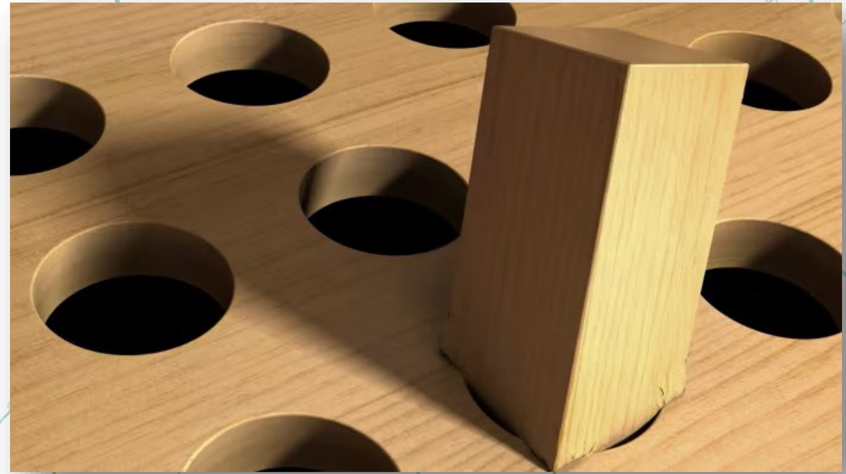
Population distribution weighted estimation of populations from ACS data

Dr. Nicholas Kruskamp
Dr. Stefany Ramos
Dr. Chris Inkpen

RTI International
Research Triangle Park, NC

Background and motivation

- "areal interpolation" is a primary approach for taking a variable from a set of source polygons and assigning that variable to target polygons
- This approach has a critical assumption: **uniform spatial distribution of variable**
- Population density weighted interpolation is an alternative that we explore here



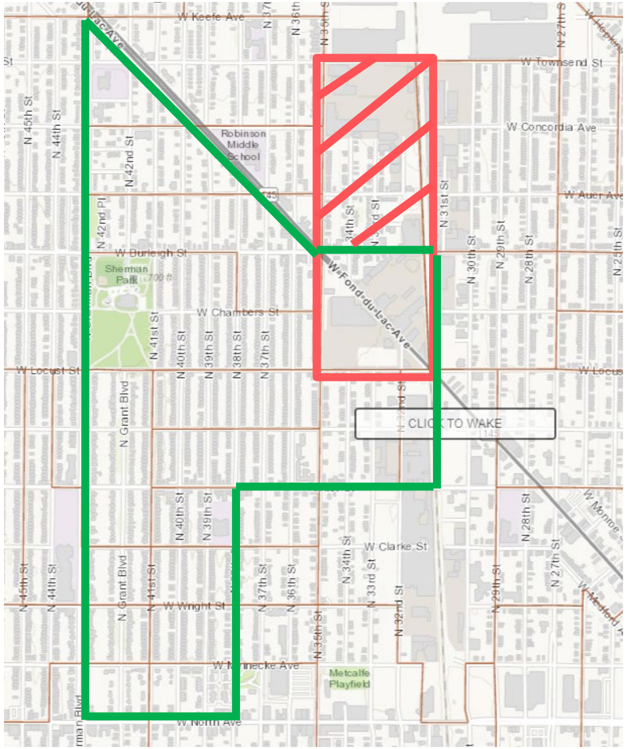
Research questions

- Does the source geographic level matter in estimating population totals and distributions of target geometries?
- Does population density weighted interpolation provide better results than simple areal interpolation?

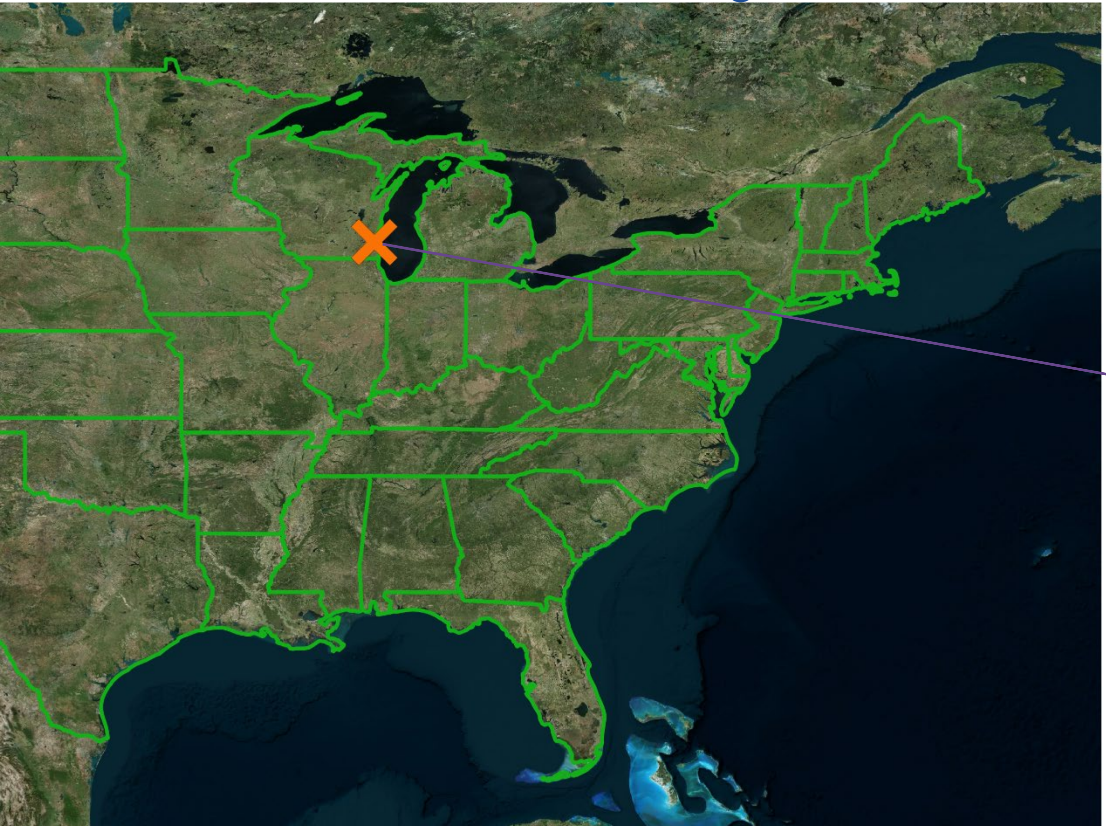
Motivation

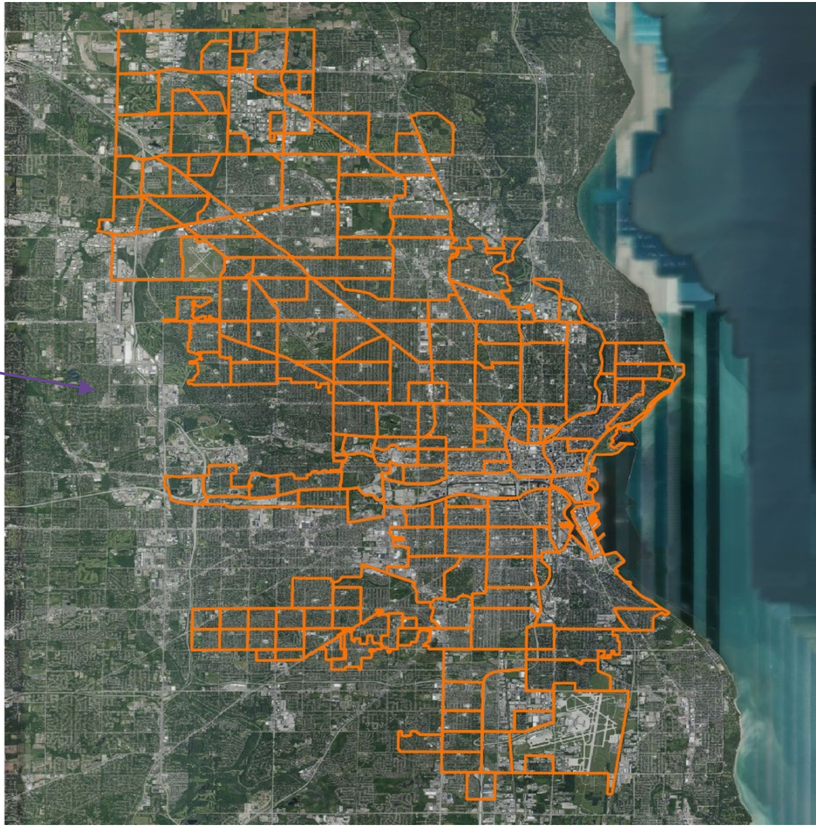
- Multiple impact evaluations of place-based community violence interventions
- Analysis calls for matching treatment and control communities on contextual data available from ACS
- "Communities" = neighborhoods in Milwaukee, WI
- Neighborhood boundaries do not align with ACS geographic boundaries
- How should we create neighborhood-level estimates?

Example: Sherman Park



Use case: Milwaukee neighborhoods

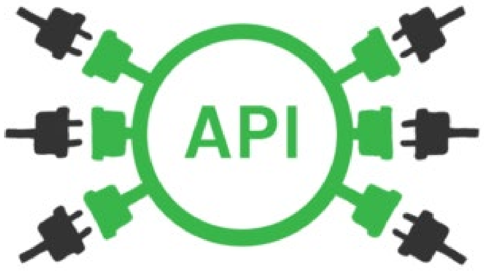




data

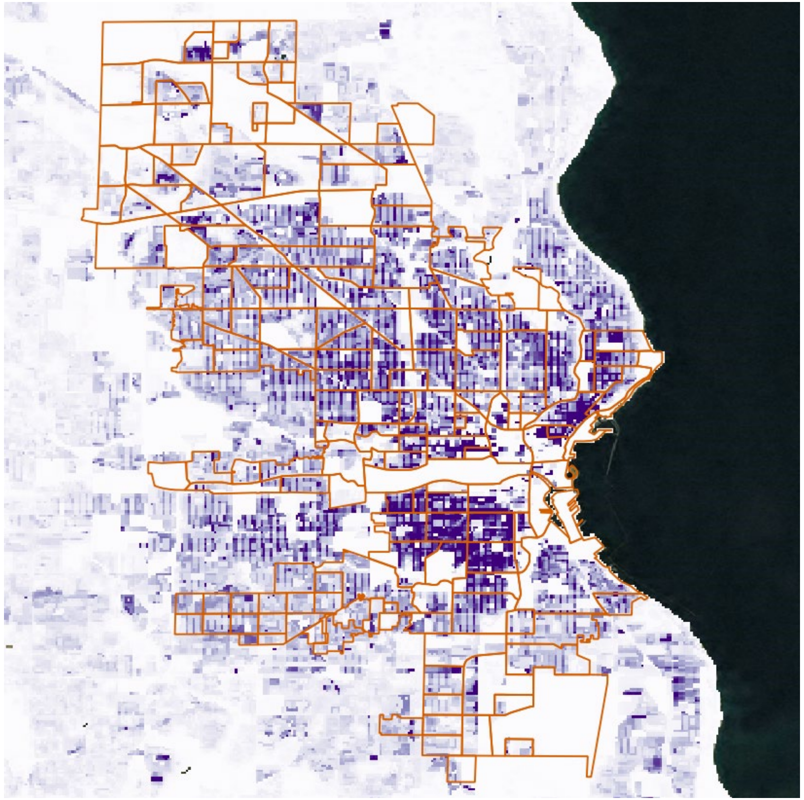
- The ACS data:
  - Survey:
    - 5 year, ending year 2010 – 2020
      - Block group available from 2013 - present
  - Geographic levels:
    - Block group, tract, and place level estimates
  - Variables:
    - B01001: Sex by age
      - *Universe: Total population*
      - B01001_001E “total”
      - B01001_001M “margin of error”

- All data was programmatically pulled using the `censusdata` python package
  - Thank you census for the API!



WorldPop

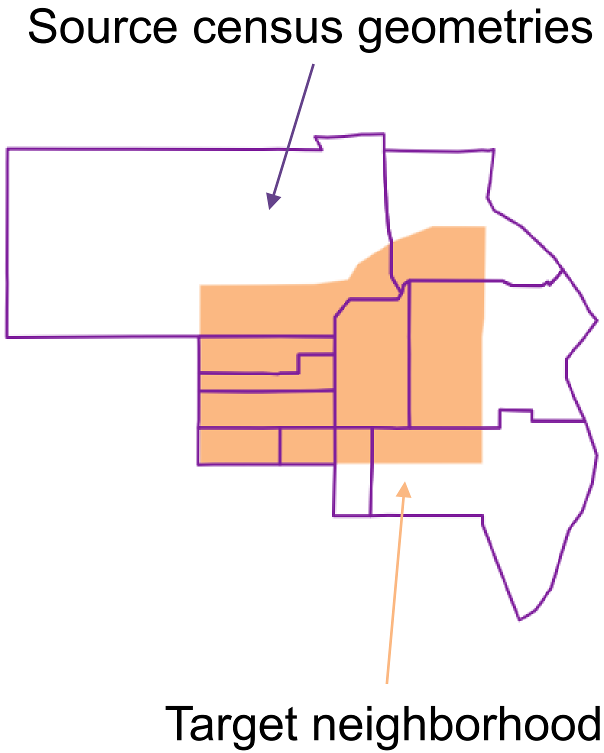
- Unconstrained population counts @ 100 m resolution
- *Only* using *relative* population counts within each polygon, not actual WorldPop cell counts



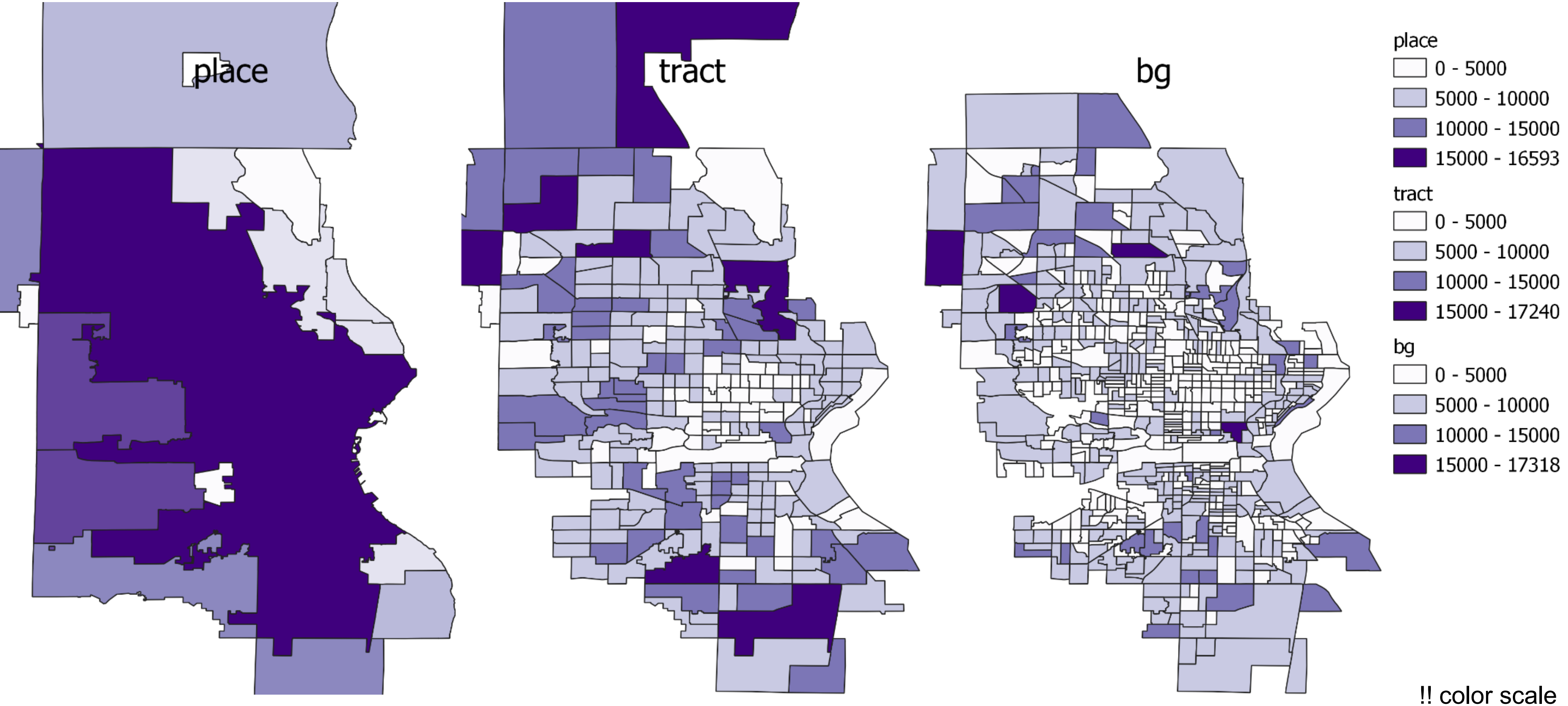
https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained/

Pop weighted vs. areal interpolation

- Areal interpolation: what percent of the total source land area falls within the target geometry?
- Population weighted interpolation: what percent of the total auxiliary variable (population) falls within the target geometry?

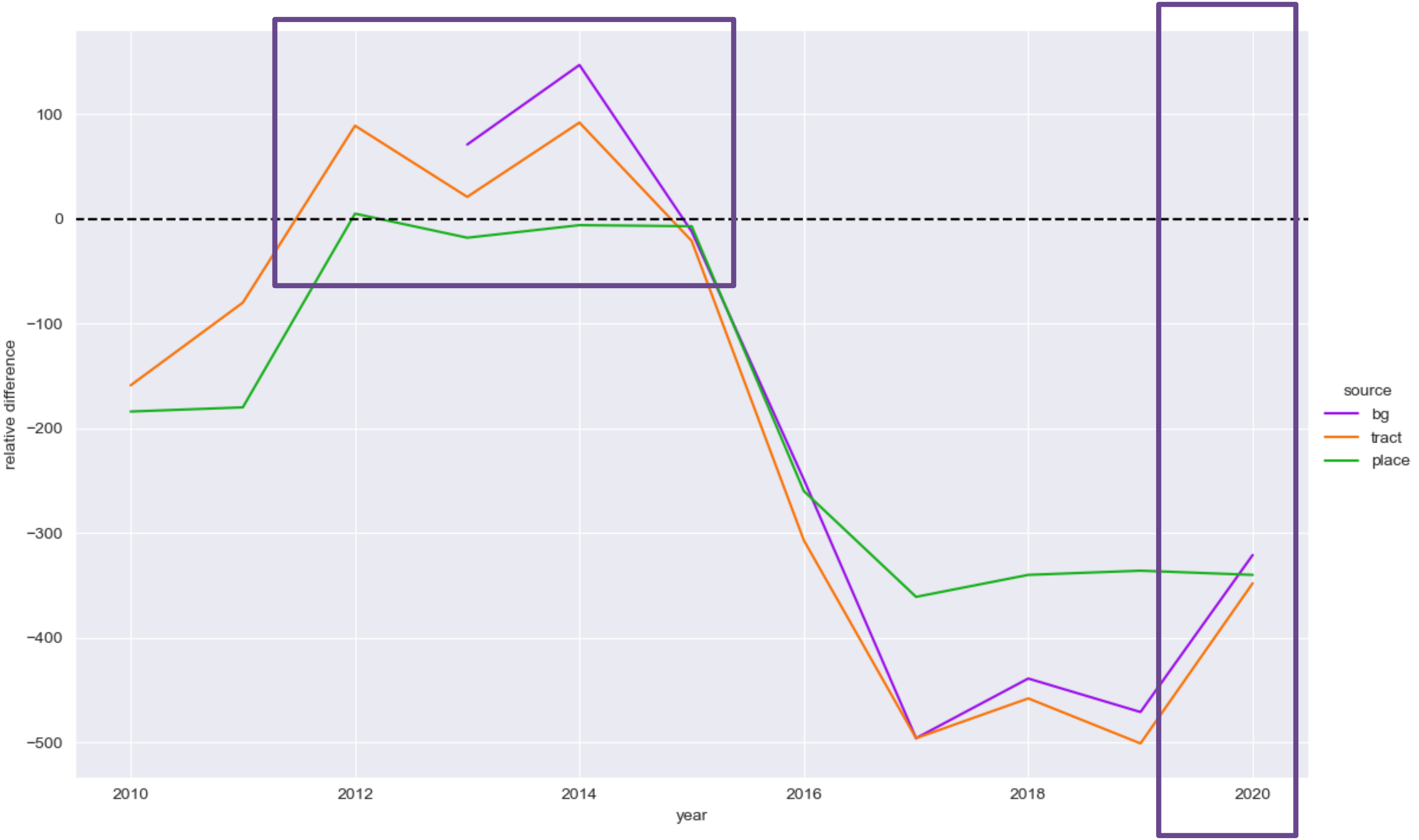


ACS data: population 2020

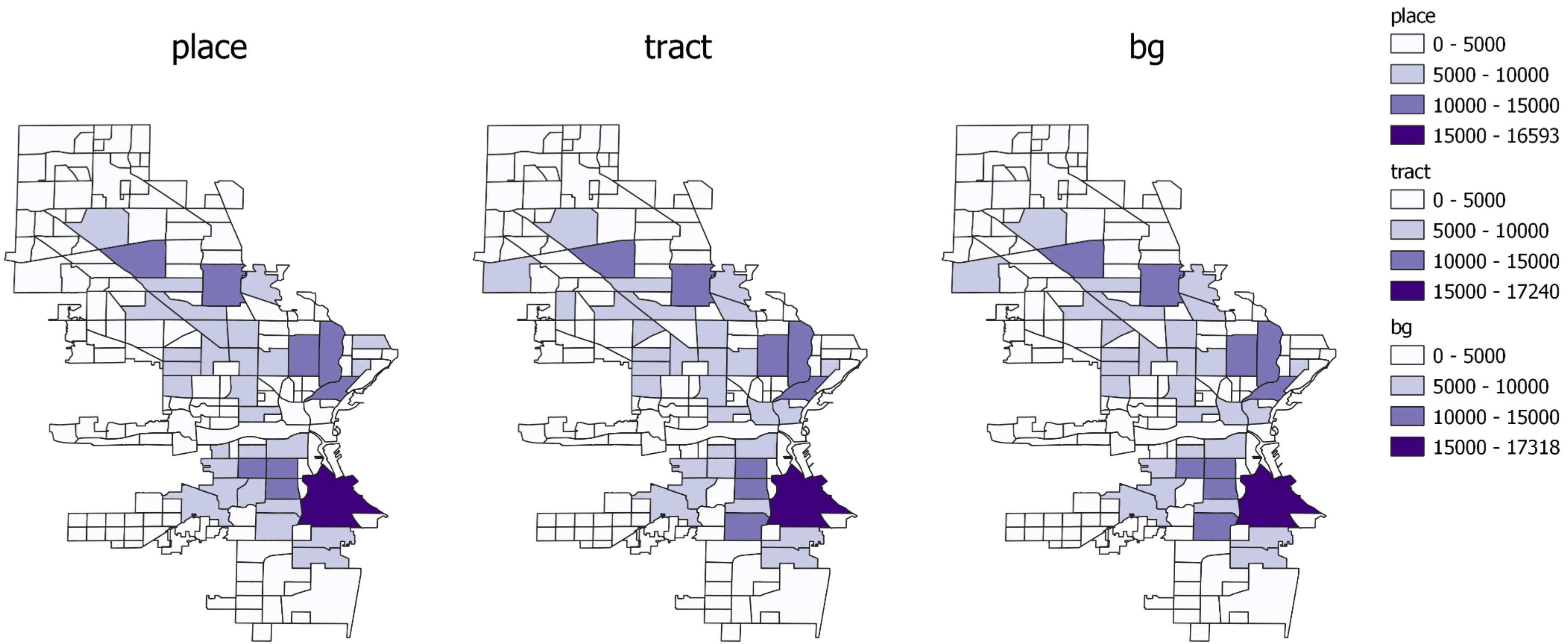


!! color scale different on each map

Estimating population totals trough time

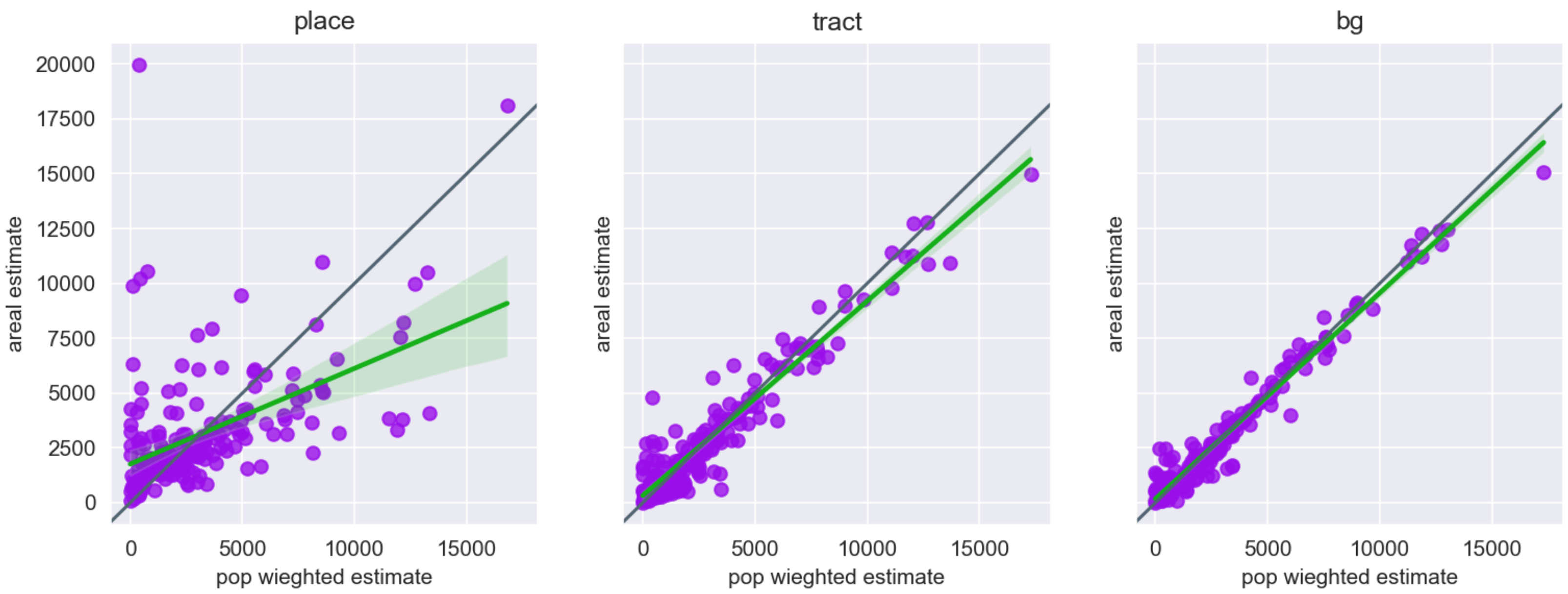


Results: neighborhood estimates

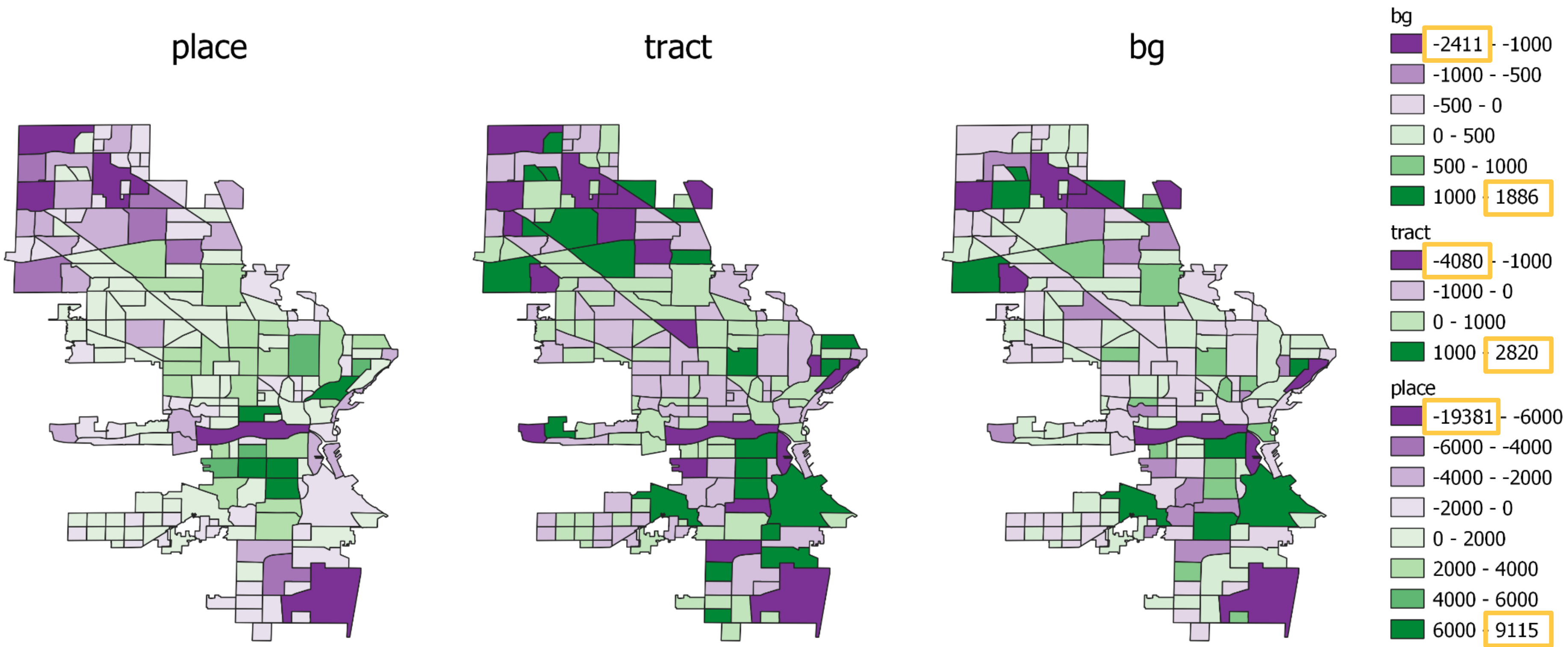


!! color scale different on each map

neighborhood estimates: population weighting vs areal interpolation



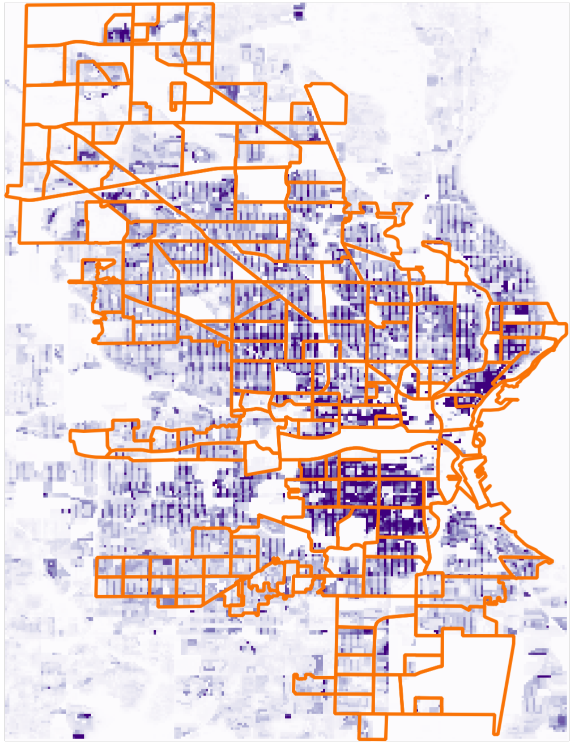
neighborhood estimates: population weighting vs areal interpolation



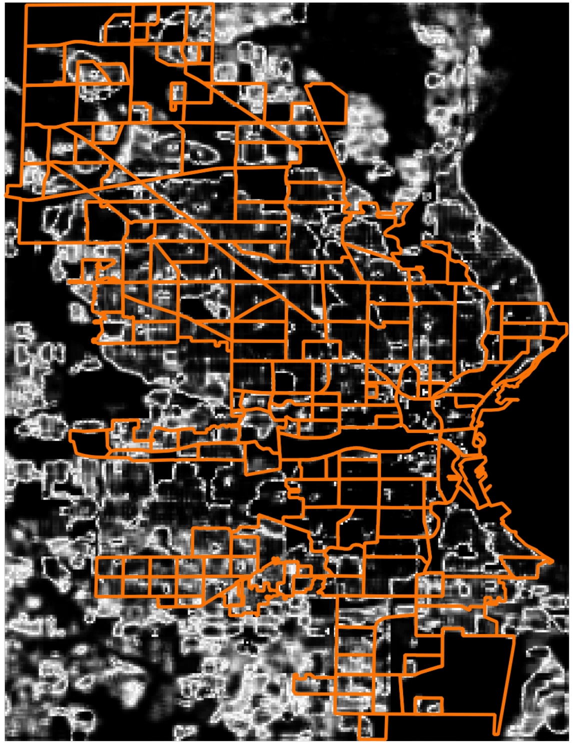
!! color scale different on each map

Population distributions: are they much different than uniform?

world pop counts



local moran's I significance



Population distributions: are they much different than uniform?

conclusion

- In our use case:
  - Level of census geometry:
    - very strong correlation among population estimates from any level of geometry considered
  - Areal vs population weighted interpolation:
    - Tract and block groups shown strong correlation between the two methods
    - Local population distributions are not different enough from uniform to have a strong affect on weights compared to the relative size of the source and target polygons
    - Future work: test true and synthetic distributions, significantly different from random, at different spatial lags, to tease out the interaction between source/polygon size and importance of weighting surface on results.

Thank you

Contact: Dr. Nicholas Kruskamp | email: nkruskamp@rti.org