

## C: Reallocating geographic variables

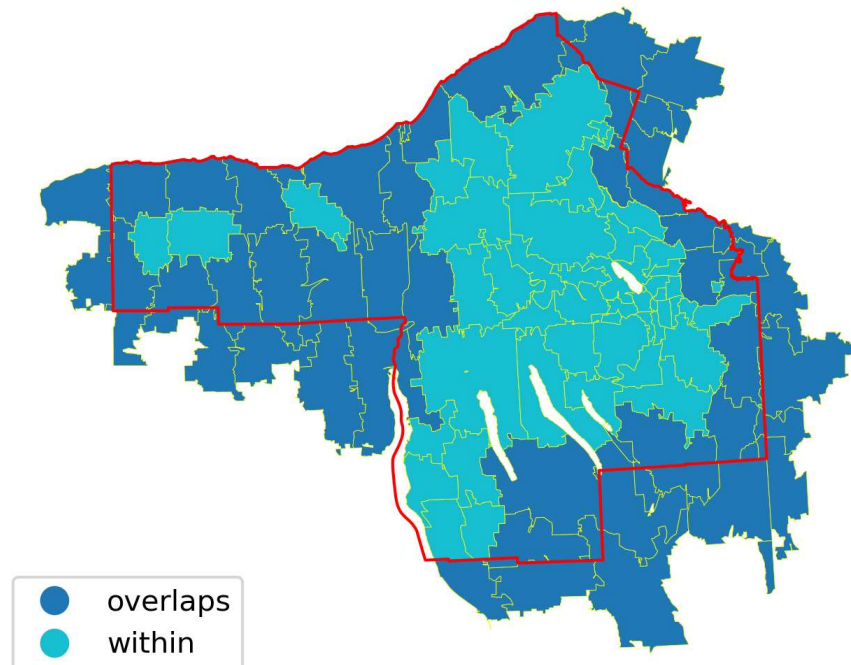
### Working with datasets with incompatible geographies

Key use of geographic intersection (slicing polygons)

Example:

- Presidential election contributions data by zip code
- Demographic data by Congressional District

New York's 24th Congressional District



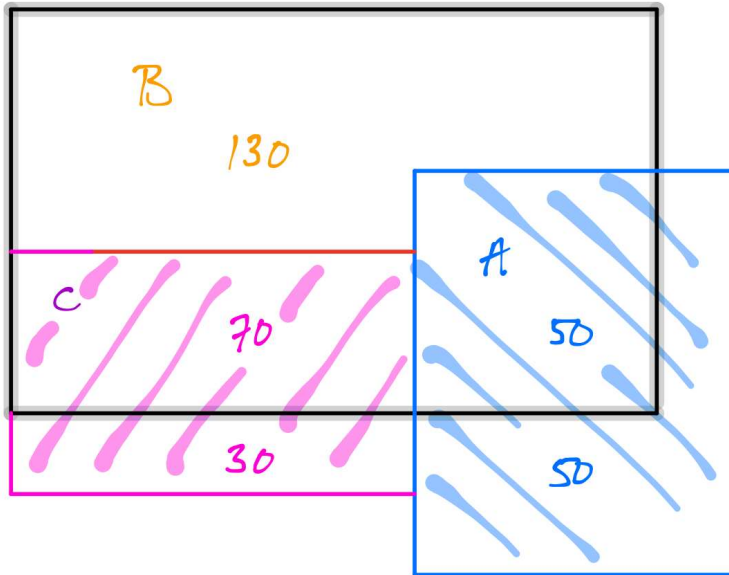
### Basic approach: impute based on areas

1. Use **geometric intersection** to **slice zip codes** by district boundary
2. Calculate each **slice's share of its zip code's area**
3. Split contributions **in proportion** to area

#### 4. Aggregate slice contributions by district

Schematically:

Suppose district CDXX is the black rectangle below  
 Consists of zones that are parts of 3 zips



Data on the zips:

Zone	Size	Alignment with District	Contributions
A	100	Overlaps	\$2000
B	130	Within	\$3000
C	100	Overlaps	\$4000
Total	330		\$9000

Geometric intersection creates five slices:

Slice	Zone	District	Size	Share	Contributions
0	A	CDXX	50	0.5	$0.5 * 2000 = 1000$

1	A	NaN	50	0.5	$0.5 * 2000 = 1000$
2	B	CDXX	130	1.0	$1.0 * 3000 = 3000$
3	C	CDXX	70	0.7	$0.7 * 4000 = 2800$
4	C	NaN	30	0.3	$0.3 * 4000 = 1200$
Total			330		9000

Aggregating to the district:

District	Size	Contributions
CDXX	$50+130+70 = 250$	$1000+3000+2800 = 6800$
NaN	$50+30 = 80$	$1000+1200 = 2200$
Total	330	9000

End result:

- **Reallocates** contributions from zips to districts
- Key assumption: donors are approximately uniformly distributed within zips
- Could refine with block-level population density