

How spatial polygons shape our world

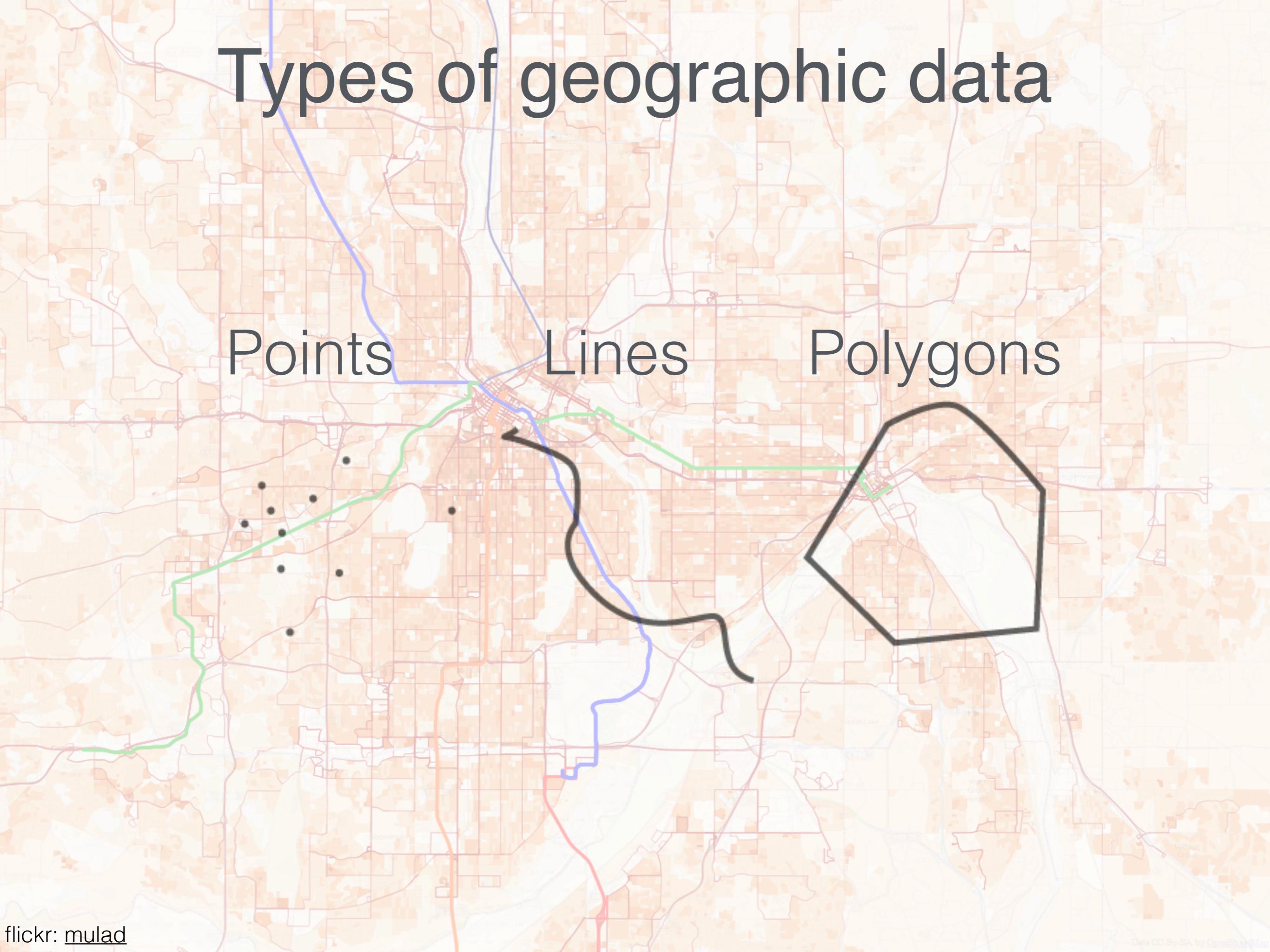
Amelia McNamara @AmeliaMN
Smith College, Northampton MA

Types of geographic data

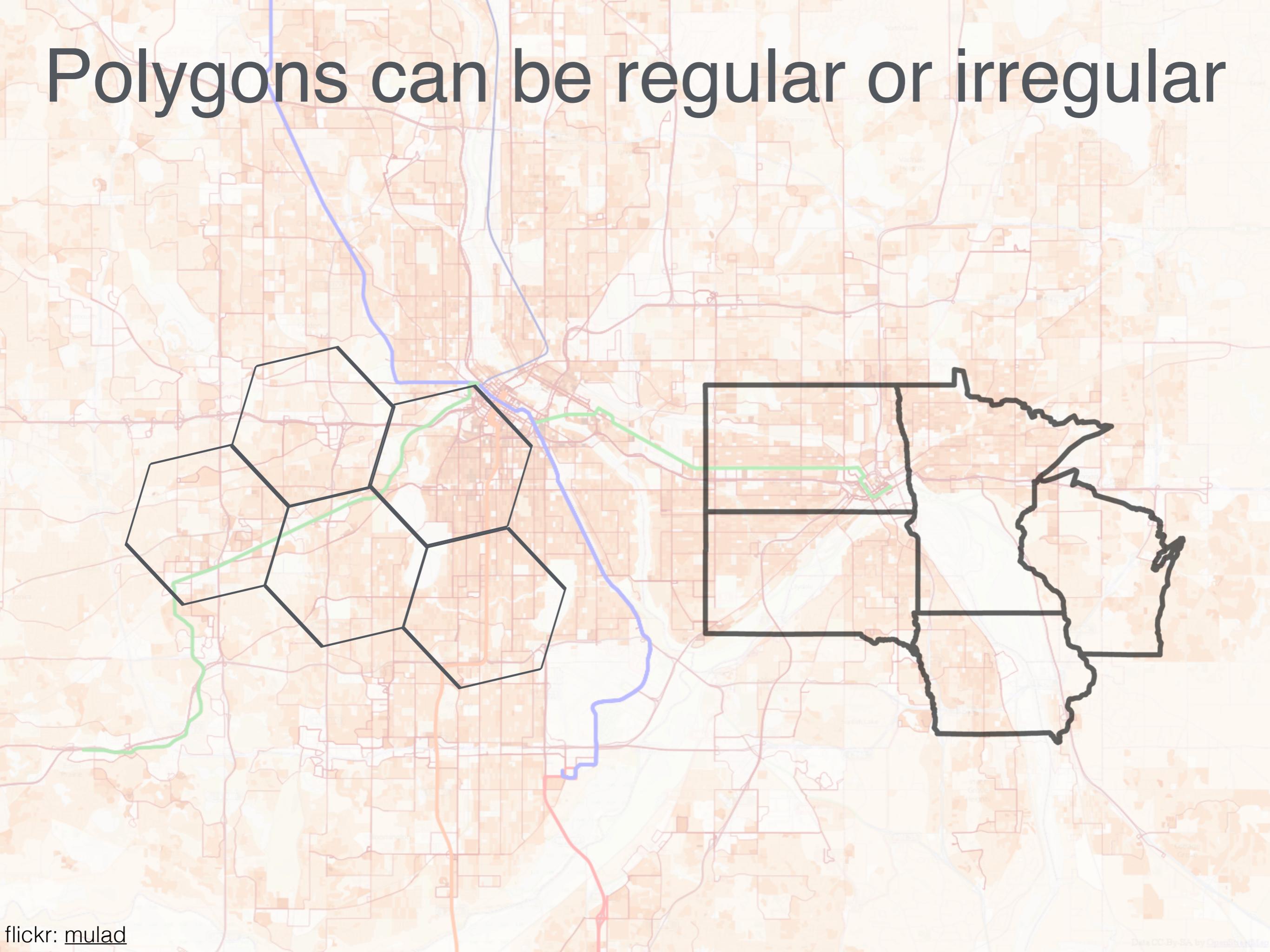
Points

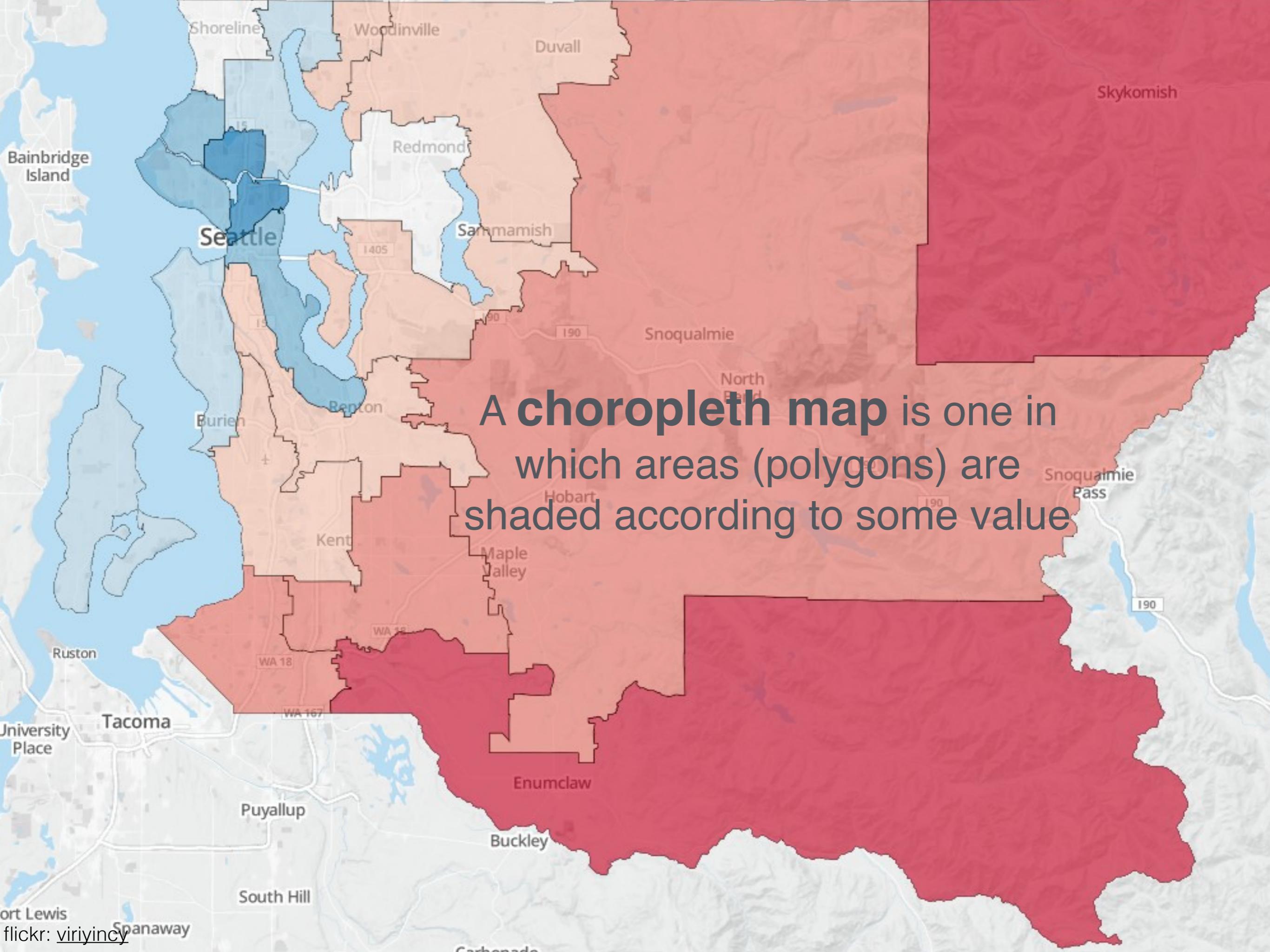
Lines

Polygons

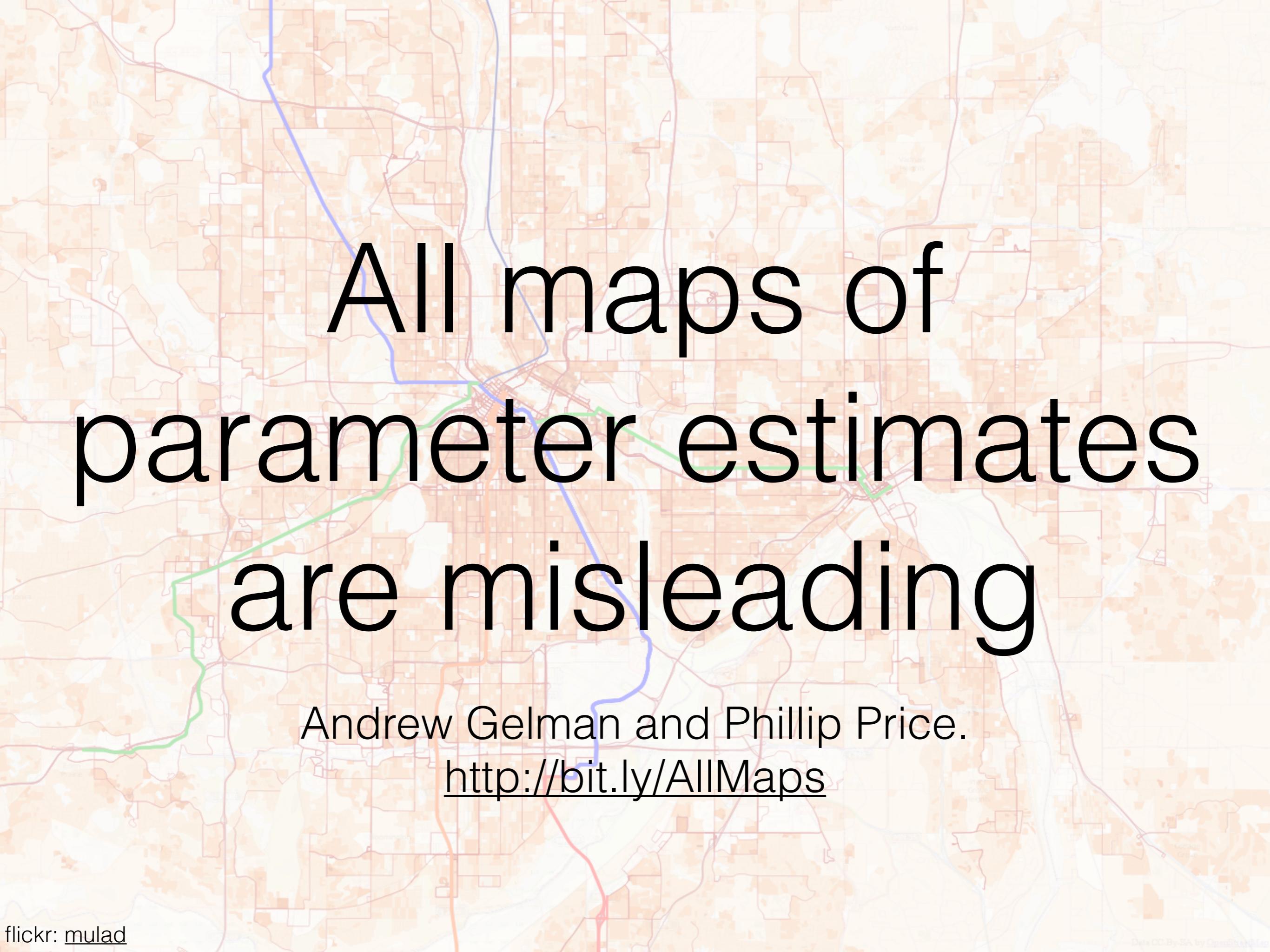


Polygons can be regular or irregular





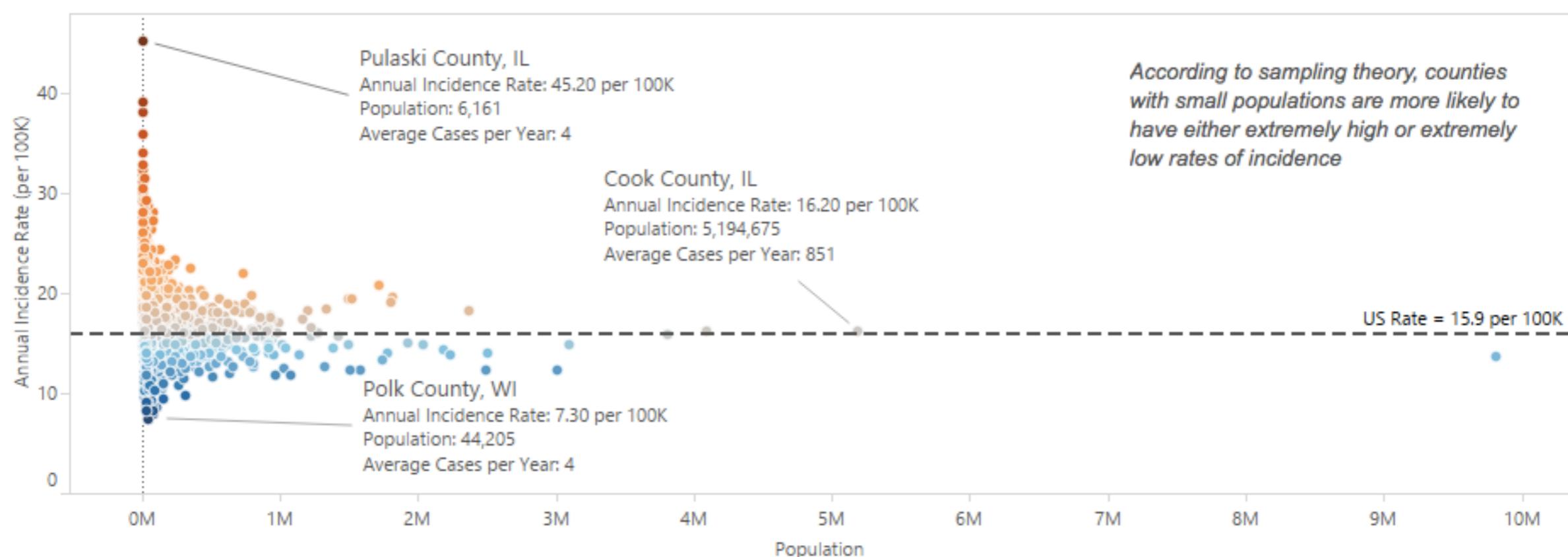
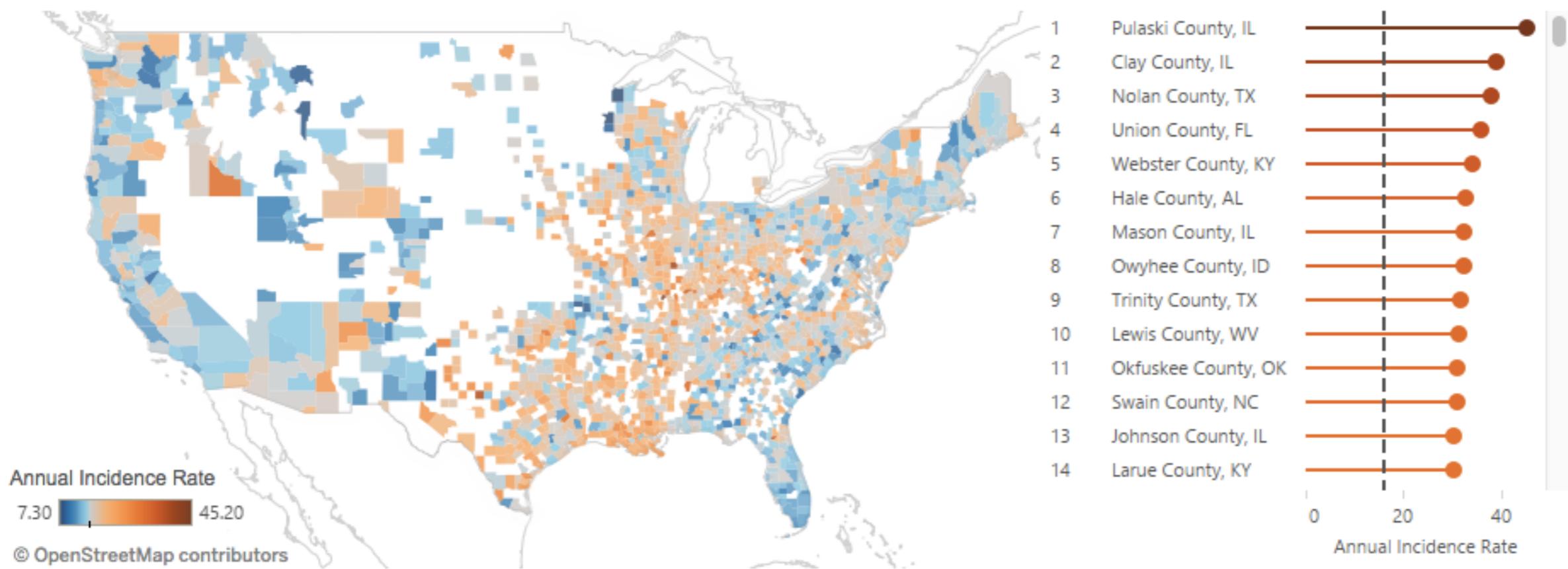
A choropleth map is one in which areas (polygons) are shaded according to some value



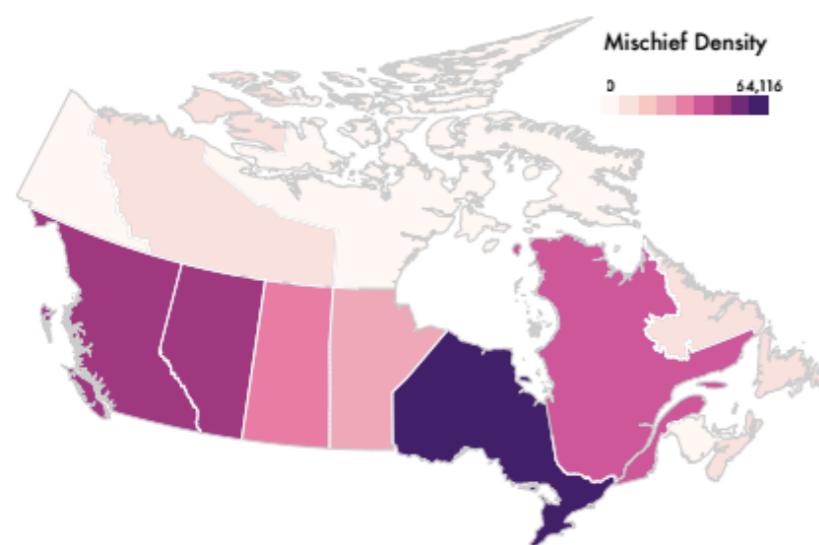
All maps of
parameter estimates
are misleading

Andrew Gelman and Phillip Price.
<http://bit.ly/AllMaps>

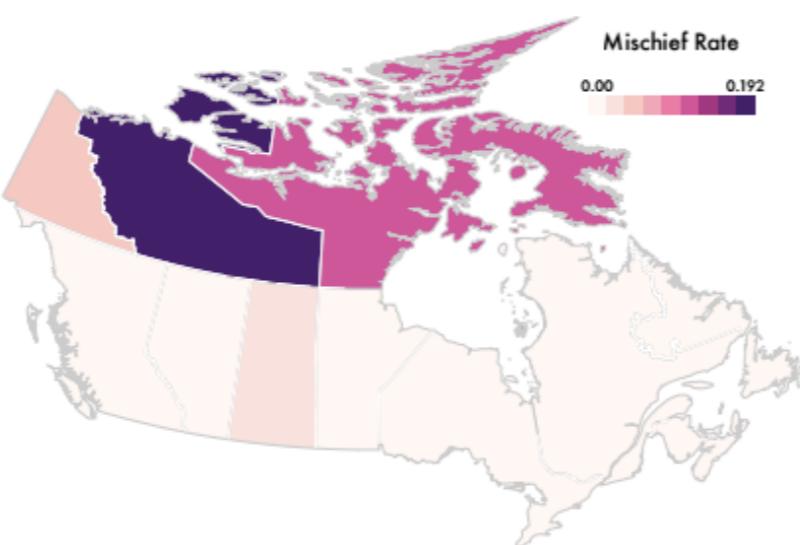
Kidney Cancer and Insensitivity to Sample Size



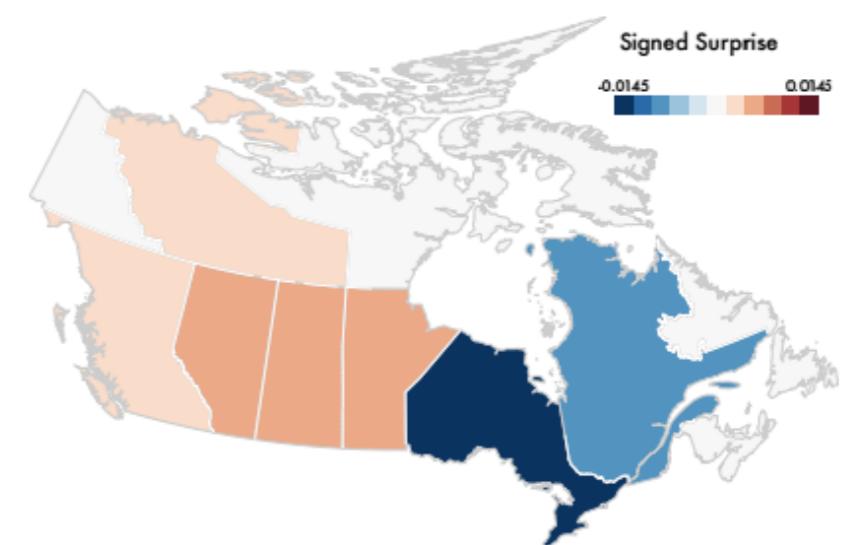
Surprise! Bayesian Weighting for De-Biasing Thematic Maps.



(a) The **Event Density** of “mischief” in Canada.



(b) The per-capita **Event Rate** of mischief.



(c) The **Surprise Map** of mischief.

Michael Correll and Jeffrey Heer
<http://bit.ly/SurpriseMaps>

Area gives a lot of visual weight

365 **Obama**
Electoral Votes
Projected Winner

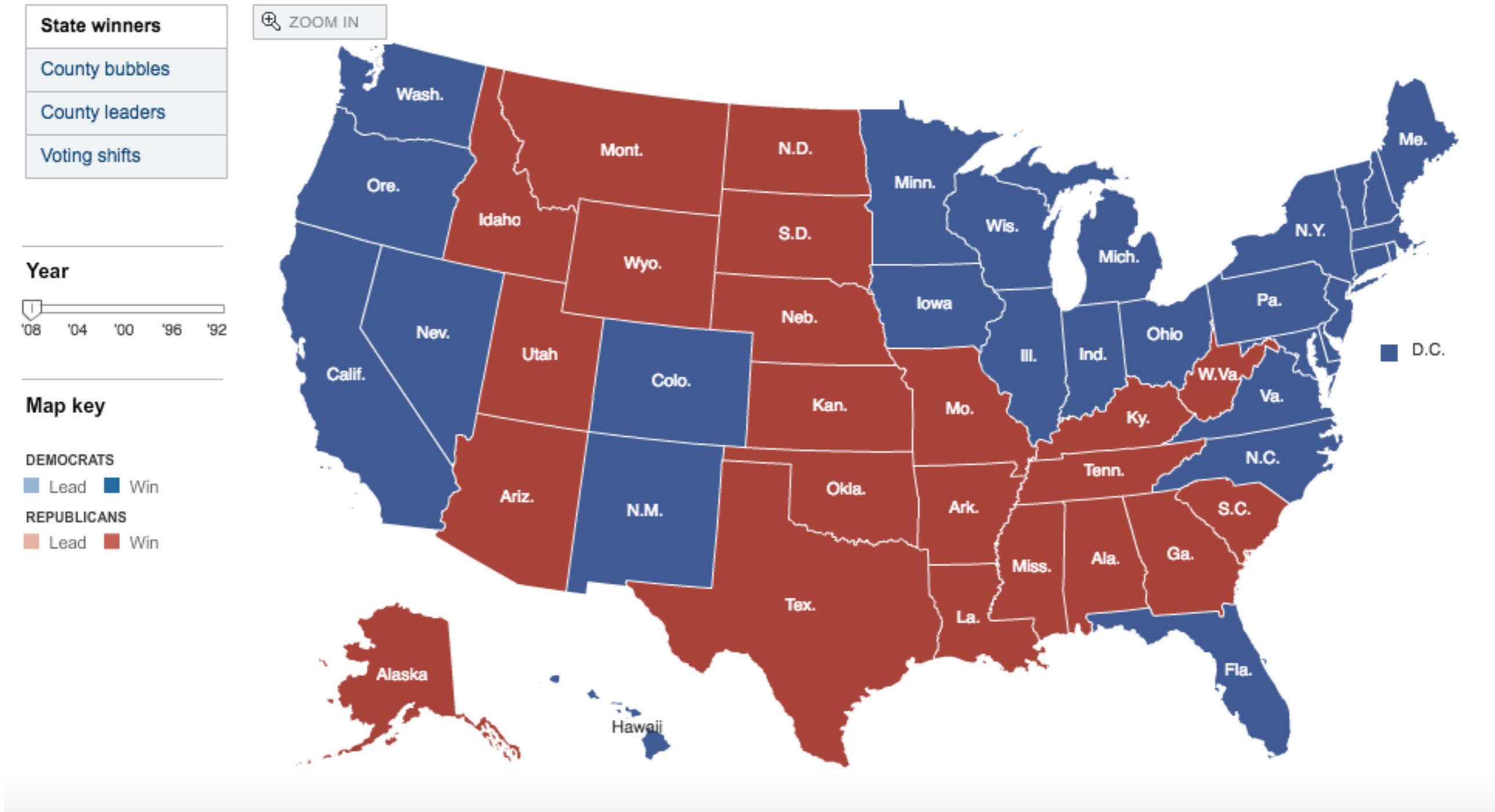
0
undecided

173 **McCain**
Electoral Votes

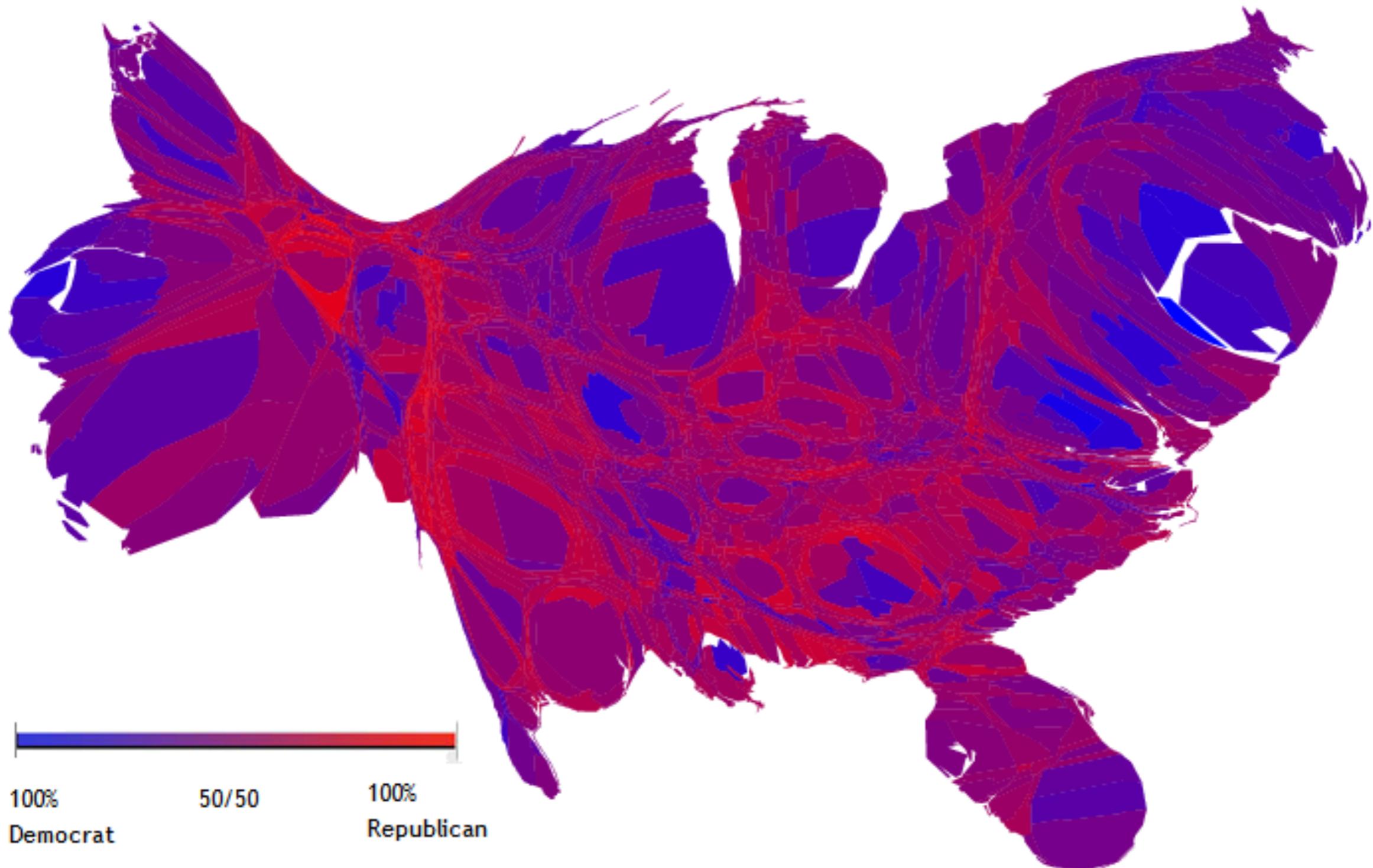
Popular vote: 66,862,039

270 needed to win

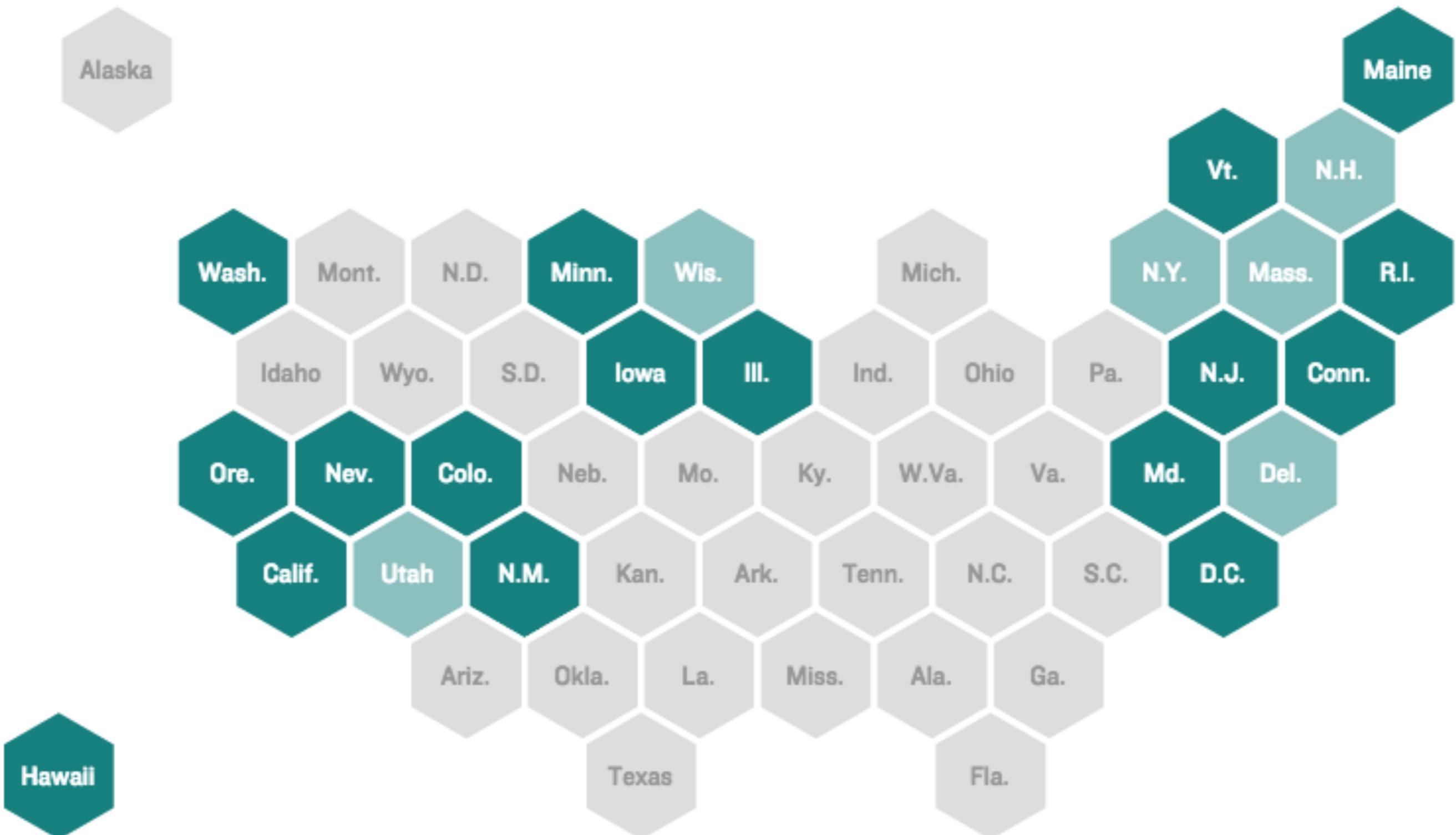
Popular vote: 58,319,442



Cartogram

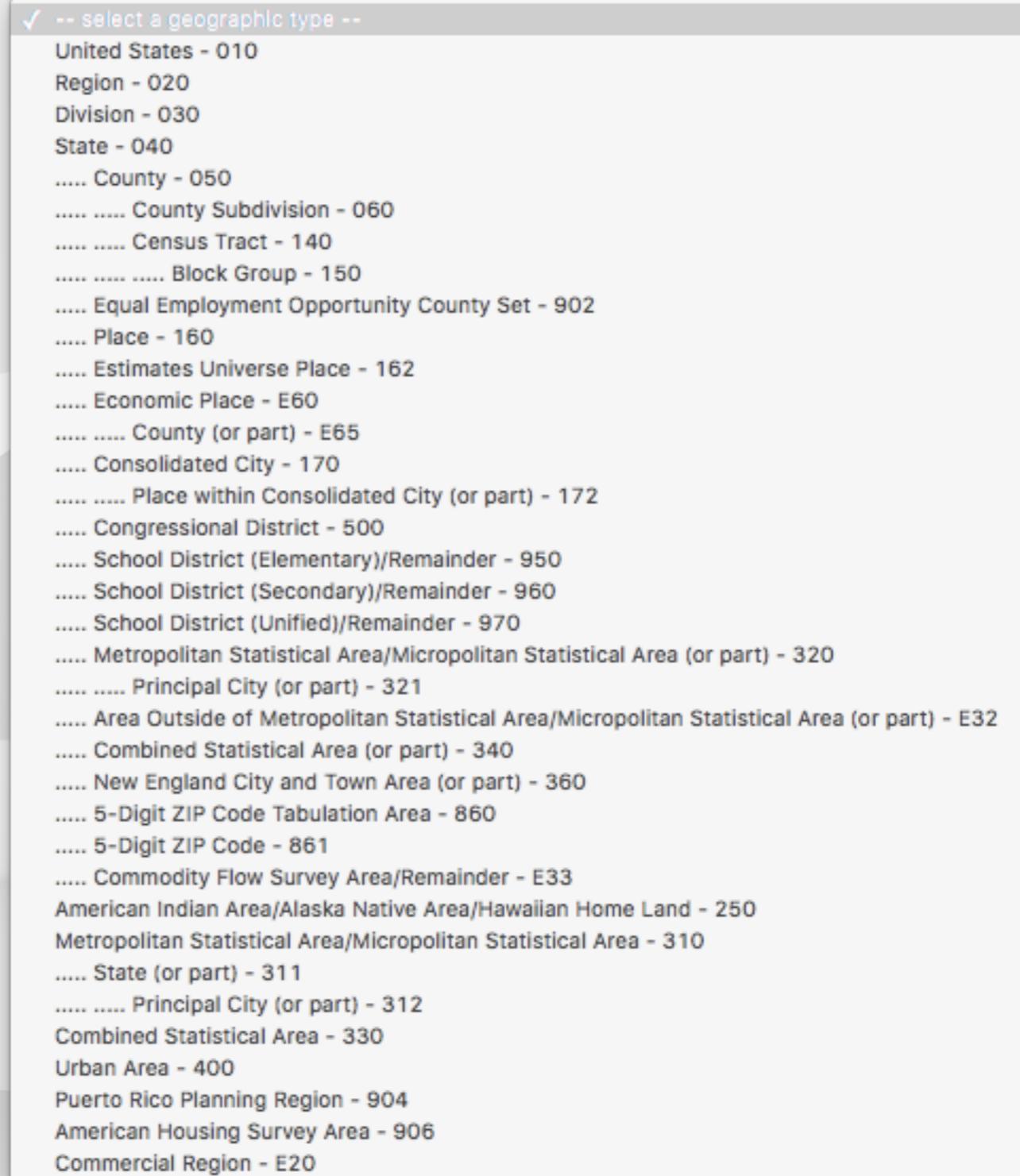


Cartogram



Some common spatial polygons

- States
- Census blocks
- Counties
- Zip codes
- School districts
- ... and many more!



The problem comes when you need to combine data at different spatial aggregation levels.

Combining tabular data

Combine Data Sets

a		b	
x1	x2	x1	x3
A	1	A	T
B	2	B	F
C	3	D	T

+

=

Mutating Joins

x1	x2	x3
A	1	T
B	2	F
C	3	NA

dplyr::left_join(a, b, by = "x1")

Join matching rows from b to a.

x1	x3	x2
A	T	1
B	F	2
D	T	NA

dplyr::right_join(a, b, by = "x1")

Join matching rows from a to b.

x1	x2	x3
A	1	T
B	2	F

dplyr::inner_join(a, b, by = "x1")

Join data. Retain only rows in both sets.

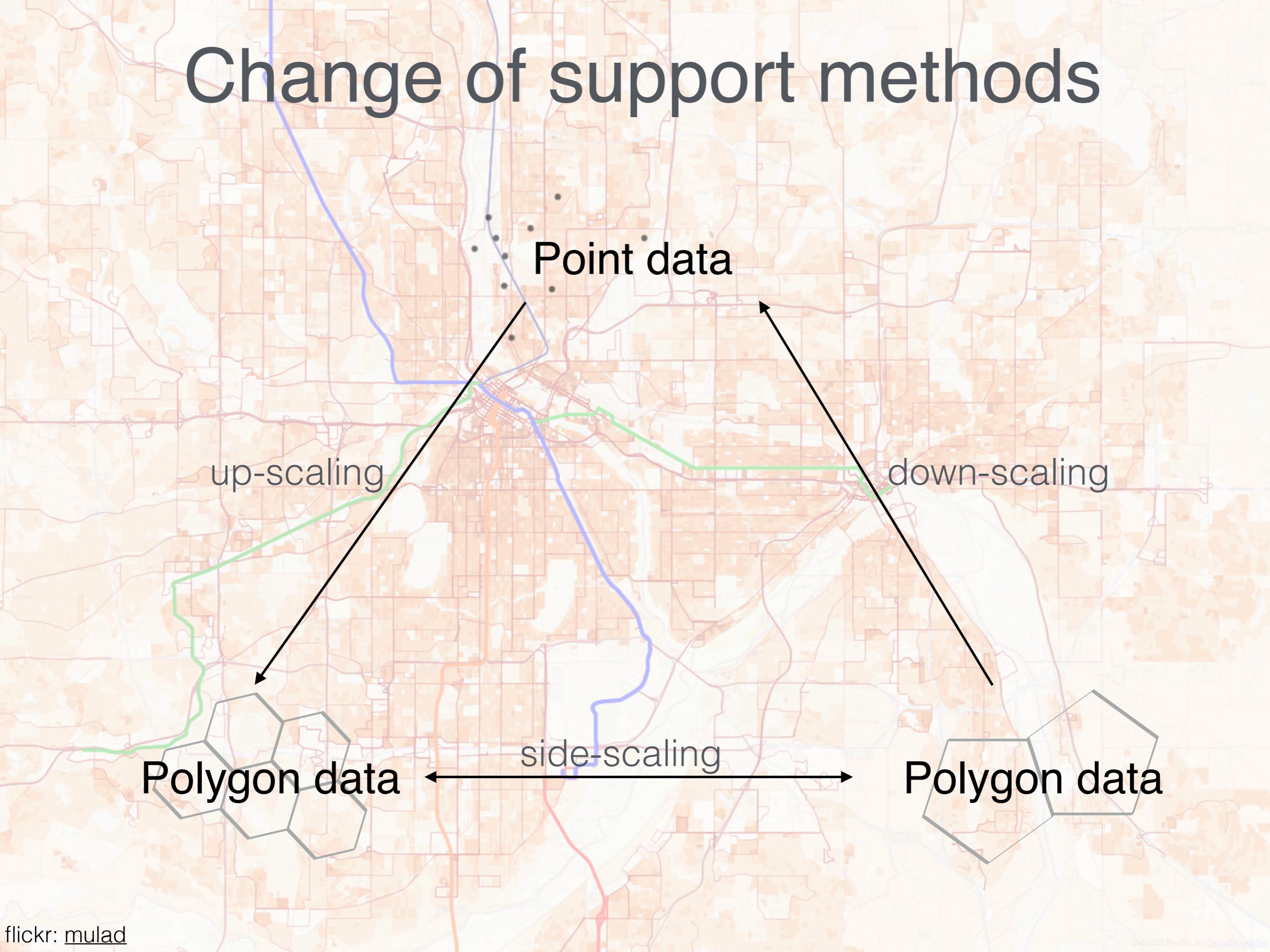
x1	x2	x3
A	1	T
B	2	F
C	3	NA
D	NA	T

dplyr::full_join(a, b, by = "x1")

Join data. Retain all values, all rows.

RStudio data wrangling cheatsheet

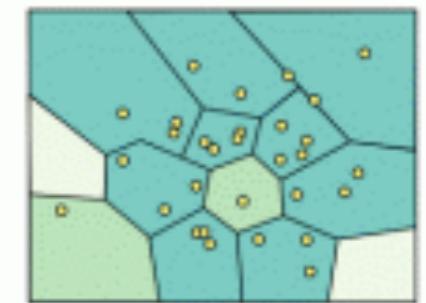
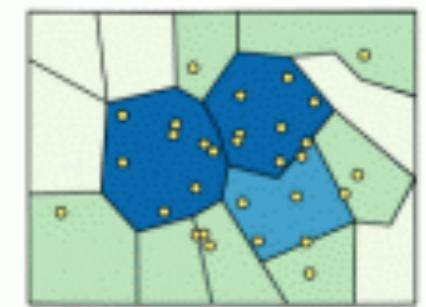
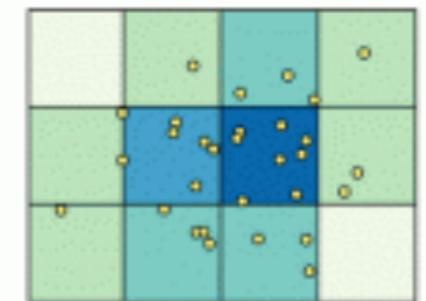
Change of support methods



Easiest first—
up-scaling

Modifiable Areal Unit Problem

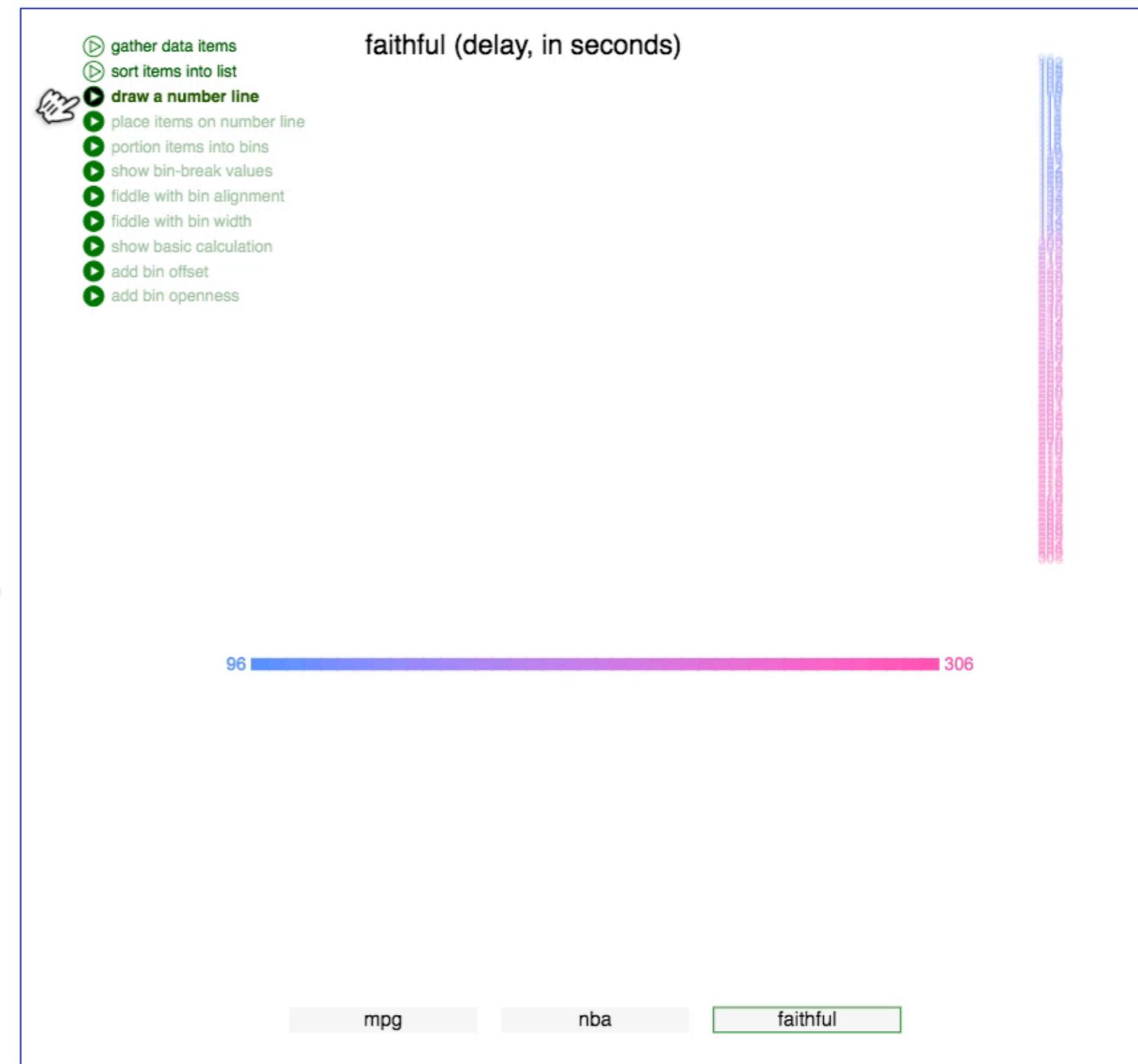
“The areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating”
- Stan Openshaw



<http://gispopsci.org/maup/>

Prepare a number line

A common convention is to use a number line, where higher values move to the right and smaller (or negative) values to the left. We can draw a line representing all possible numbers between the minimum and maximum data values.



Gerrymandering



Gerrymandering: Last Week Tonight with John Oliver (HBO)

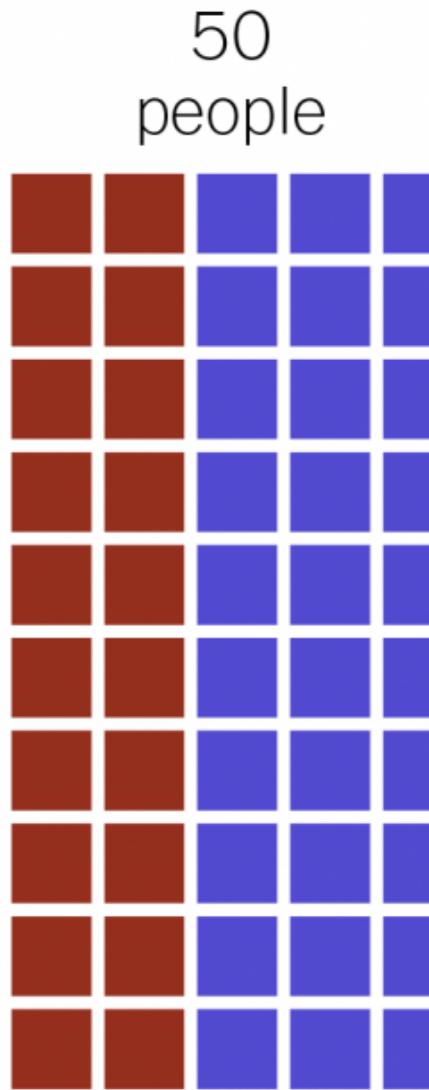


▶ ▶ 🔊 1:11 / 19:33

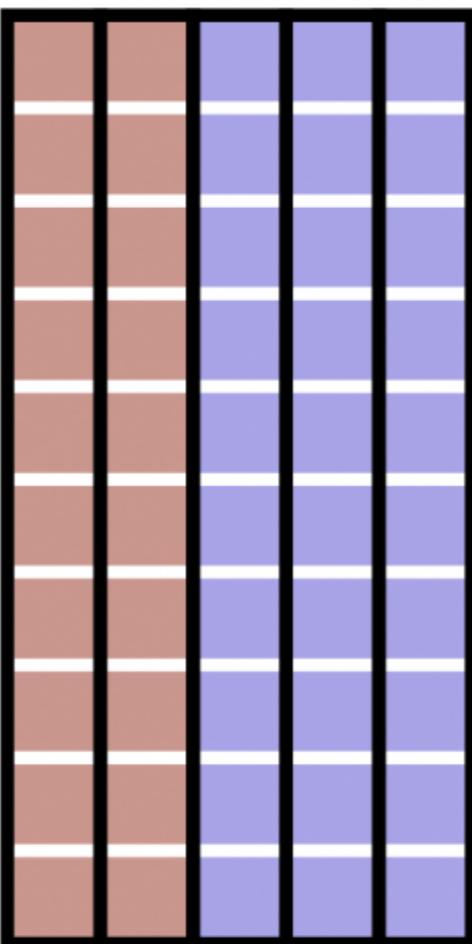


Gerrymandering, explained

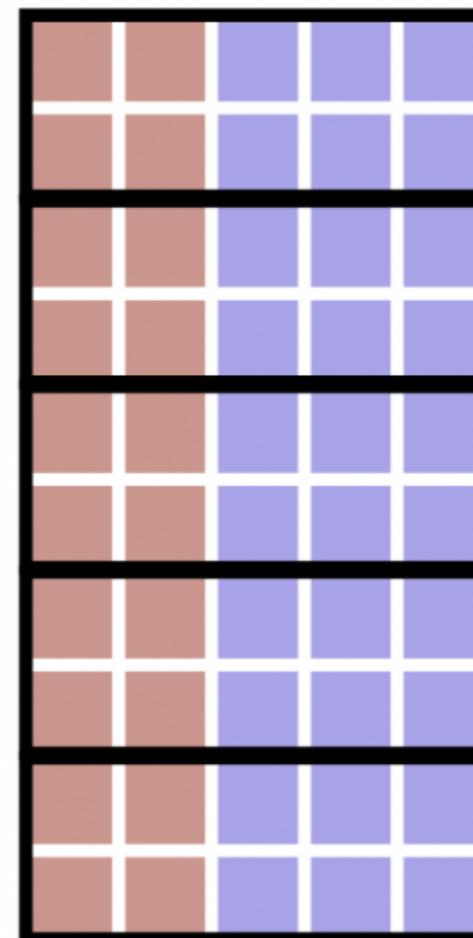
Three different ways to divide 50 people into five districts



1. Perfect representation

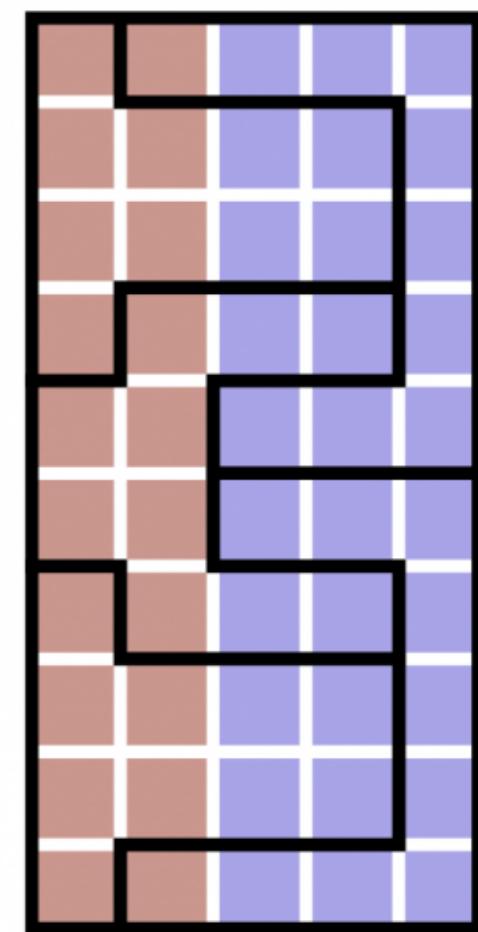


2. Compact, but unfair



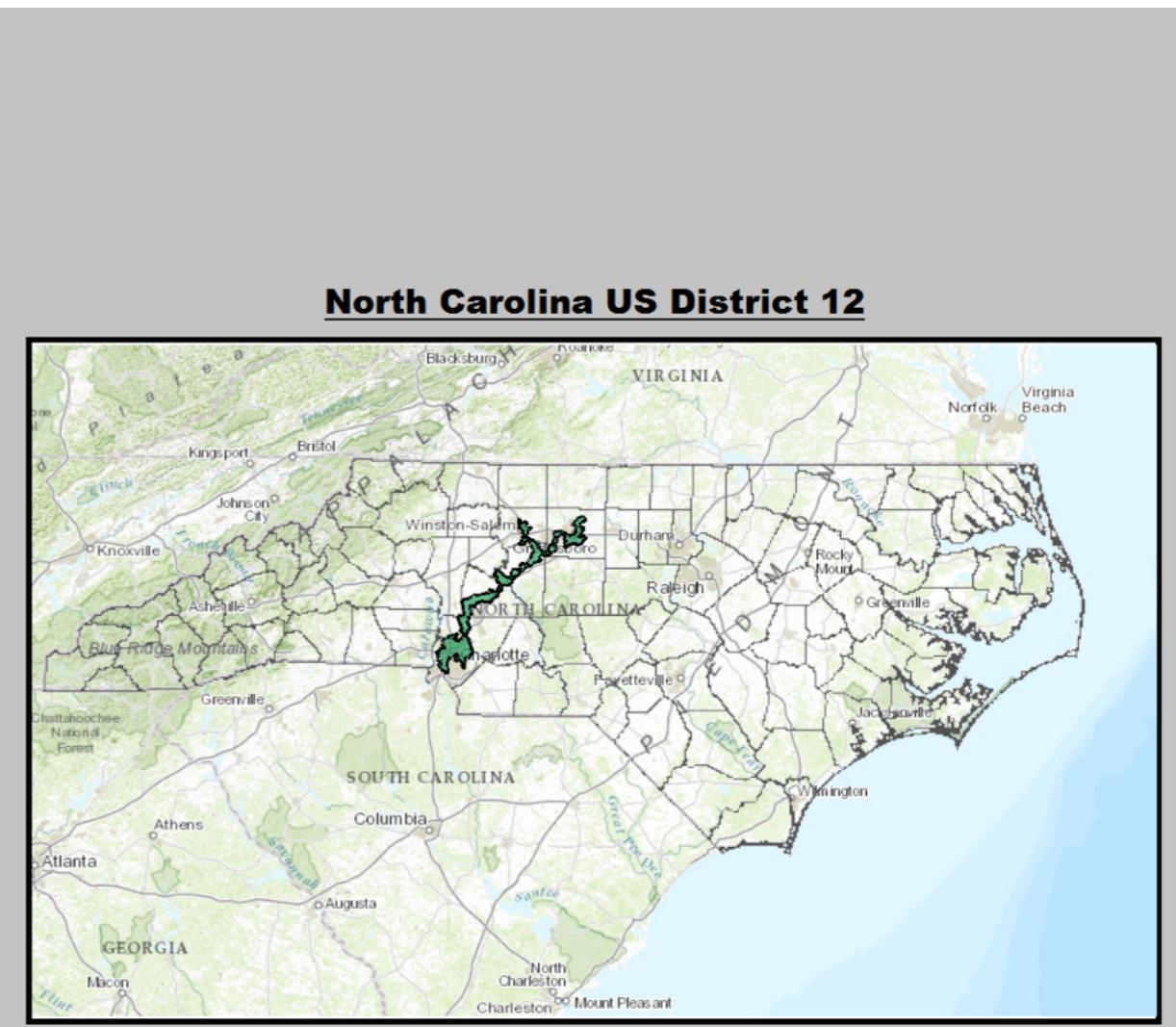
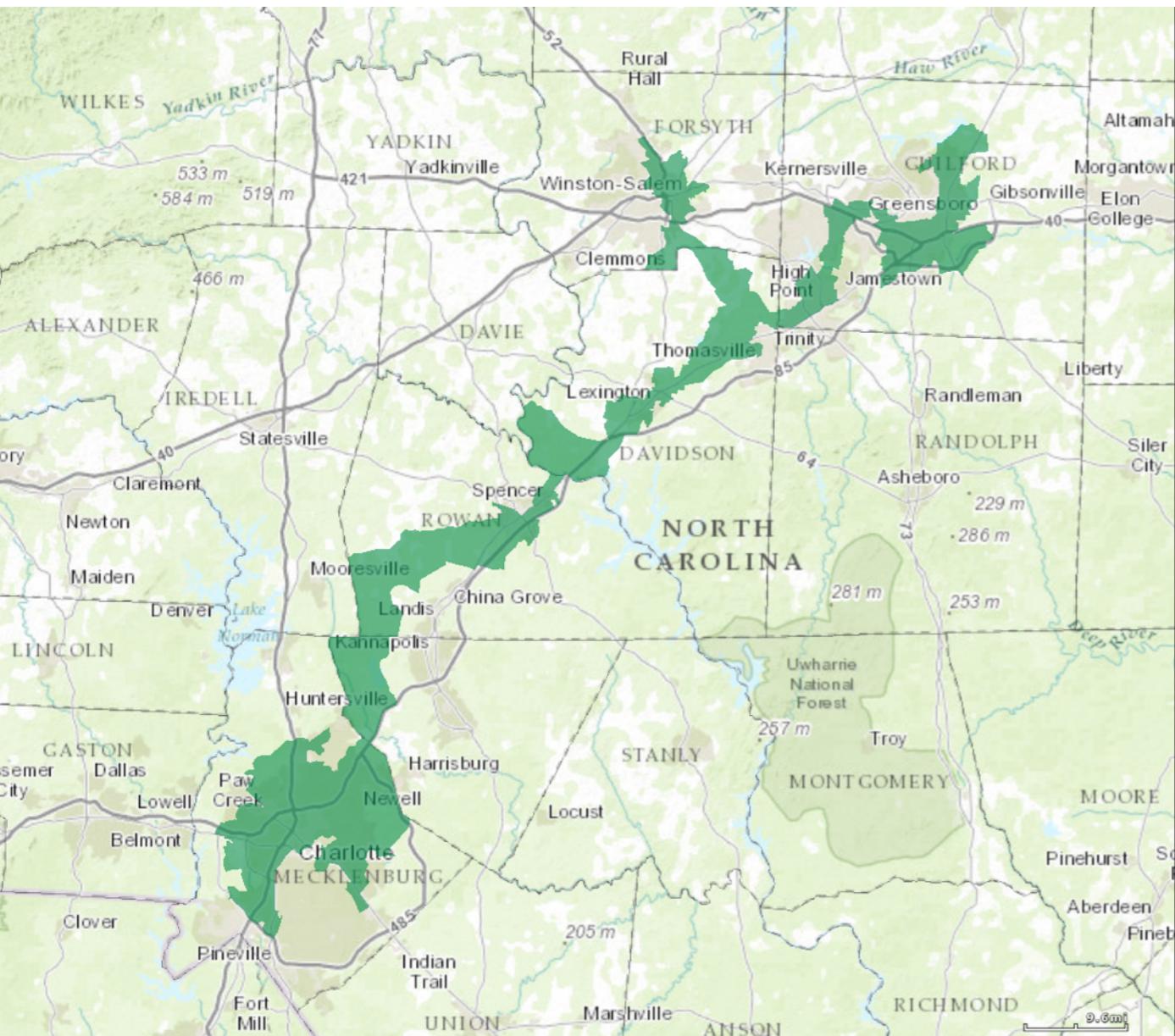
BLUE WINS

3. Neither compact nor fair



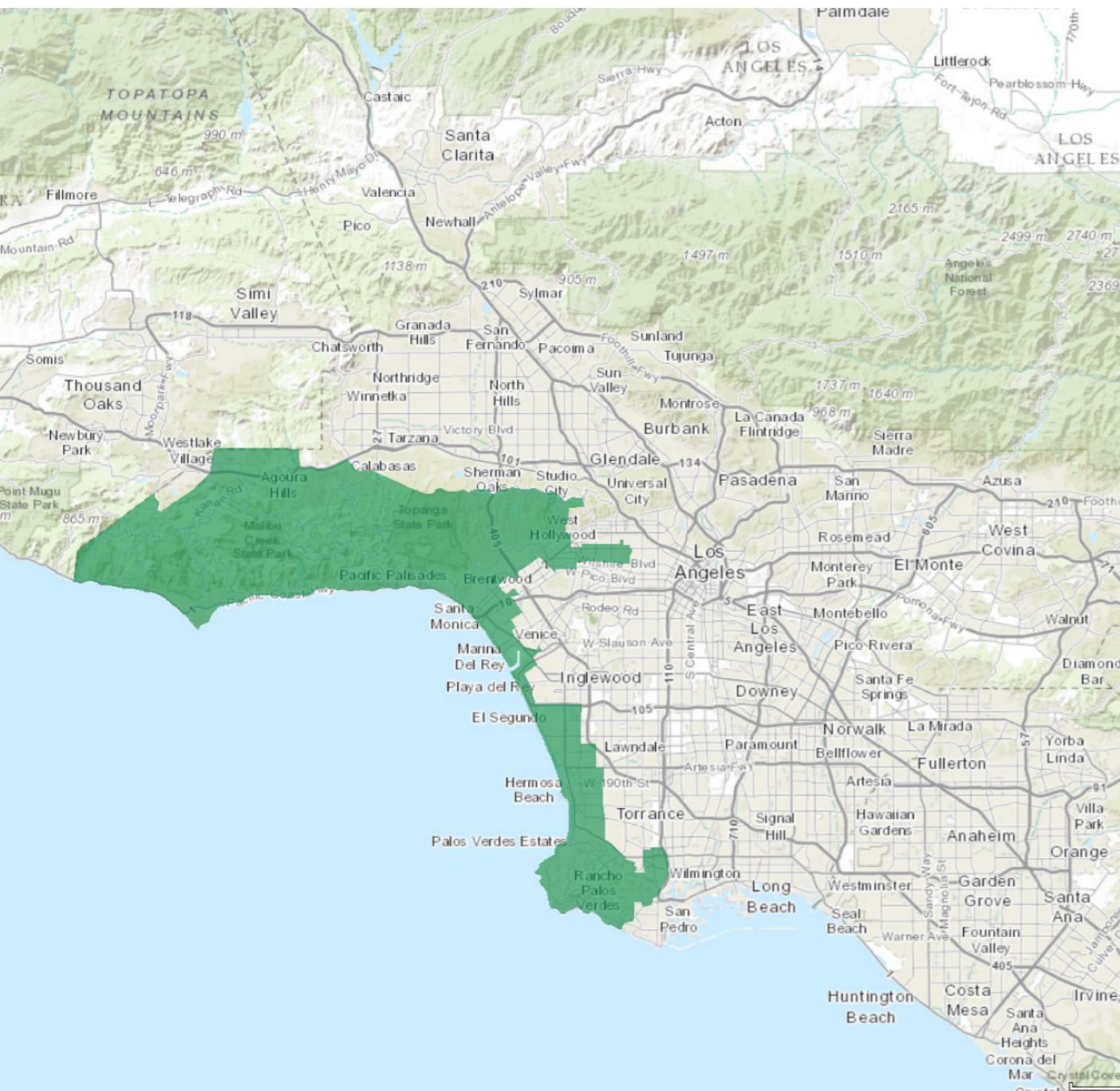
RED WINS

North Carolina's 12th district

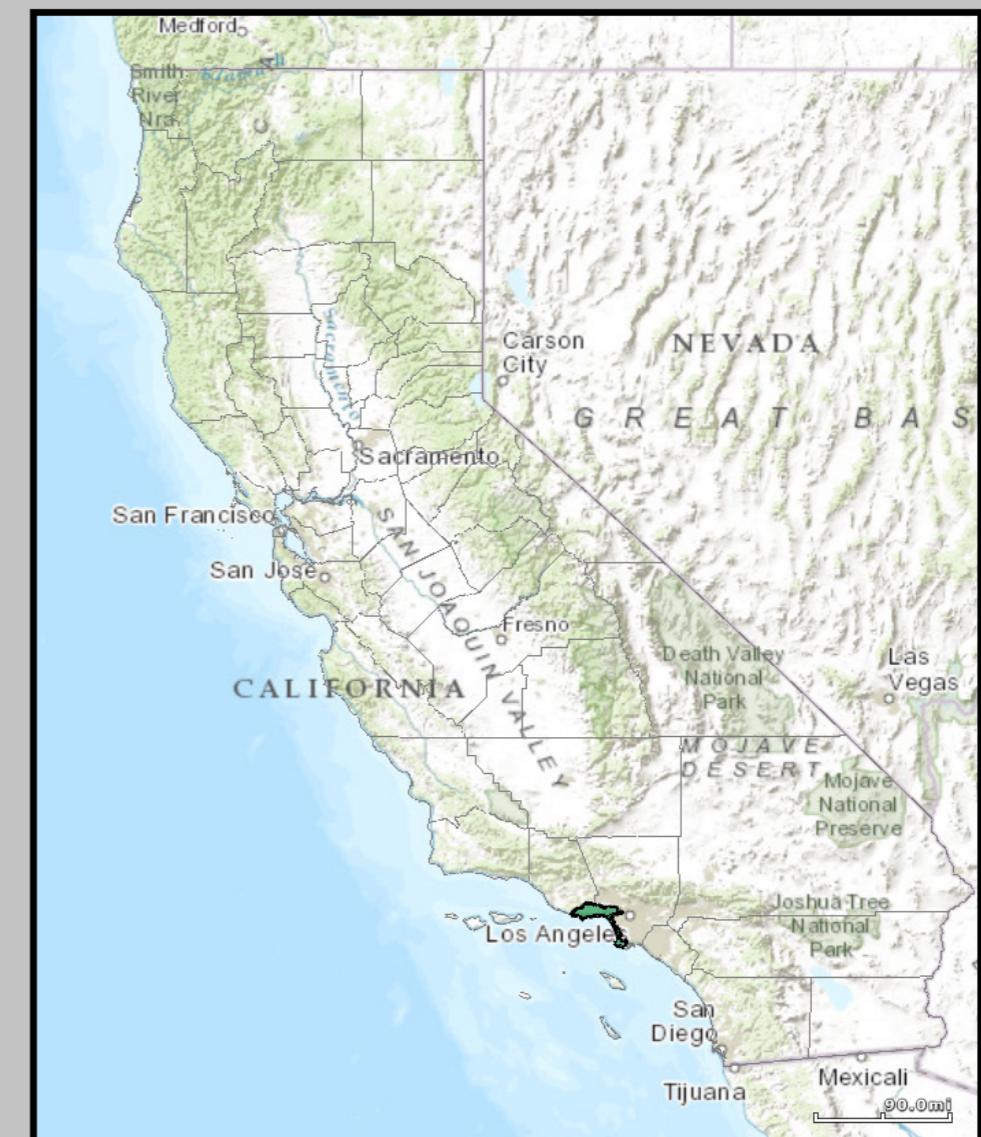


US Congressional districts since 2013
Source: <http://nationalatlas.gov>, 1 Million Scale project.

California's 33rd district



California US District 33



US Congressional districts since 2013
Source: <http://nationalatlas.gov>, 1 Million Scale project.

▼ mission

Perform a Partisan Gerrymander
Gain a third Democratic district

Ver. 1.0.1

THE ReDISTRICTING GAME

1 DRAW &
REDRAW
MAP

2 GET
FEED
BACK

3 SUBMIT
FOR
APPROVAL

(17,14)

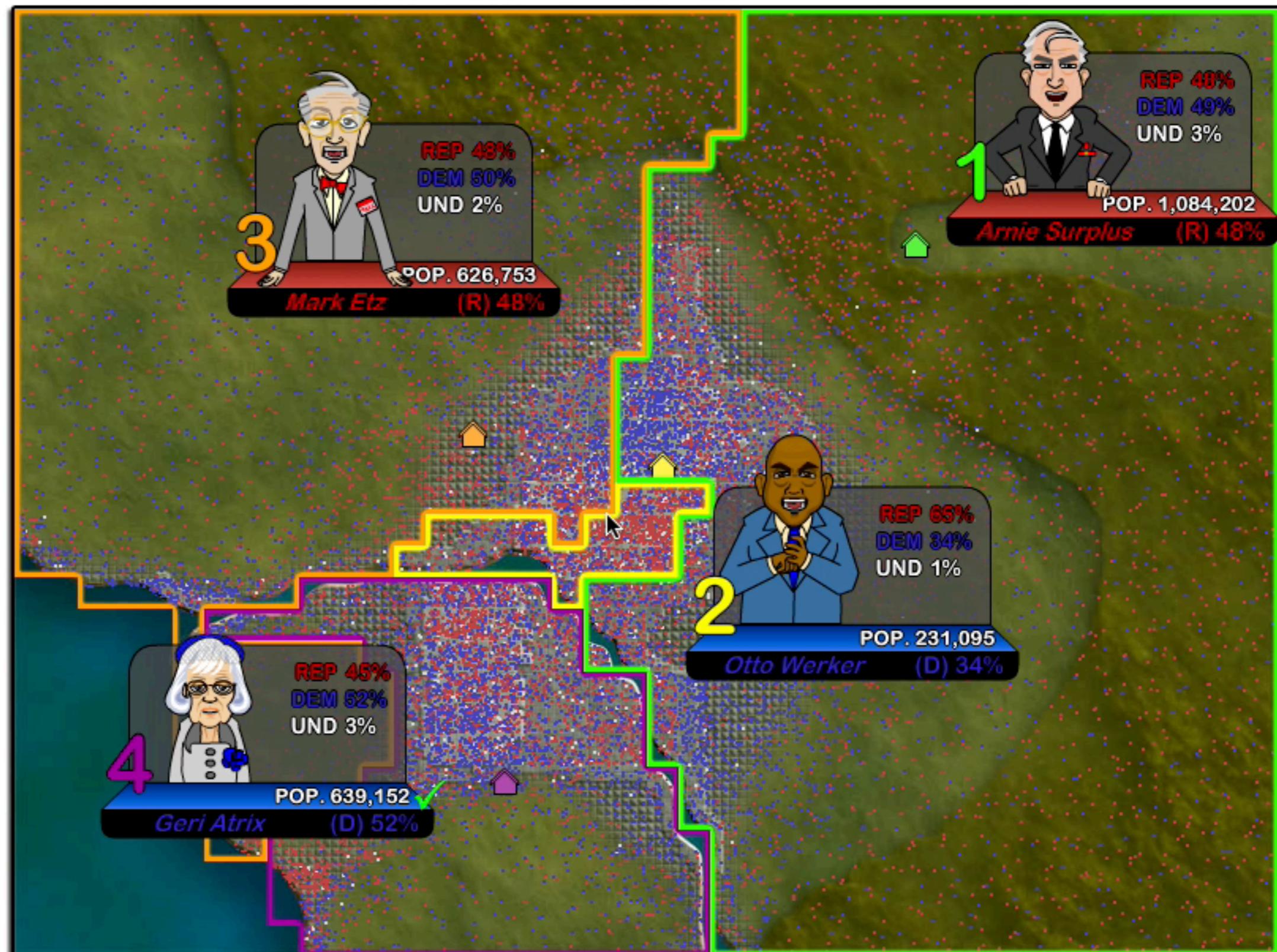
Pop: 12001

DEM: 50%

REP: 50%

UND: 0%

MISSION GOAL



PARTY

TERRAIN

HELP

UNDO

Toward a Talismanic Redistricting Tool: A Computational Method for Identifying Extreme Redistricting Plans.

Wendy Tam Cho and Yan Liu
<http://bit.ly/TalismanicMaps>



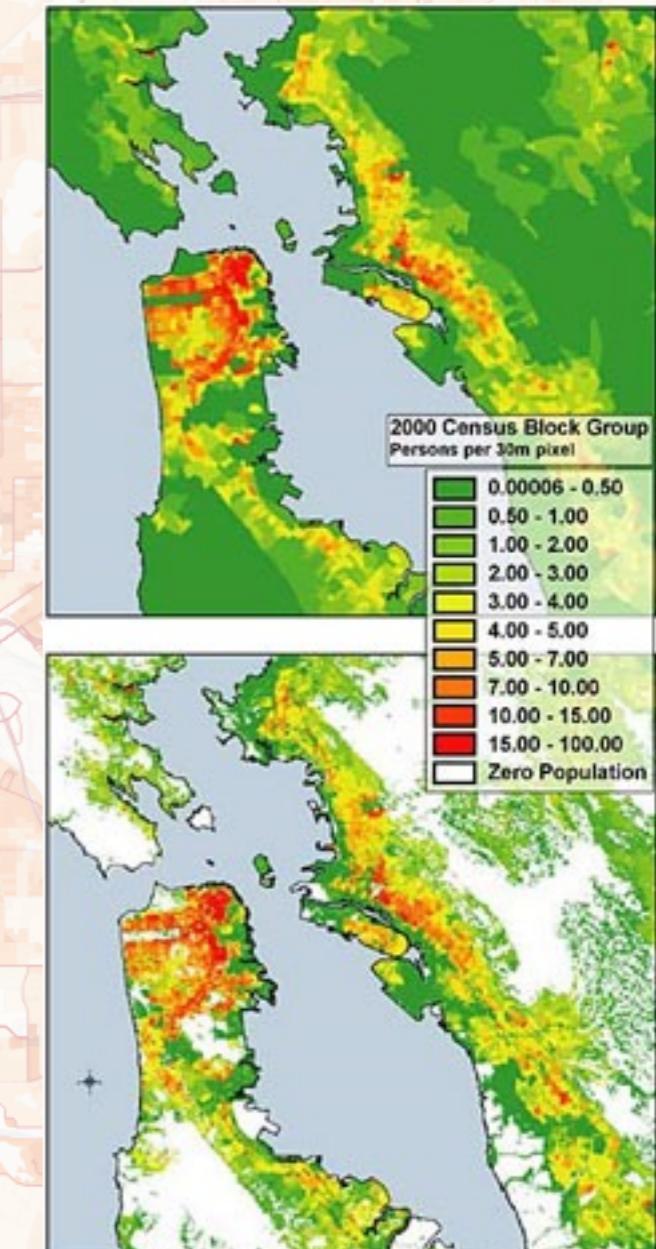
down-scaling



Dasymetric map

"on which population density, irrespective of any administrative boundaries, is shown as it is distributed in reality, i.e. by natural spots of concentration and rarefaction."

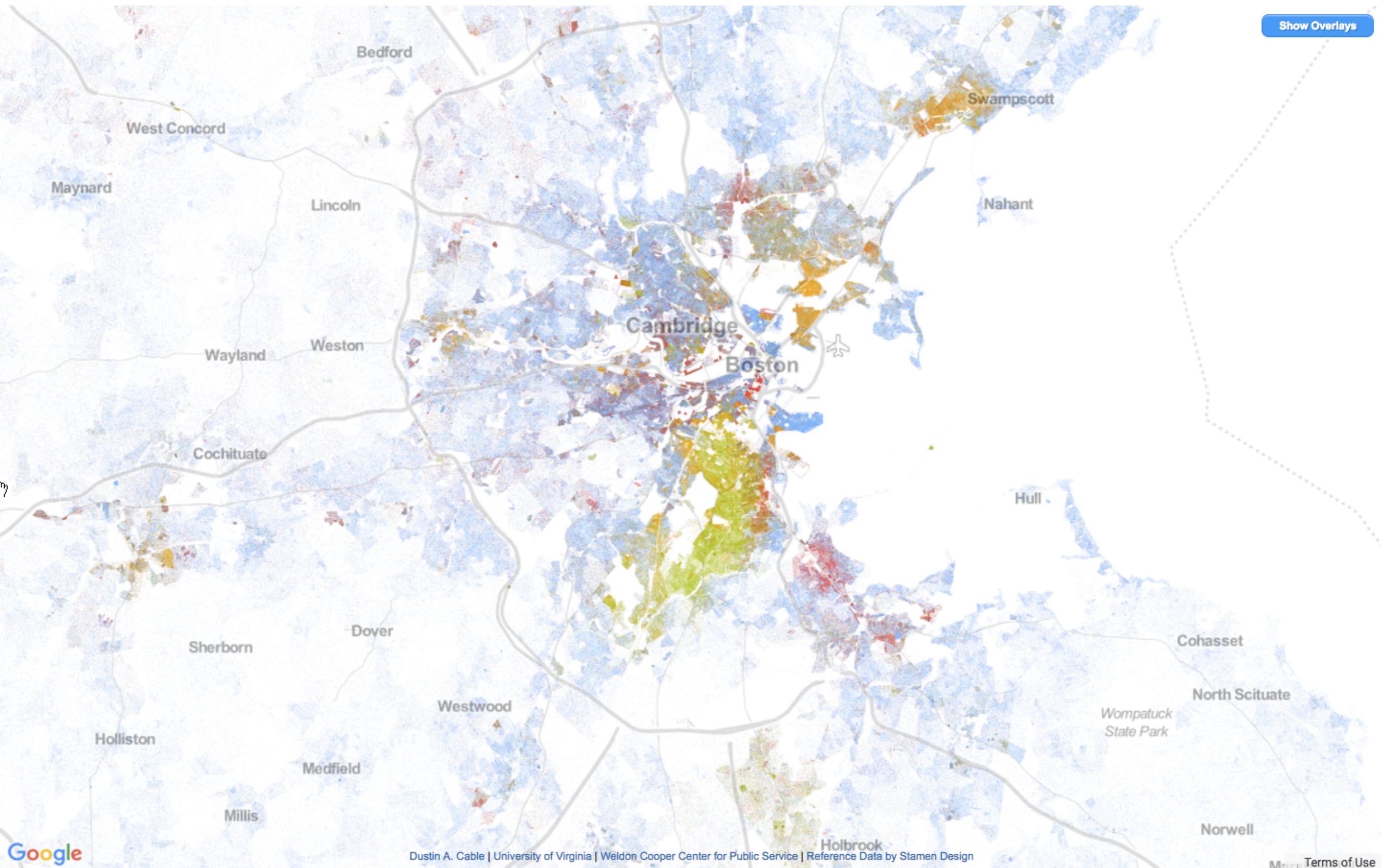
-Semenov-Tyan-Shansky



wikipedia: dasymetric map

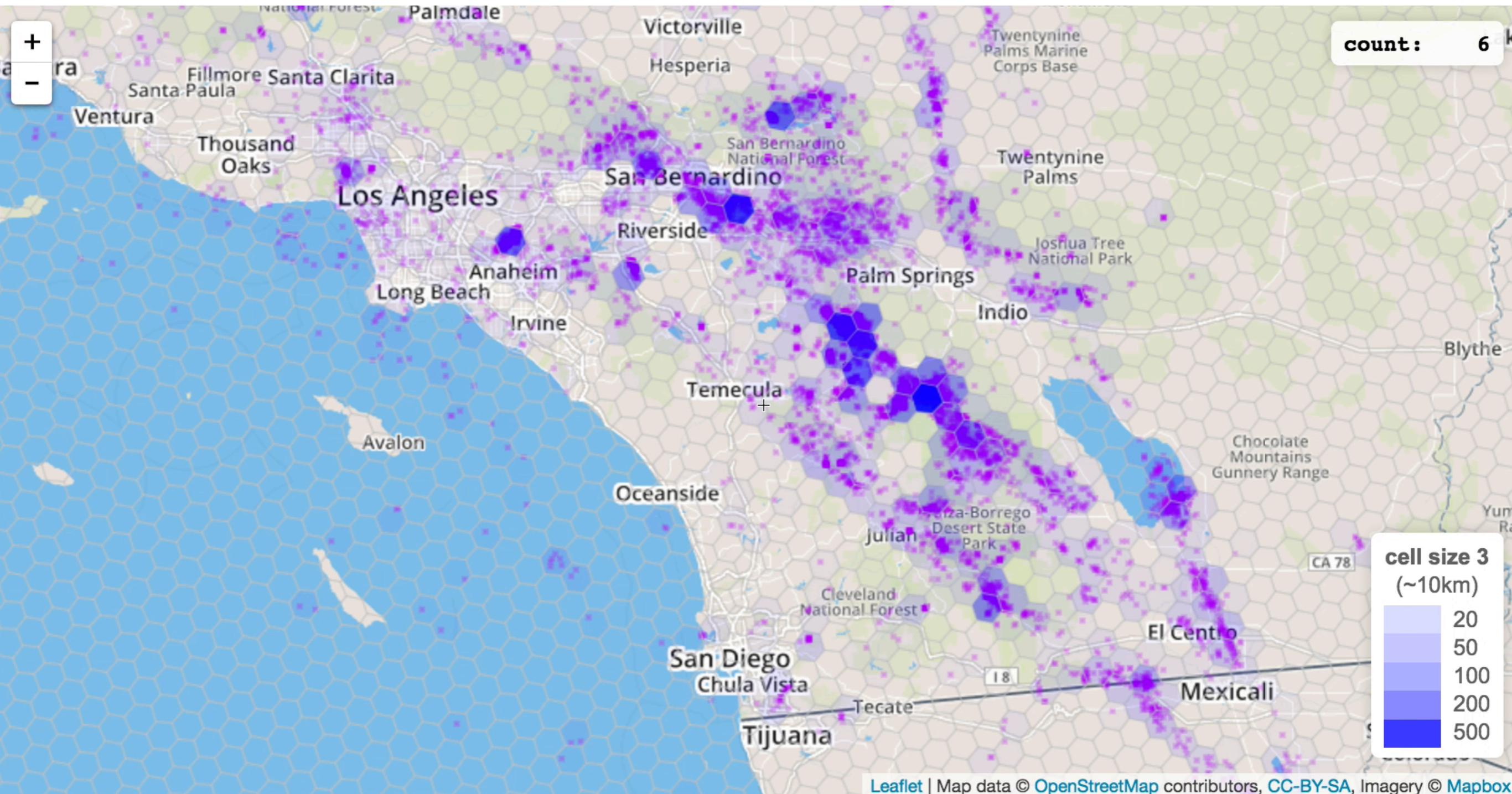








side-scaling
(the hardest)

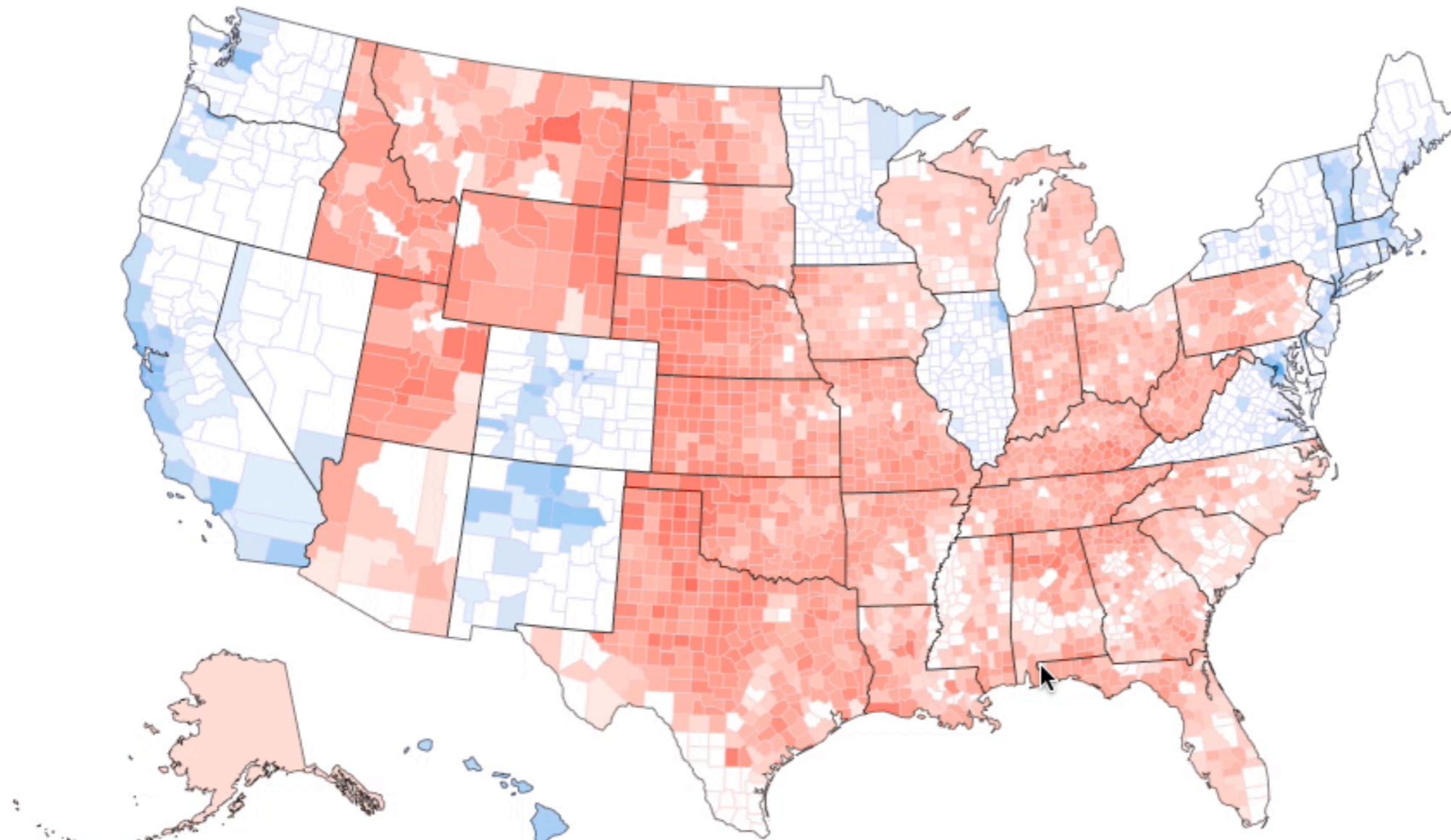


The background of the slide features a subtle, abstract geometric pattern composed of nested polygons, primarily triangles and quadrilaterals, in shades of light gray.

with nested polygons,
not so bad

233

305



Instructions

How few counties can you move to make Hillary Clinton win the 2016 election?

Choose a county (or several) to move to a new state. Then click the **Move** button and the state you want to move your counties to.

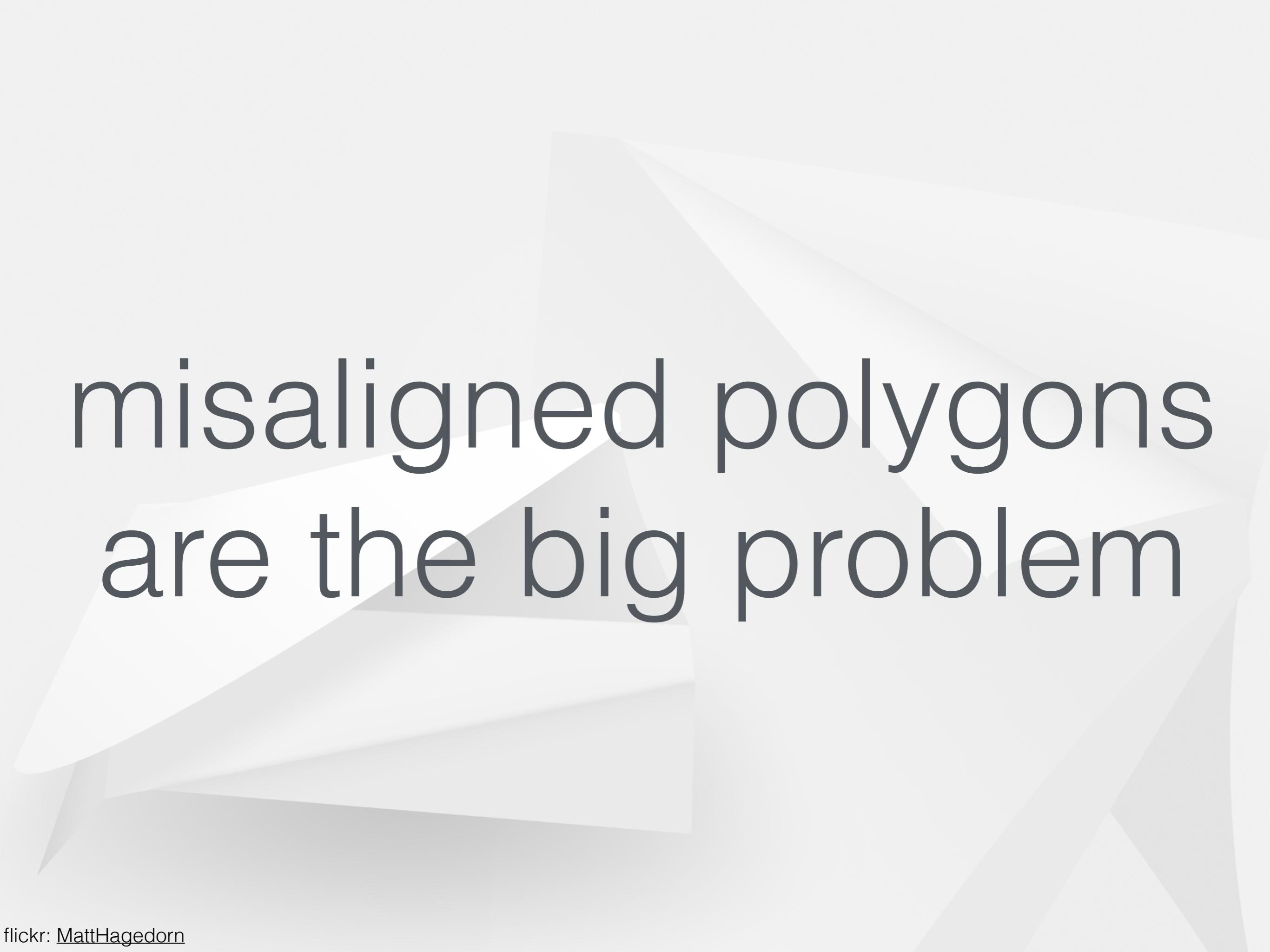
We'll automatically recompute the number of electoral votes the state would get with their new counties, and update the electoral vote. However, we don't account for Maine and Nebraska's splitting of votes by congressional district.

Weep at how arbitrary our electoral system is.

[Move](#)[Hide Counties](#)[Share](#)

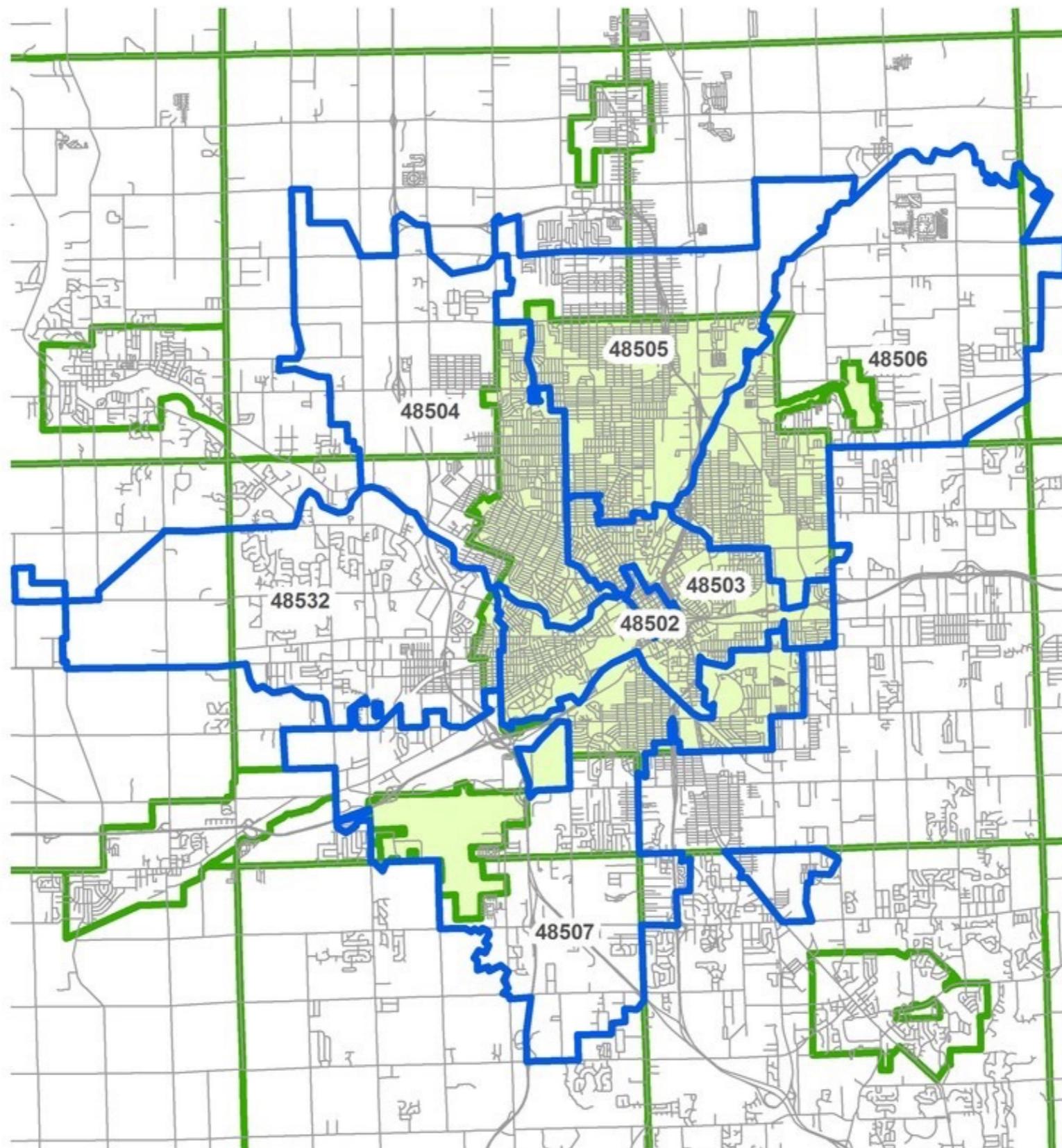
Year:

2016

The background of the slide features a subtle, abstract geometric pattern composed of numerous overlapping light gray triangles of varying sizes and orientations, creating a sense of depth and movement.

misaligned polygons
are the big problem

Misalignment between Flint ZIP Codes and City of Flint



Flint ZIP Codes

City of Flint

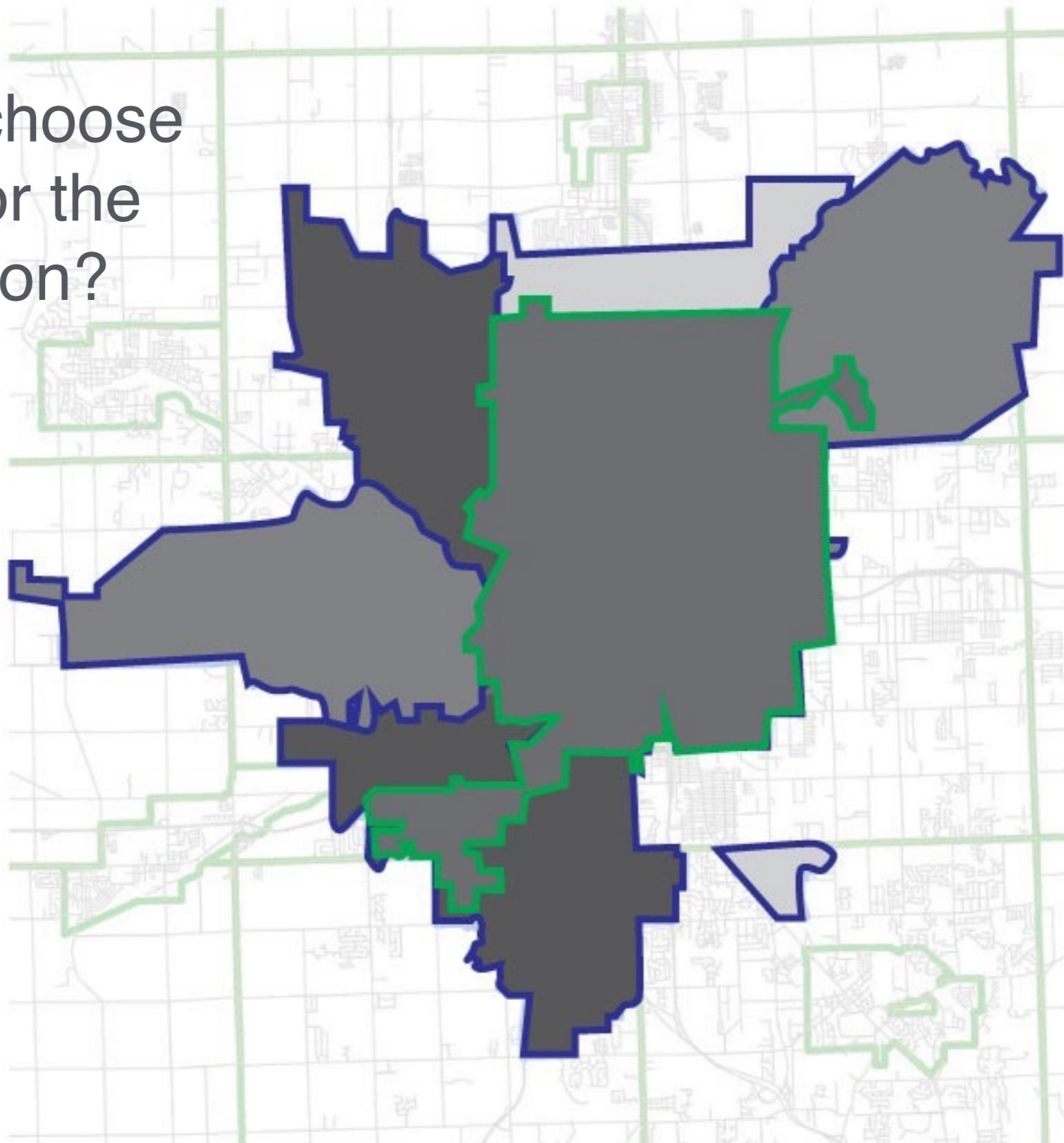
Other Municipalities

0 1 2 4 Miles



Misalignment between Flint ZIP Codes and City of Flint

How do we choose
the value for the
green region?



Flint ZIP Codes

City of Flint

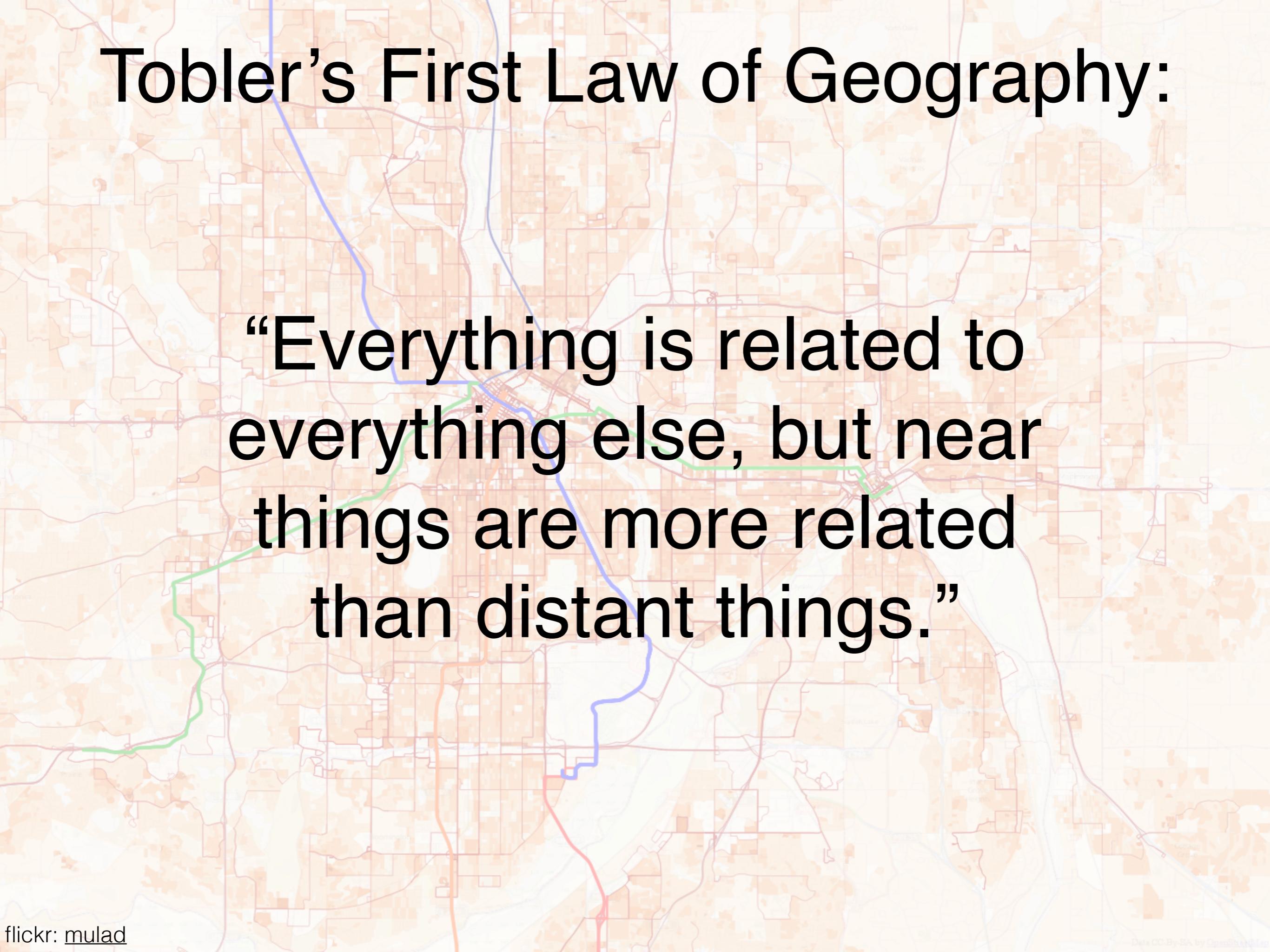
Other Municipalities

0 1 2 4 Miles



Tobler's First Law of Geography:

“Everything is related to everything else, but near things are more related than distant things.”

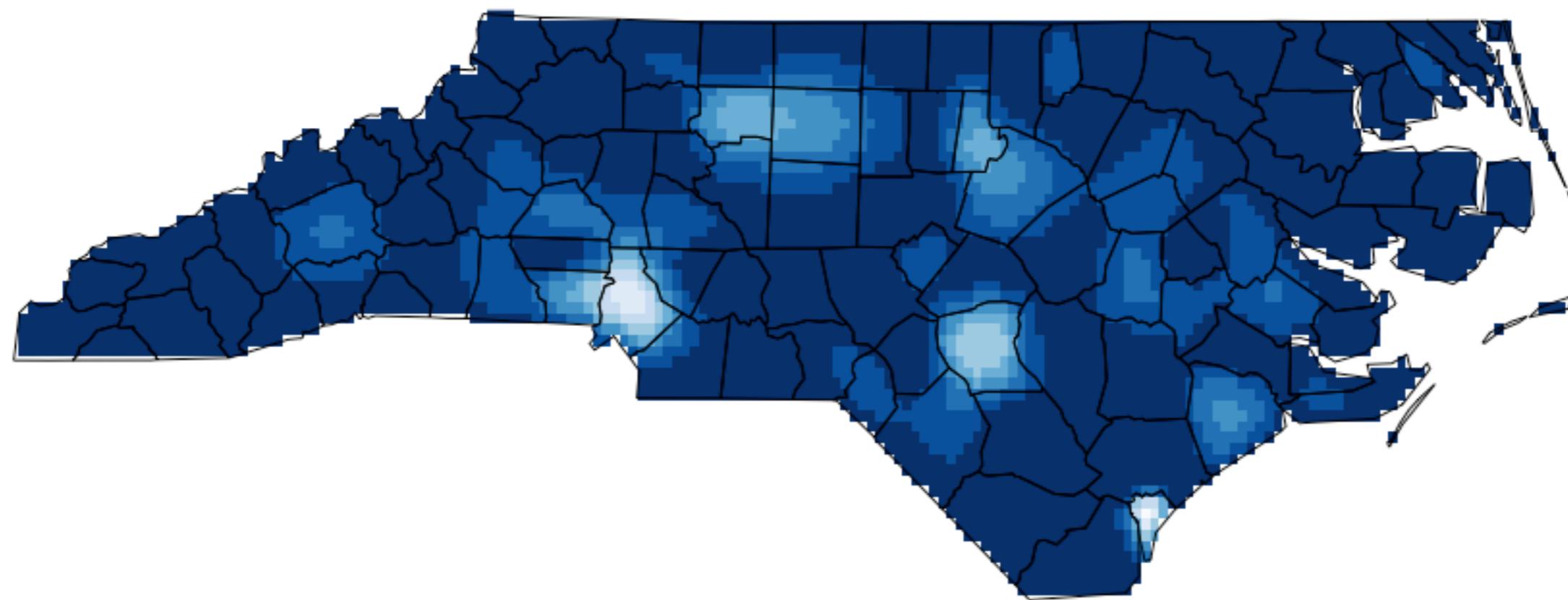


Tobler's pycnophylactic property:

$$\int_{A_i} \lambda(s) ds = |A|$$

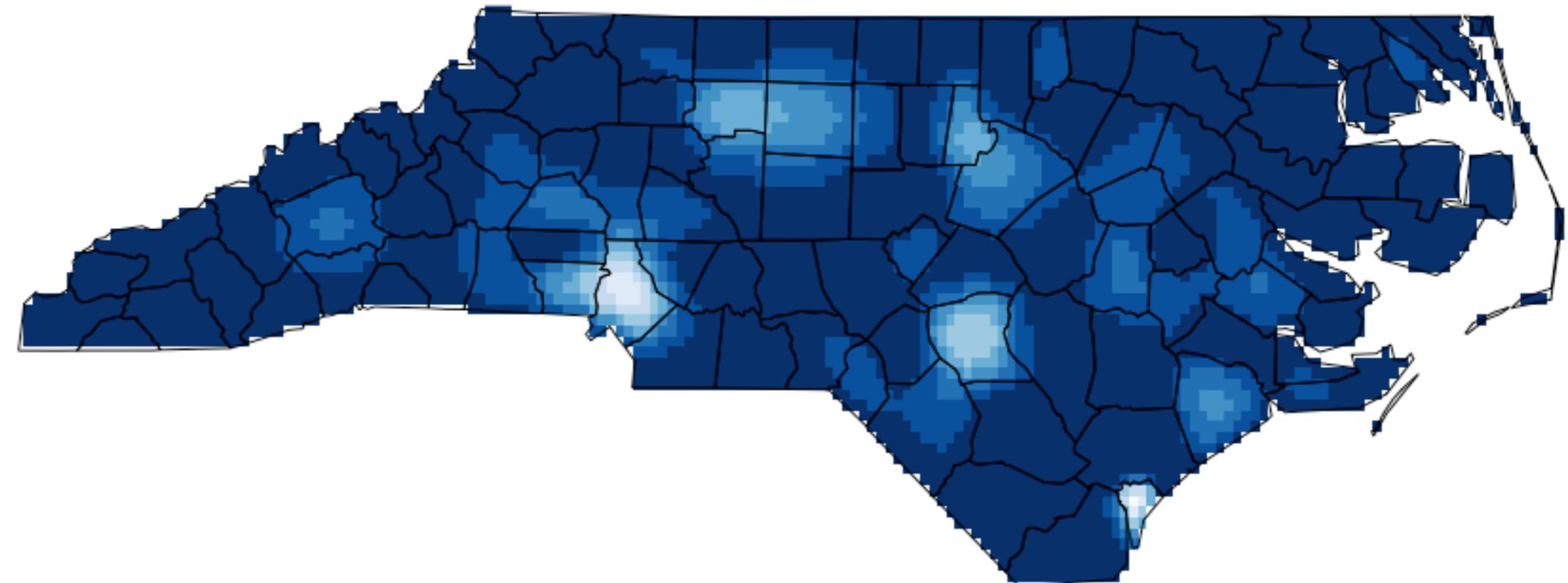
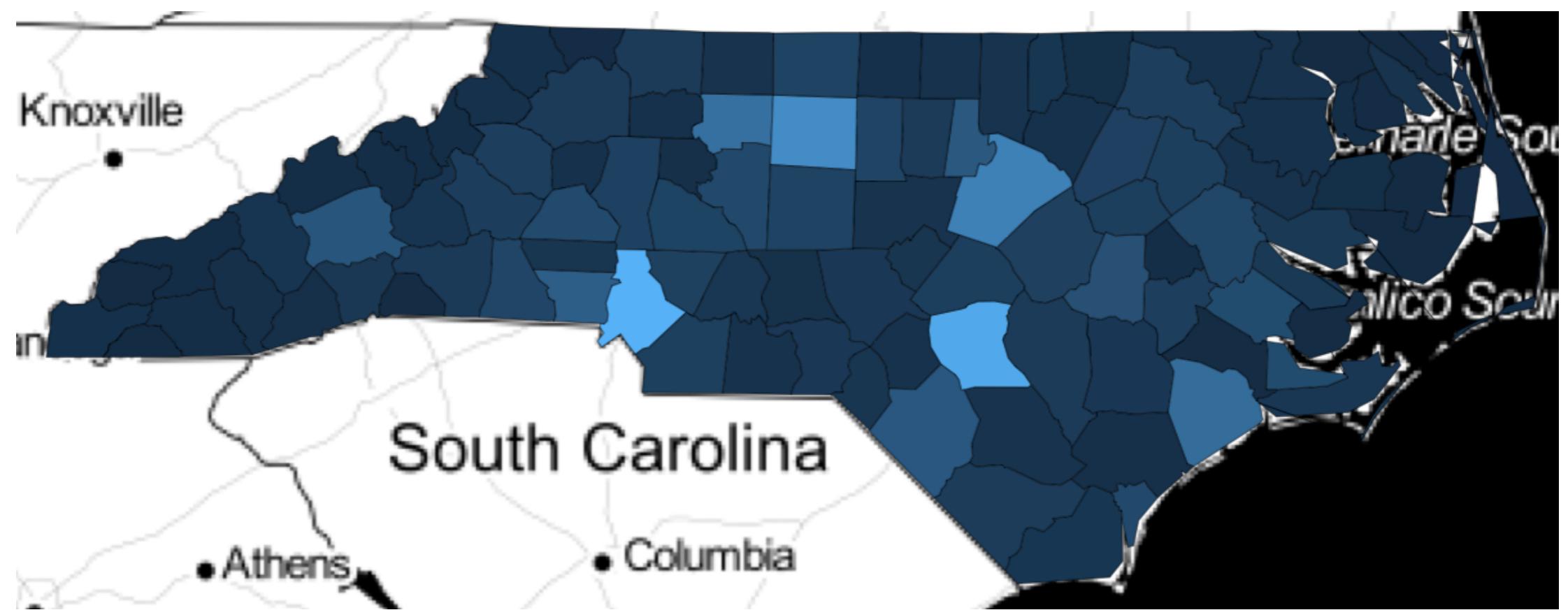
basically, you want your interpolation to be reversible

Working with the pycno package in R



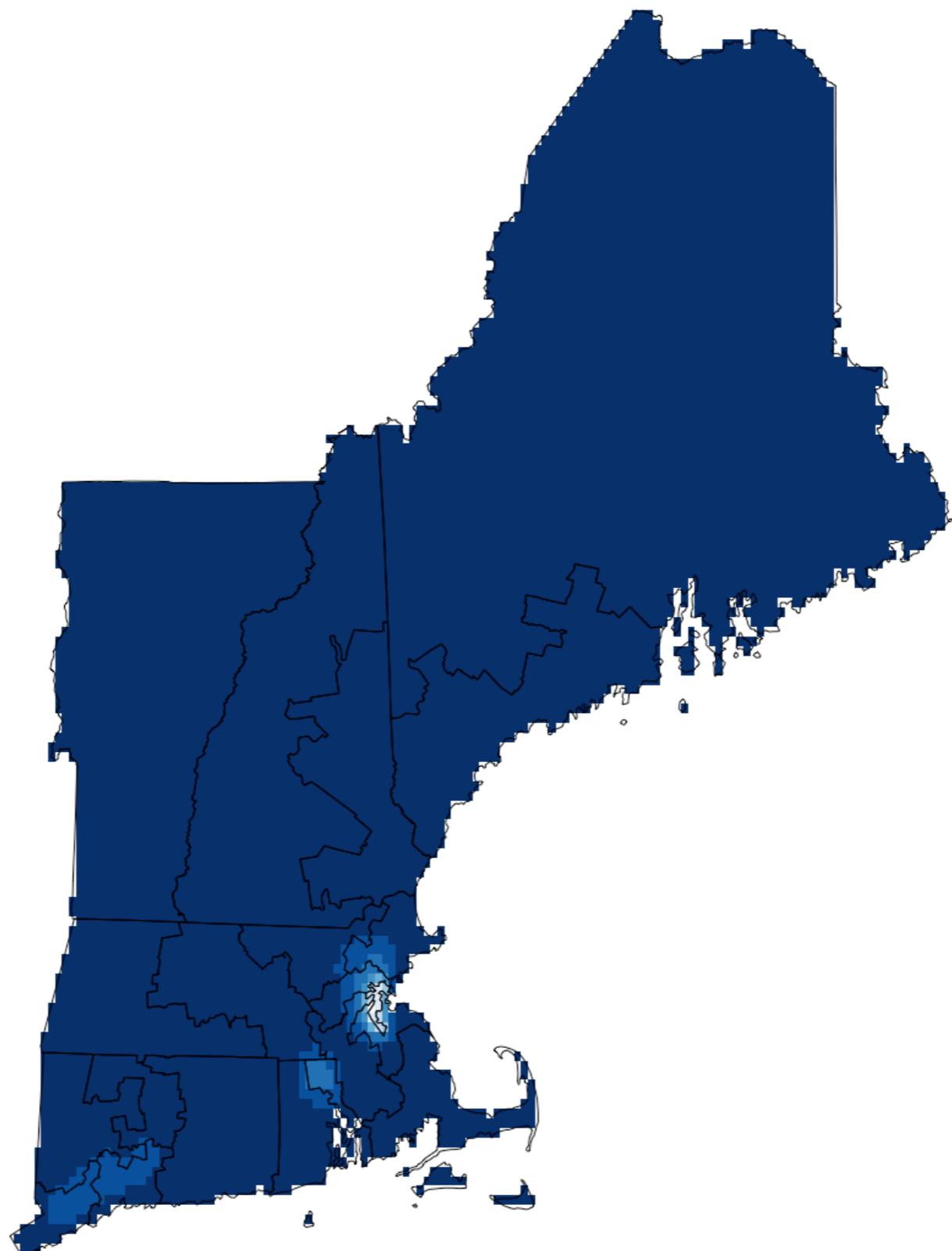
Joint work with students Jessica Mao and MyVan Vo.

Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems

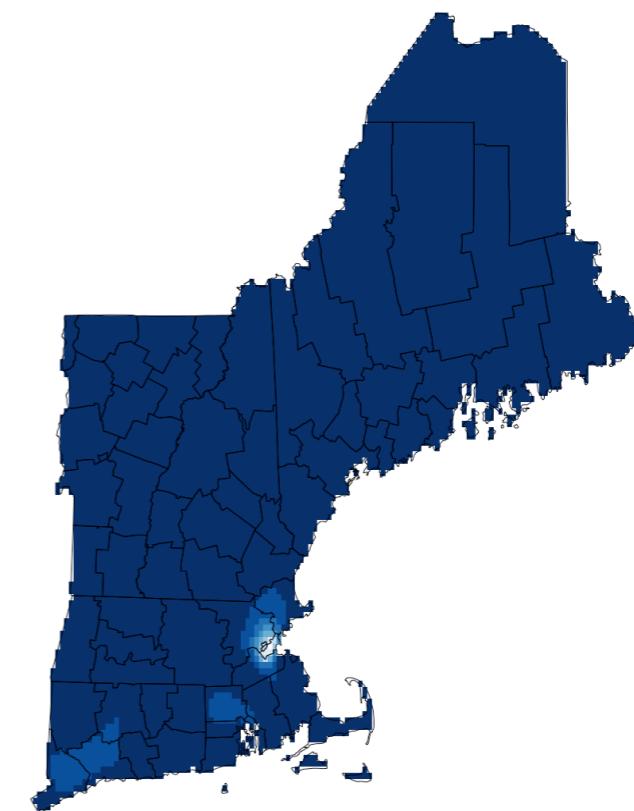
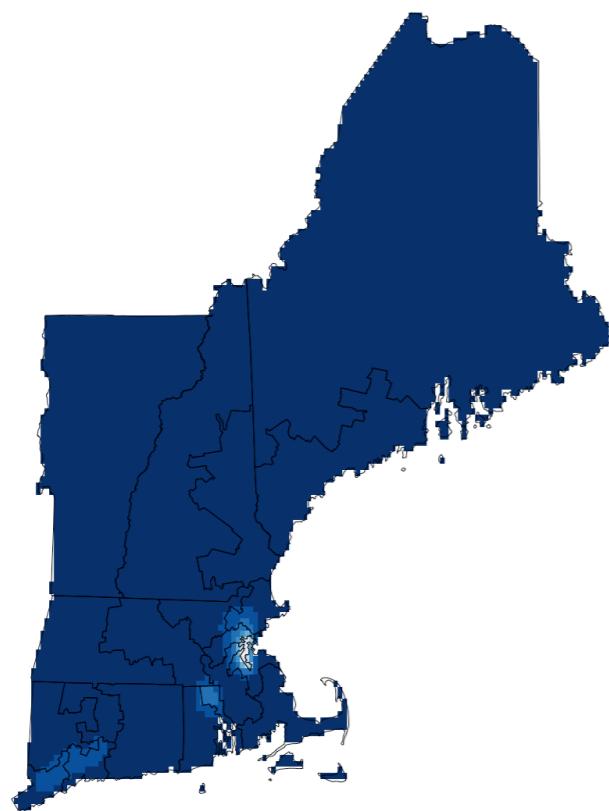
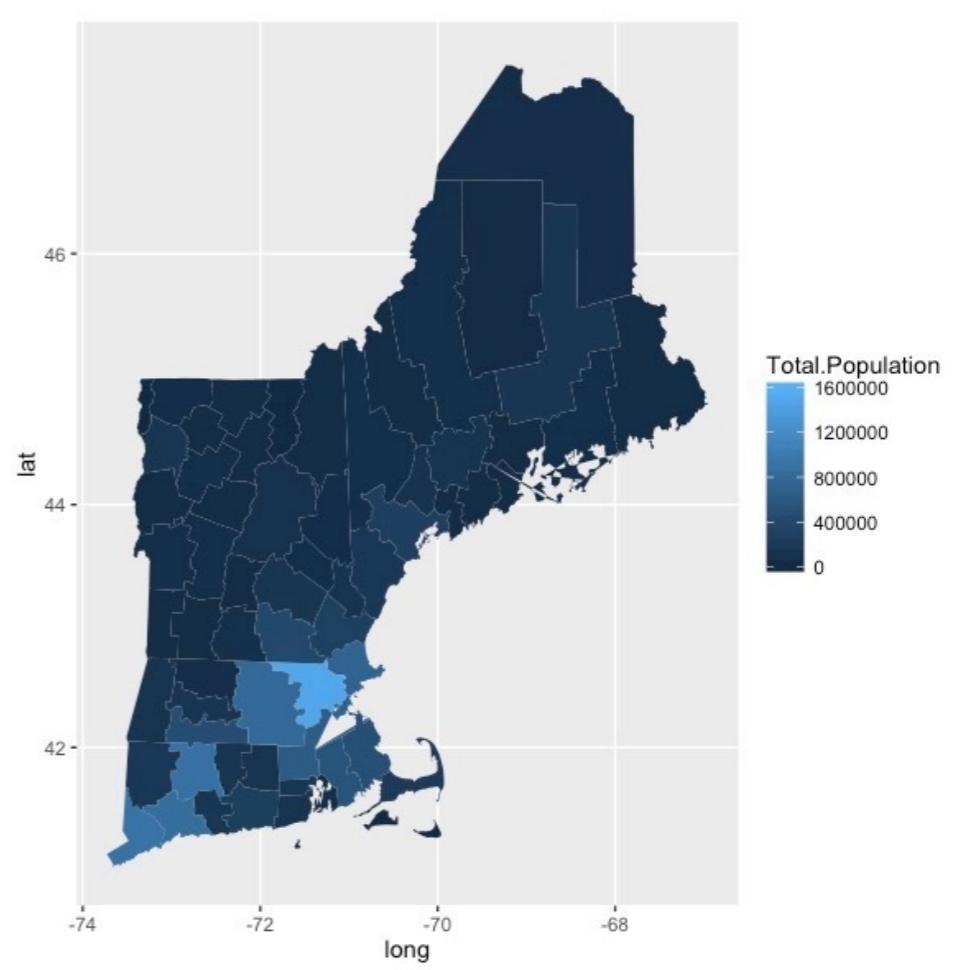
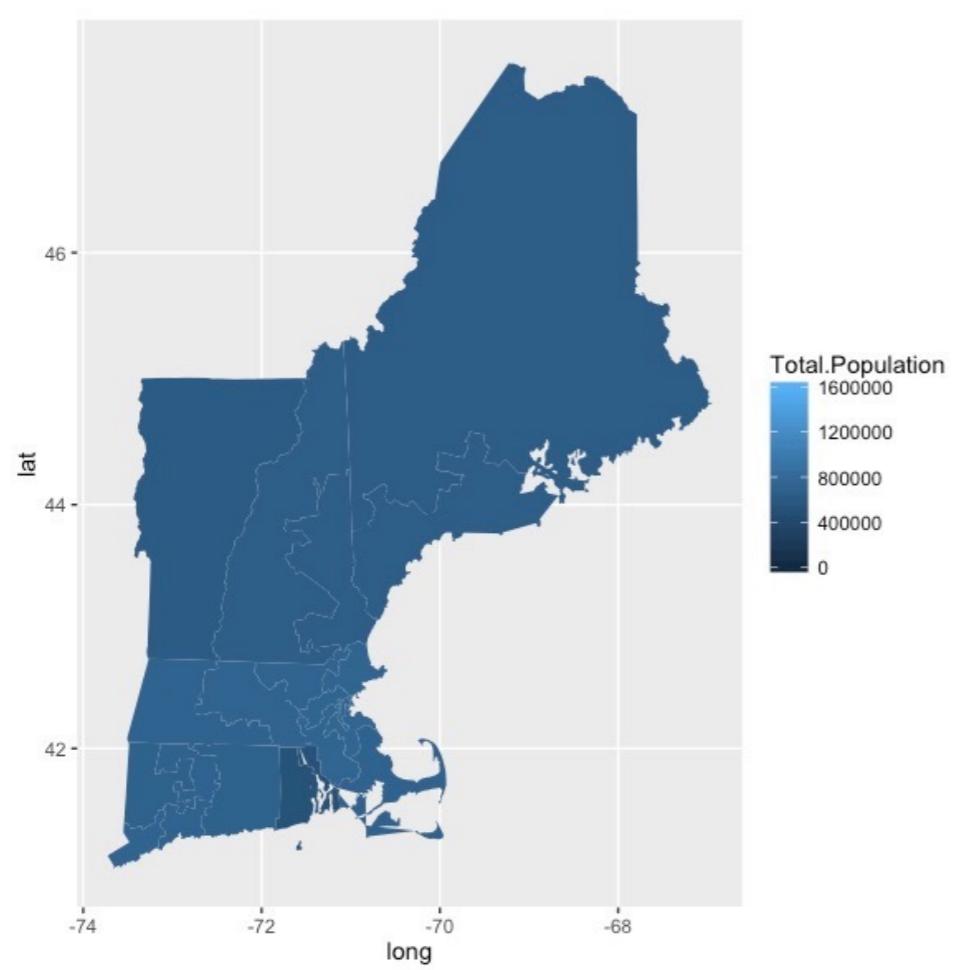


Joint work with students Jessica Mao and MyVan Vo.

Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems



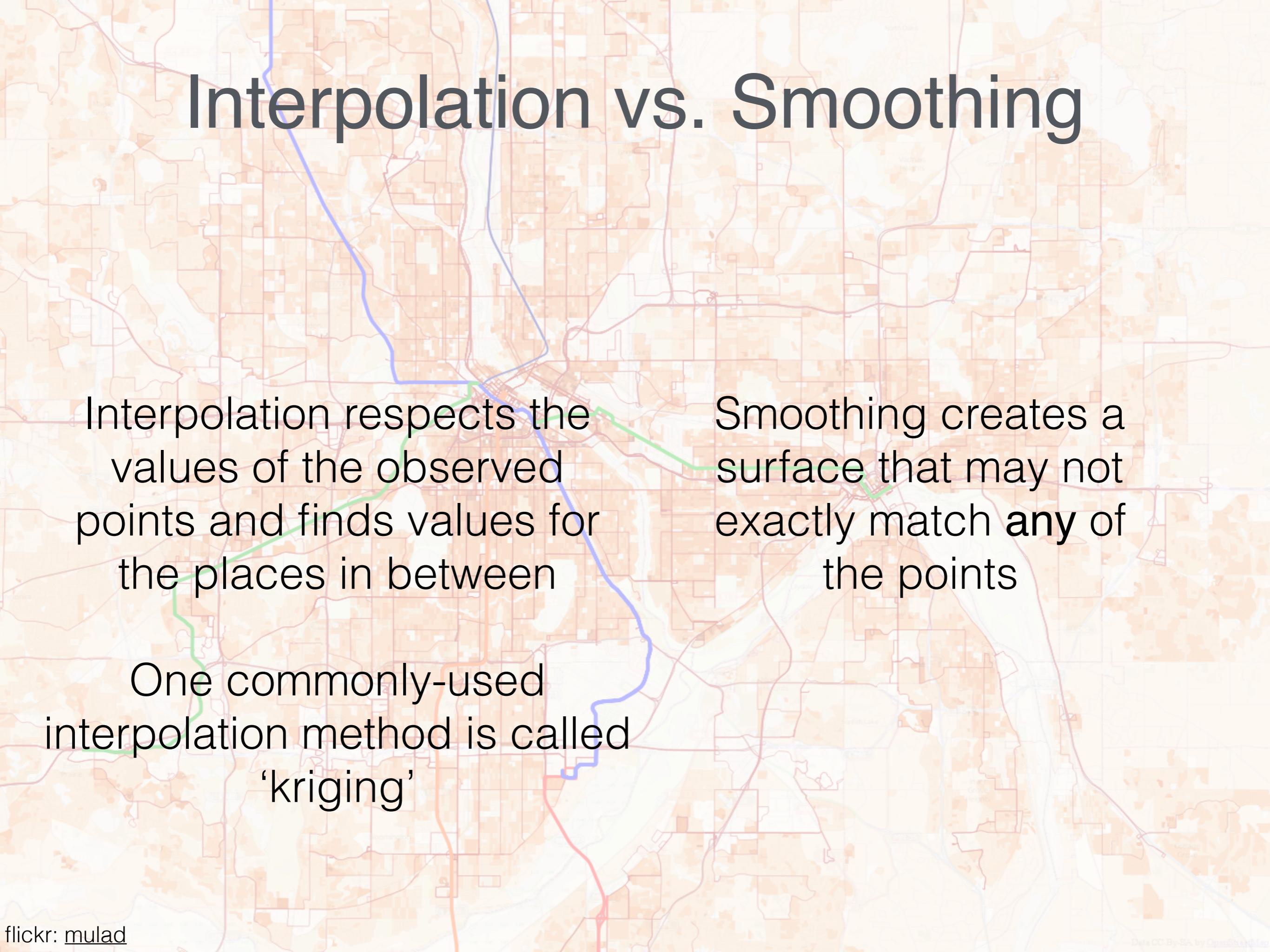
Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems



Joint work with students Jessica Mao and MyVan Vo.

Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems

Interpolation vs. Smoothing



Interpolation respects the values of the observed points and finds values for the places in between

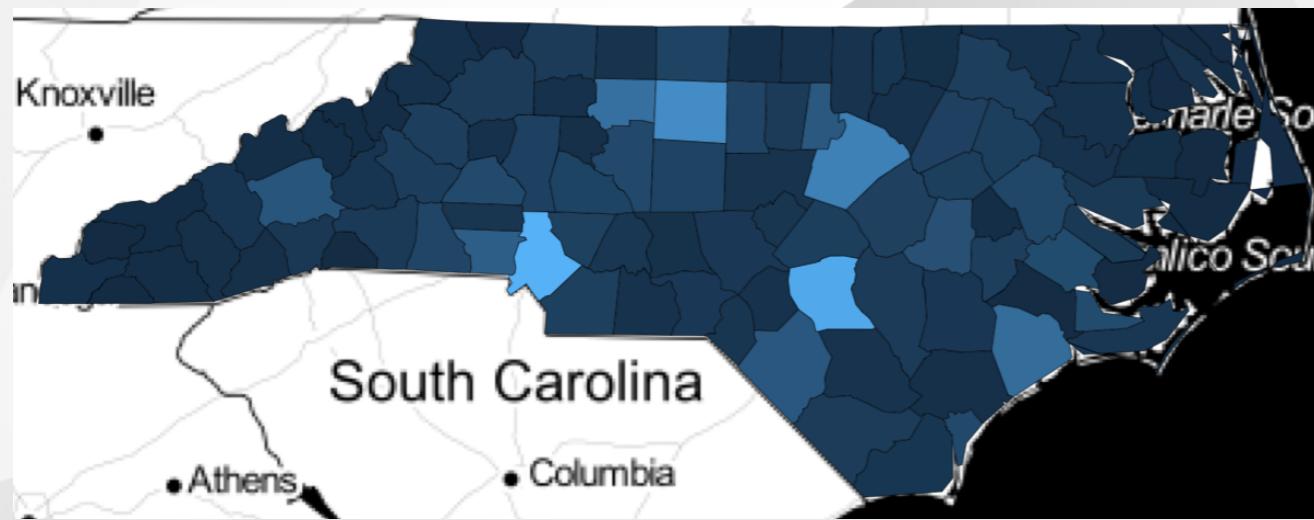
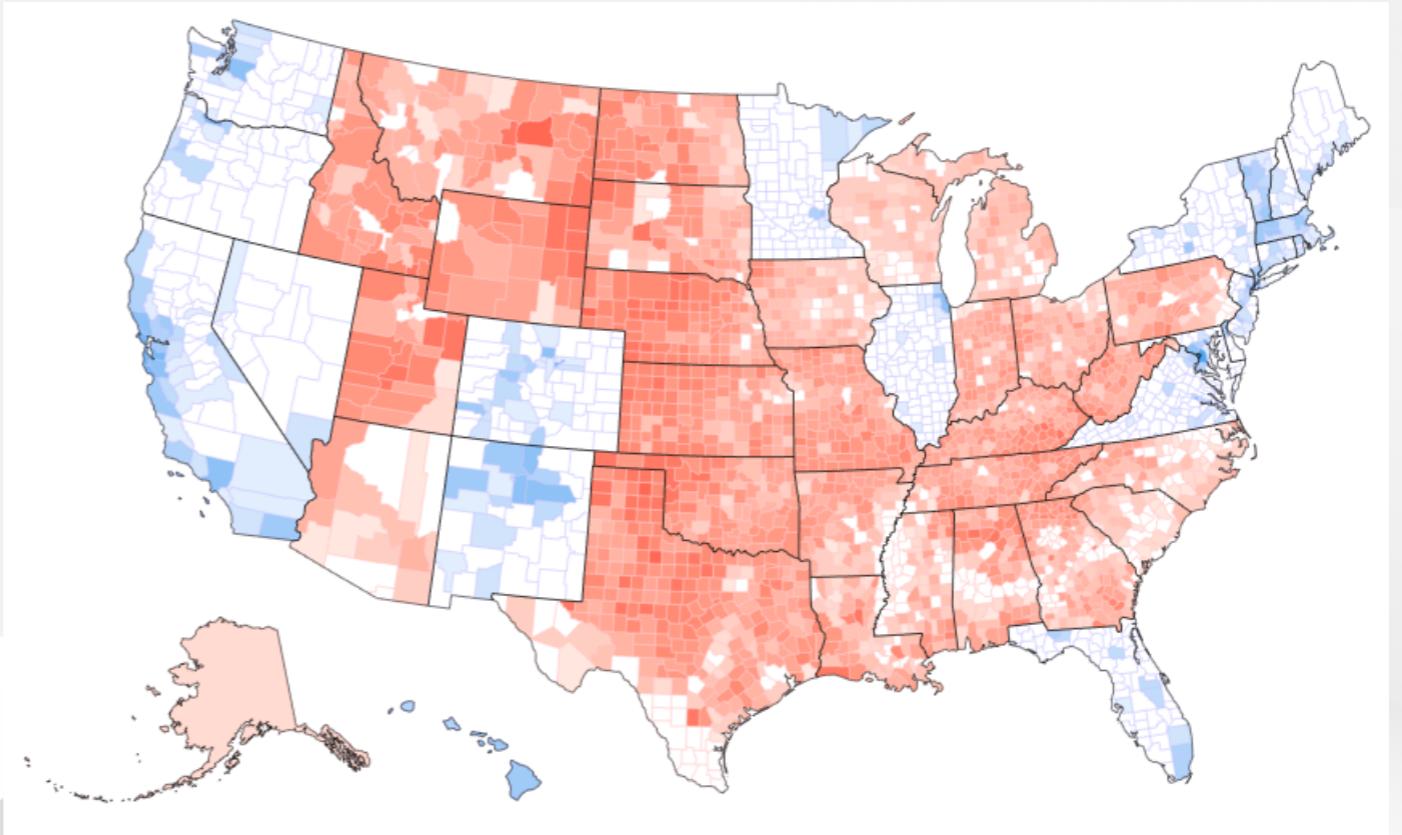
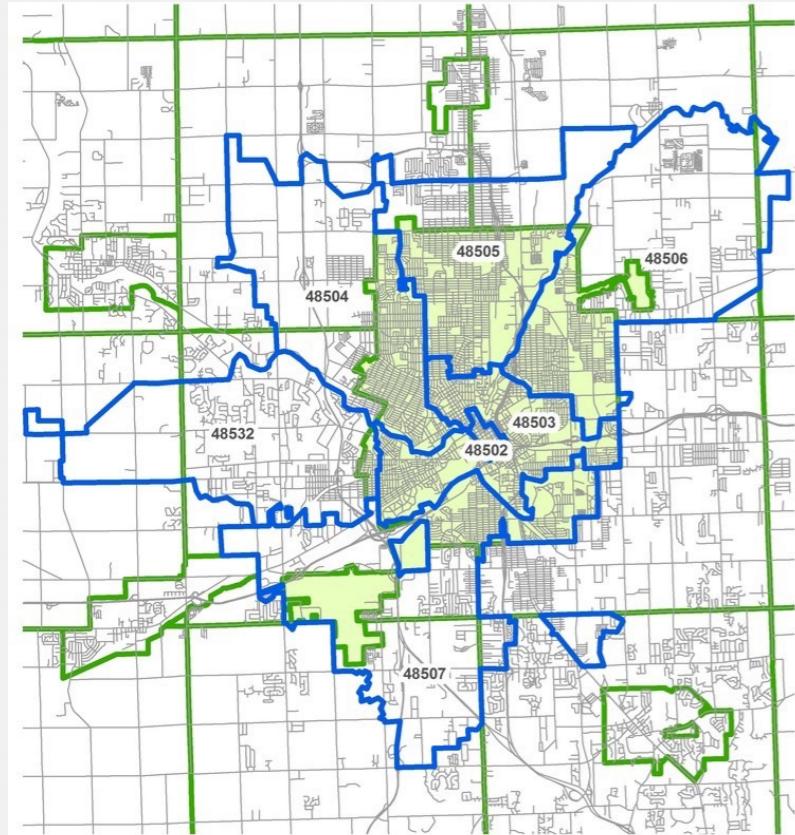
Smoothing creates a surface that may not exactly match **any** of the points

One commonly-used interpolation method is called 'kriging'

takeaways

- don't aggregate if you don't have to
- pay attention to your spatial polygons
- use auxiliary information if you have it

since spatial polygons shape our world



it's important to keep trying

An aerial photograph of a coastal area featuring a distinct grid pattern of yellowish-green paths or tracks on a dark, textured terrain. The paths appear to follow the contours of the land, possibly indicating a network of trails or agricultural fields. Small blue spots, likely water bodies or artificial structures, are scattered across the landscape.

Thank you