

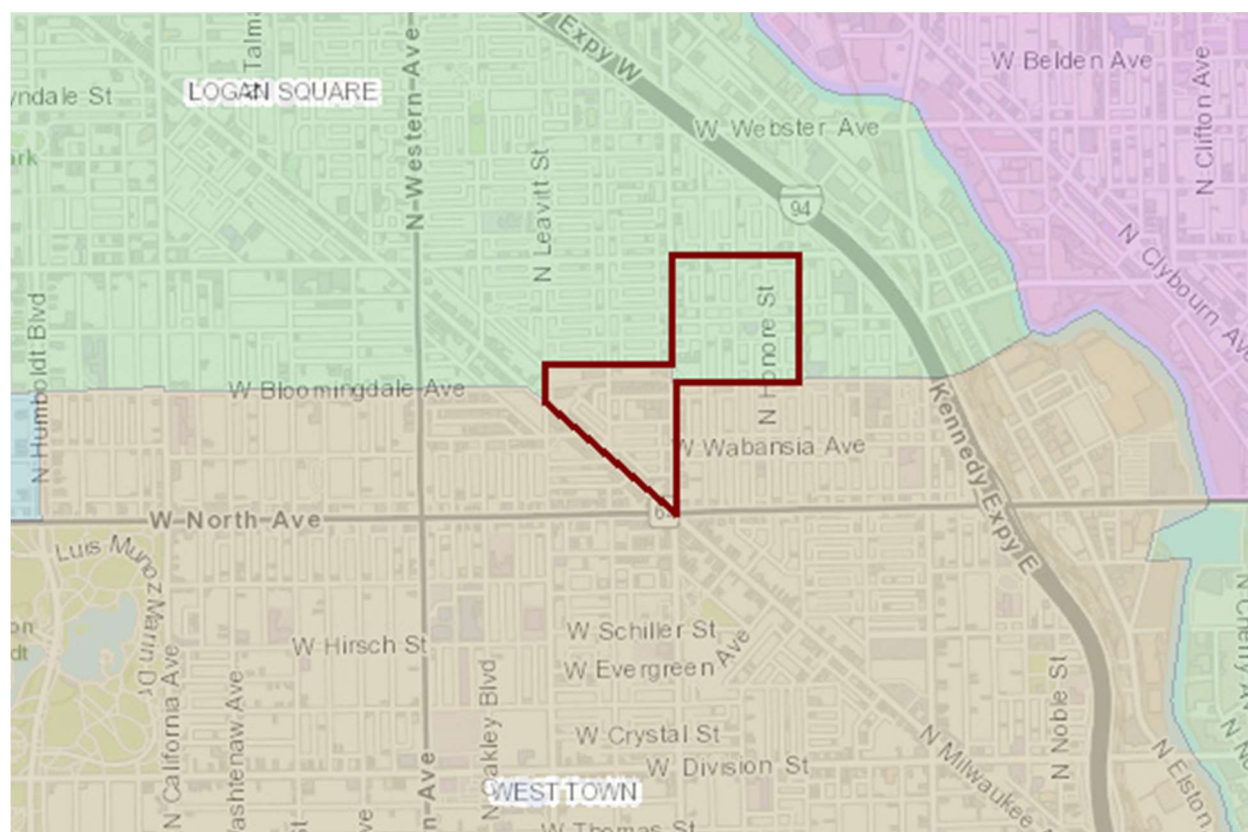
Tract-to-Neighborhood Aggregation

THE CHALLENGE

The neighborhood is often the most important level of analysis for policy questions facing Chicago. However, census data collected on Chicago's residents does not always lend itself easily to a neighborhood-level analysis. This is because census tracts do not always share borders with Chicago's 77 neighborhoods (a.k.a. community areas, or CCAs), and sometimes overlap severely (see below a map of Census Tract 8310, which overlaps with West Town and Logan Square). Neighborhood-level analysis in these cases is not straightforward, and if done incorrectly can lead to improper estimates of neighborhood-level statistics, misleading policymakers.

Prior methods for tract-to-neighborhood aggregation include splitting tract statistics evenly between overlapping neighborhoods, and splitting tracts in proportion to the area of the tract that overlaps with each neighborhood. However, even the latter approach can lead to inaccurate estimates. This is because the population of a census tract is not distributed evenly throughout the tract. Take for example a census tract that includes a large park; attributing part of the tract's employment or housing statistics to that park, simply because it occupies a large amount of that tract's area, would not make sense.

OUR APPROACH



Census Tract 8310

We allocate census tract statistics to their corresponding neighborhoods based on the proportion of individuals and households in a tract that belong to that neighborhood. This is done by leveraging block-level information published by the census. Census blocks are the smallest geographic unit with data published by the census. The census does not provide any demographic statistics at the block-level for privacy reasons, but they do provide the number of individuals and households within each block. With this information, we are able to calculate the number of individuals and households from a given tract that belong to each neighborhood. We then allocate a tracts' statistics, e.g. the number of individuals employed, or the mean household income, into the corresponding neighborhood in proportion to the neighborhood's actual population – not a proxy.

TECHNICAL DETAILS

- Statistics are first split up into tract/neighborhood pairs, and then aggregated by neighborhood. R code used to aggregate is below, where `est` is the estimated statistic and `pct` is the percent of said estimate in a given CCA.

```
df.aggEst <- df %>%  
  dplyr::mutate(est = est * pct) %>%  
  dplyr::group_by(CCA) %>%  
  dplyr::summarise(est = sum(est, na.rm = T))
```

- Standard errors are aggregated according to the methods described in ACS documentation.¹ When a tract is split across neighborhoods, the standard errors are allocated according to the following rule:
 - When $\geq 90\%$ of a tract's population falls within one neighborhood, all of its error is assumed by that neighborhood.
 - When a tract is split into two neighborhoods (where no neighborhood has $\geq 90\%$ of a tract's population), the tract's error is assumed by both neighborhoods.

¹ A-14 in <https://www.census.gov/content/dam/Census/library/publications/2009/acs/ACSRResearch.pdf>