY_FLAG equals 1
Areas excluded: 1 of 21
6. Merge type: Closest population-weighted centroid
7. Population file: hfblock
Population variable: Pop_tot
8. Rate calculation: $\text{pop_dens} = 10,000 * \text{TOTAL_POP} / \text{AREALAND}$
Color scheme: Greens
9. Save KML file? No
10. Save location: C:/Users/AStamm/Documents/GAT/hftown_agg6k15k_popwt_2

Instructions

To modify a setting, choose it from the list and click 'Next >'.
After you modify most settings, you will return to this dialog.
If you modify setting 1, GAT will start over.

Select the setting you wish to modify:



Aggregation methods

1. Closest geographic centroid
2. Closest population-weighted centroid
3. Neighbor with the lowest count
4. Most similar neighbor

Step 6: Merging method

Instructions

1. Select your merging method.
2. If you select the first or third option, also select your choice(s) from the drop-down menu(s).

Merge options

closest area by **population-weighted** centroid
(note: selecting population weighting will open a dialog to select a population shapefile)

area with least TOTAL_POP

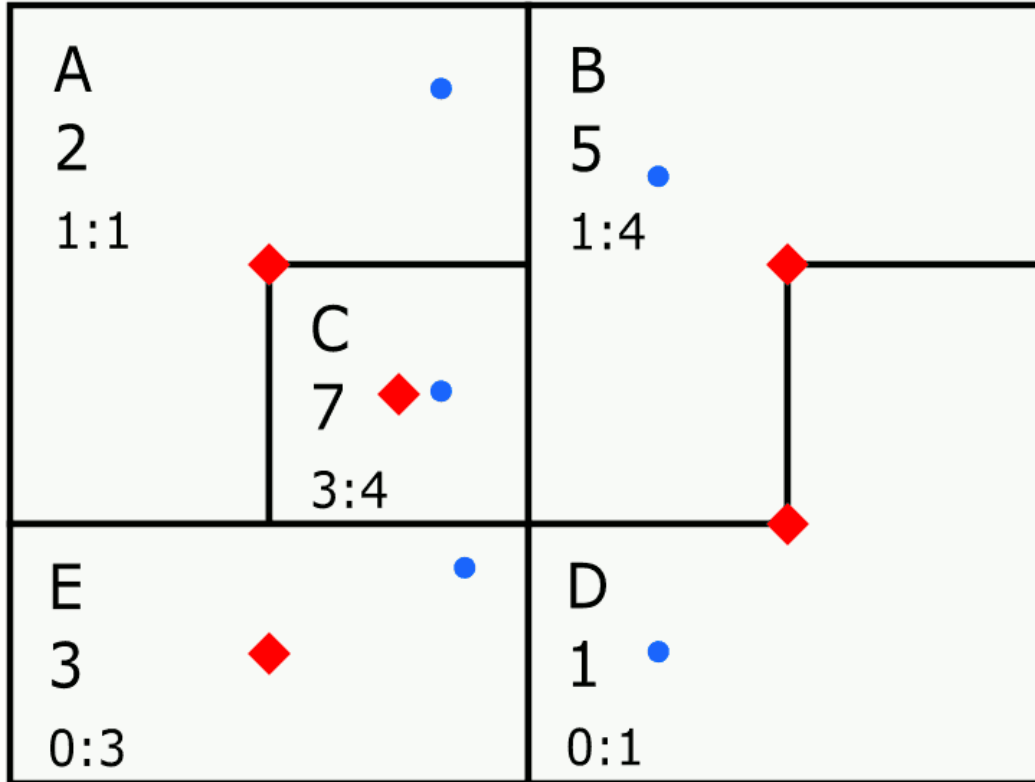
area with most similar ratio of **AREALAND** to **AREALAND**
(note: the numerator and denominator must be different; variables with 0 or missings cannot be in the denominator)

< Back Next > Cancel GAT Help

Closest geographic centroid

Minimum desired
value: 5

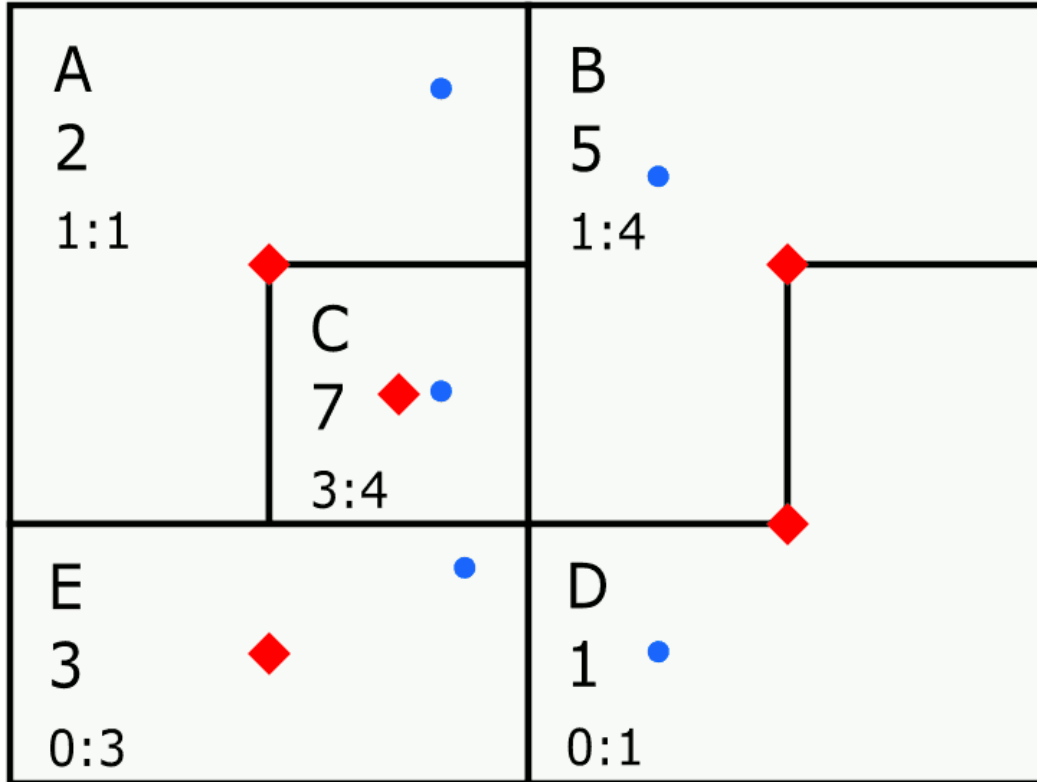
- ◆ Geographic centroid
- Population-weighted centroid



Closest geographic centroid

Minimum desired
value: 5

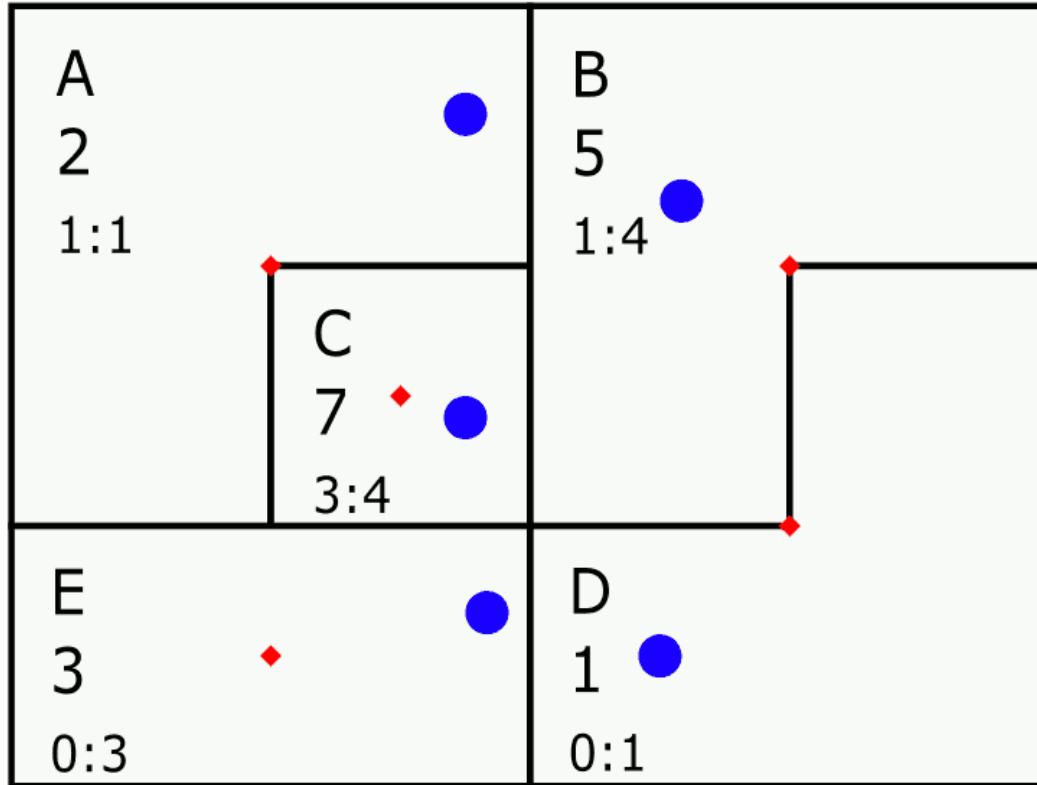
- ◆ Geographic centroid
- Population-weighted centroid



Closest population-weighted centroid

Minimum desired
value: 5

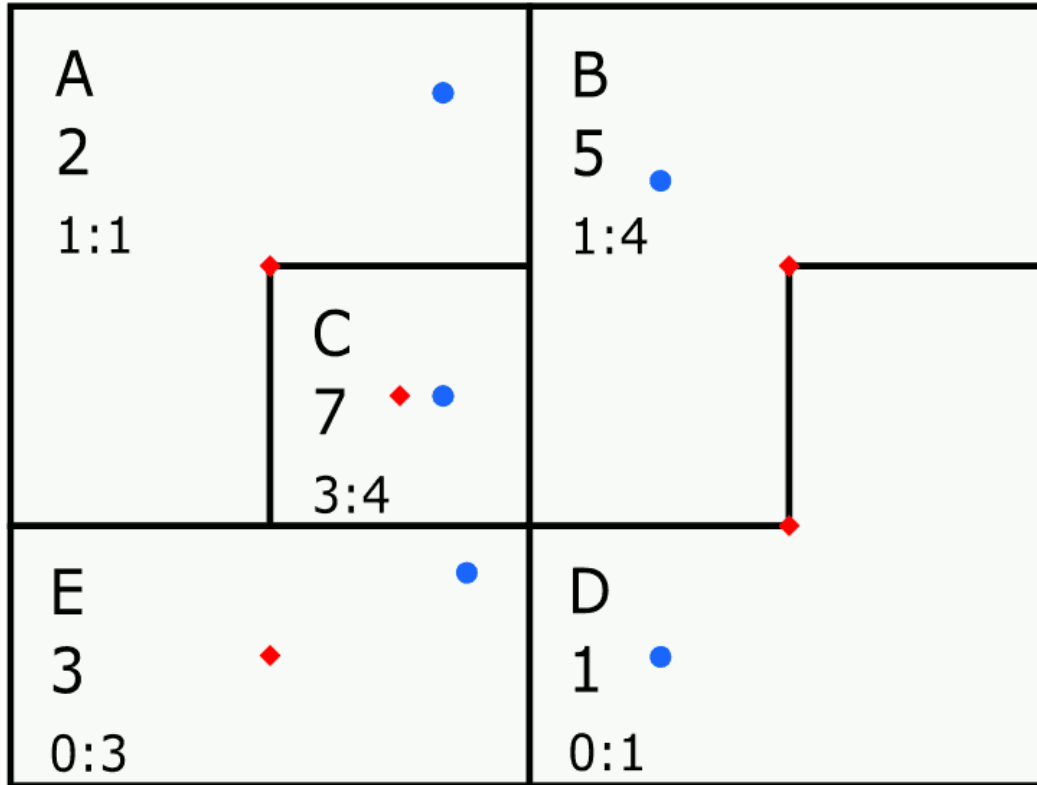
- ◆ Geographic centroid
- Population-weighted centroid



Neighbor with the lowest count

Minimum desired
value: 5

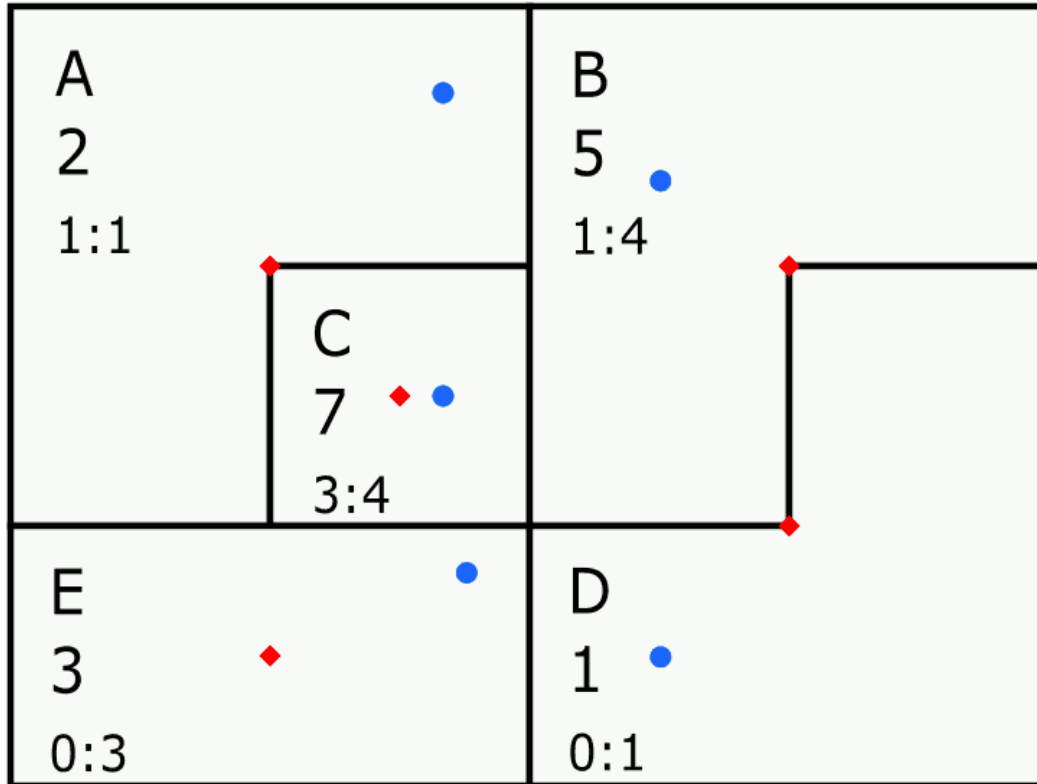
- ◆ Geographic centroid
- Population-weighted centroid



Most similar neighbor

Minimum desired
value: 5

- ◆ Geographic centroid
- Population-weighted centroid



Differences between GAT 2015 and GAT 2020

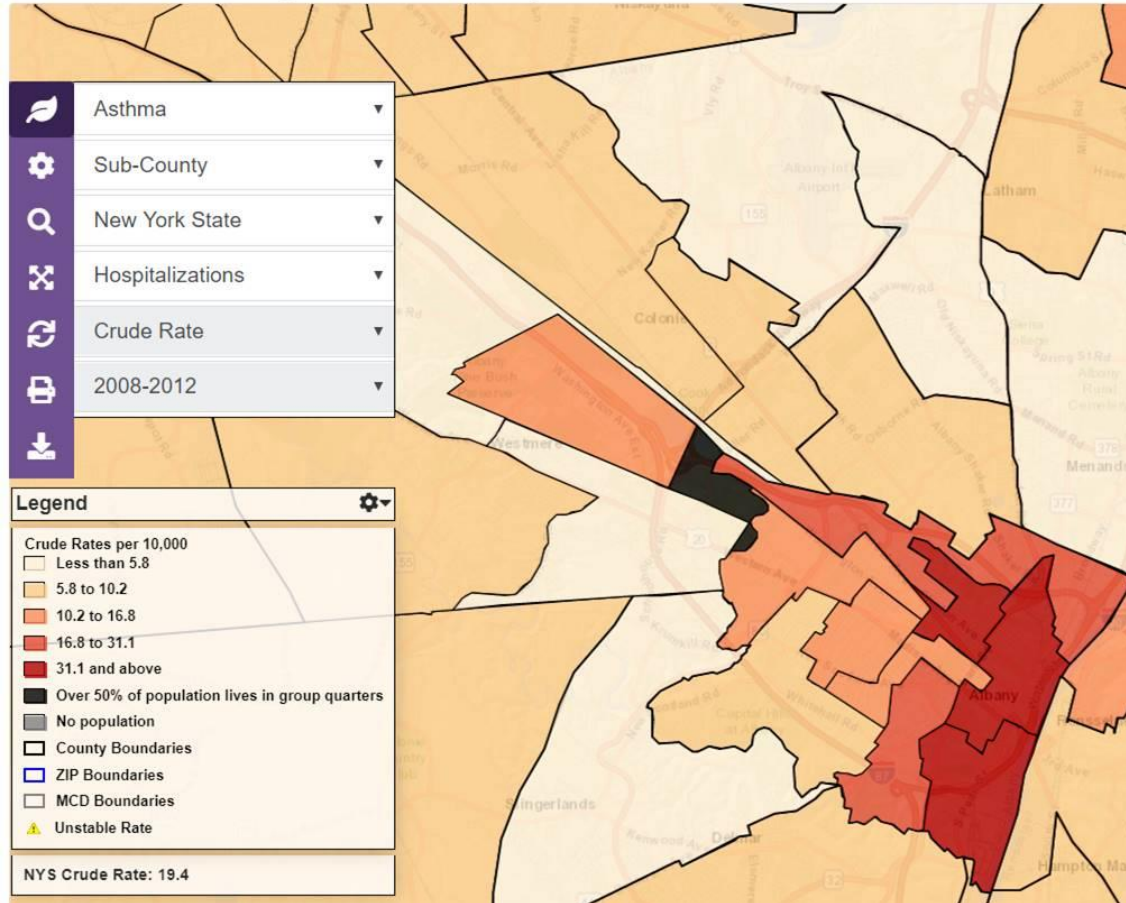
	GAT 2015	GAT 2020
Format	SAS and R scripts	R package
Log	Minimal	Comprehensive
Maps	Simple, not saved	Detailed, saved to PDF
Change settings dialog	No	Yes
Population weighting	SAS yes, R no	Yes
Exclusion criteria	No	Yes
Maximum values	No	Yes



Applying GAT: disease

- aggregation by population
- closest population-weighted centroid

Asthma Crude Hospitalization Rate per 10,000, New York State, 2008-2012



Applying GAT: mortality

- aggregation by number of deaths
- closest geographic centroid

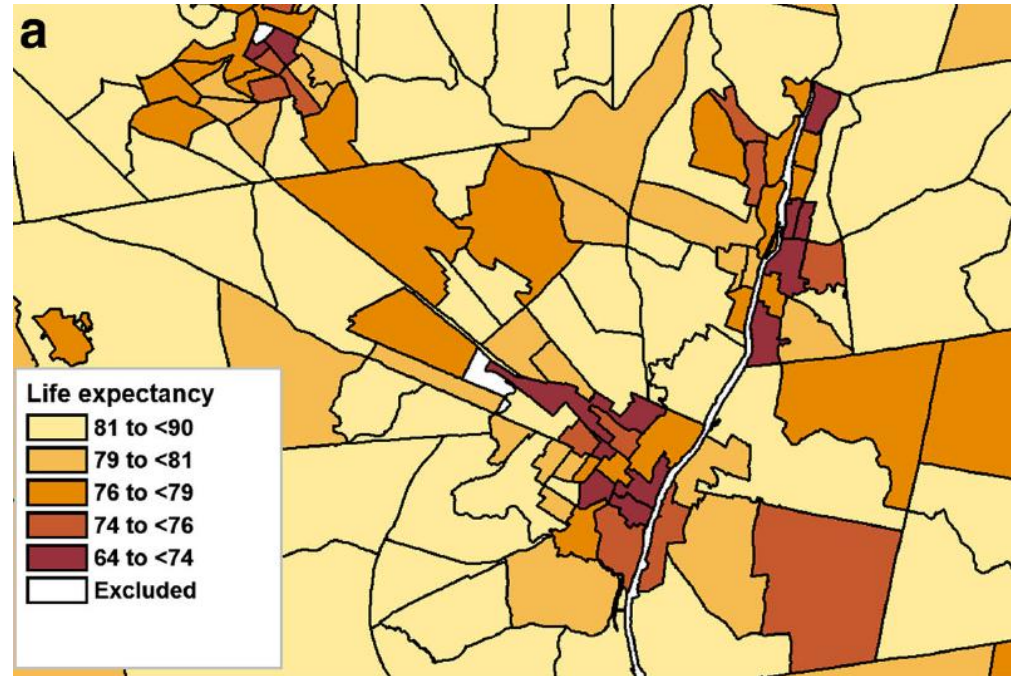


Fig. 6 Thematic Maps of the New York State Capital District after aggregation. a by life expectancy (image from Talbot et al. Population Health Metrics (2018) 16:1)

Takeaways

How GAT can help you

- Small areas with stable rates
- Standardization and documentation
- Customization

Acknowledgements

CDC for funding

Gwen LaSelva for code and testing

NYS DOH EPHT team for testing and feedback

Email me at abigail.stamm@health.ny.gov

Projects that have cited GAT

Sherman RL, Henry KA, Tannenbaum SL, Feaster DJ, Kobetz E, Lee DJ. *Prev Chronic Dis* 2014;11:130264. DOI: <http://dx.doi.org/10.5888/pcd11.130264> (referenced R v1.2)

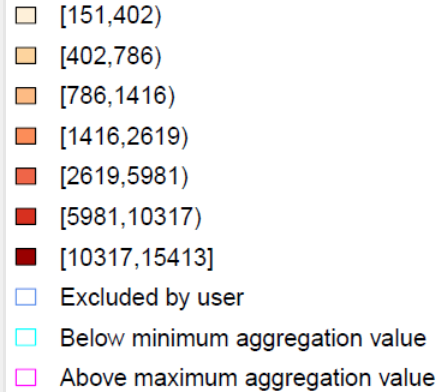
Werner AK, Strosnider HM. *Spatial and Spatio-temporal Epidemiology* 2020;33. DOI: <https://doi.org/10.1016/j.sste.2020.100339> ((used SAS v1.31)

Werner AK, Strosnider H, Kassinger C, Shin M. *J Public Health Manag Pract*. 2018;24(5):E20-E27. doi:10.1097/PHH.0000000000000686 ((used SAS v1.31)

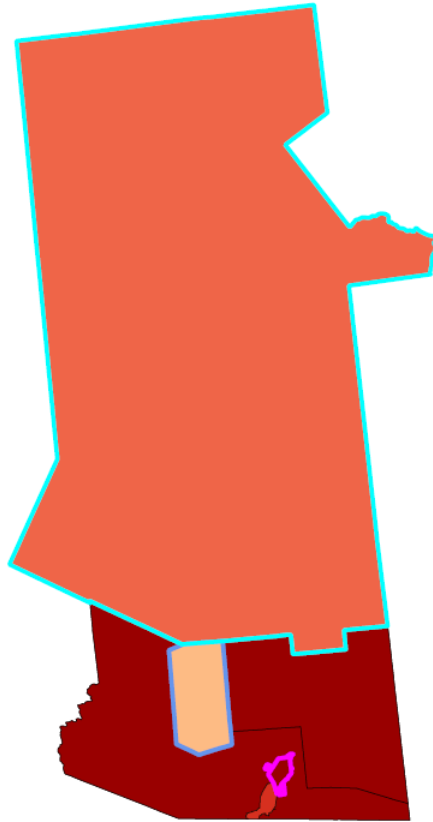
Boscoe FP, Talbot TO, Kulldorff M. *Geospat Health*. 2016;11(1):304. Published 2016 Apr 18. doi:10.4081/gh.2016.304 (used SAS v1.31)

Boothe VL, Fierro LA, Laurent A, Shih M. *Global Diaspora News*. Published 3/28/2020. <https://www.globaldiasporanews.com/a-tool-to-improve-community-health-and-advance-health-equity/> (used R v1.33)

TOTAL_POP After Merging



Summary stats for
TOTAL_POP :
Minimum: 1,407
Median: 11,671
Maximum: 15,413



30 mi

Aggregation values: 6,000 to 15,000 TOTAL_POP
Exclusion criteria: MY_FLAG equals 1

Example assessment map



Department
of Health

Example log excerpt

```
NYSDOH Geographic Aggregation Tool (GAT) Log
Version & date: 1.52 2020-07-14
Date run: 2020-07-22
Time GAT took to run: 5.73 minutes |

Input file:          C:/Users/AStamm/Documents/R/
Projection:          +proj=longlat +datum=NAD27 +
Field names:         TOWN, ID, COUNTY, AREALAND,
Identifier:          ID
Boundary variable:   COUNTY
  You chose to require the aggregation to respect

Output file: C:/Users/AStamm/Documents/GAT/hftown_
Number of input areas:    21
Number of output areas:   6
Number of aggregations:  15
Number of excluded areas: 1

Merge type: closest population-weighted centroid
Population file: C:/Users/AStamm/Documents/R/win
Population variable: Pop_tot

Exclusion criteria:
1. MY_FLAG equals 1

First aggregation variable: TOTAL_POP
Minimum value: 6,000
Maximum value: 15,000
```

