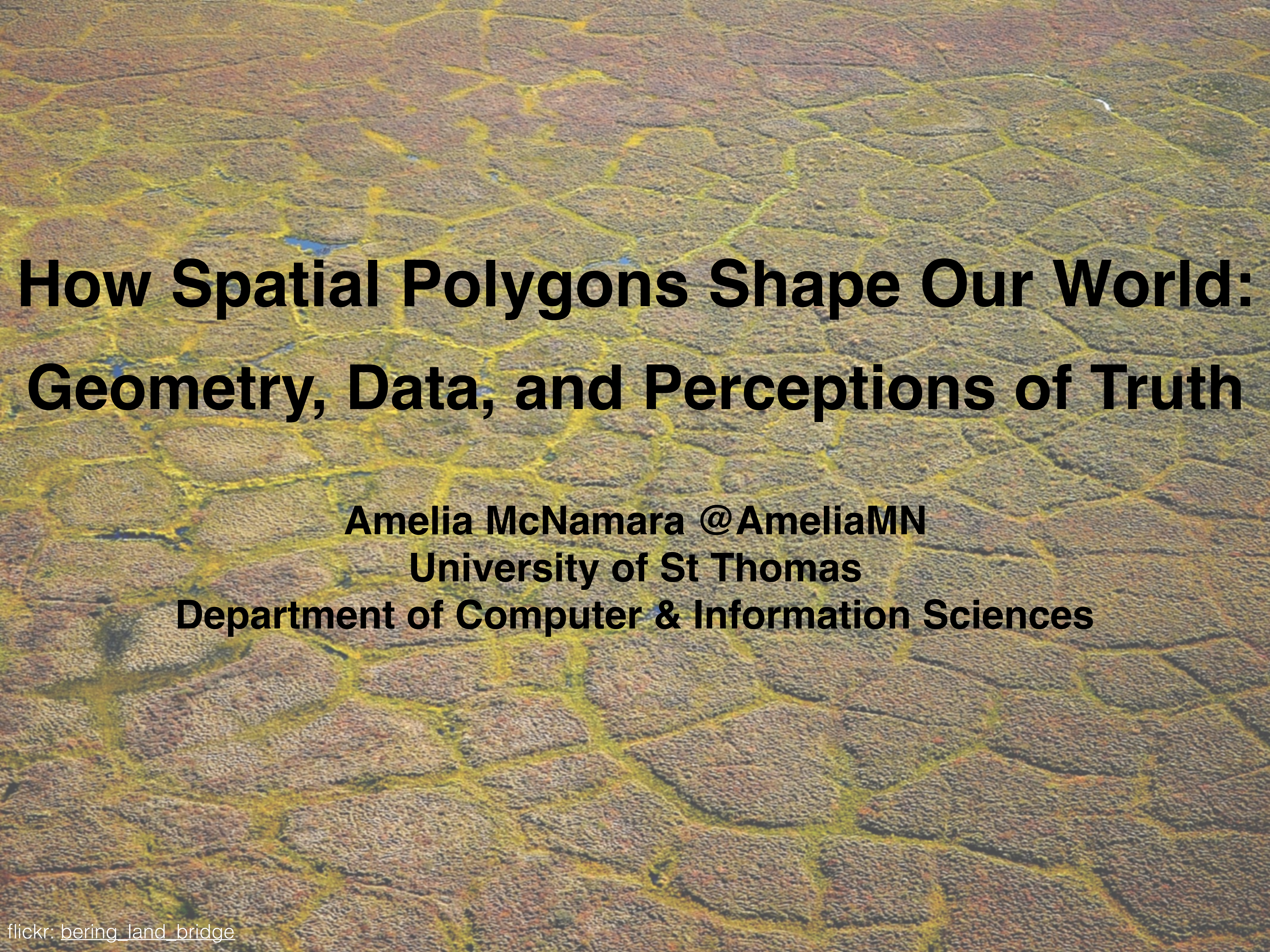
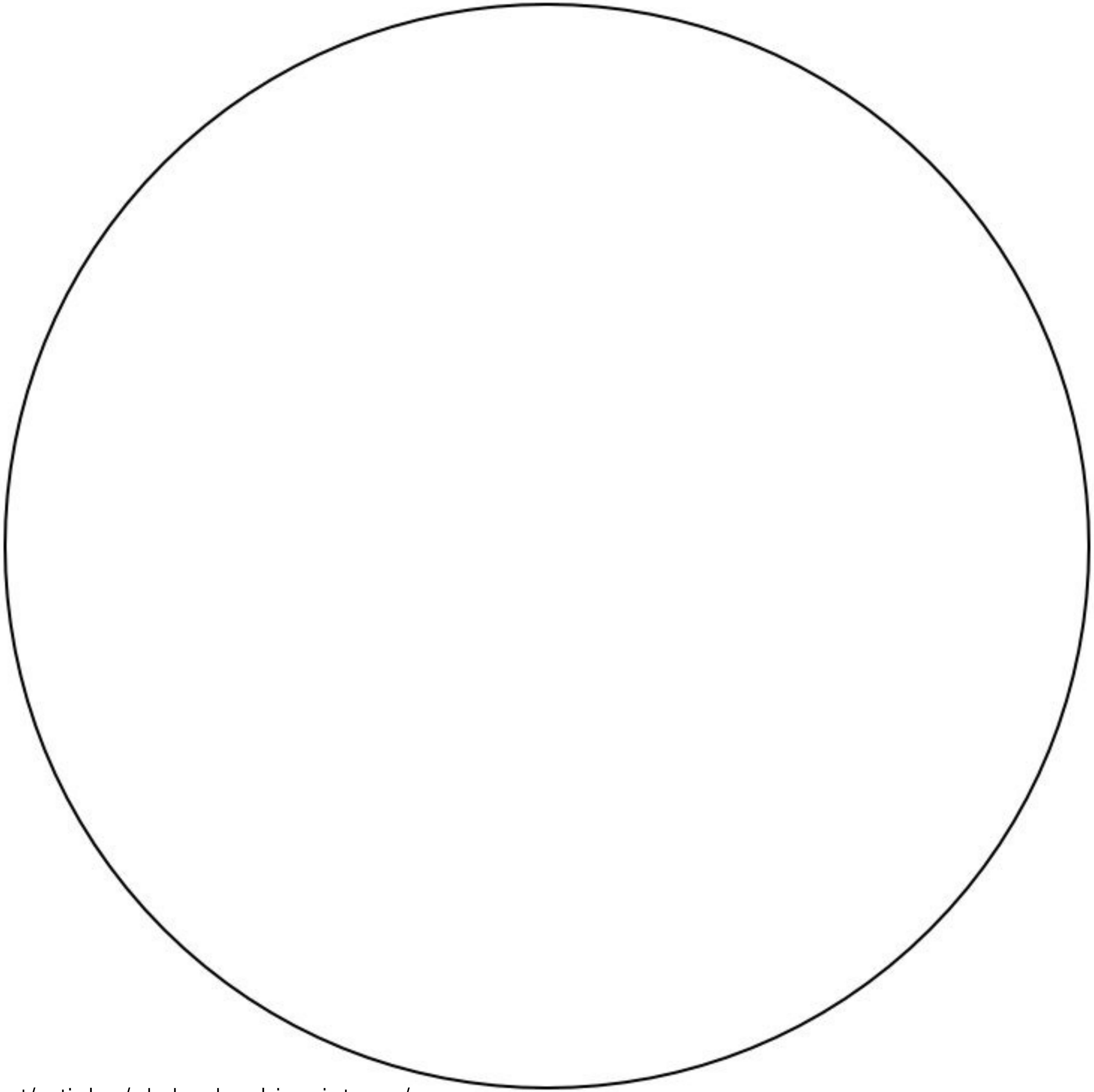


How Spatial Polygons Shape Our World: Geometry, Data, and Perceptions of Truth

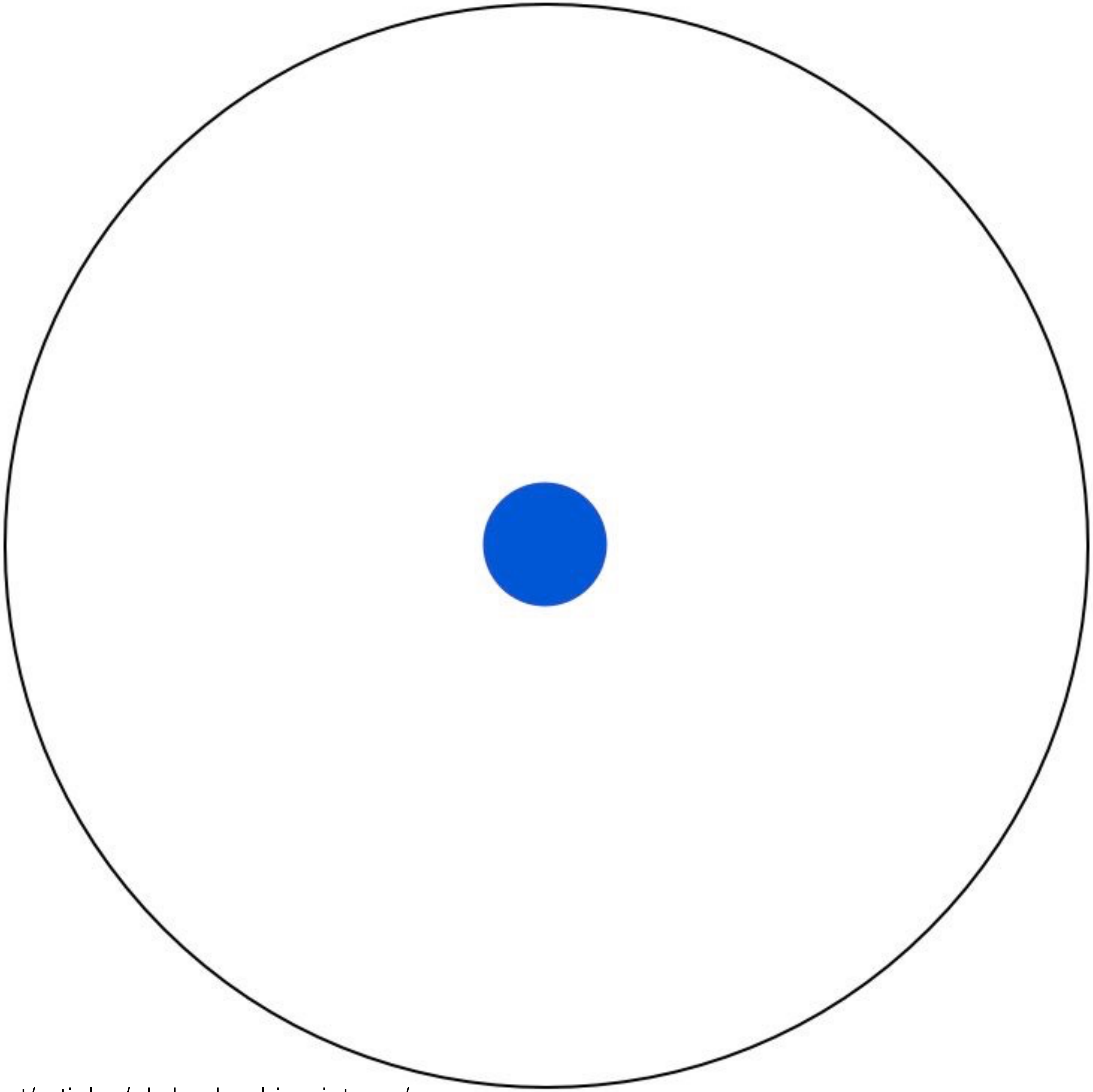
Amelia McNamara @AmeliaMN
University of St Thomas
Department of Computer & Information Sciences

flickr: bering_land_bridge

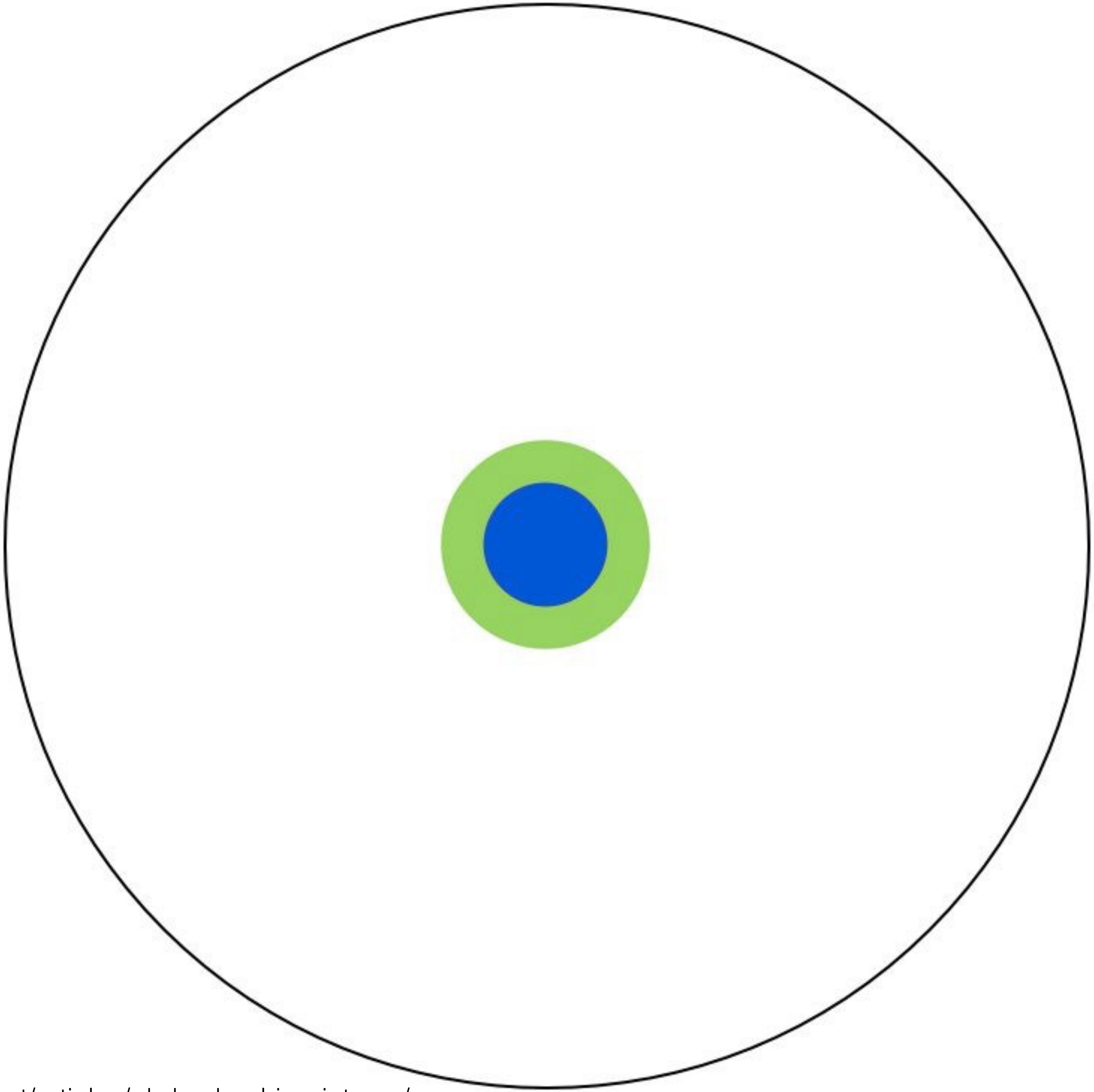
http://matt.might.net/articles/phd-school-in-pictures/



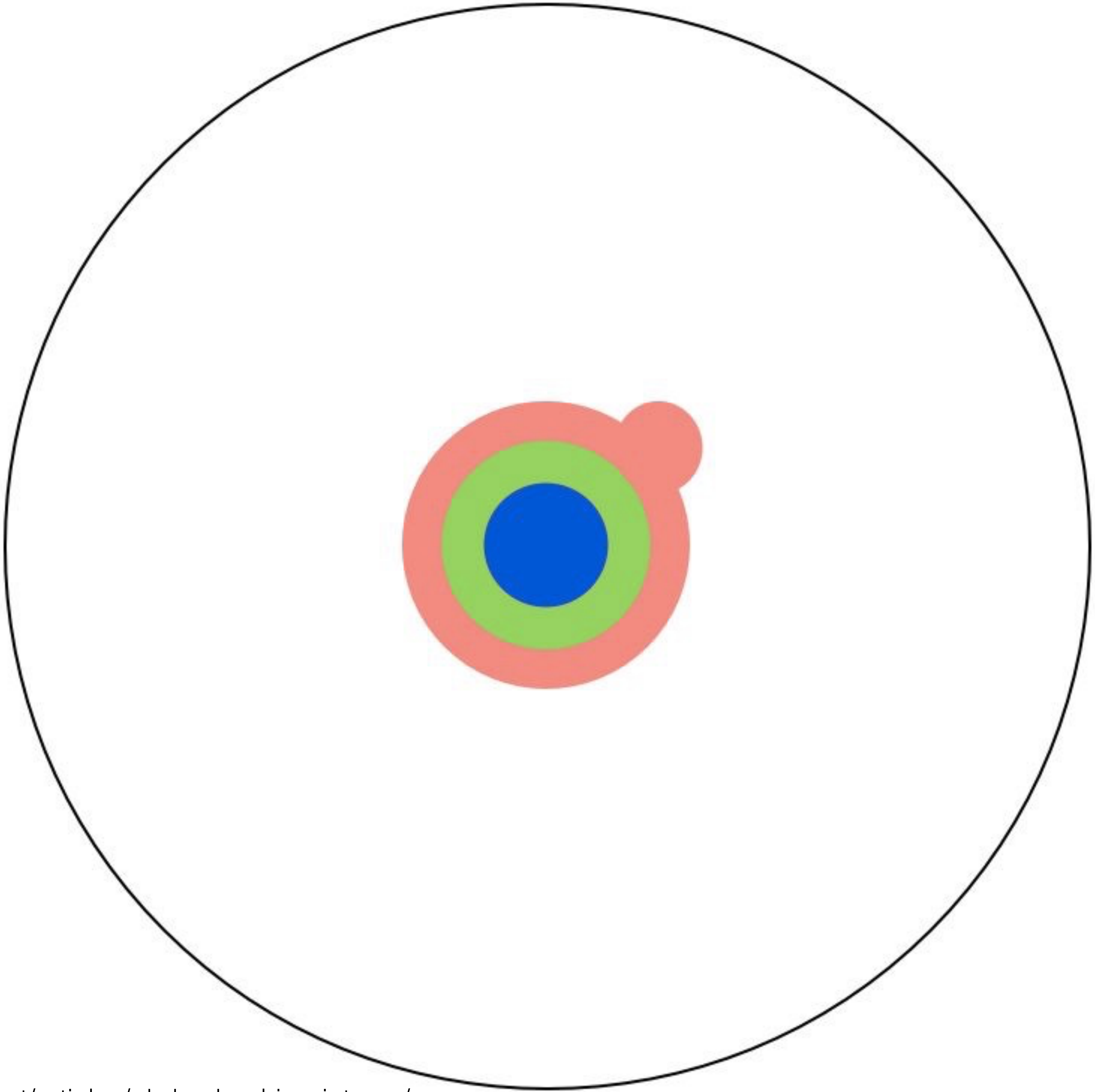
http://matt.might.net/articles/phd-school-in-pictures/



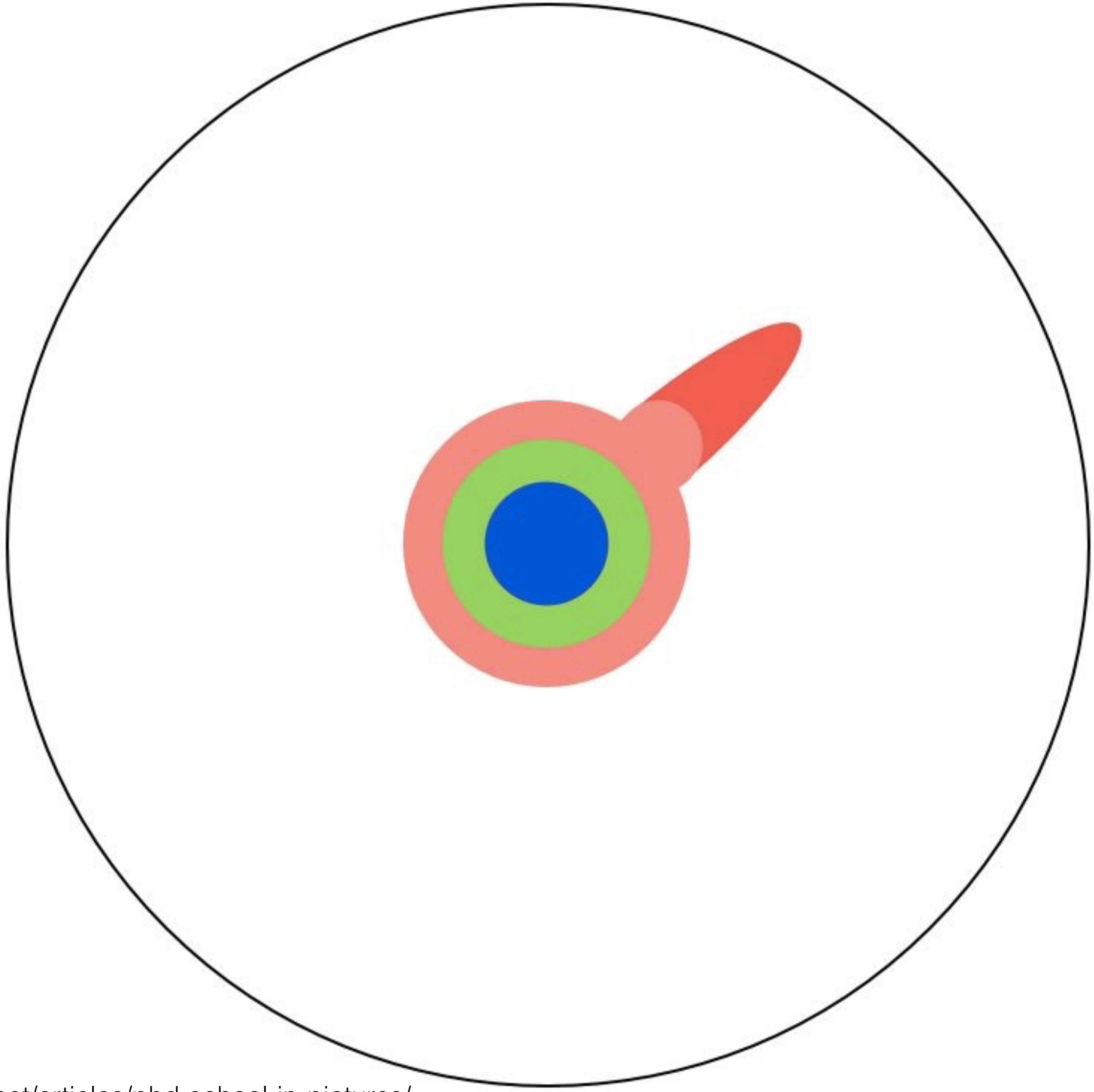
http://matt.might.net/articles/phd-school-in-pictures/



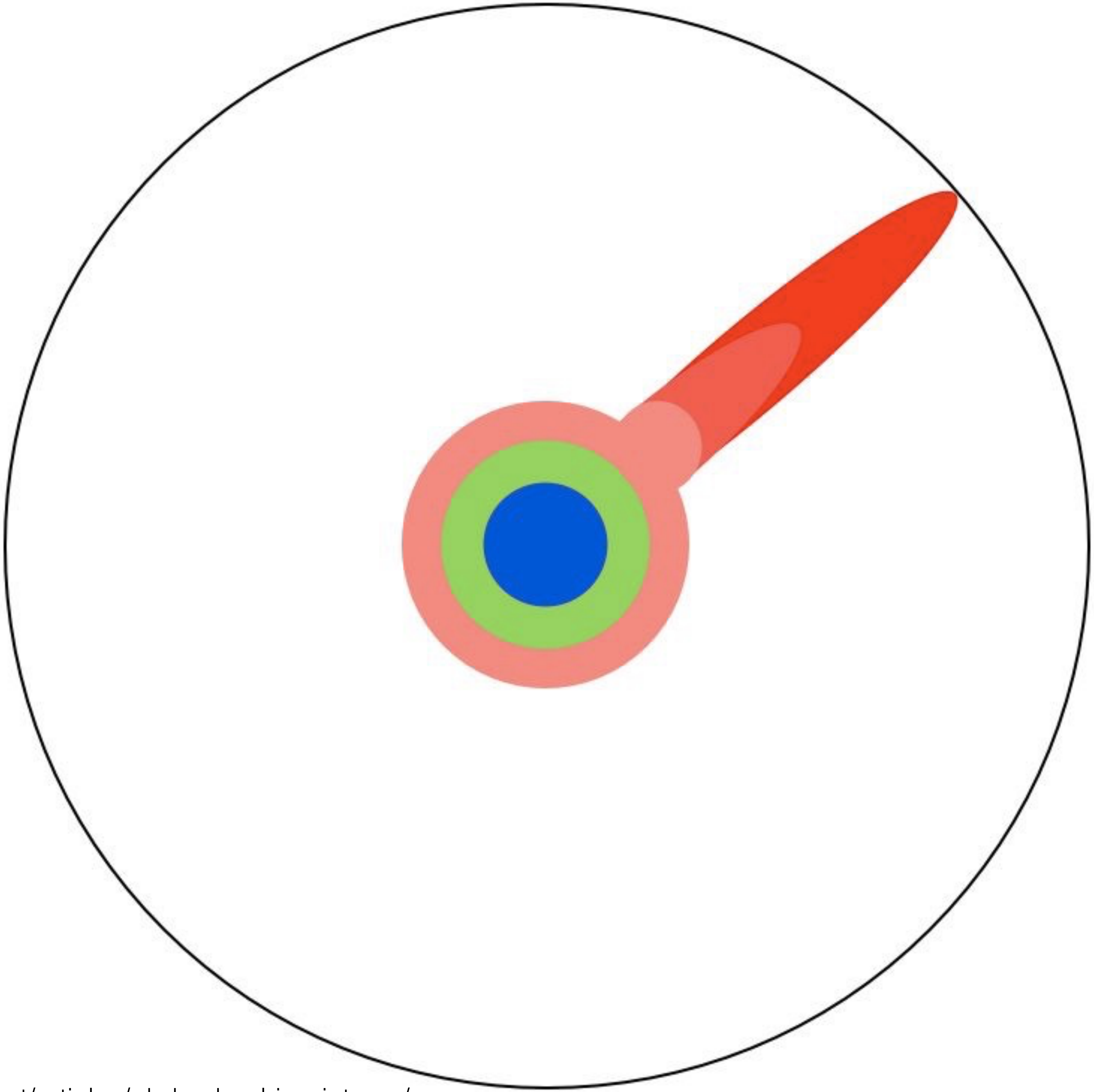
http://matt.might.net/articles/phd-school-in-pictures/



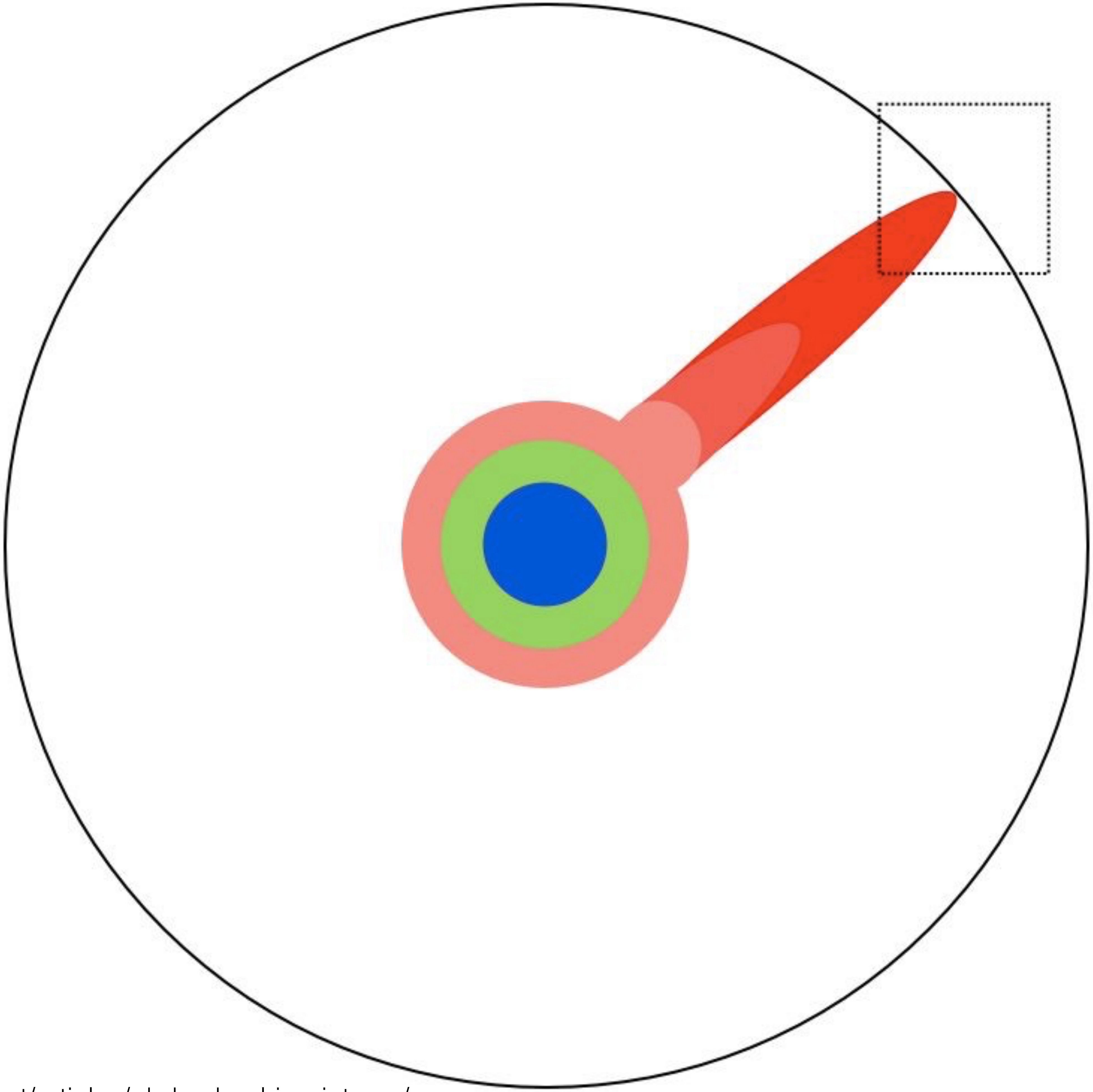
http://matt.might.net/articles/phd-school-in-pictures/



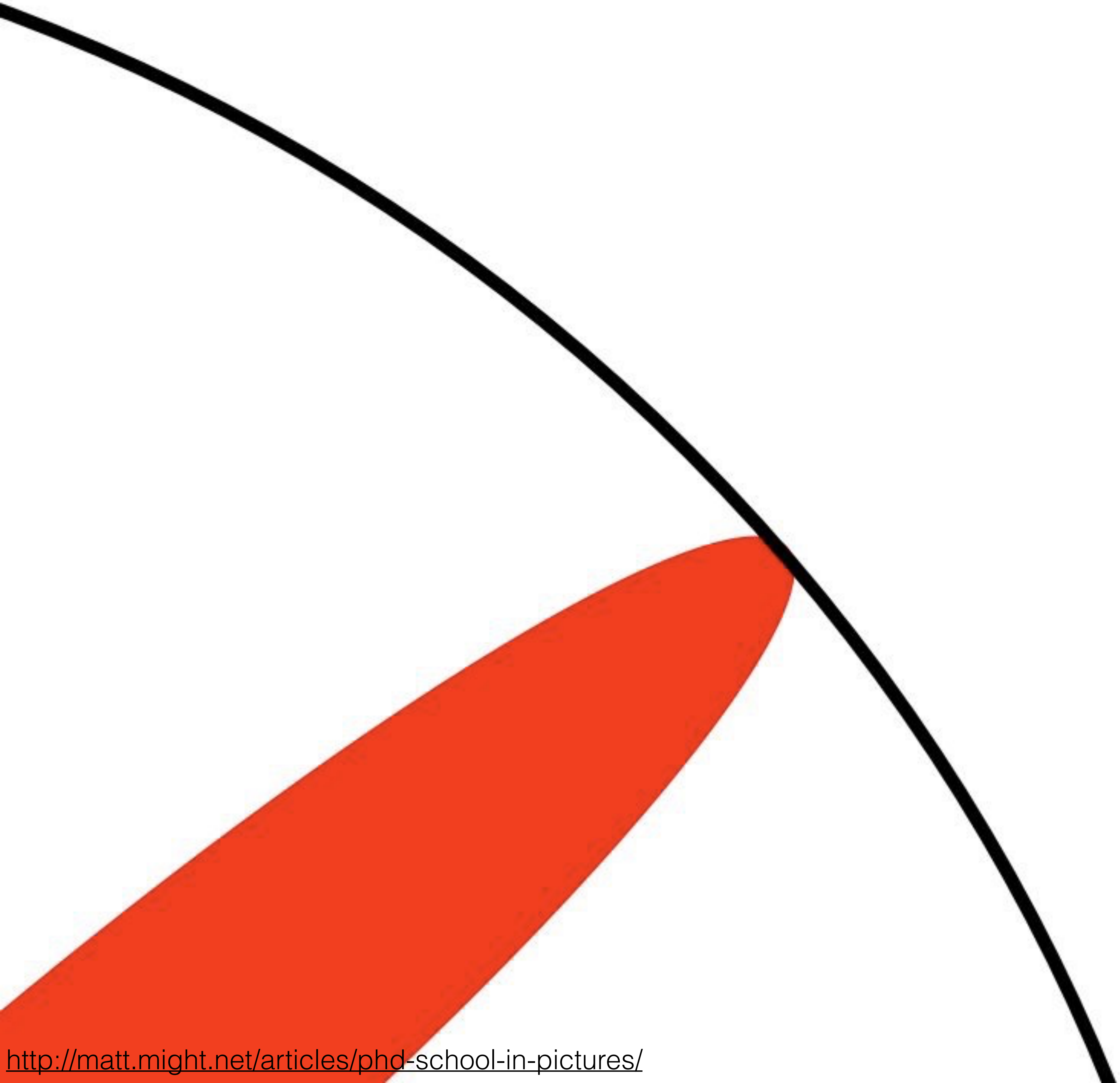
http://matt.might.net/articles/phd-school-in-pictures/



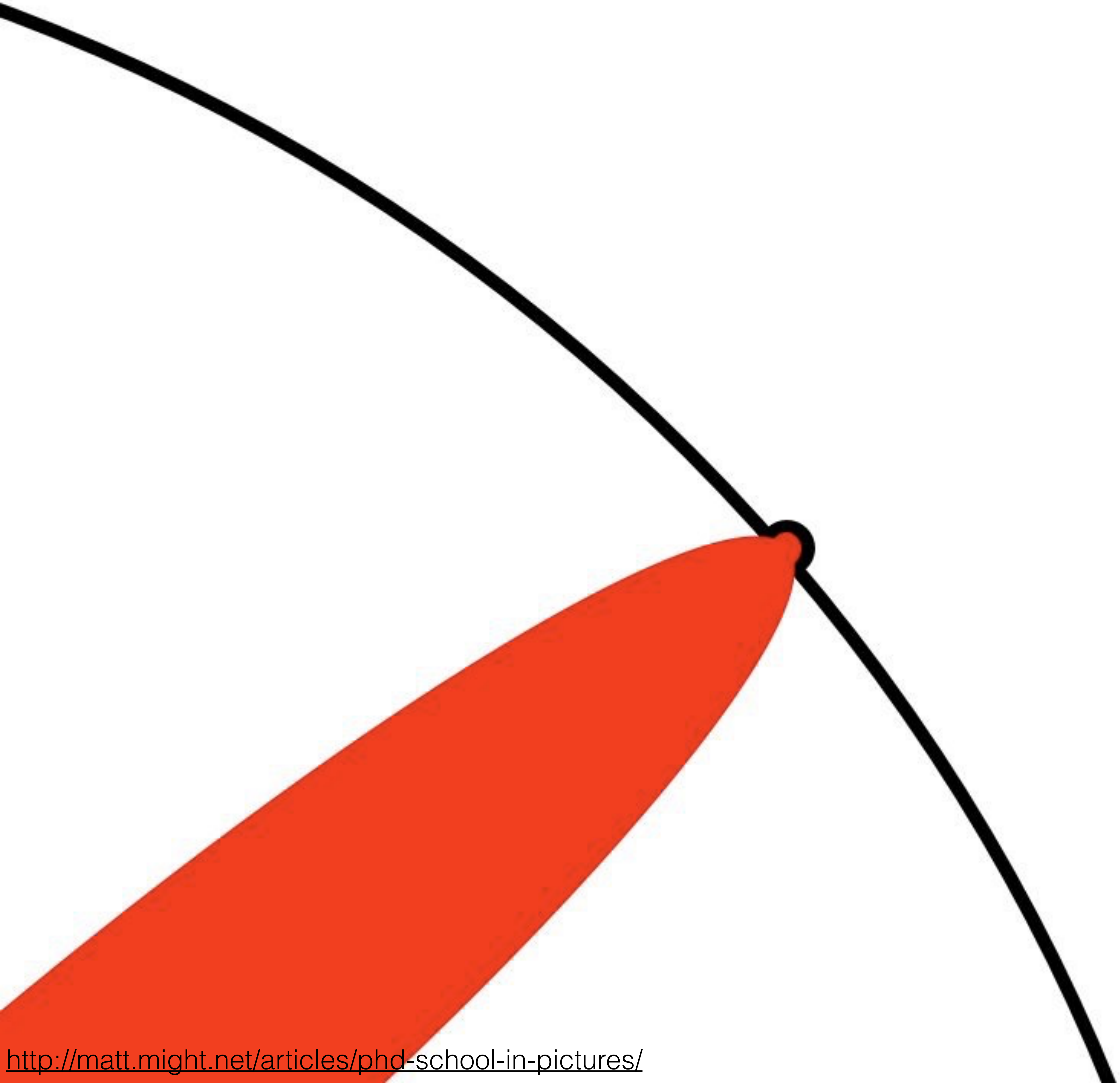
http://matt.might.net/articles/phd-school-in-pictures/



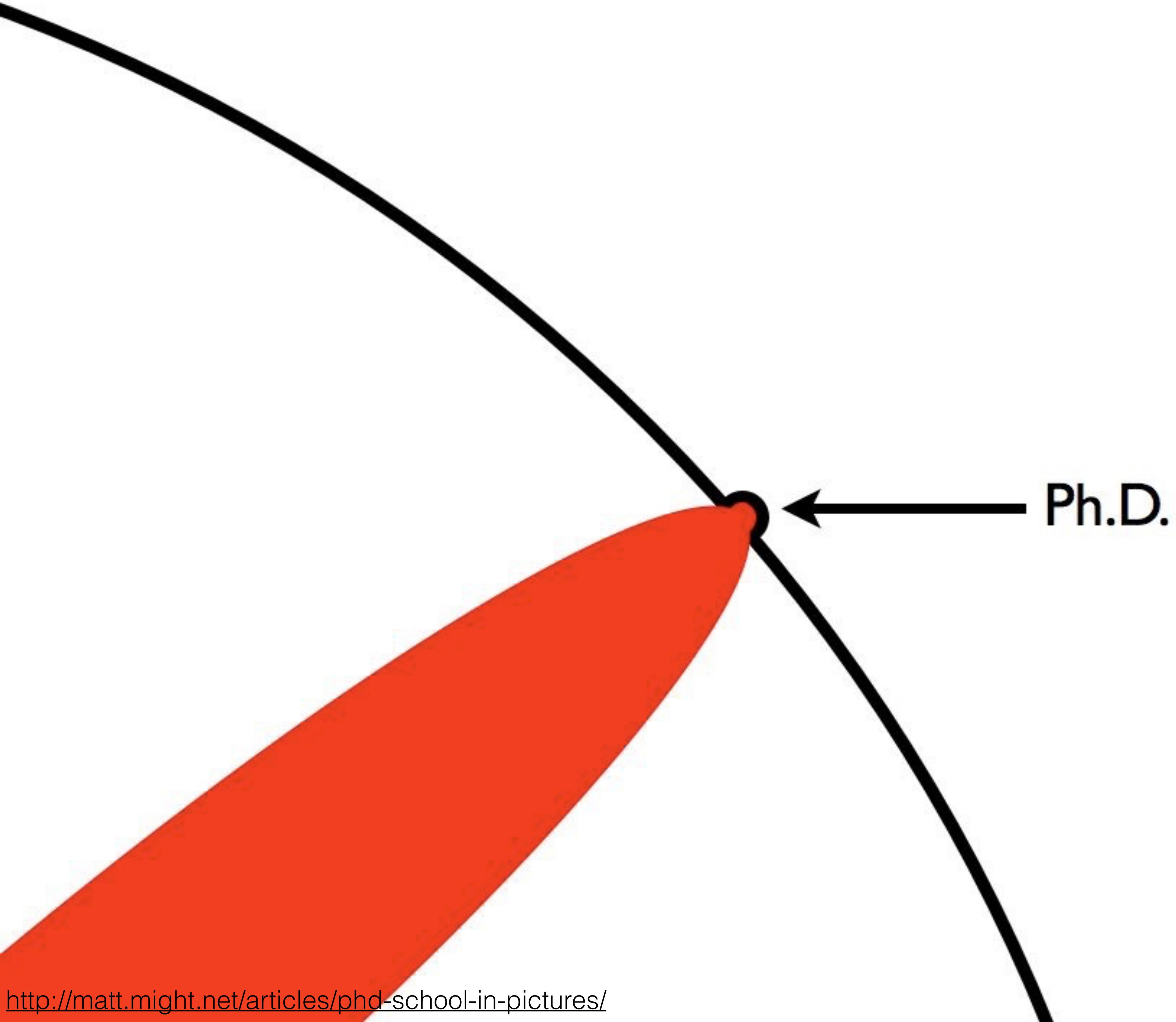
http://matt.might.net/articles/phd-school-in-pictures/



http://matt.might.net/articles/phd-school-in-pictures/

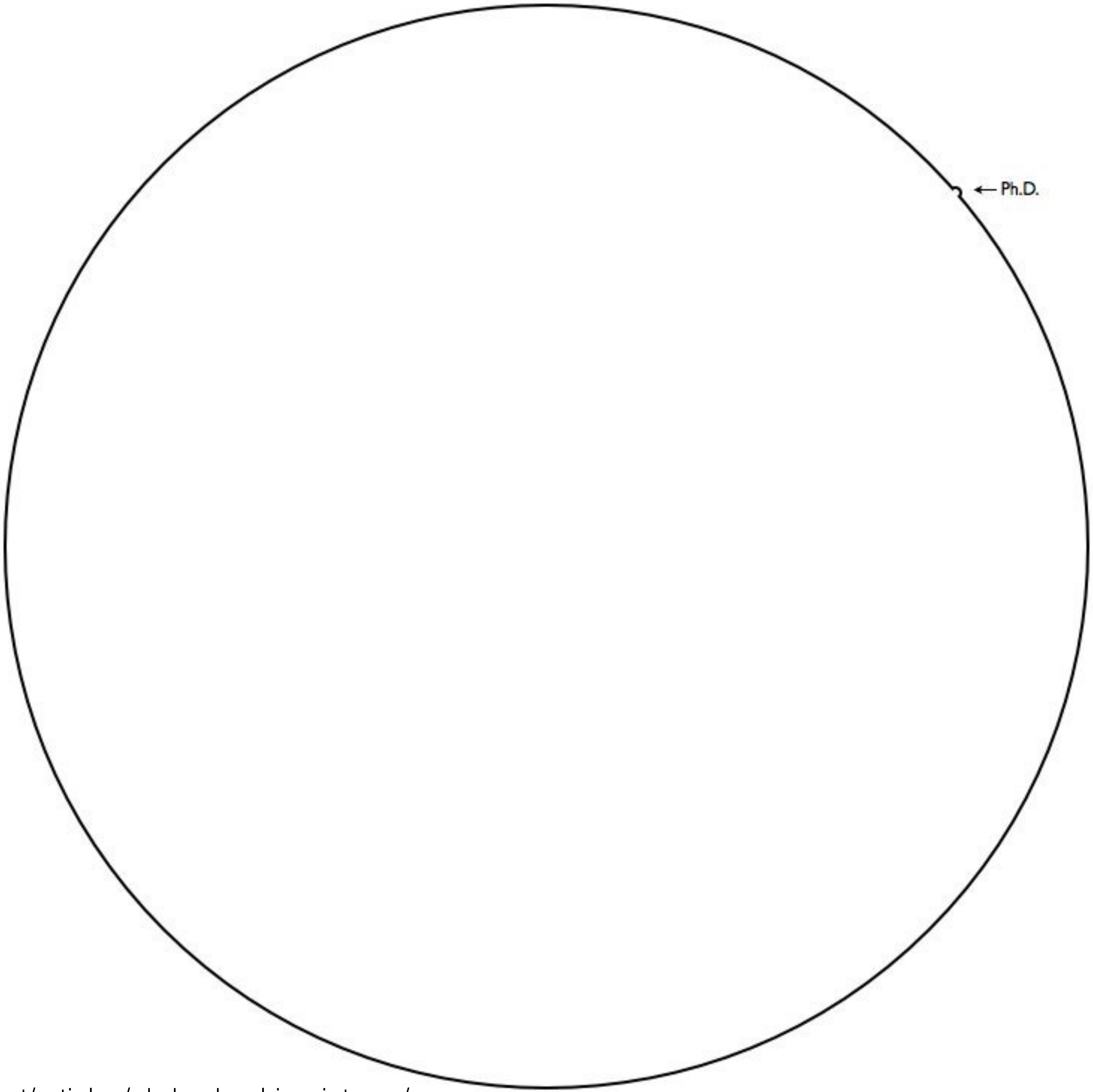


http://matt.might.net/articles/phd-school-in-pictures/



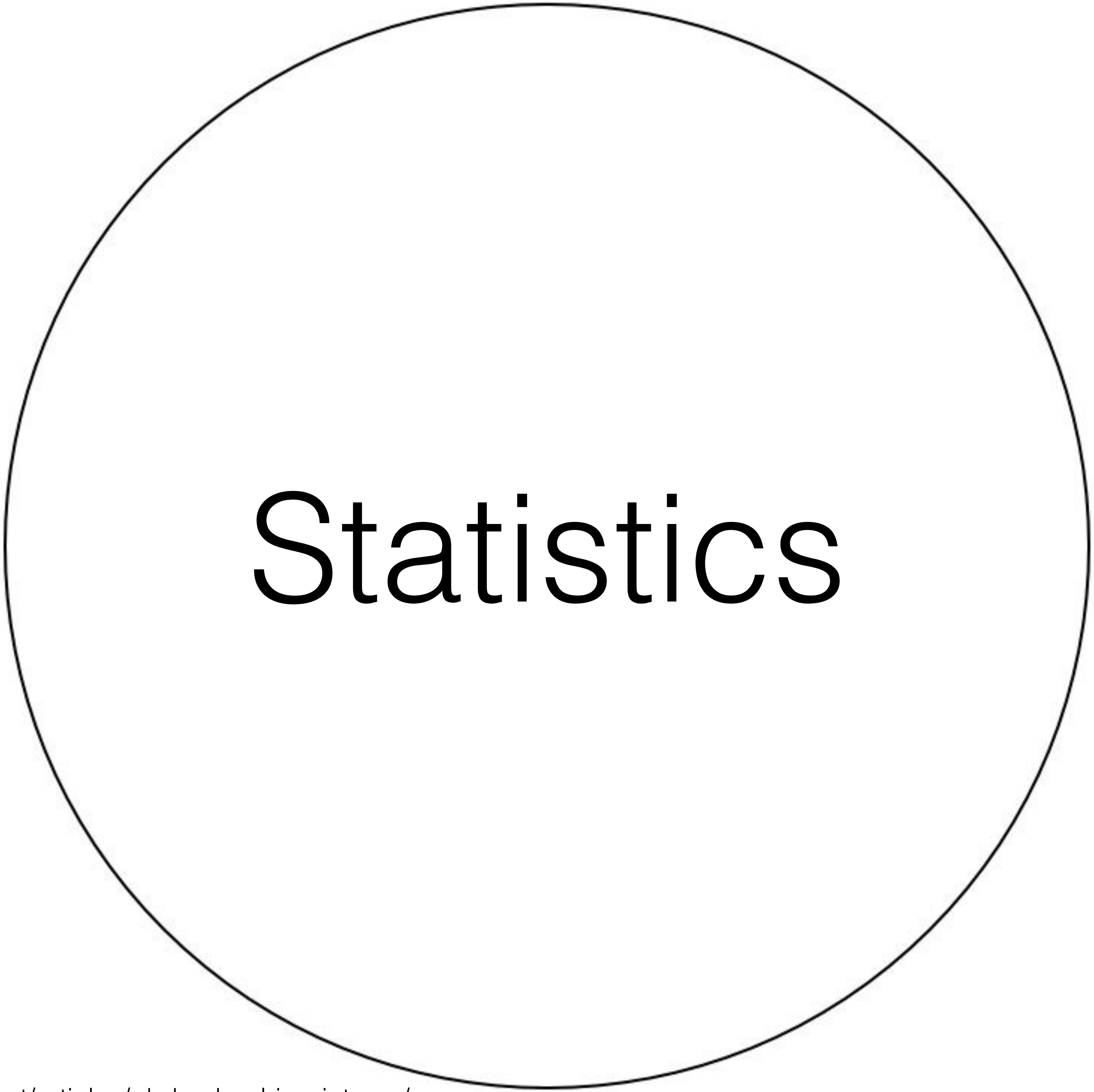
http://matt.might.net/articles/phd-school-in-pictures/

http://matt.might.net/articles/phd-school-in-pictures/



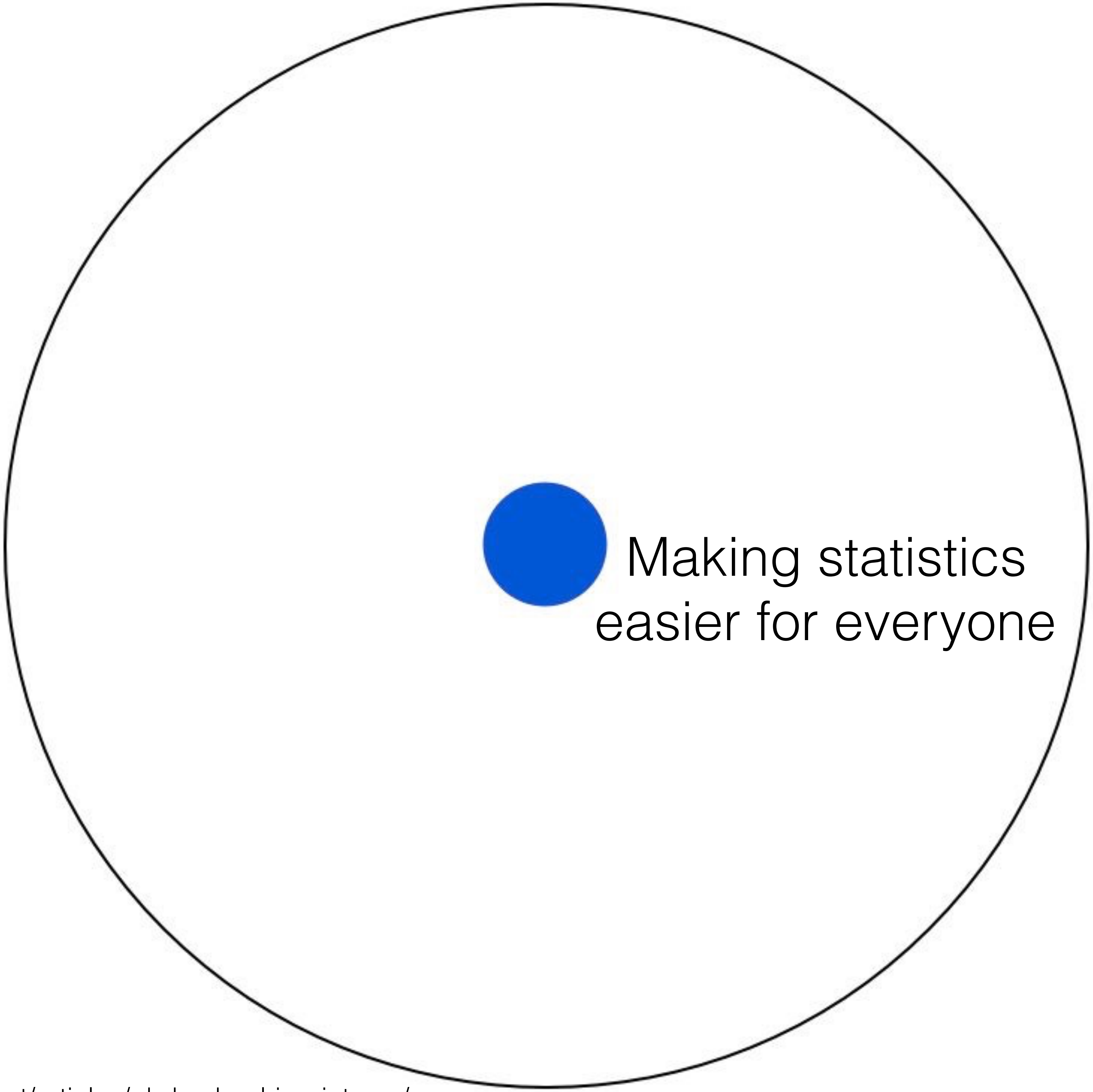
http://matt.might.net/articles/phd-school-in-pictures/

http://matt.might.net/articles/phd-school-in-pictures/



Statistics

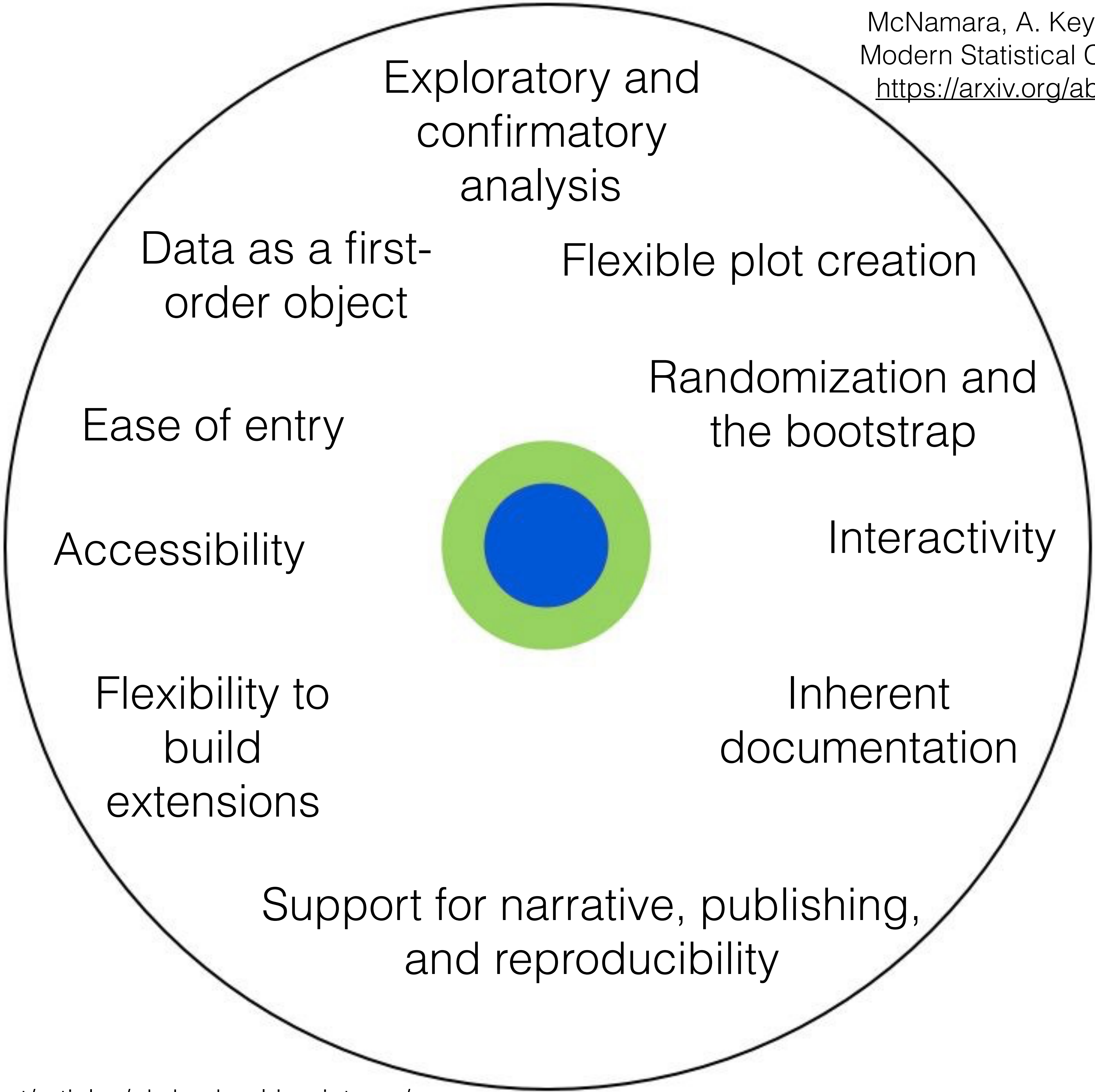
http://matt.might.net/articles/phd-school-in-pictures/



Making statistics easier for everyone

http://matt.might.net/articles/phd-school-in-pictures/

McNamara, A. Key Attributes of a Modern Statistical Computing Tool. https://arxiv.org/abs/1610.00985



Exploratory and confirmatory analysis

Data as a first-order object

Flexible plot creation

Randomization and the bootstrap

Ease of entry

Accessibility

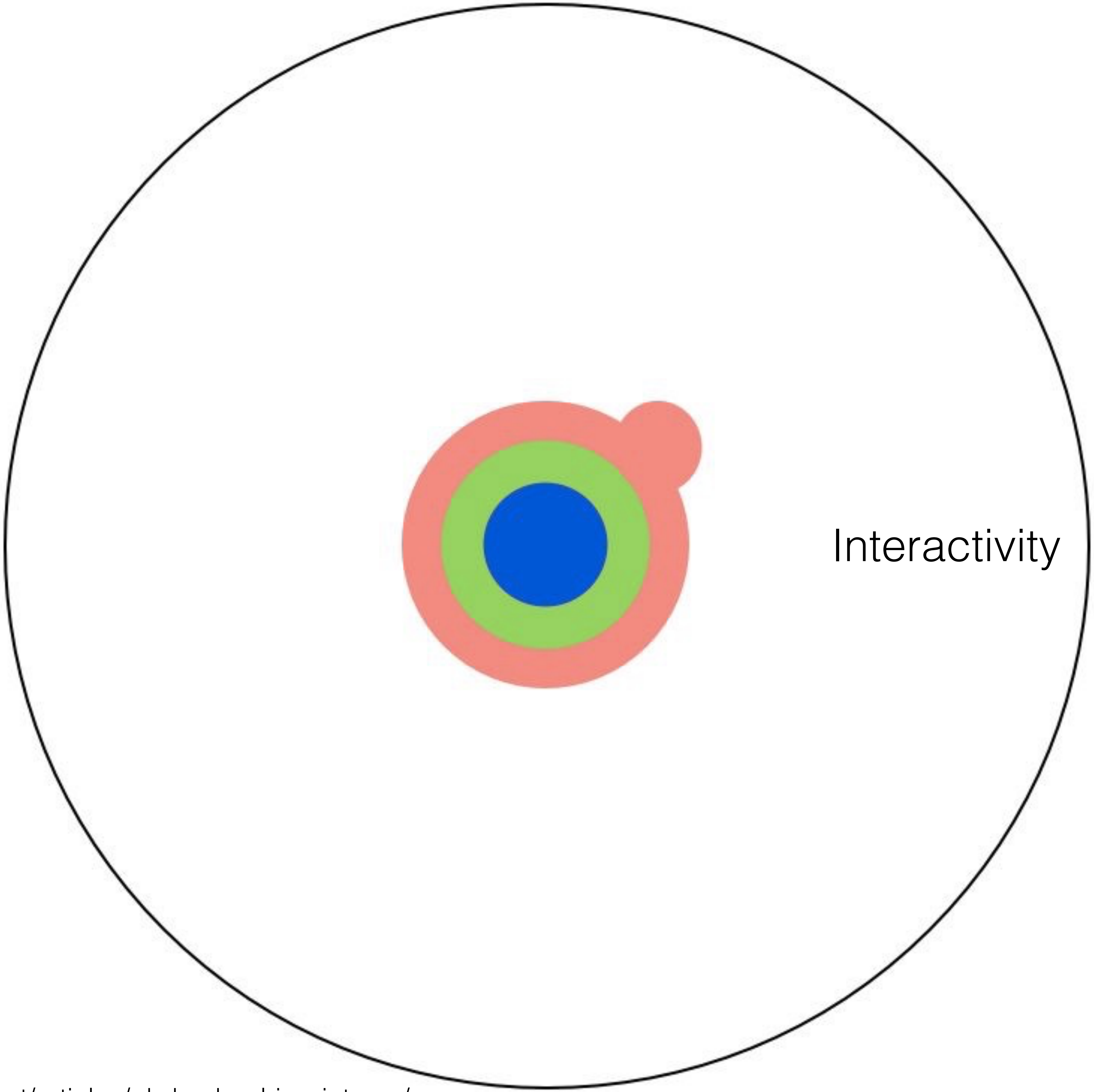
Interactivity

Flexibility to build extensions

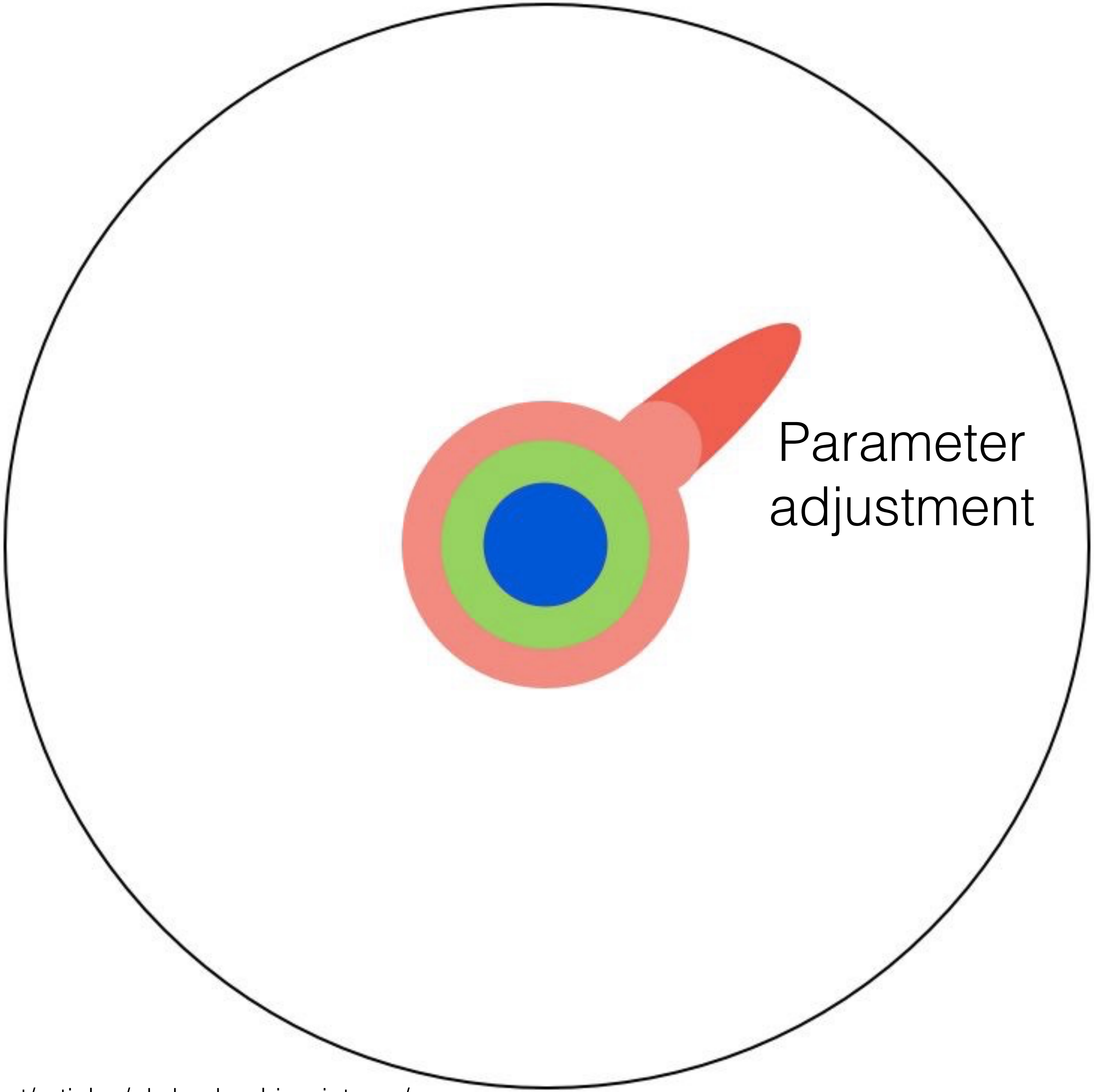
Inherent documentation

Support for narrative, publishing, and reproducibility

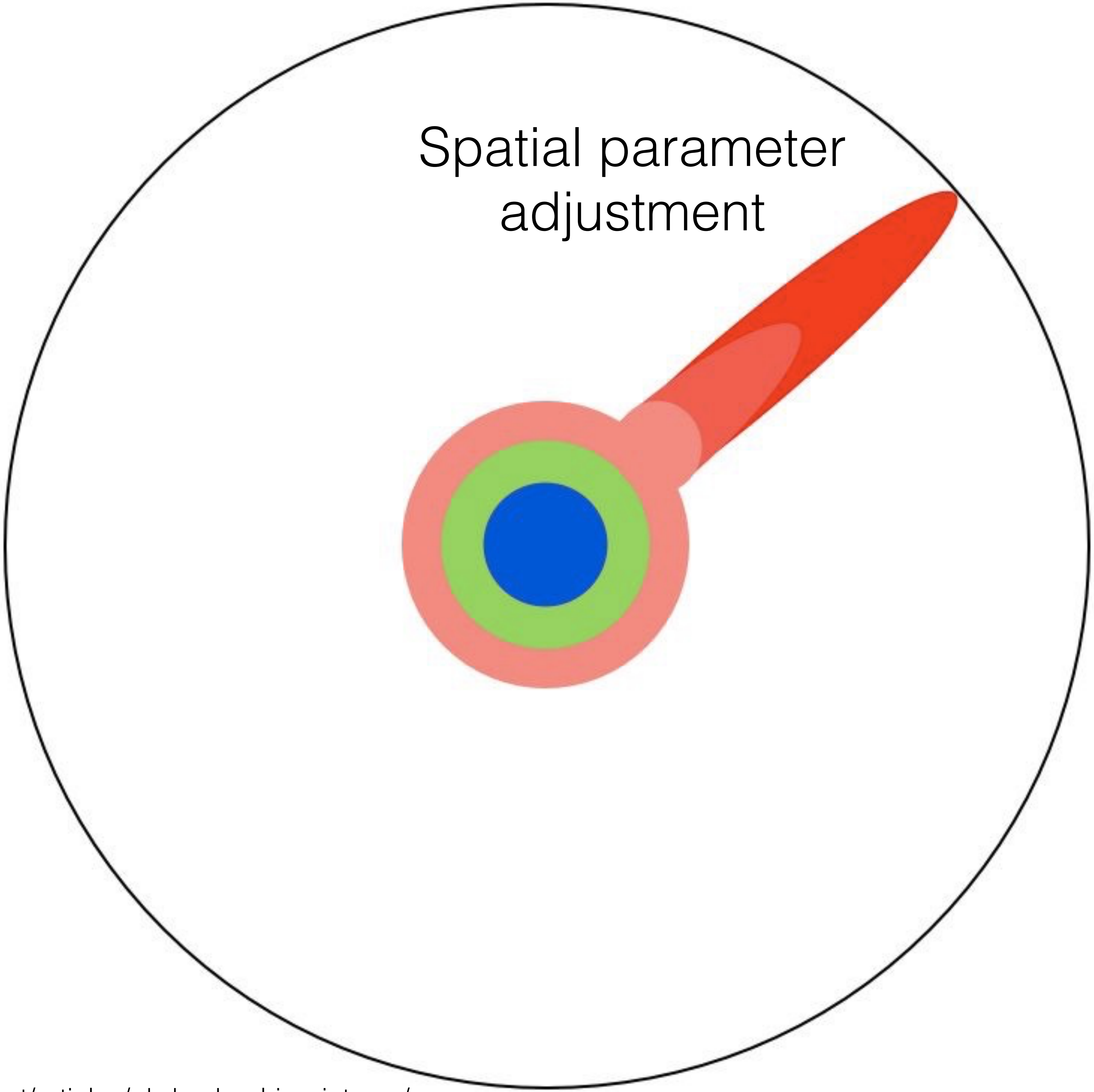
http://matt.might.net/articles/phd-school-in-pictures/



http://matt.might.net/articles/phd-school-in-pictures/

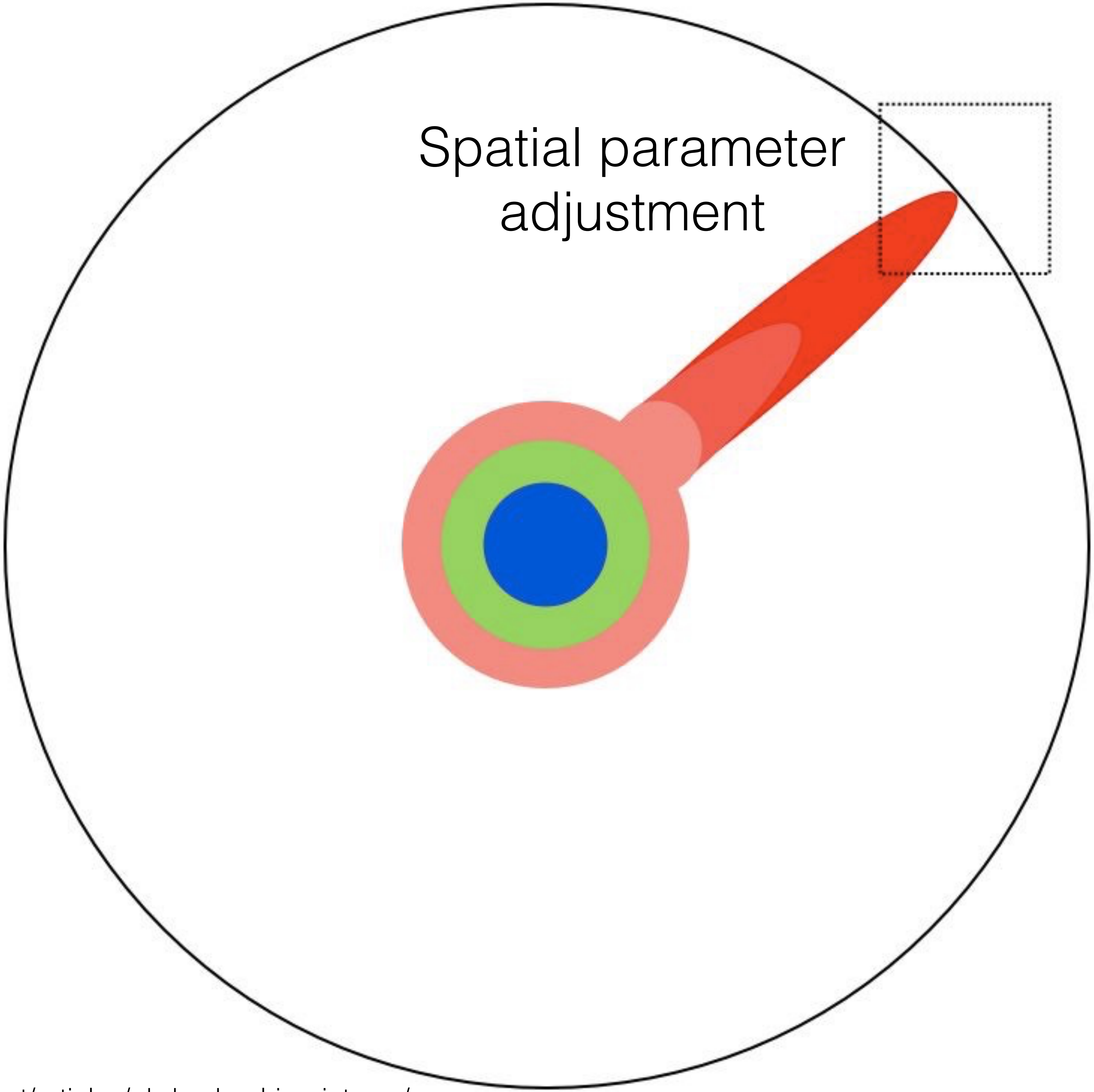


http://matt.might.net/articles/phd-school-in-pictures/



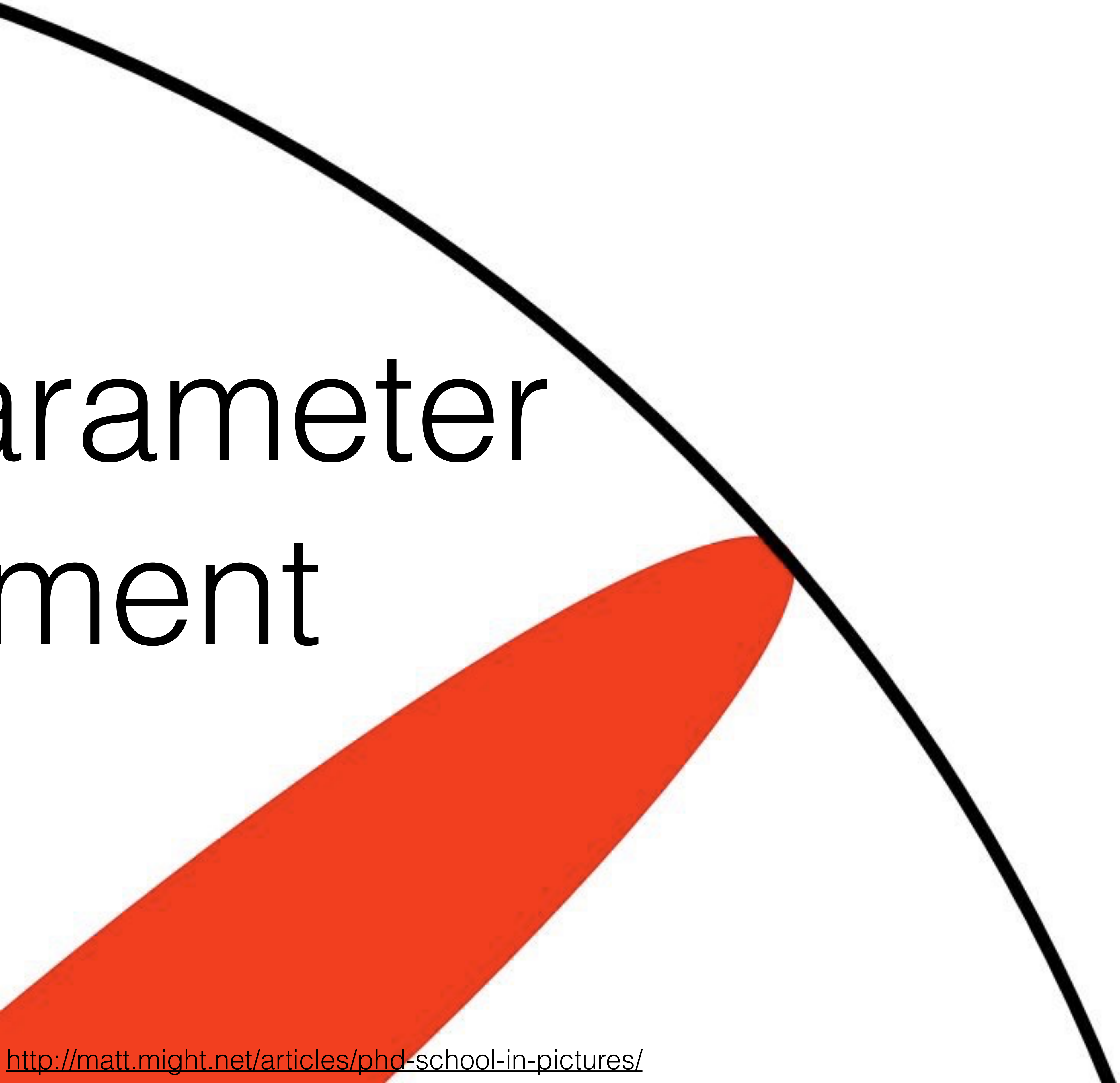
Spatial parameter adjustment

http://matt.might.net/articles/phd-school-in-pictures/



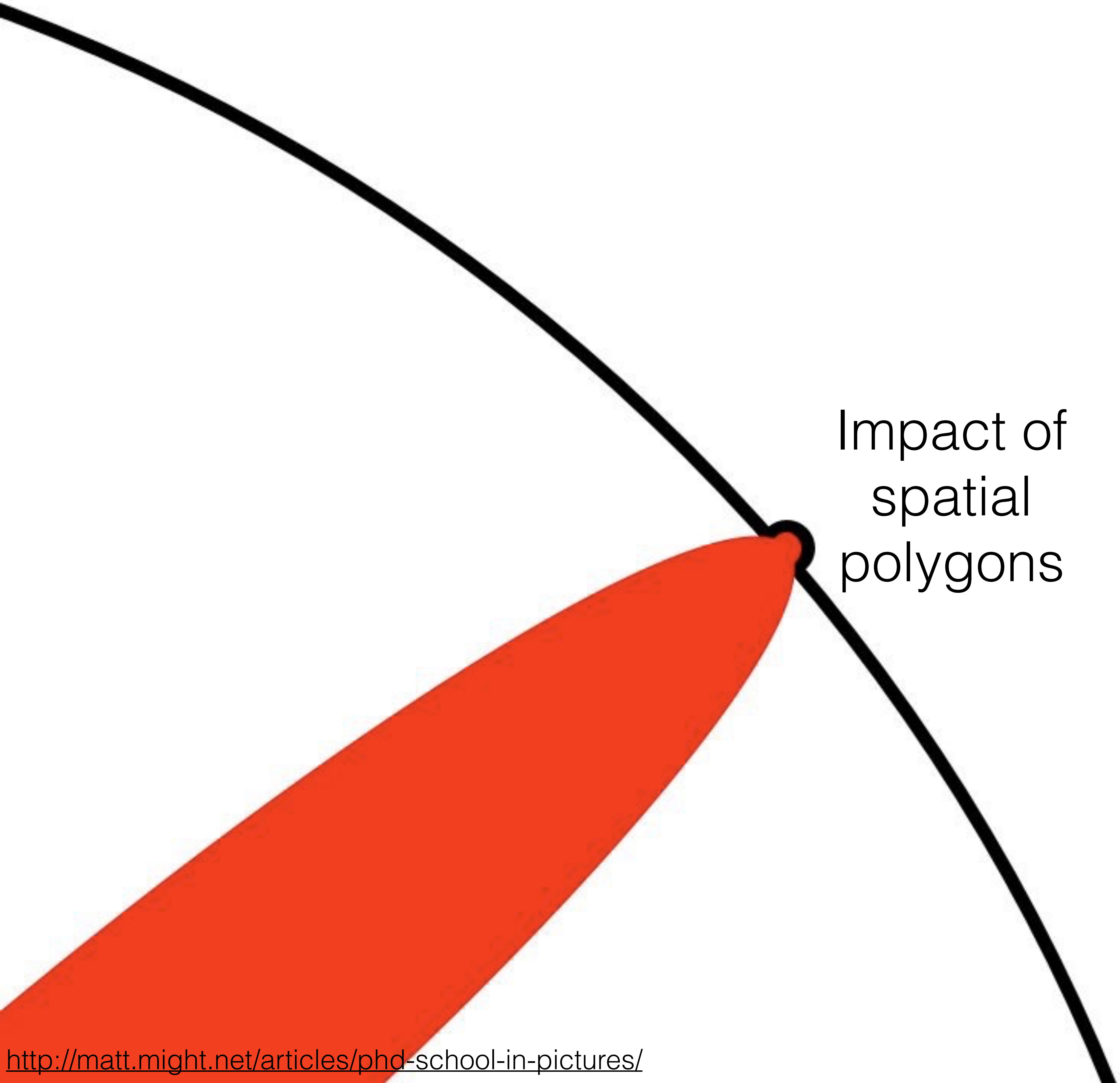
Spatial parameter adjustment

http://matt.might.net/articles/phd-school-in-pictures/



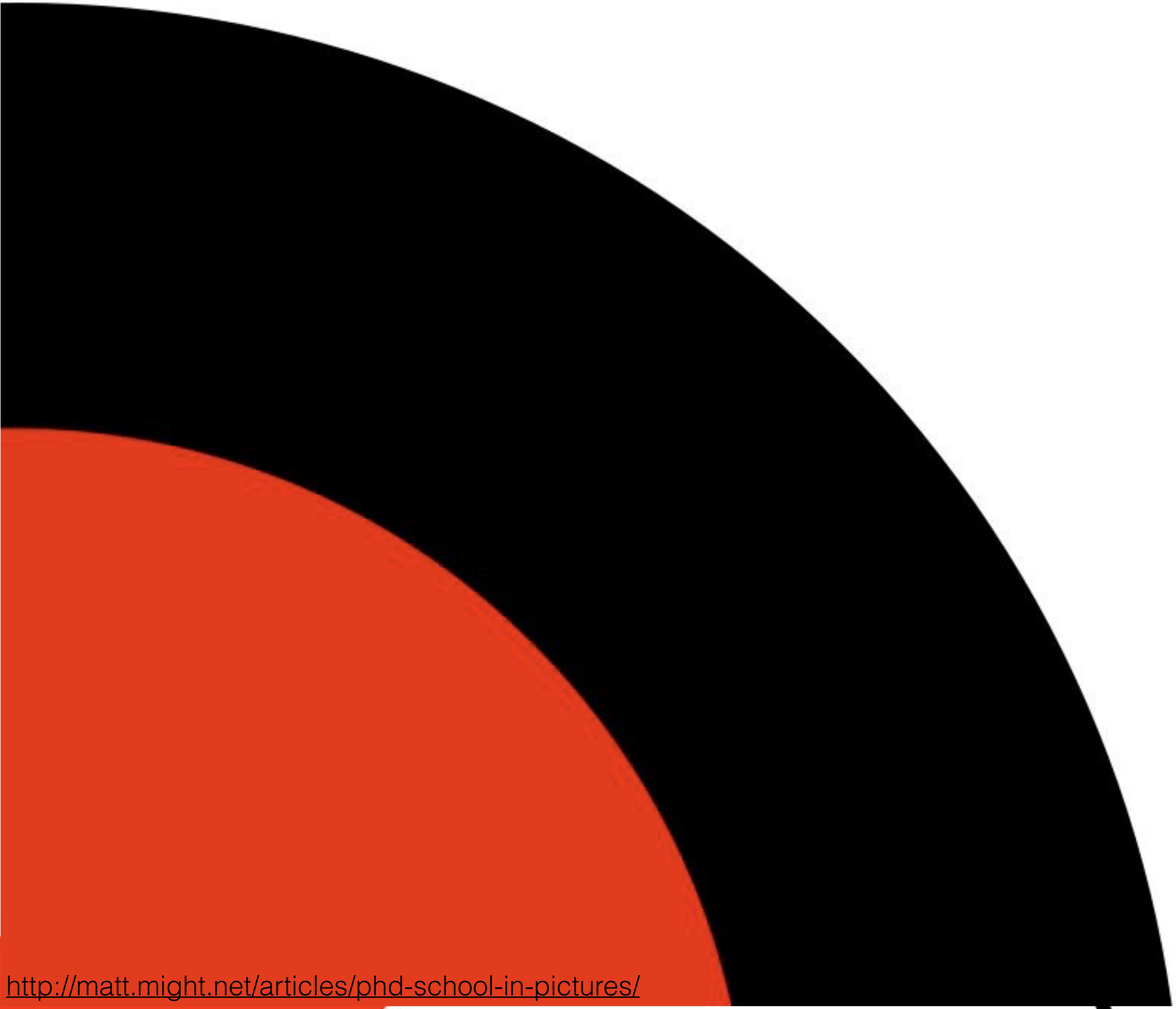
arameter
ment

http://matt.might.net/articles/phd-school-in-pictures/



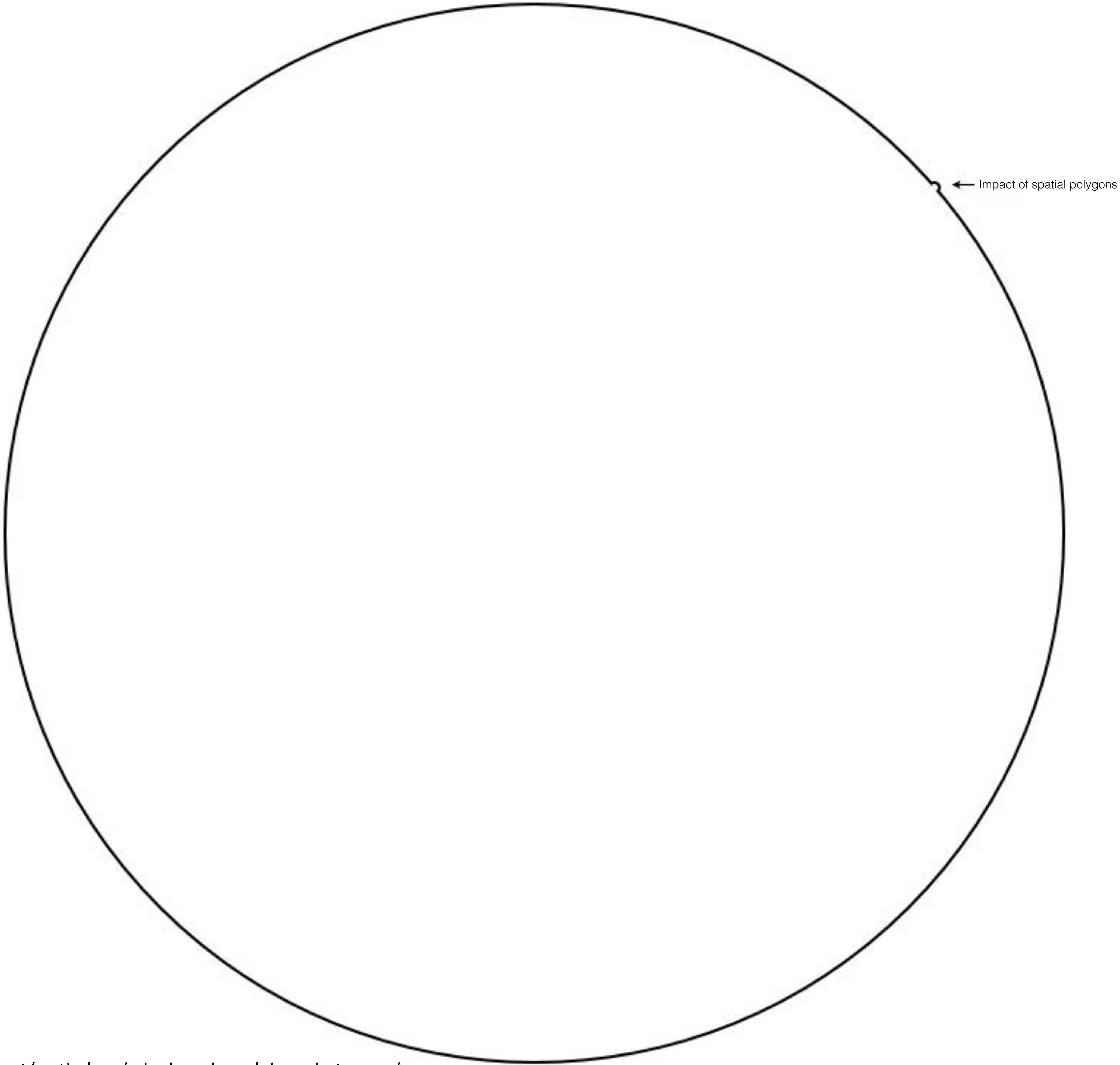
Impact of
spatial
polygons

http://matt.might.net/articles/phd-school-in-pictures/

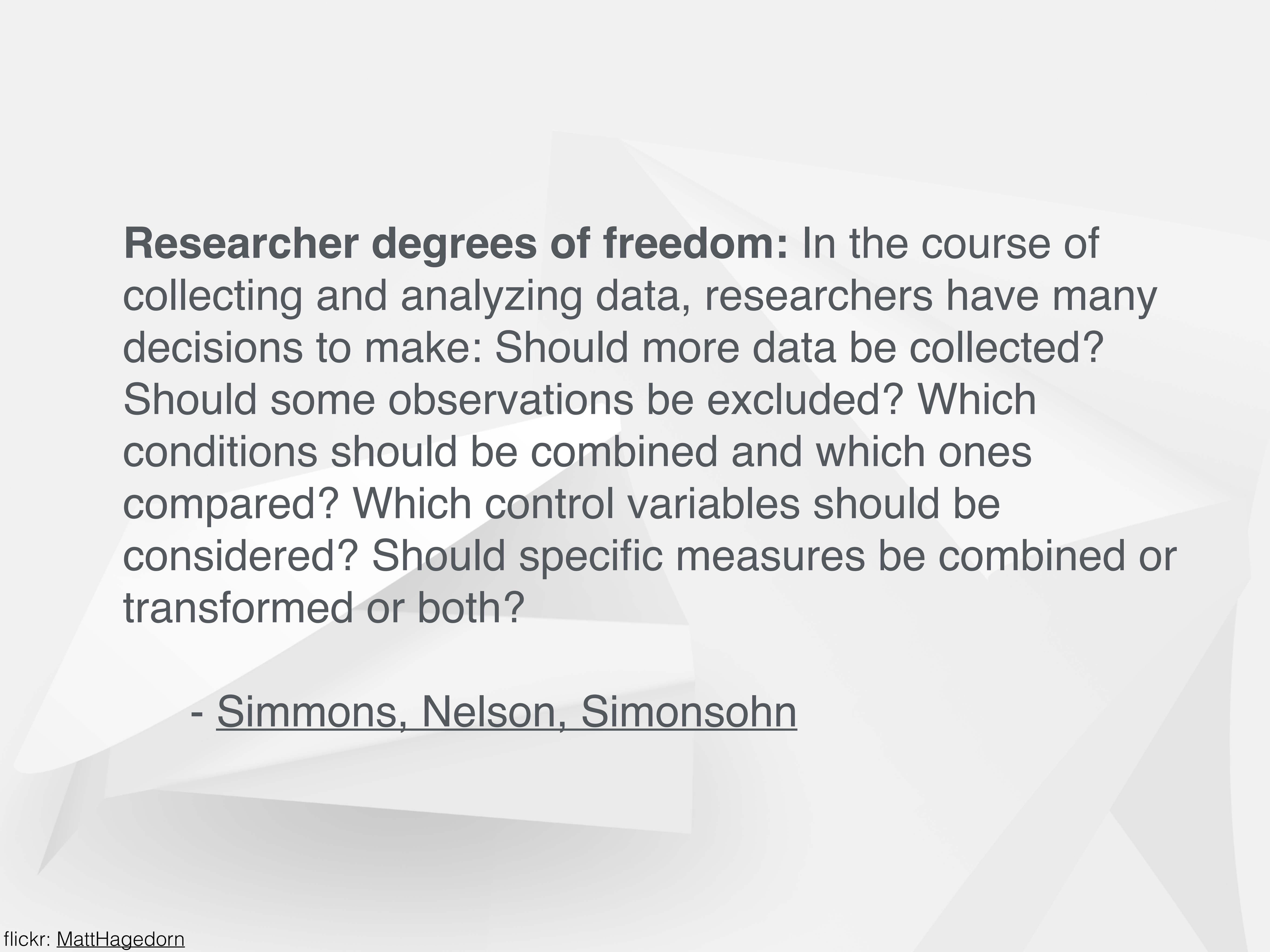


p

http://matt.might.net/articles/phd-school-in-pictures/



http://matt.might.net/articles/phd-school-in-pictures/

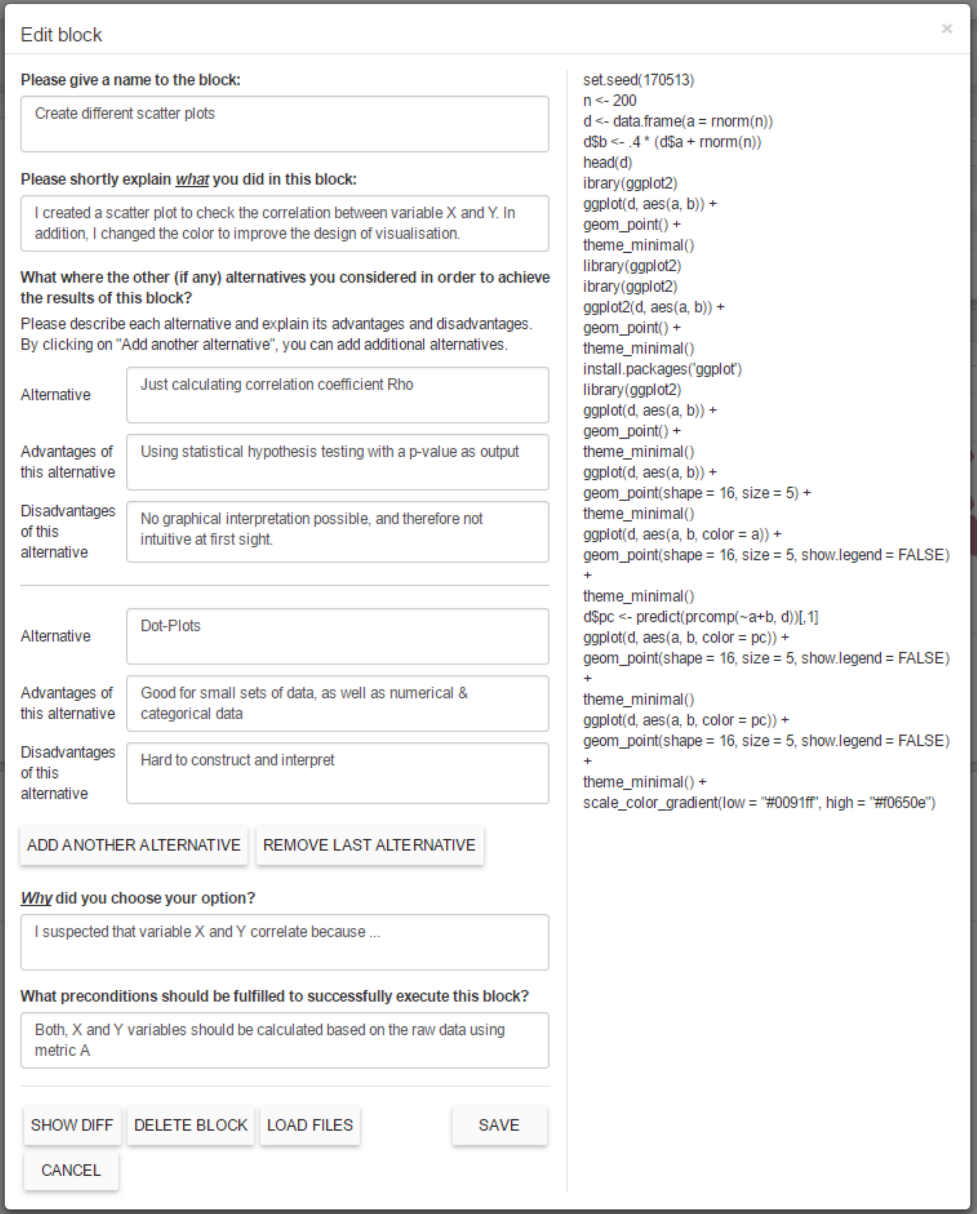


Researcher degrees of freedom: In the course of collecting and analyzing data, researchers have many decisions to make: Should more data be collected? Should some observations be excluded? Which conditions should be combined and which ones compared? Which control variables should be considered? Should specific measures be combined or transformed or both?

- Simmons, Nelson, Simonsohn

flickr: MattHagedorn

Crowdsourcing Data Analysis, Martin Schweinsberg et al



Edit block

Please give a name to the block:

Create different scatter plots

Please shortly explain *what* you did in this block:

I created a scatter plot to check the correlation between variable X and Y. In addition, I changed the color to improve the design of visualisation.

What where the other (if any) alternatives you considered in order to achieve the results of this block?

Please describe each alternative and explain its advantages and disadvantages. By clicking on "Add another alternative", you can add additional alternatives.

Alternative	Just calculating correlation coefficient Rho
Advantages of this alternative	Using statistical hypothesis testing with a p-value as output
Disadvantages of this alternative	No graphical interpretation possible, and therefore not intuitive at first sight.
Alternative	Dot-Plots
Advantages of this alternative	Good for small sets of data, as well as numerical & categorical data
Disadvantages of this alternative	Hard to construct and interpret

***Why* did you choose your option?**

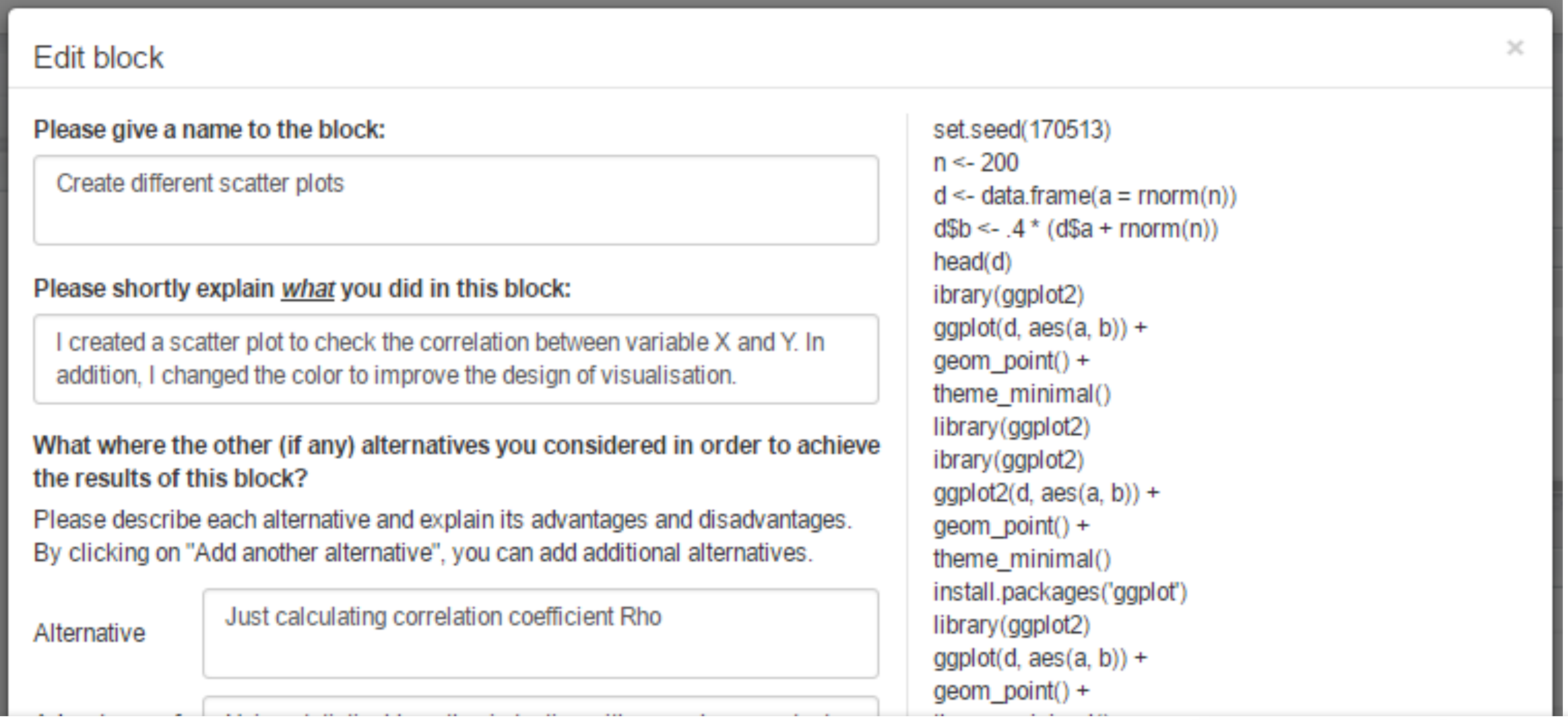
I suspected that variable X and Y correlate because ...

What preconditions should be fulfilled to successfully execute this block?

Both, X and Y variables should be calculated based on the raw data using metric A

```
set.seed(170513)
n <- 200
d <- data.frame(a = rnorm(n))
d$b <- .4 * (d$a + rnorm(n))
head(d)
ibrary(ggplot2)
ggplot(d, aes(a, b)) +
geom_point() +
theme_minimal()
library(ggplot2)
ibrary(ggplot2)
ggplot2(d, aes(a, b)) +
geom_point() +
theme_minimal()
install.packages('ggplot')
library(ggplot2)
ggplot(d, aes(a, b)) +
geom_point() +
theme_minimal()
ggplot(d, aes(a, b)) +
geom_point(shape = 16, size = 5) +
theme_minimal()
ggplot(d, aes(a, b, color = a)) +
geom_point(shape = 16, size = 5, show.legend = FALSE)
+
theme_minimal()
d$pc <- predict(prcomp(~a+b, d))[,1]
ggplot(d, aes(a, b, color = pc)) +
geom_point(shape = 16, size = 5, show.legend = FALSE)
+
theme_minimal()
ggplot(d, aes(a, b, color = pc)) +
geom_point(shape = 16, size = 5, show.legend = FALSE)
+
theme_minimal() +
scale_color_gradient(low = "#0091ff", high = "#f0650e")
```

Crowdsourcing Data Analysis, Martin Schweinsberg et al



Please give a name to the block:

Create different scatter plots

Please shortly explain *what* you did in this block:

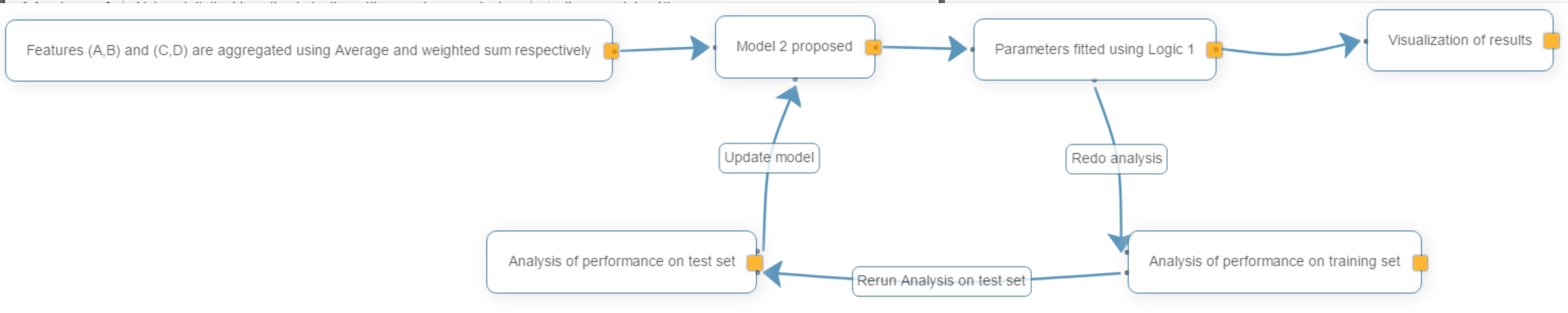
I created a scatter plot to check the correlation between variable X and Y. In addition, I changed the color to improve the design of visualisation.

What where the other (if any) alternatives you considered in order to achieve the results of this block?

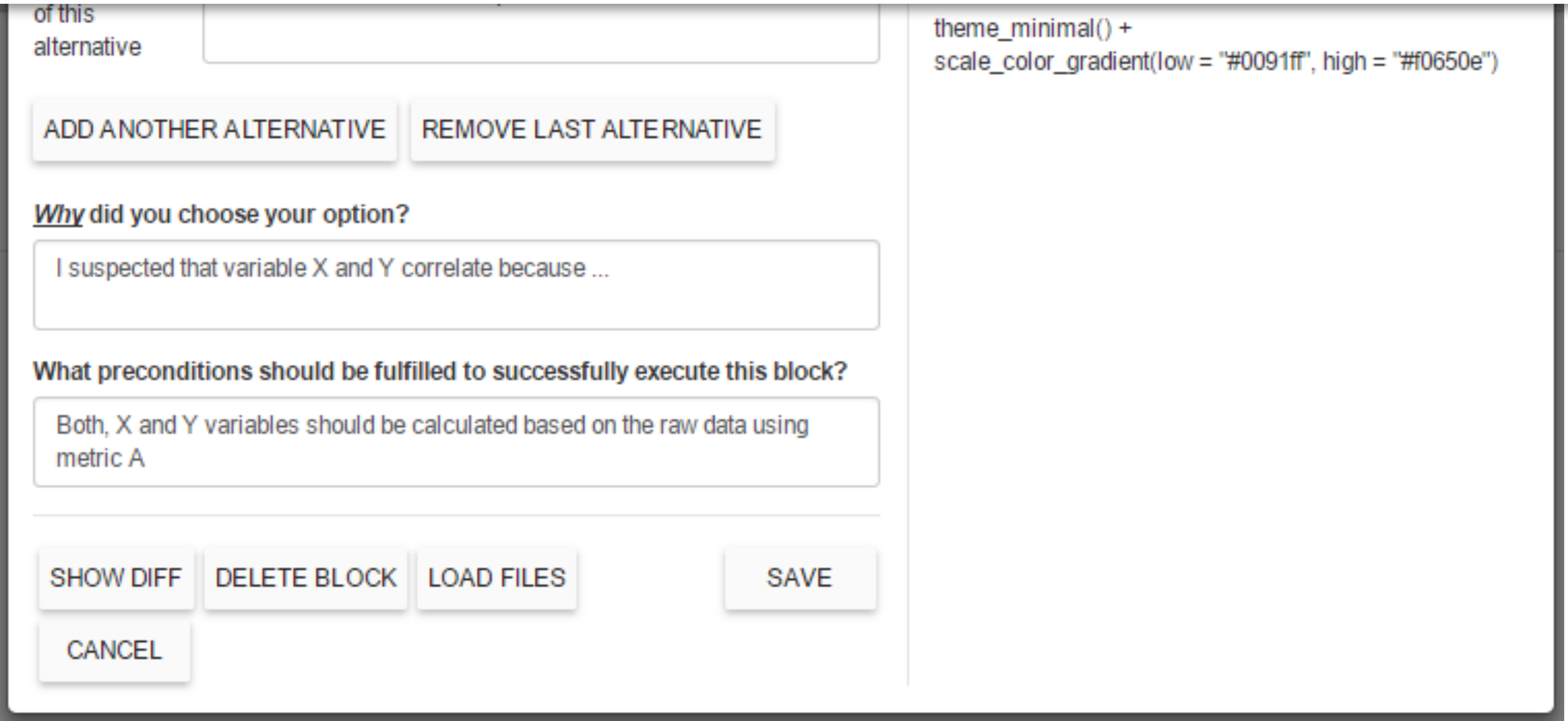
Please describe each alternative and explain its advantages and disadvantages. By clicking on "Add another alternative", you can add additional alternatives.

Alternative: Just calculating correlation coefficient Rho

```
set.seed(170513)
n <- 200
d <- data.frame(a = rnorm(n))
d$b <- .4 * (d$a + rnorm(n))
head(d)
ibrary(ggplot2)
ggplot(d, aes(a, b)) +
geom_point() +
theme_minimal()
library(ggplot2)
ibrary(ggplot2)
ggplot2(d, aes(a, b)) +
geom_point() +
theme_minimal()
install.packages('ggplot')
library(ggplot2)
ggplot(d, aes(a, b)) +
geom_point() +
```



Features (A,B) and (C,D) are aggregated using Average and weighted sum respectively
Model 2 proposed
Parameters fitted using Logic 1
Visualization of results
Update model
Redo analysis
Analysis of performance on test set
Rerun Analysis on test set
Analysis of performance on training set



of this alternative

```
theme_minimal() +
scale_color_gradient(low = "#0091ff", high = "#f0650e")
```

***Why* did you choose your option?**

I suspected that variable X and Y correlate because ...

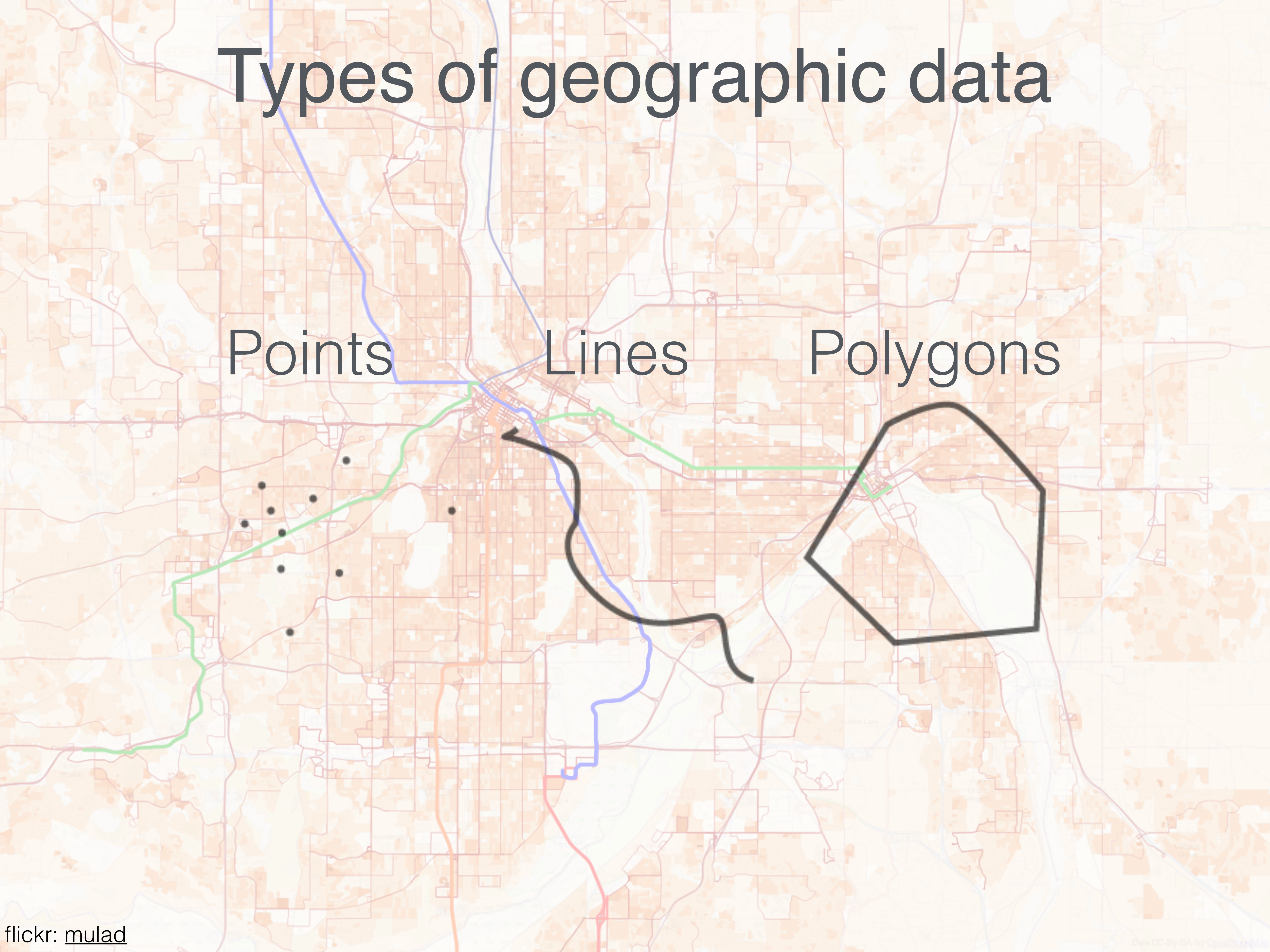
What preconditions should be fulfilled to successfully execute this block?

Both, X and Y variables should be calculated based on the raw data using metric A

Spatial background

flickr: MattHagedorn

Types of geographic data



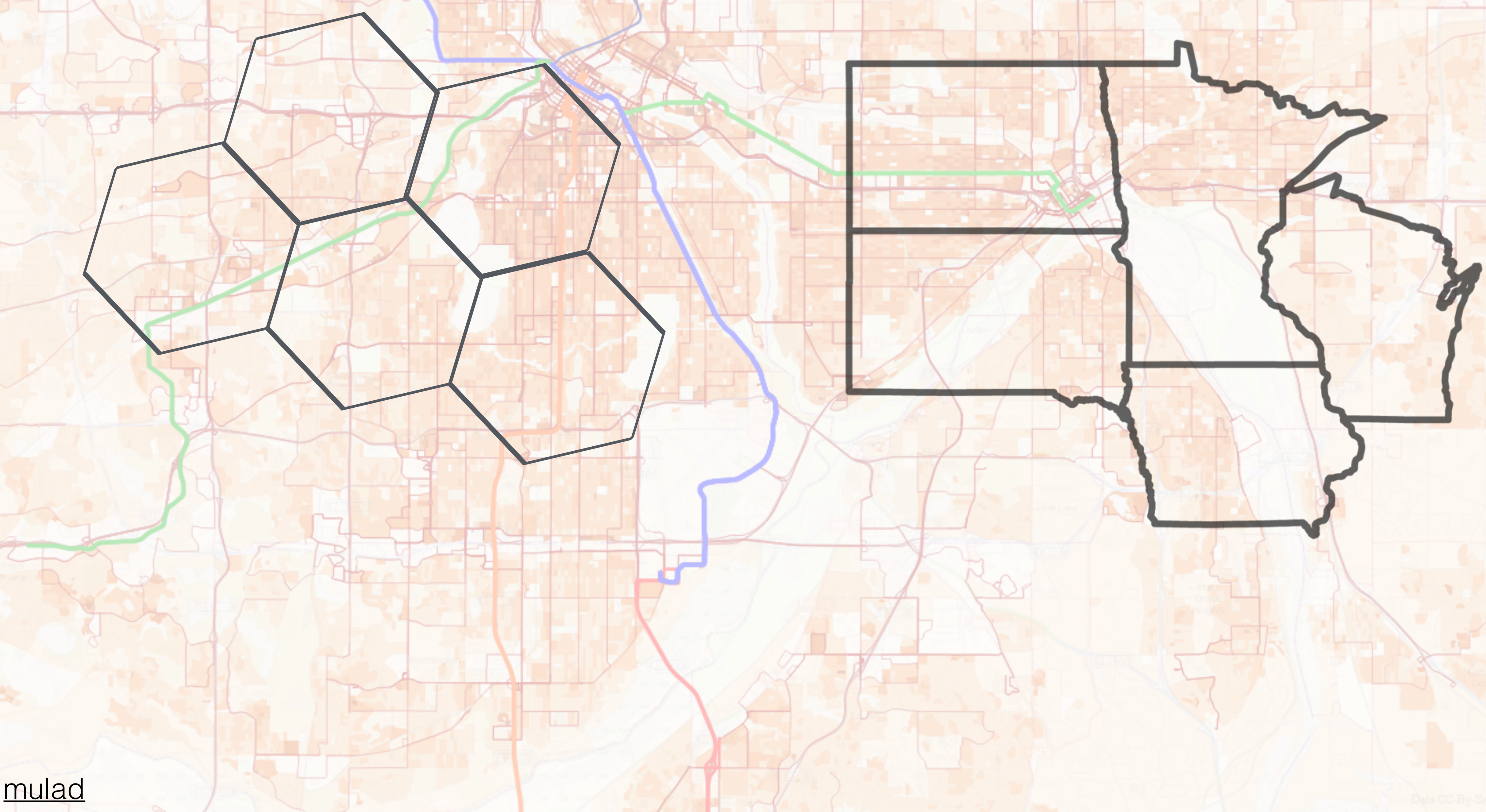
Points

Lines

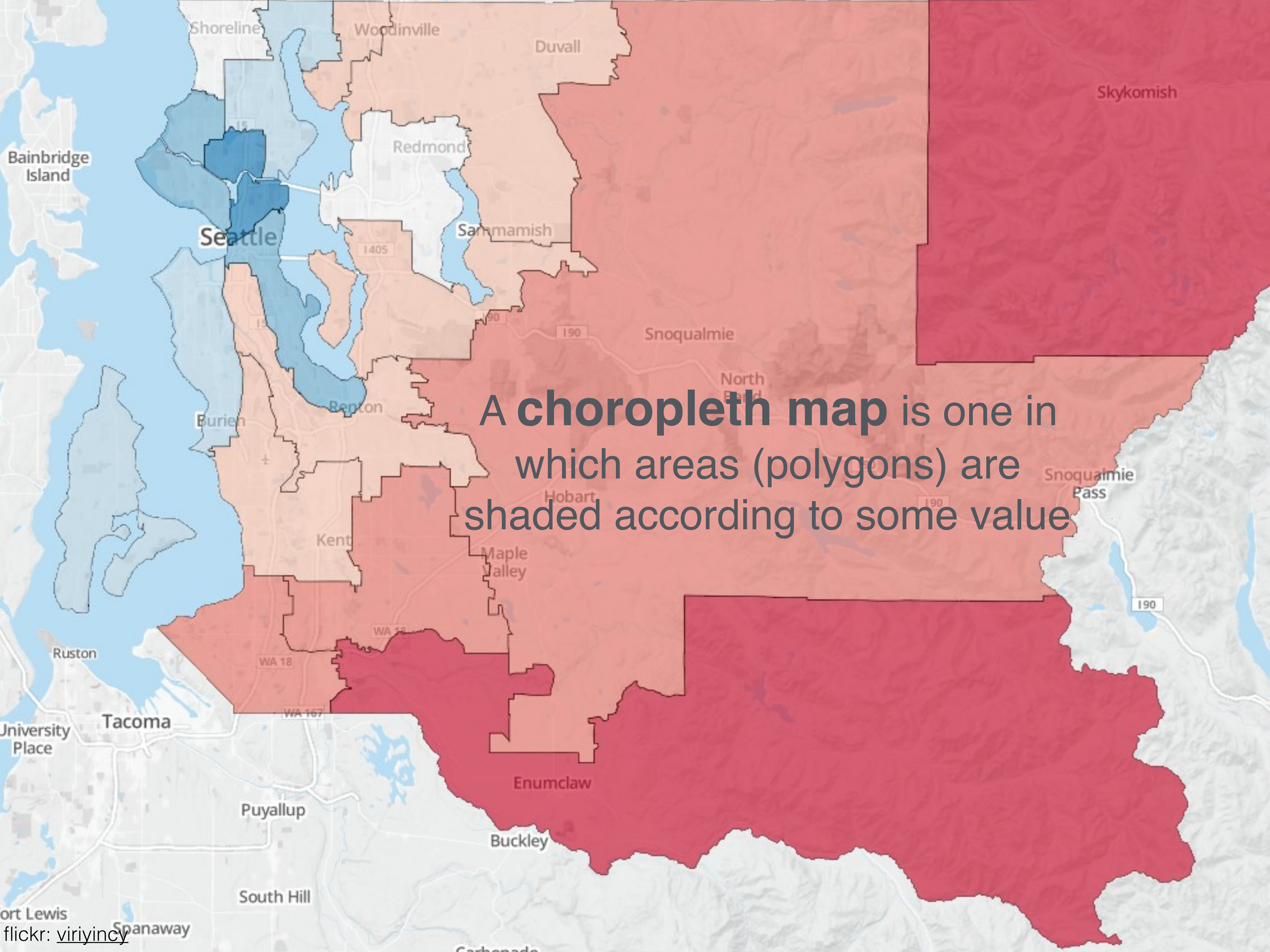
Polygons

flickr: mulad

Polygons can be regular or irregular



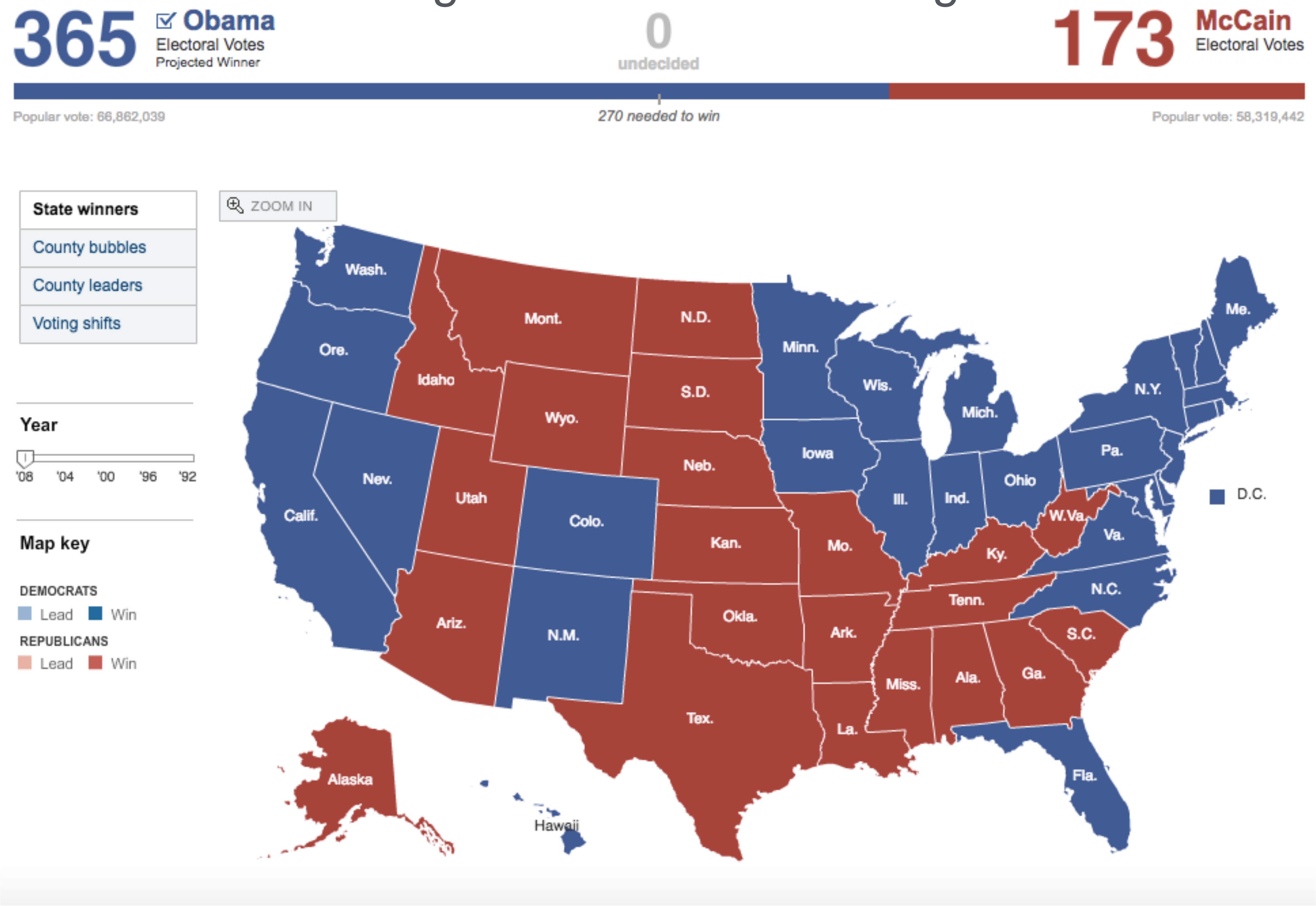
flickr: mulad



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

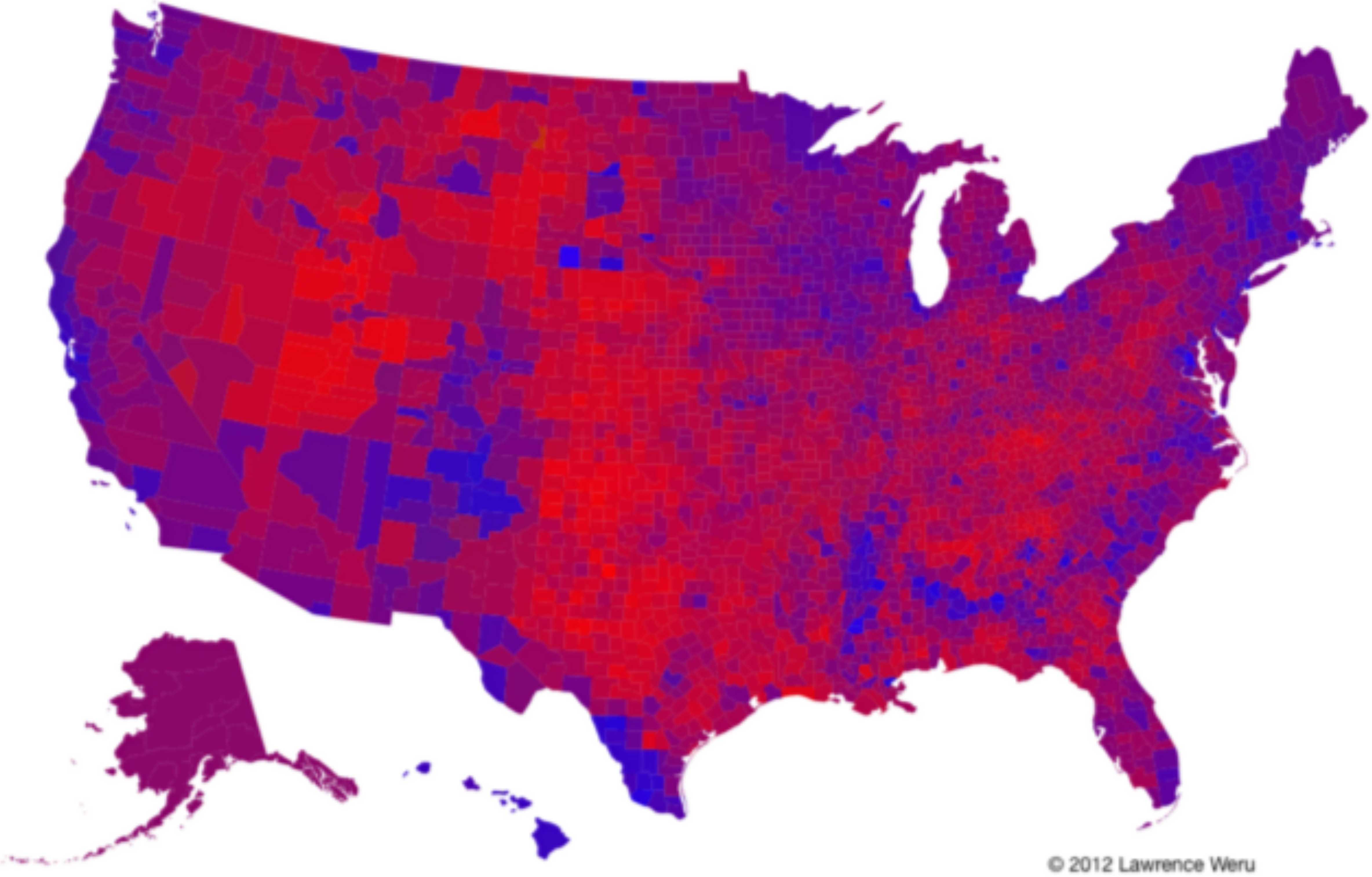
flickr: viriyincy

Area gives a lot of visual weight



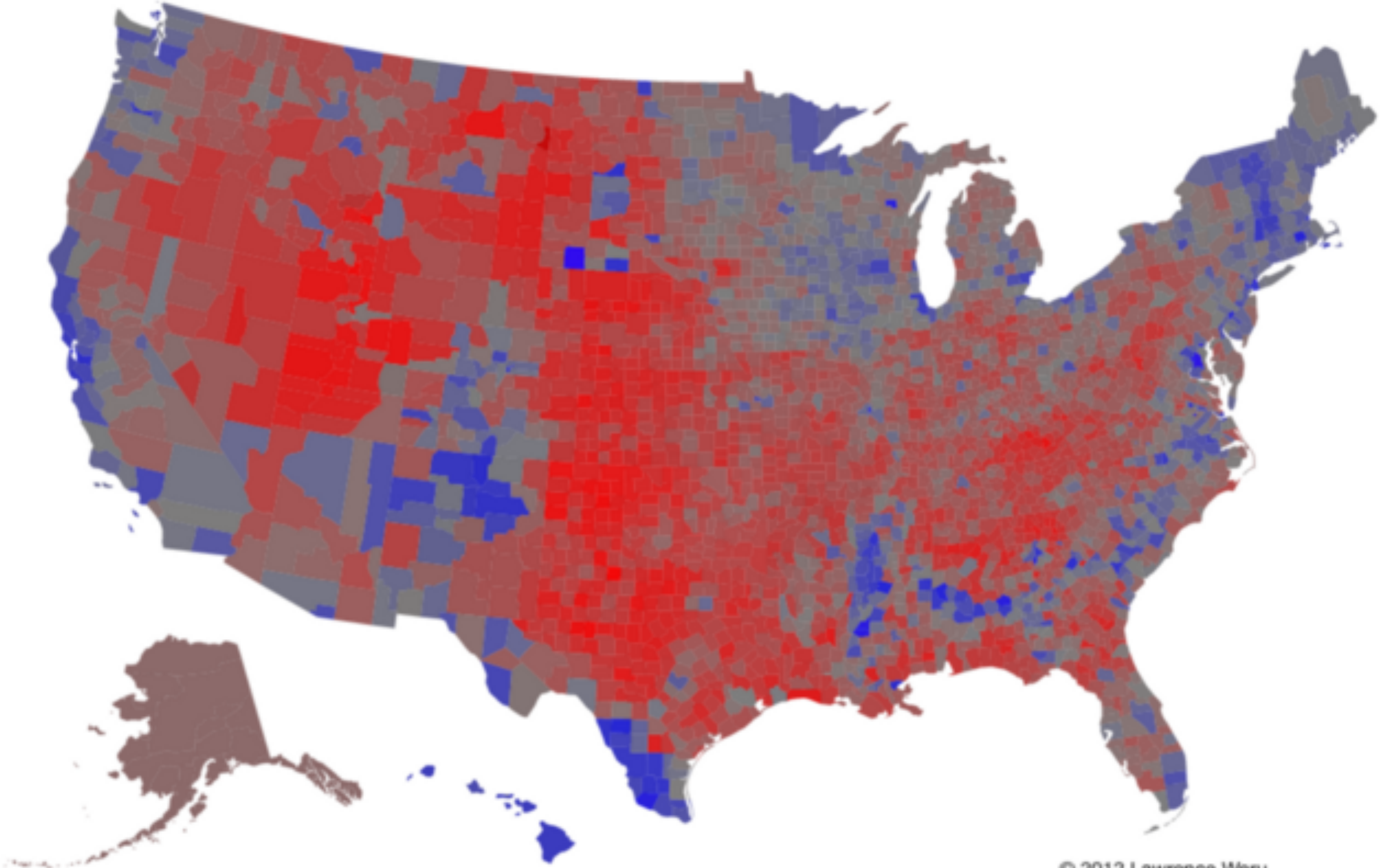
http://www.nytimes.com/elections/2008/results/president/map.html

Purpling the map



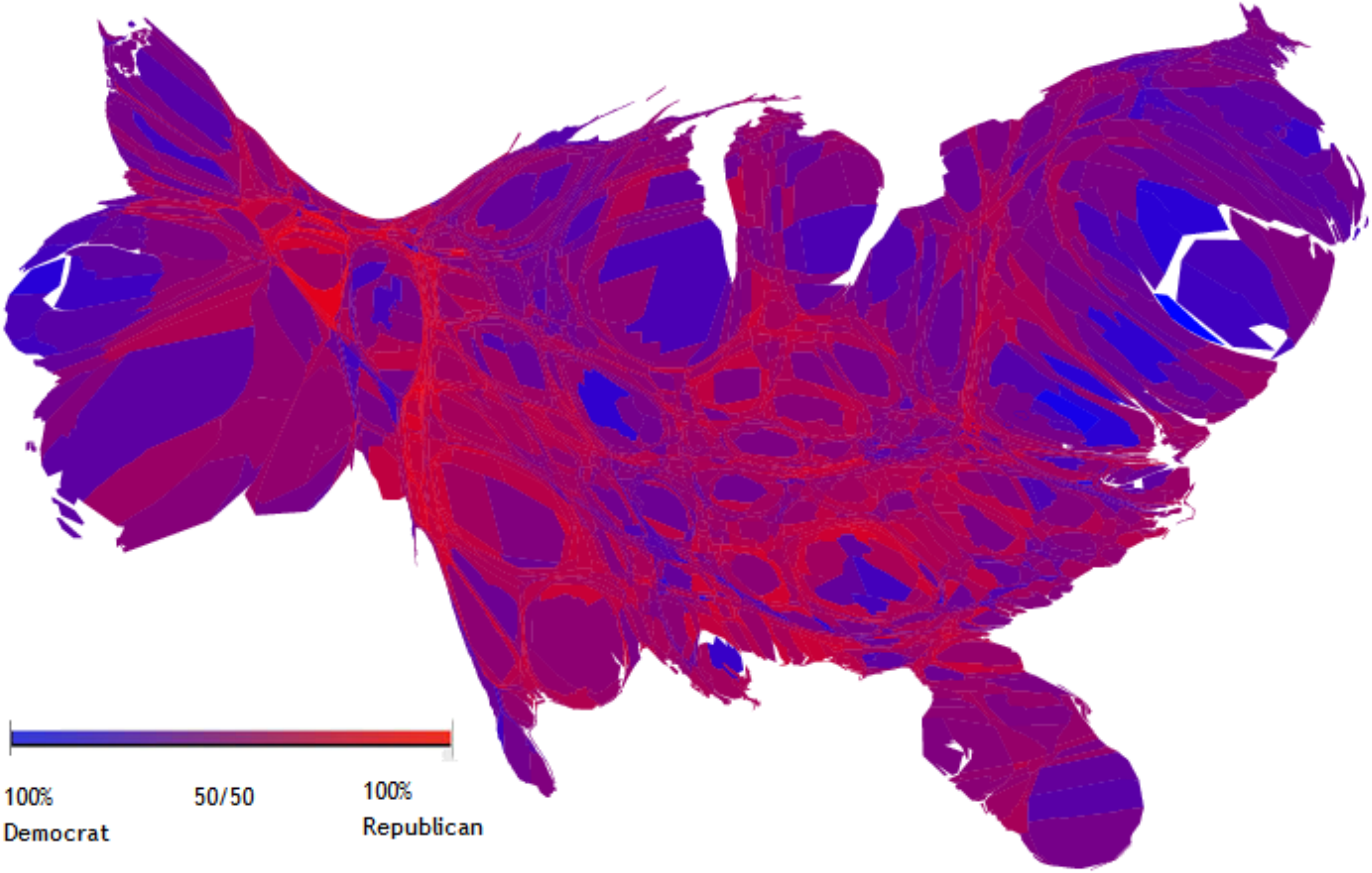
https://medium.com/matter/the-trouble-with-the-purple-election-map-31e6cb9f1827

Purpling the map



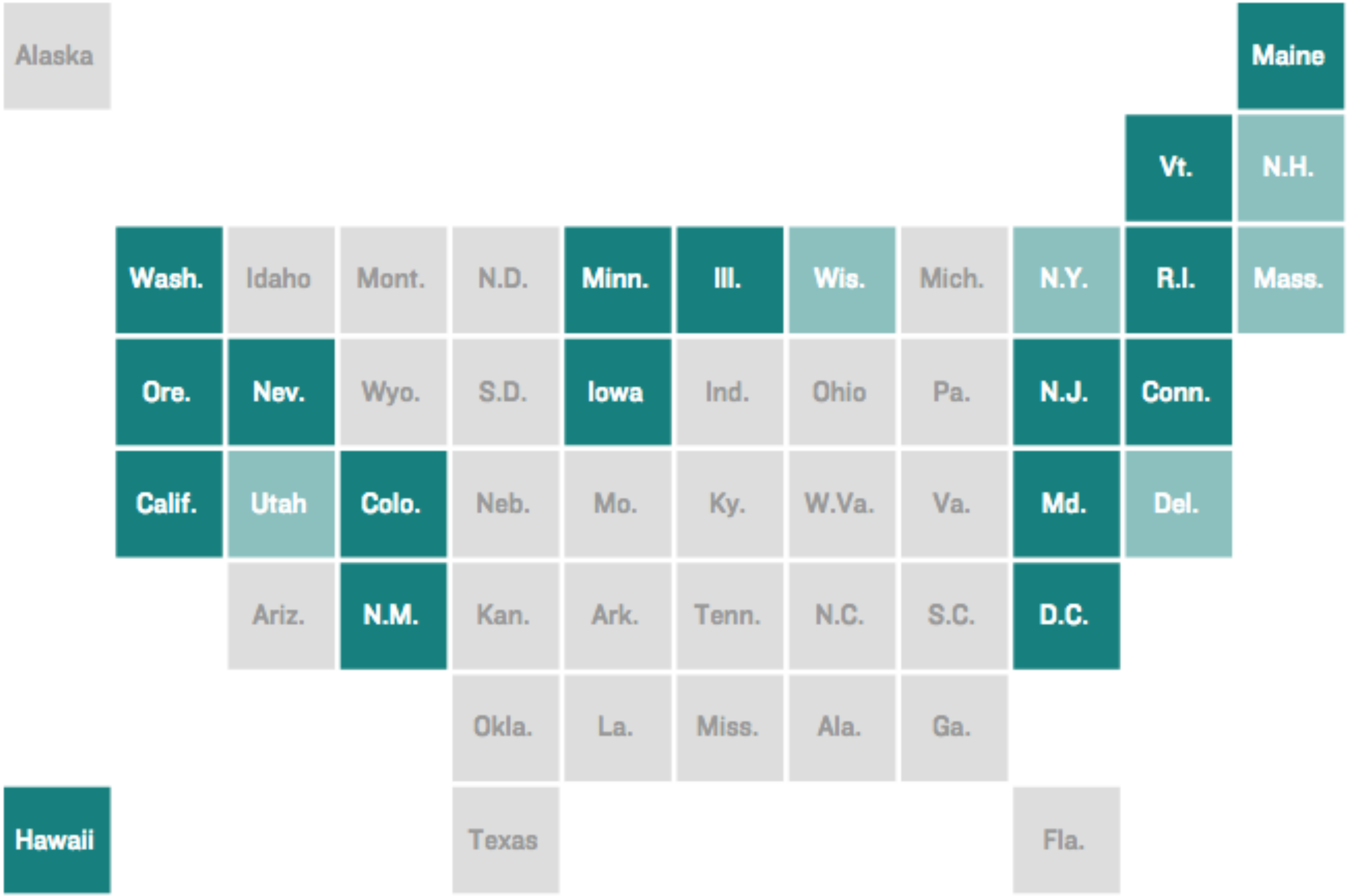
https://medium.com/matter/the-trouble-with-the-purple-election-map-31e6cb9f1827

Cartogram



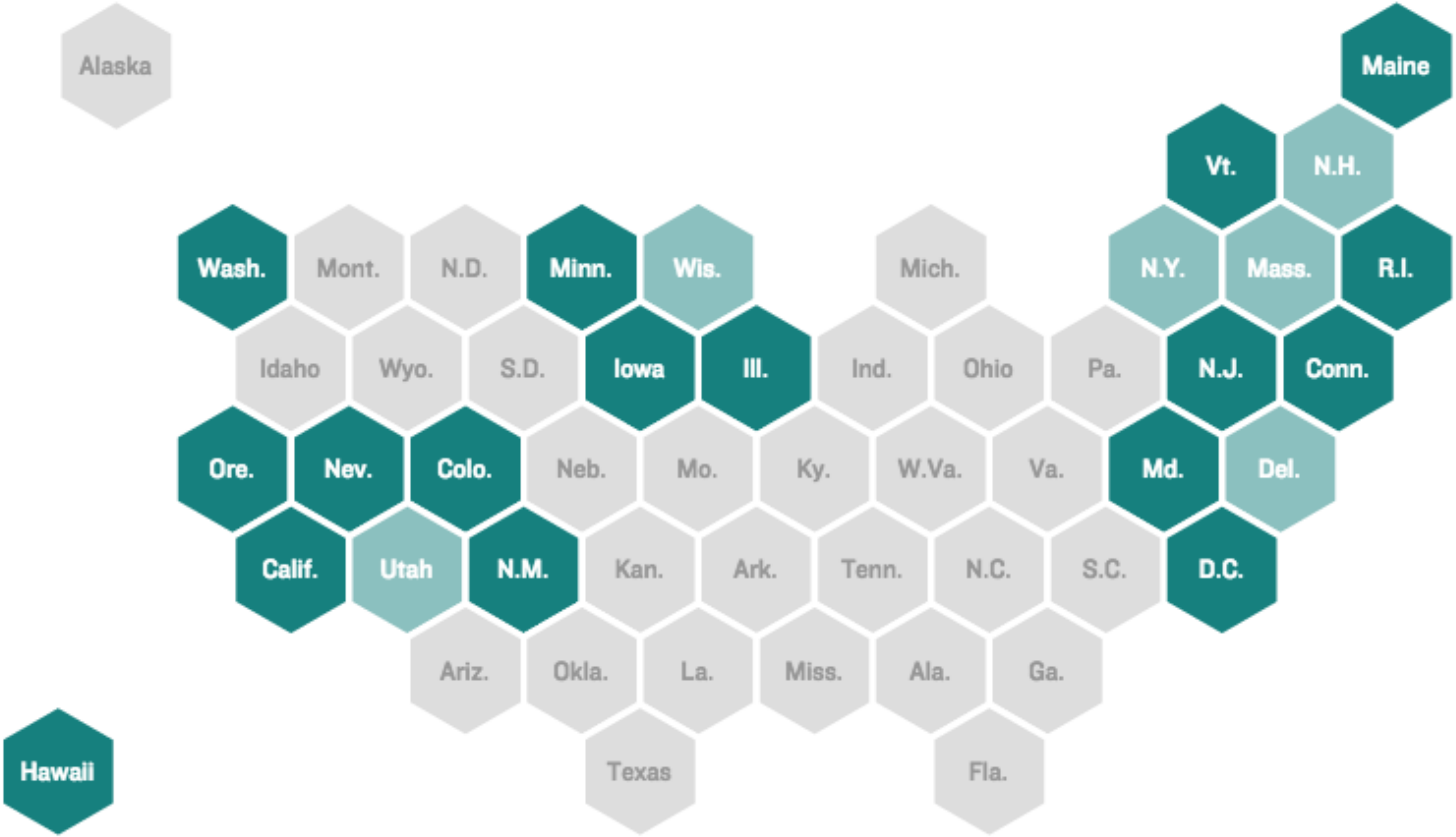
wikipedia: M. E. J. Newman

Cartogram



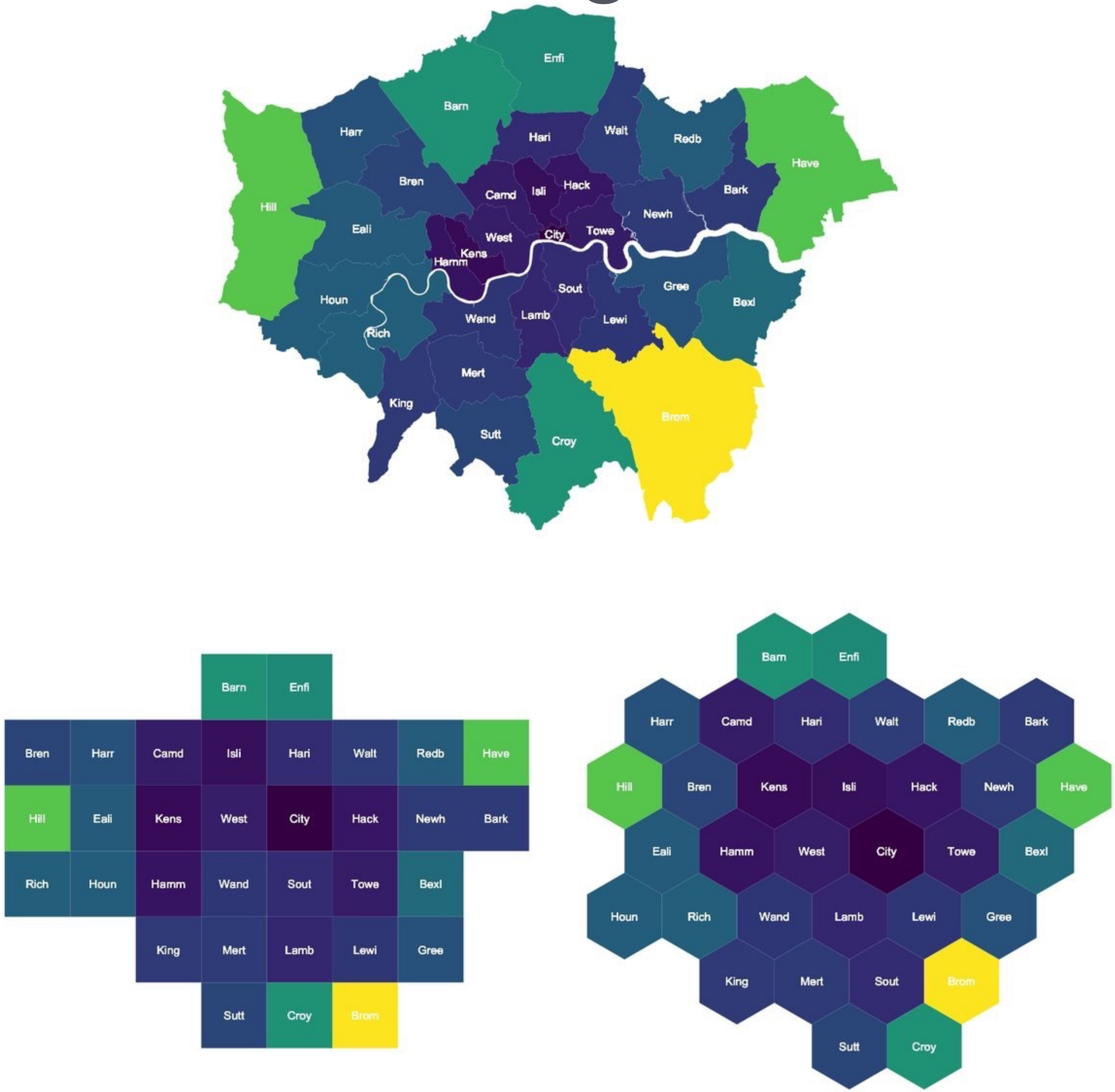
http://blog.apps.npr.org/2015/05/11/hex-tile-maps.html

Cartogram

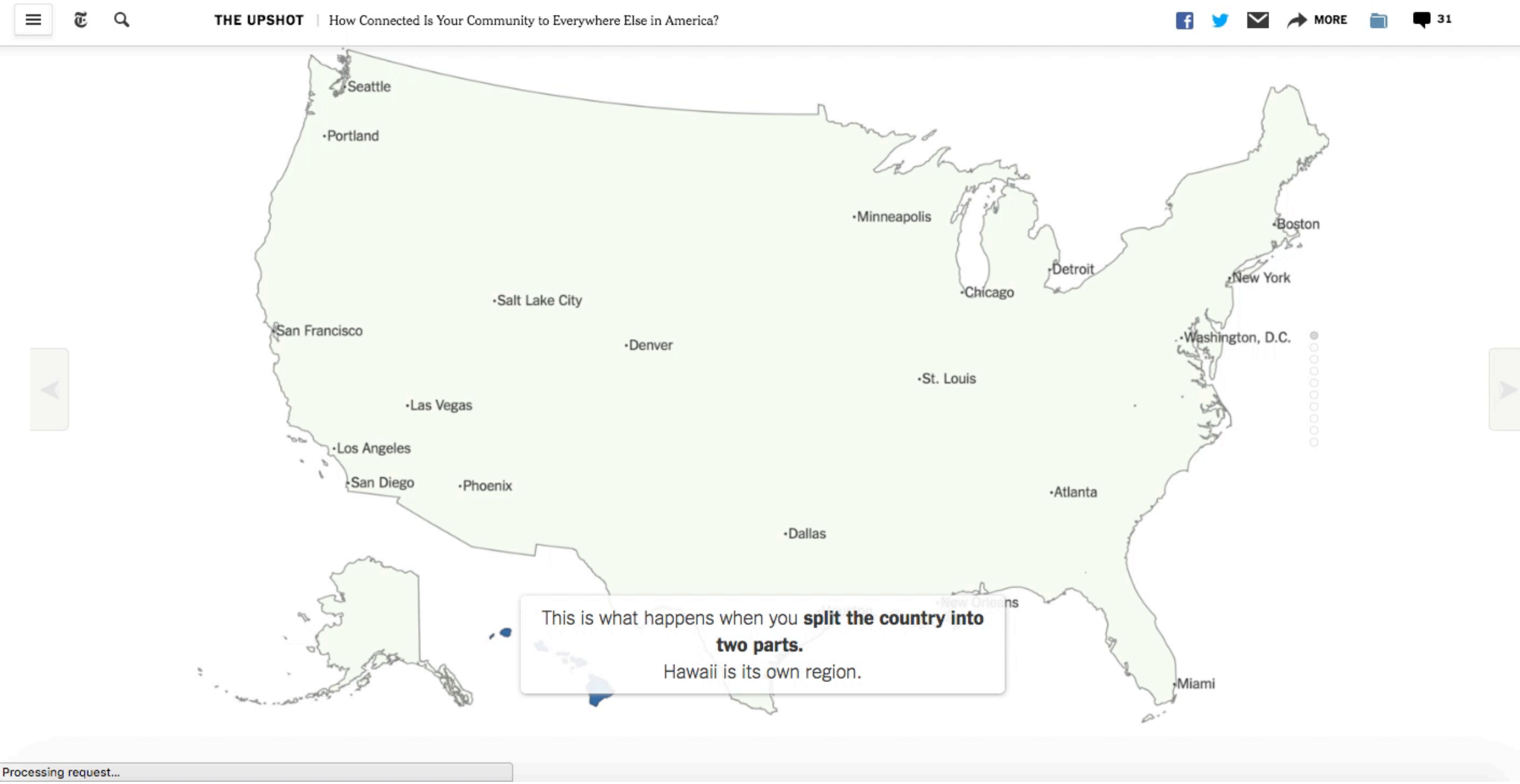


http://blog.apps.npr.org/2015/05/11/hex-tile-maps.html

Cartogram



https://twitter.com/iammrbailey/status/925346870381240320/photo/1



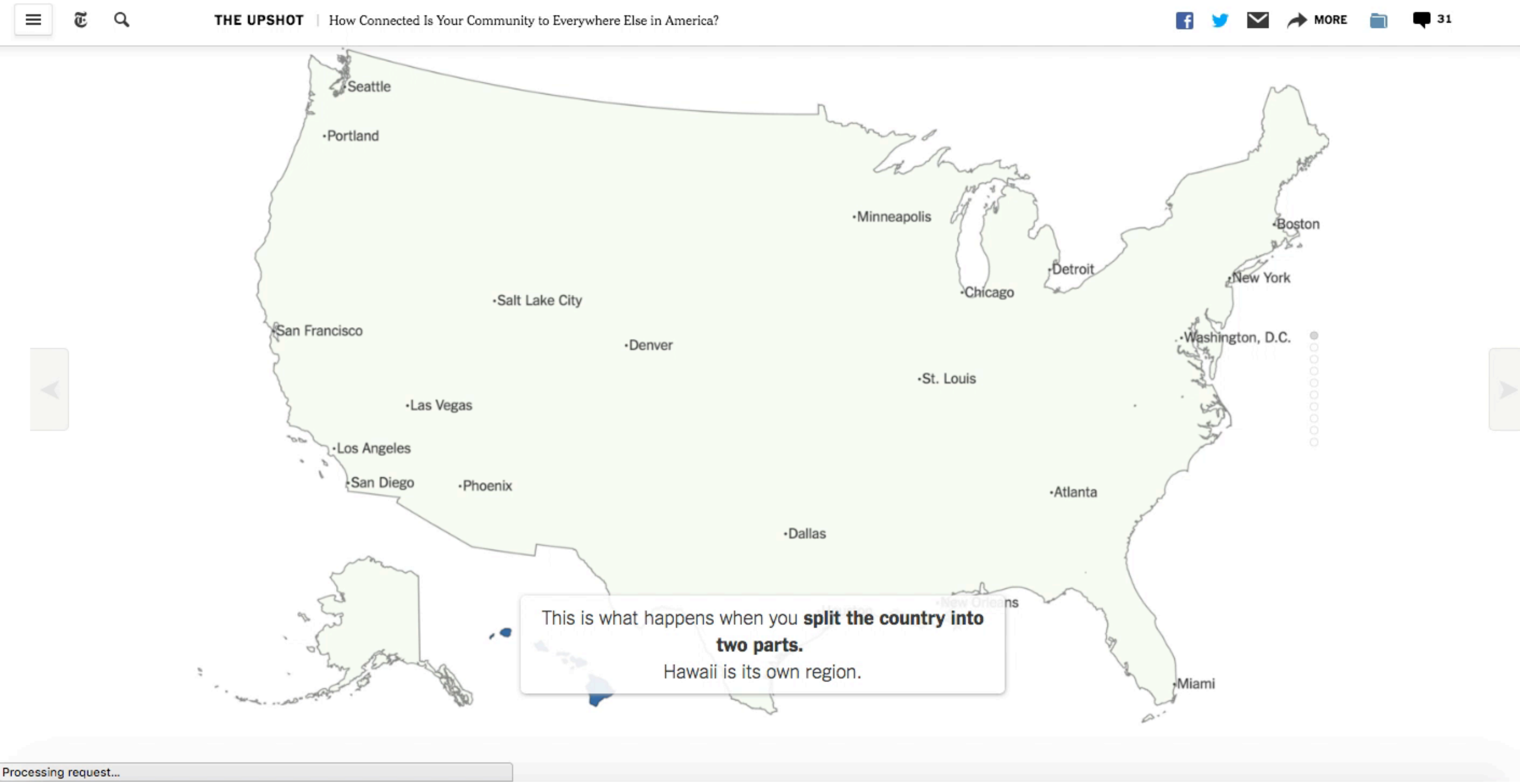
THE UPSHOT | How Connected Is Your Community to Everywhere Else in America?

This is what happens when you **split the country into two parts.**

Hawaii is its own region.

Processing request...

https://www.nytimes.com/interactive/2018/09/19/upshot/facebook-county-friendships.html



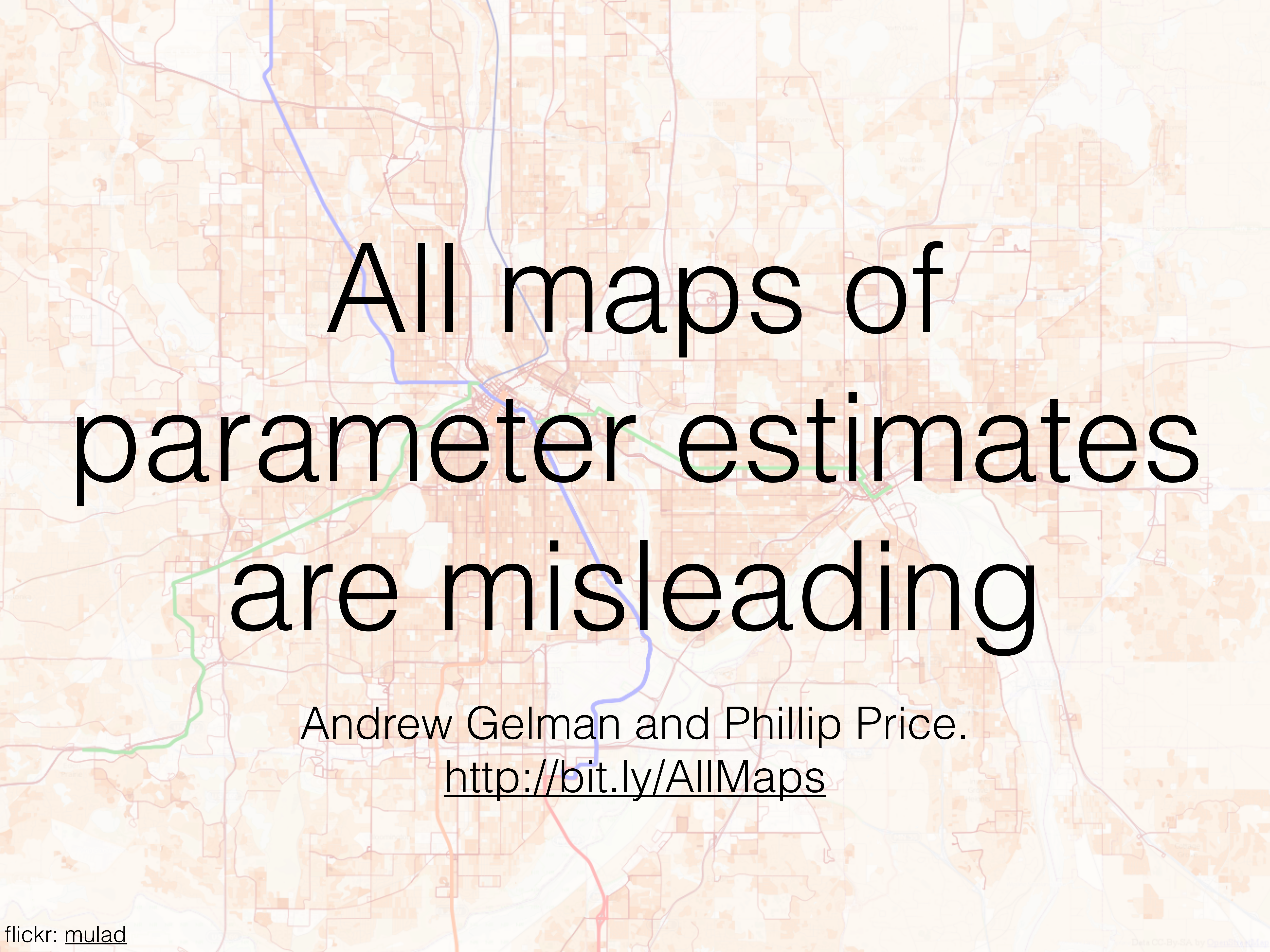
This is what happens when you **split the country into two parts.**
Hawaii is its own region.

Processing request...

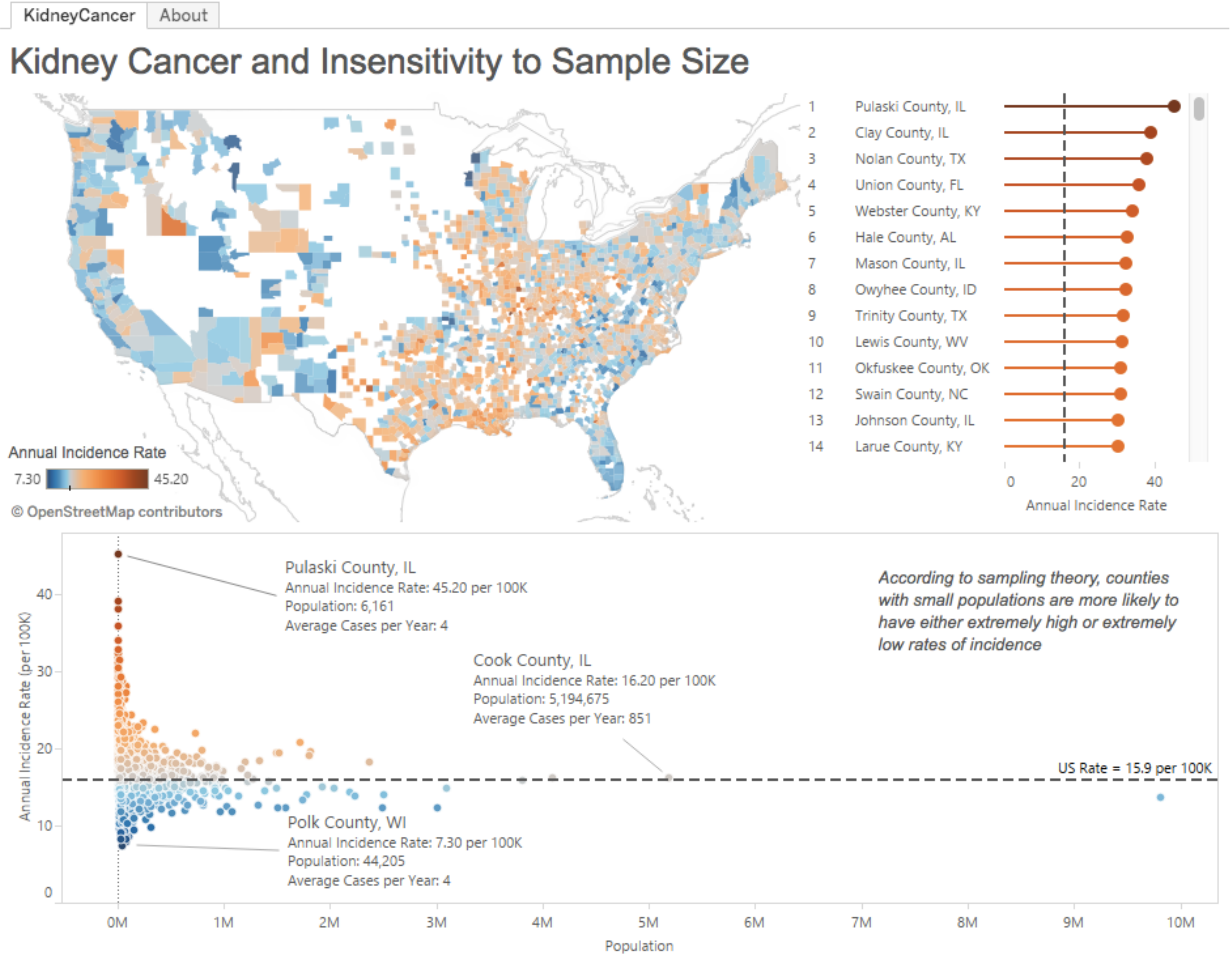
https://www.nytimes.com/interactive/2018/09/19/upshot/facebook-county-friendships.html

All maps of parameter estimates are misleading

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



flickr: mulad



Kidney Cancer and Insensitivity to Sample Size

1	Pulaski County, IL
2	Clay County, IL
3	Nolan County, TX
4	Union County, FL
5	Webster County, KY
6	Hale County, AL
7	Mason County, IL
8	Owyhee County, ID
9	Trinity County, TX
10	Lewis County, WV
11	Okfuskee County, OK
12	Swain County, NC
13	Johnson County, IL
14	Larue County, KY

Pulaski County, IL
Annual Incidence Rate: 45.20 per 100K
Population: 6,161
Average Cases per Year: 4

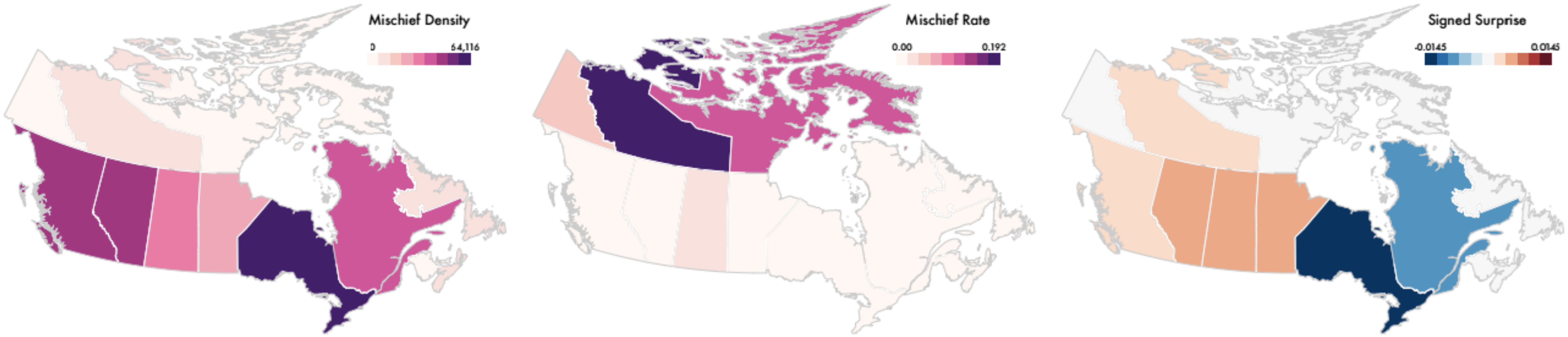
Cook County, IL
Annual Incidence Rate: 16.20 per 100K
Population: 5,194,675
Average Cases per Year: 851

Polk County, WI
Annual Incidence Rate: 7.30 per 100K
Population: 44,205
Average Cases per Year: 4

According to sampling theory, counties with small populations are more likely to have either extremely high or extremely low rates of incidence

http://dataremixed.com/2015/01/avoiding-data-pitfalls-part-2/

Surprise! Bayesian Weighting for De-Biasing Thematic Maps.



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

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

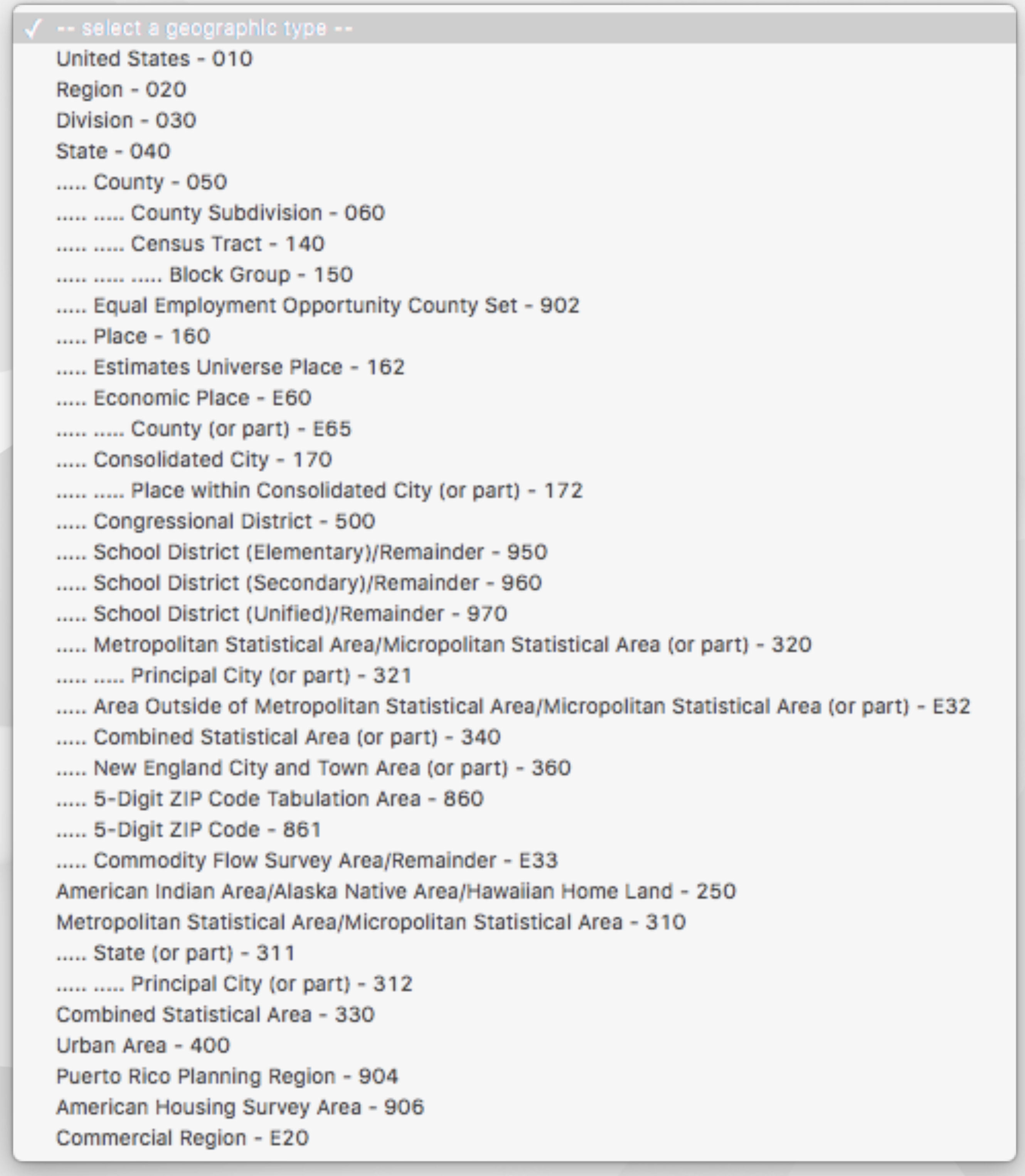
(c) The **Surprise Map** of mischief.

Michael Correll and Jeffrey Heer

http://bit.ly/SurpriseMaps

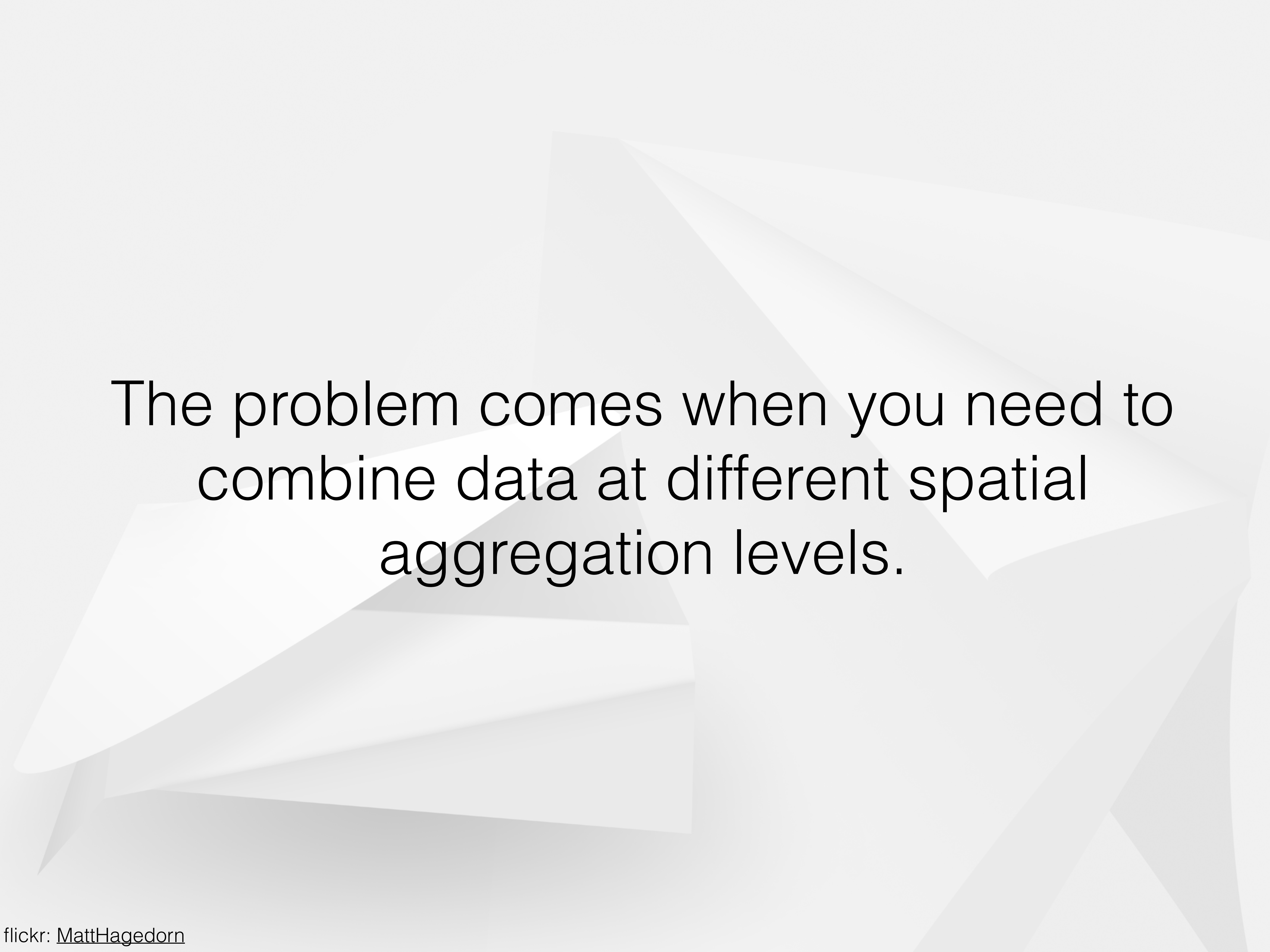
Some common spatial polygons

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



```
✓ -- select a geographic type --
United States - 010
Region - 020
Division - 030
State - 040
..... County - 050
..... ..... County Subdivision - 060
..... ..... Census Tract - 140
..... ..... ..... Block Group - 150
..... Equal Employment Opportunity County Set - 902
..... Place - 160
..... Estimates Universe Place - 162
..... Economic Place - E60
..... ..... County (or part) - E65
..... Consolidated City - 170
..... ..... Place within Consolidated City (or part) - 172
..... Congressional District - 500
..... School District (Elementary)/Remainder - 950
..... School District (Secondary)/Remainder - 960
..... School District (Unified)/Remainder - 970
..... Metropolitan Statistical Area/Micropolitan Statistical Area (or part) - 320
..... ..... Principal City (or part) - 321
..... Area Outside of Metropolitan Statistical Area/Micropolitan Statistical Area (or part) - E32
..... Combined Statistical Area (or part) - 340
..... New England City and Town Area (or part) - 360
..... 5-Digit ZIP Code Tabulation Area - 860
..... 5-Digit ZIP Code - 861
..... Commodity Flow Survey Area/Remainder - E33
American Indian Area/Alaska Native Area/Hawaiian Home Land - 250
Metropolitan Statistical Area/Micropolitan Statistical Area - 310
..... State (or part) - 311
..... ..... Principal City (or part) - 312
Combined Statistical Area - 330
Urban Area - 400
Puerto Rico Planning Region - 904
American Housing Survey Area - 906
Commercial Region - E20
```

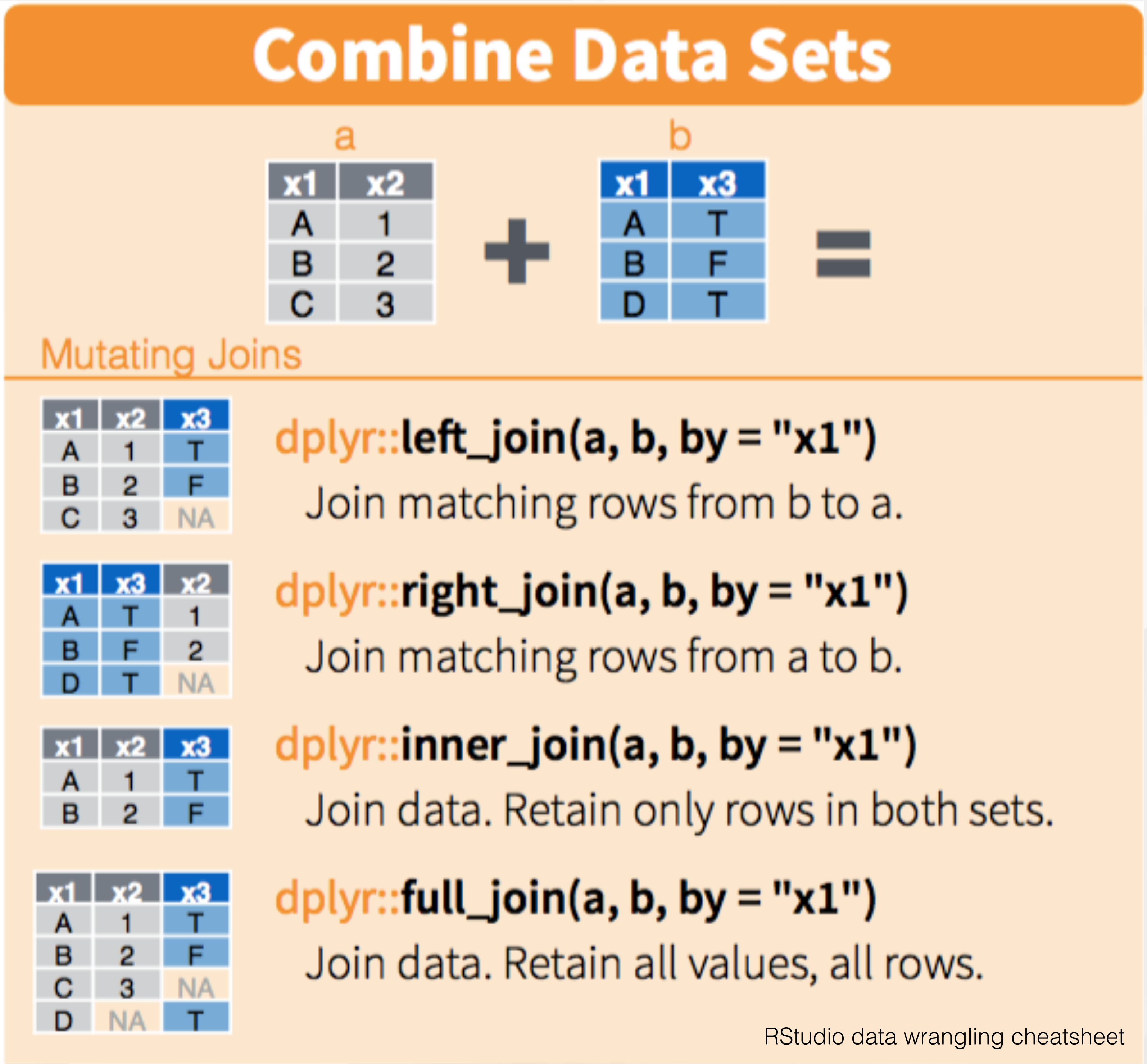
flickr: MattHagedorn



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

flickr: MattHagedorn

Combining tabular data



Combine Data Sets

a

x1	x2
A	1
B	2
C	3

+

b

x1	x3
A	T
B	F
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

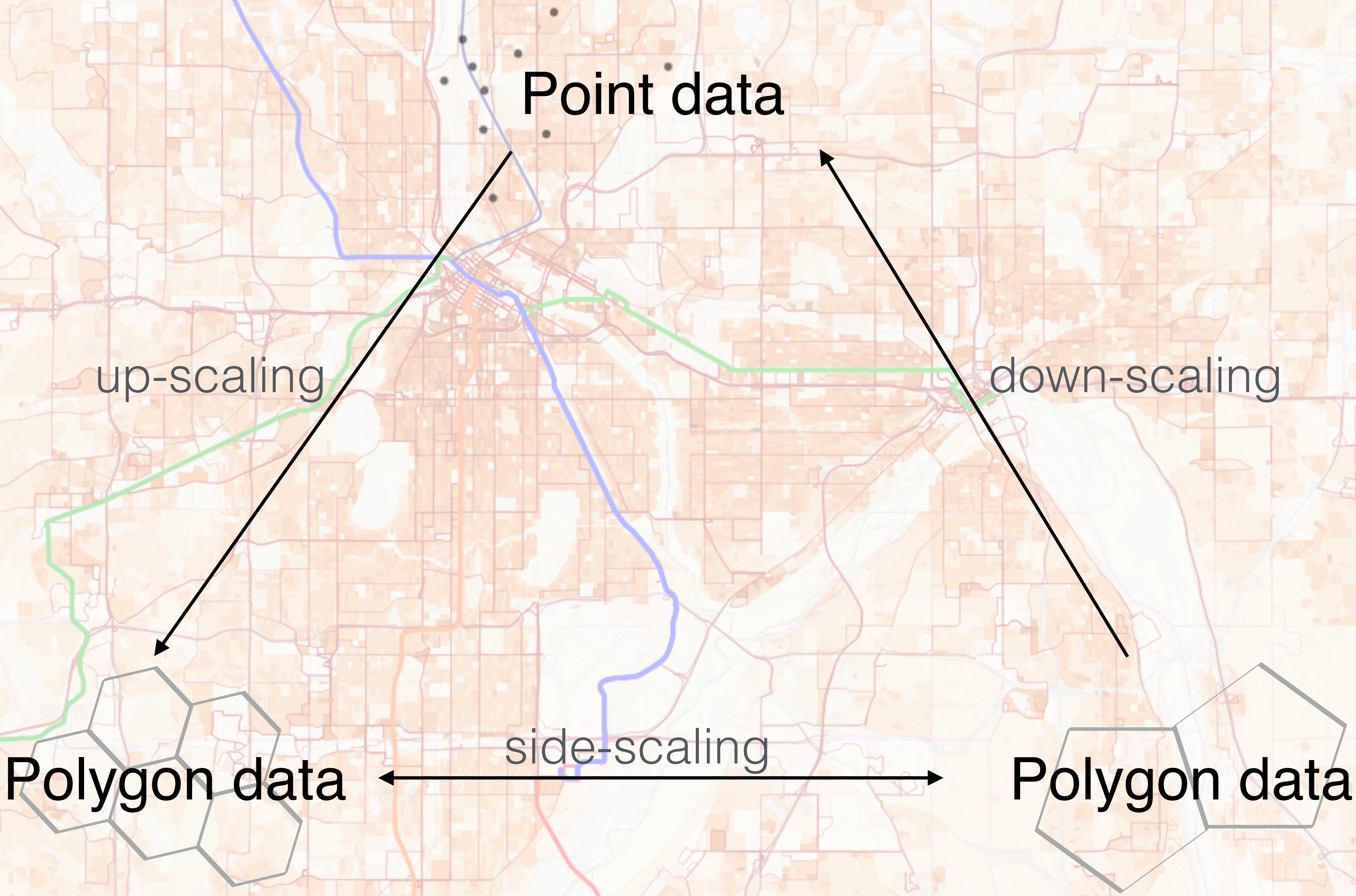
dplyr::**full_join(a, b, by = "x1")**

Join data. Retain all values, all rows.

RStudio data wrangling cheatsheet

flickr: MattHagedorn

Change of support methods



Point data

up-scaling

down-scaling

Polygon data

side-scaling

Polygon data

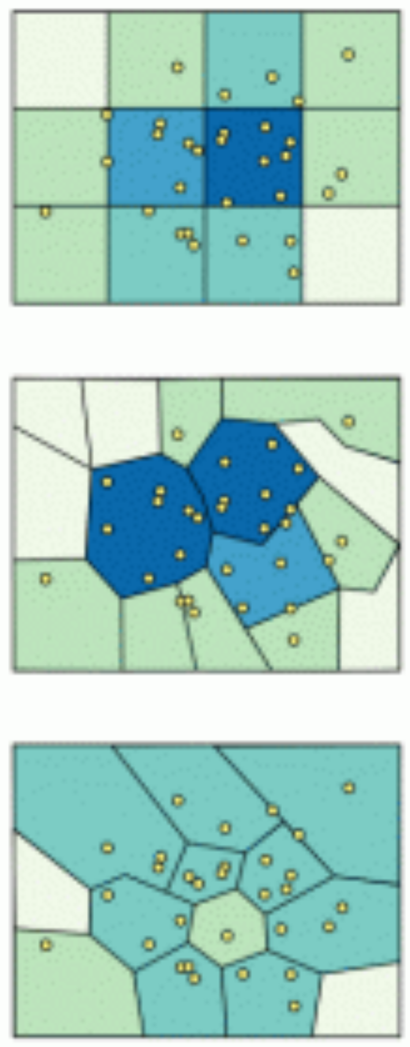
flickr: mulad

Easiest first—
up-scaling

flickr: MattHagedorn

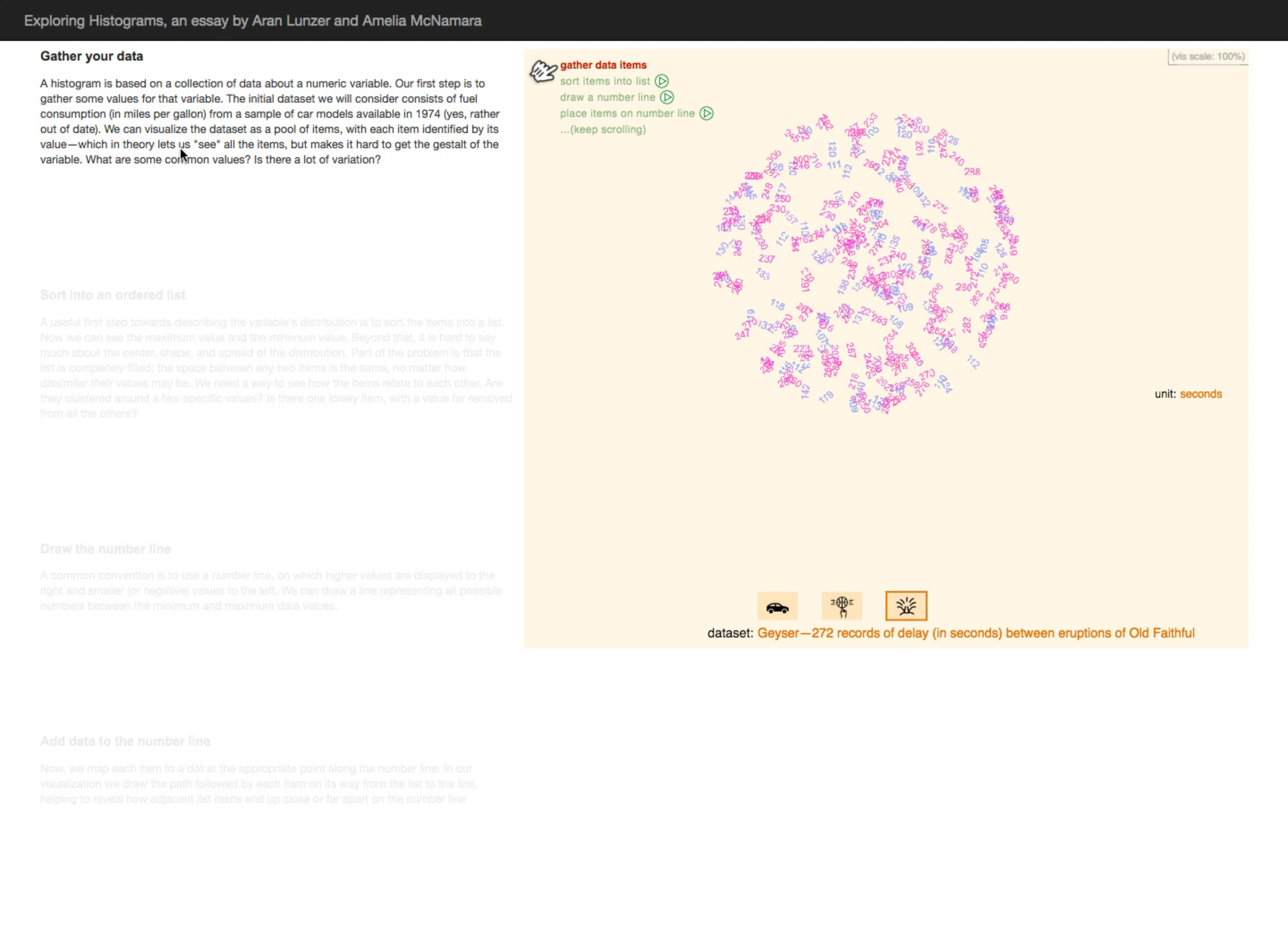
Modifiable Areal Unit Problem

“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/

flickr: mulad


Exploring Histograms, an essay by Aran Lunzer and Amelia McNamara

Gather your data

A histogram is based on a collection of data about a numeric variable. Our first step is to gather some values for that variable. The initial dataset we will consider consists of fuel consumption (in miles per gallon) from a sample of car models available in 1974 (yes, rather out of date). We can visualize the dataset as a pool of items, with each item identified by its value—which in theory lets us "see" all the items, but makes it hard to get the gestalt of the variable. What are some common values? Is there a lot of variation?

Sort into an ordered list

A useful first step towards describing the variable's distribution is to sort the items into a list. Now we can see the maximum value and the minimum value. Beyond that, it is hard to say much about the center, shape, and spread of the distribution. Part of the problem is that the list is completely filled; the space between any two items is the same, no matter how dissimilar their values may be. We need a way to see how the items relate to each other. Are they clustered around a few specific values? Is there one lonely item, with a value far removed from all the others?

Draw the number line

A common convention is to use a number line, on which higher values are displayed 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.

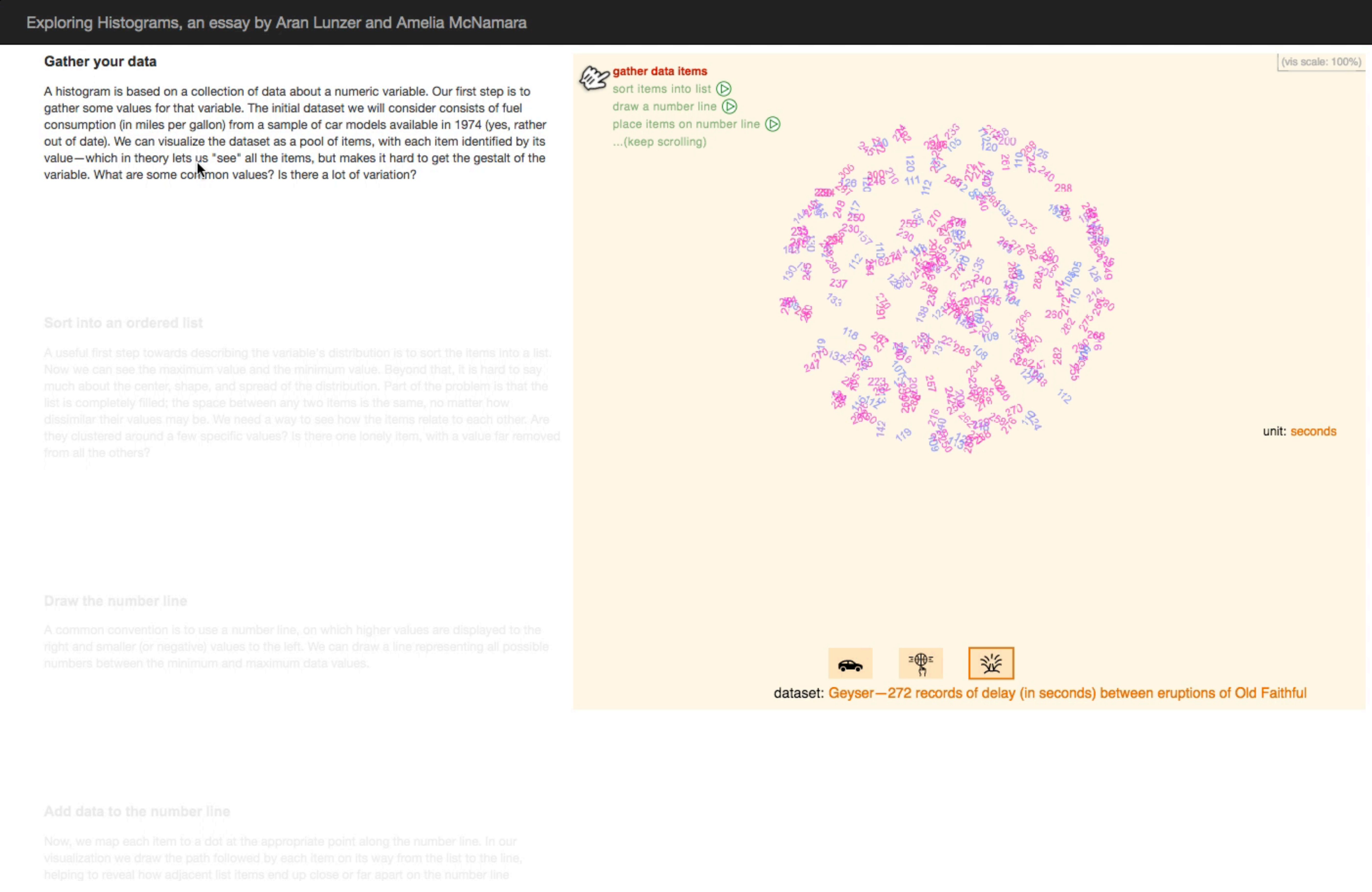
Add data to the number line

Now, we map each item to a dot at the appropriate point along the number line. In our visualization we draw the path followed by each item on its way from the list to the line, helping to reveal how adjacent list items end up close or far apart on the number line

unit: seconds

dataset: Geyser—272 records of delay (in seconds) between eruptions of Old Faithful

Lunzer and McNamara, tinlizzie.org/histograms



Exploring Histograms, an essay by Aran Lunzer and Amelia McNamara

Gather your data

A histogram is based on a collection of data about a numeric variable. Our first step is to gather some values for that variable. The initial dataset we will consider consists of fuel consumption (in miles per gallon) from a sample of car models available in 1974 (yes, rather out of date). We can visualize the dataset as a pool of items, with each item identified by its value—which in theory lets us "see" all the items, but makes it hard to get the gestalt of the variable. What are some common values? Is there a lot of variation?

unit: seconds

dataset: Geyser—272 records of delay (in seconds) between eruptions of Old Faithful

Sort into an ordered list

A useful first step towards describing the variable's distribution is to sort the items into a list. Now we can see the maximum value and the minimum value. Beyond that, it is hard to say much about the center, shape, and spread of the distribution. Part of the problem is that the list is completely filled; the space between any two items is the same, no matter how dissimilar their values may be. We need a way to see how the items relate to each other. Are they clustered around a few specific values? Is there one lonely item, with a value far removed from all the others?

Draw the number line

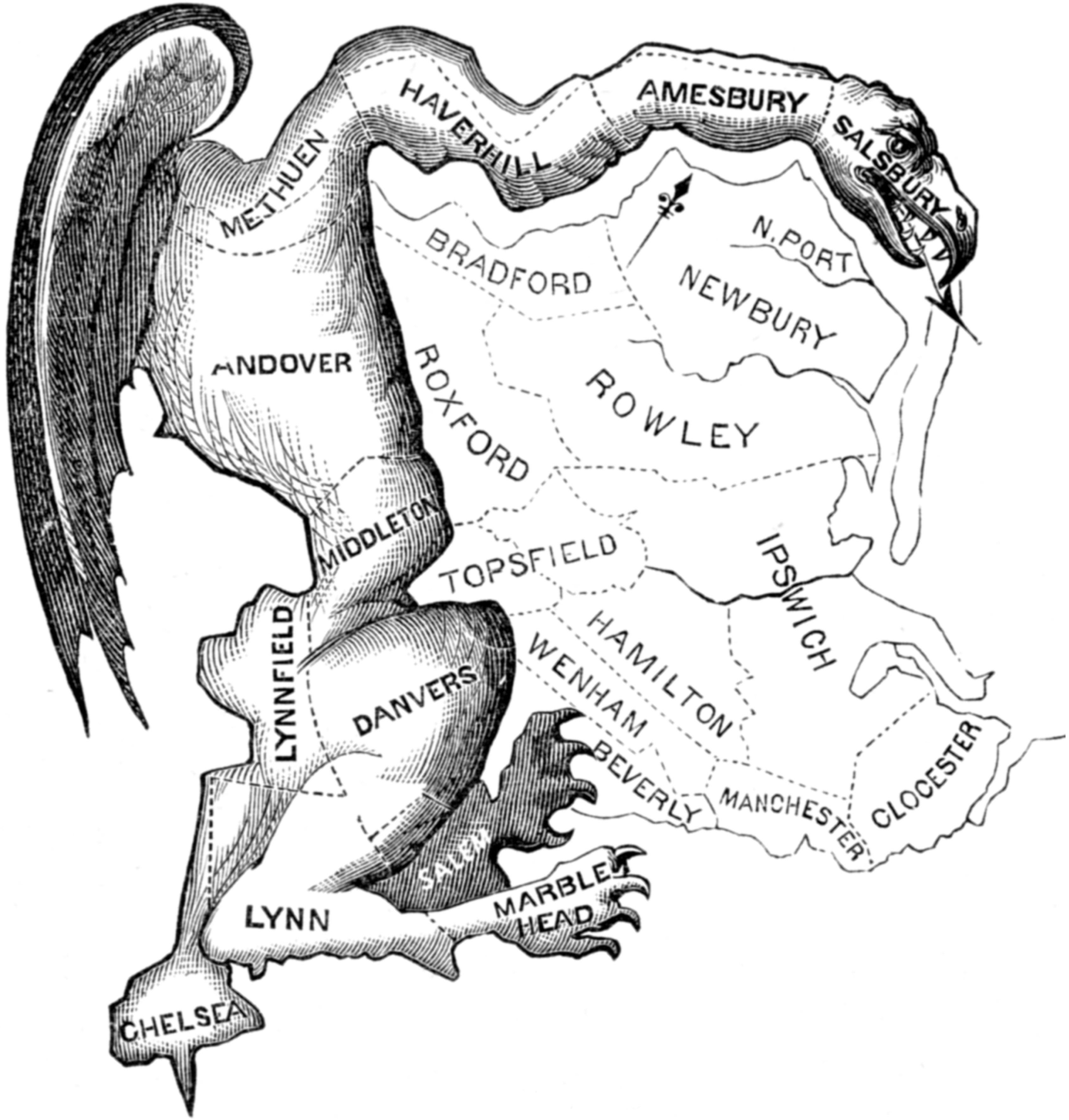
A common convention is to use a number line, on which higher values are displayed to the right and smaller (or negative) values to the left. We can draw a line representing all possible numbers between the minimum and maximum data values.

Add data to the number line

Now, we map each item to a dot at the appropriate point along the number line. In our visualization we draw the path followed by each item on its way from the list to the line, helping to reveal how adjacent list items end up close or far apart on the number line

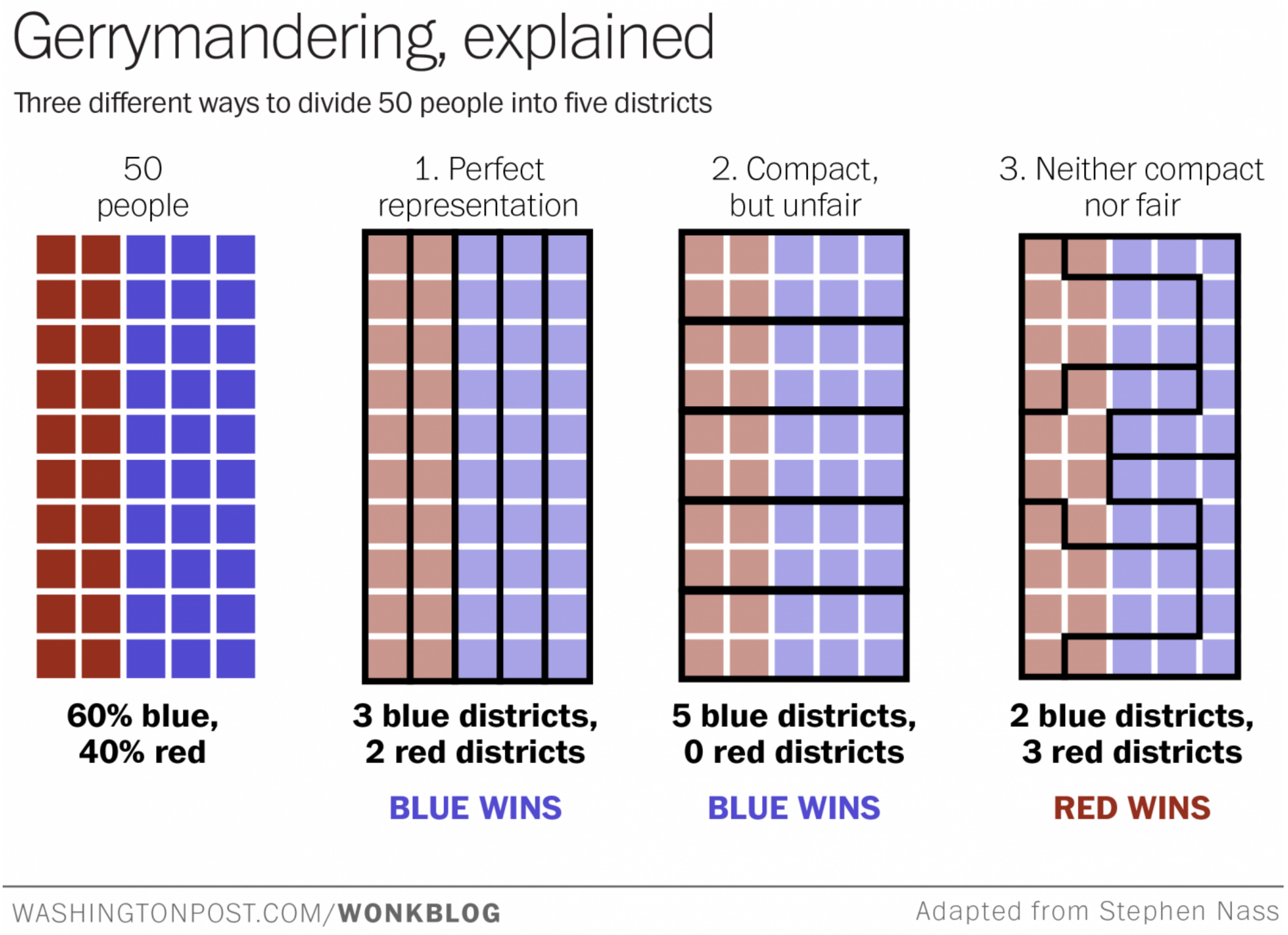
Lunzer and McNamara, tinlizzie.org/histograms

Gerrymandering



wikipedia: Elkanah Tisdale (1771-1835) (often falsely attributed to Gilbert Stuart)[1] - Originally published in the Boston Centinel, 1812.

http://bit.ly/LWT_gerrymandering



Gerrymandering, explained

Three different ways to divide 50 people into five districts

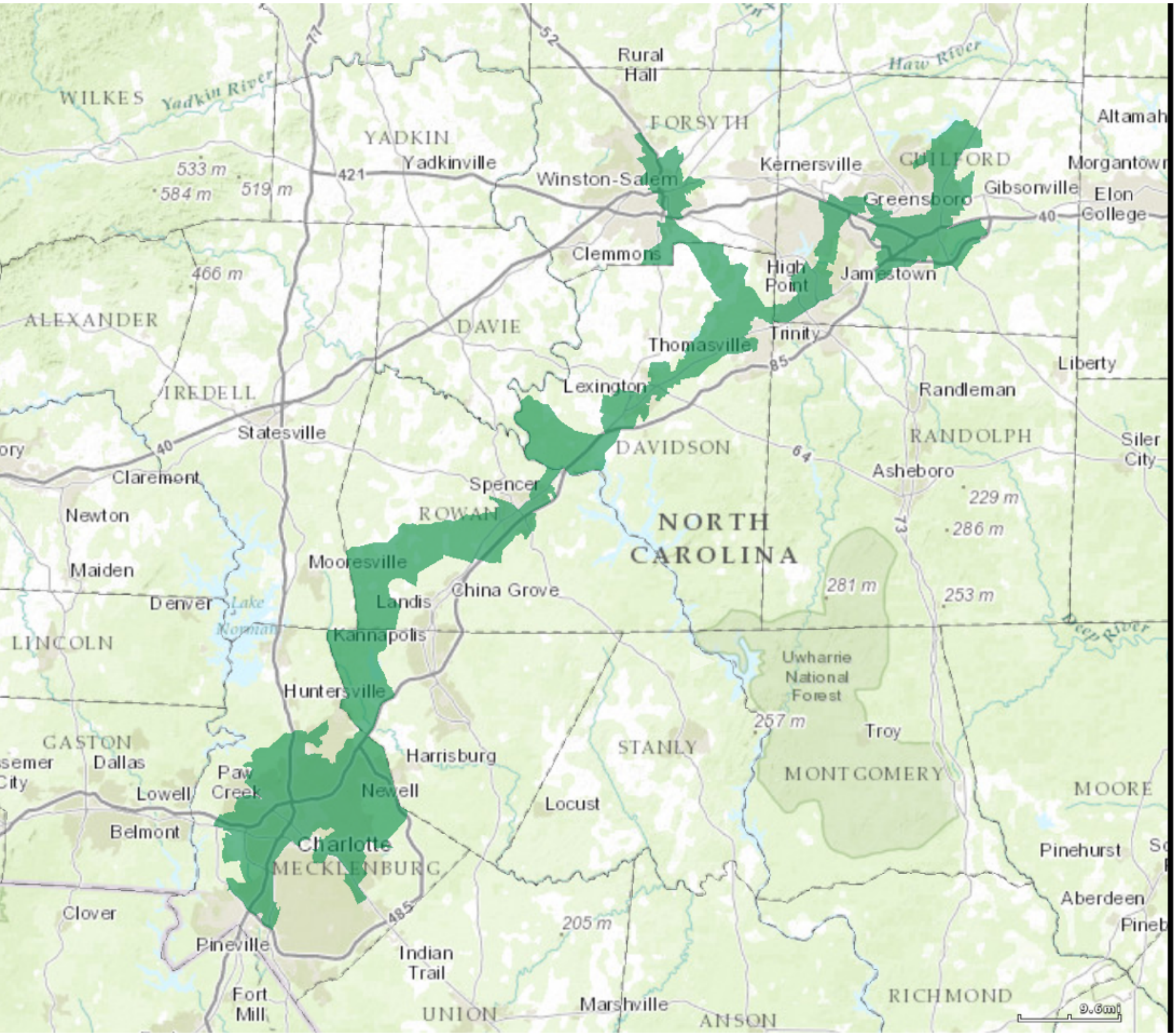
50 people	1. Perfect representation	2. Compact, but unfair	3. Neither compact nor fair
60% blue, 40% red	**3 blue districts, 2 red districts**	**5 blue districts, 0 red districts**	**2 blue districts, 3 red districts**
	BLUE WINS	**BLUE WINS**	**RED WINS**

WASHINGTONPOST.COM/**WONKBLOG**

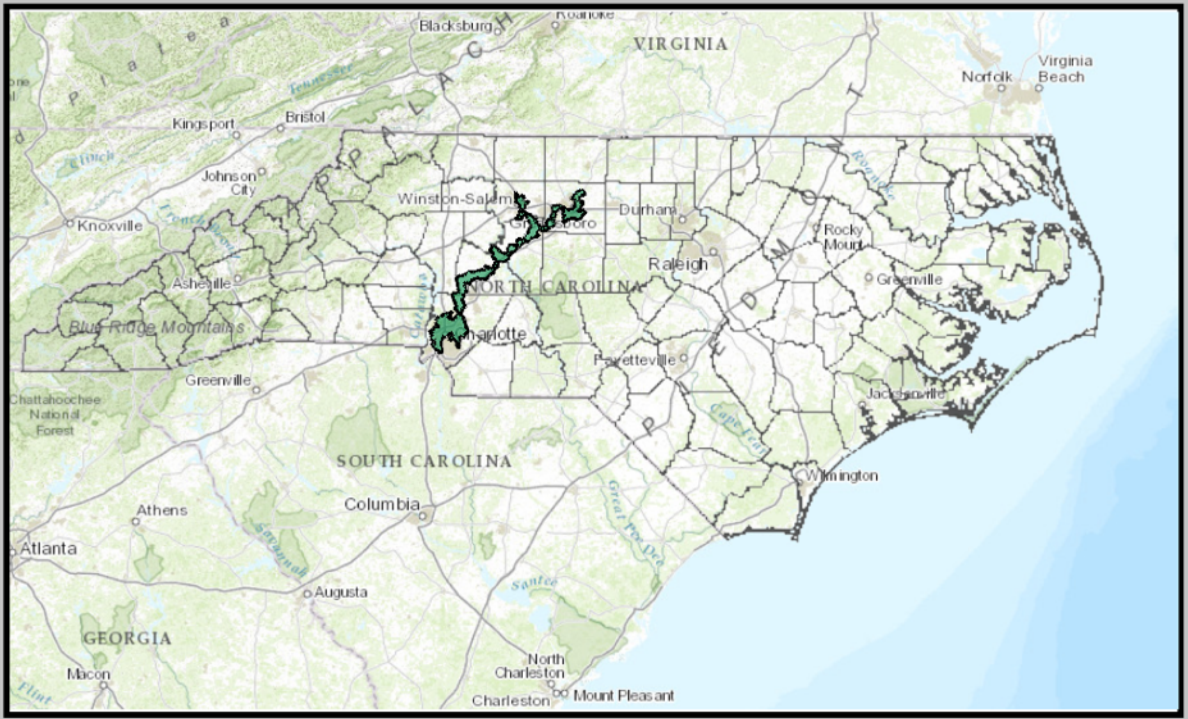
Adapted from Stephen Nass

https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see

North Carolina’s 12th district



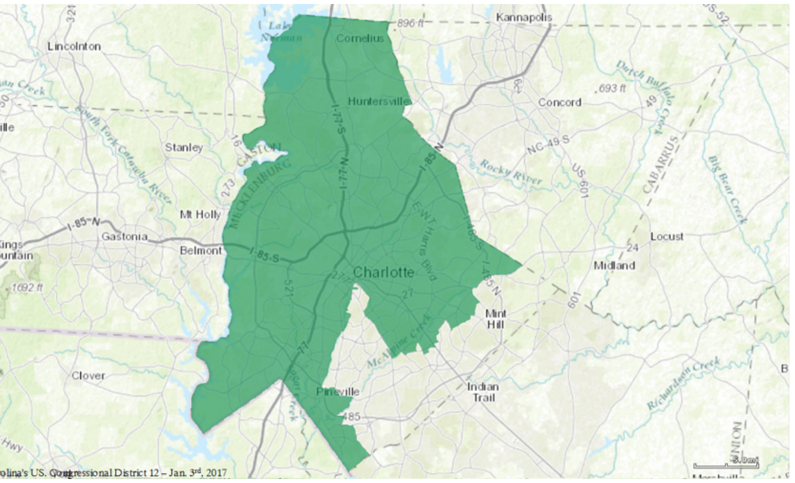
North Carolina US District 12

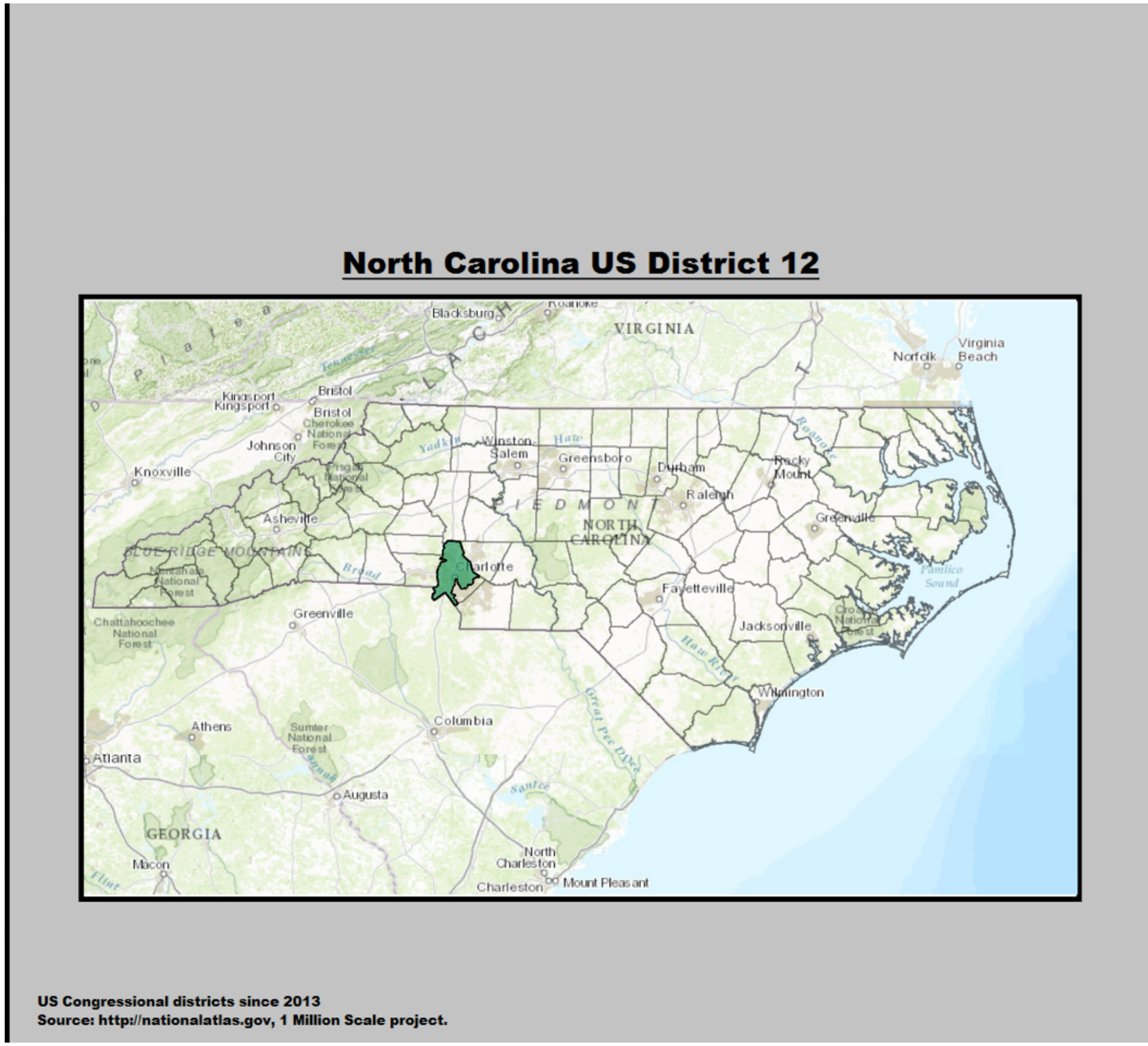


US Congressional districts since 2013
Source: http://nationalatlas.gov, 1 Million Scale project.

https://en.wikipedia.org/wiki/North_Carolina%27s_12th_congressional_district

North Carolina’s 12th district

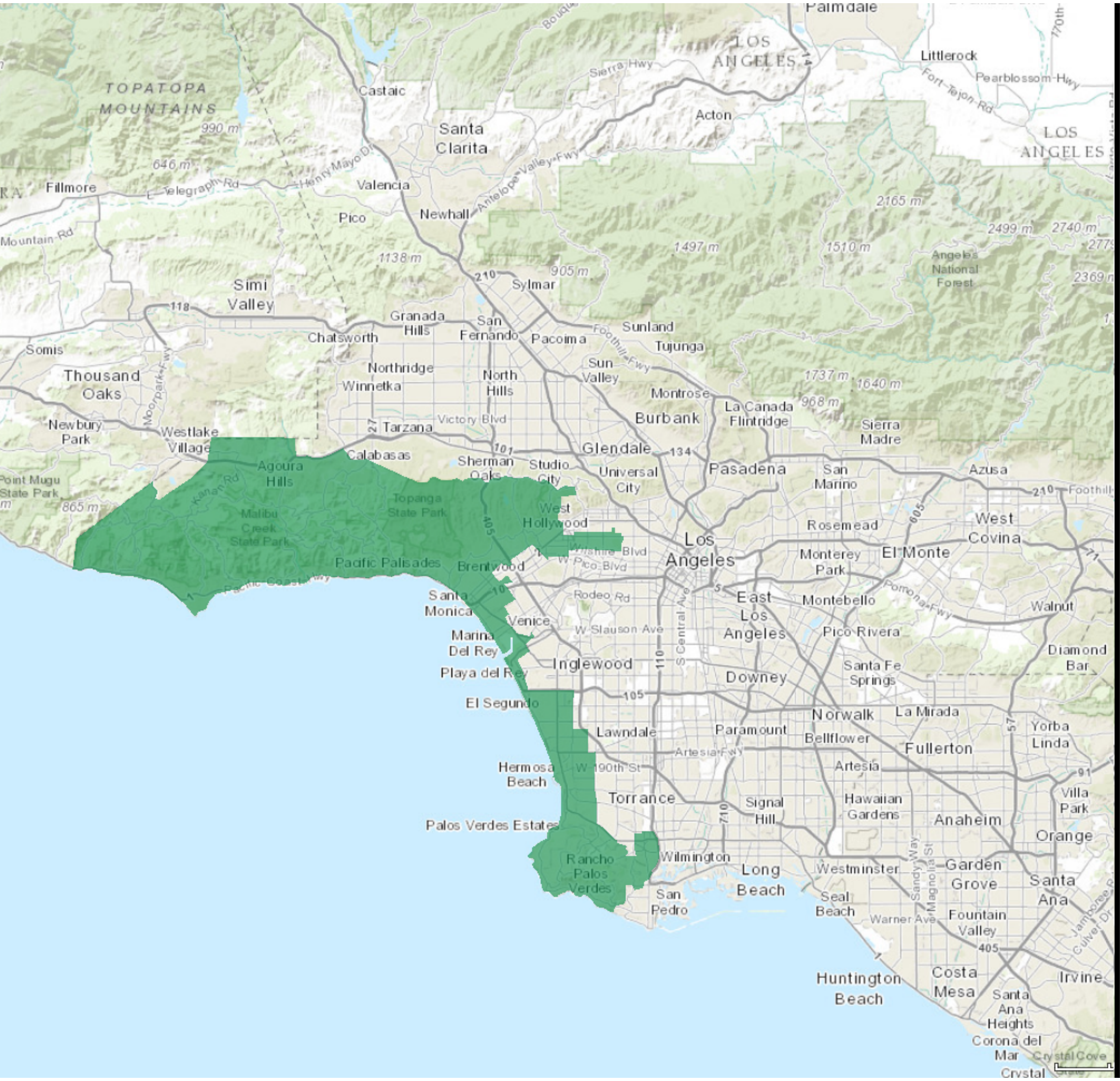



North Carolina US District 12

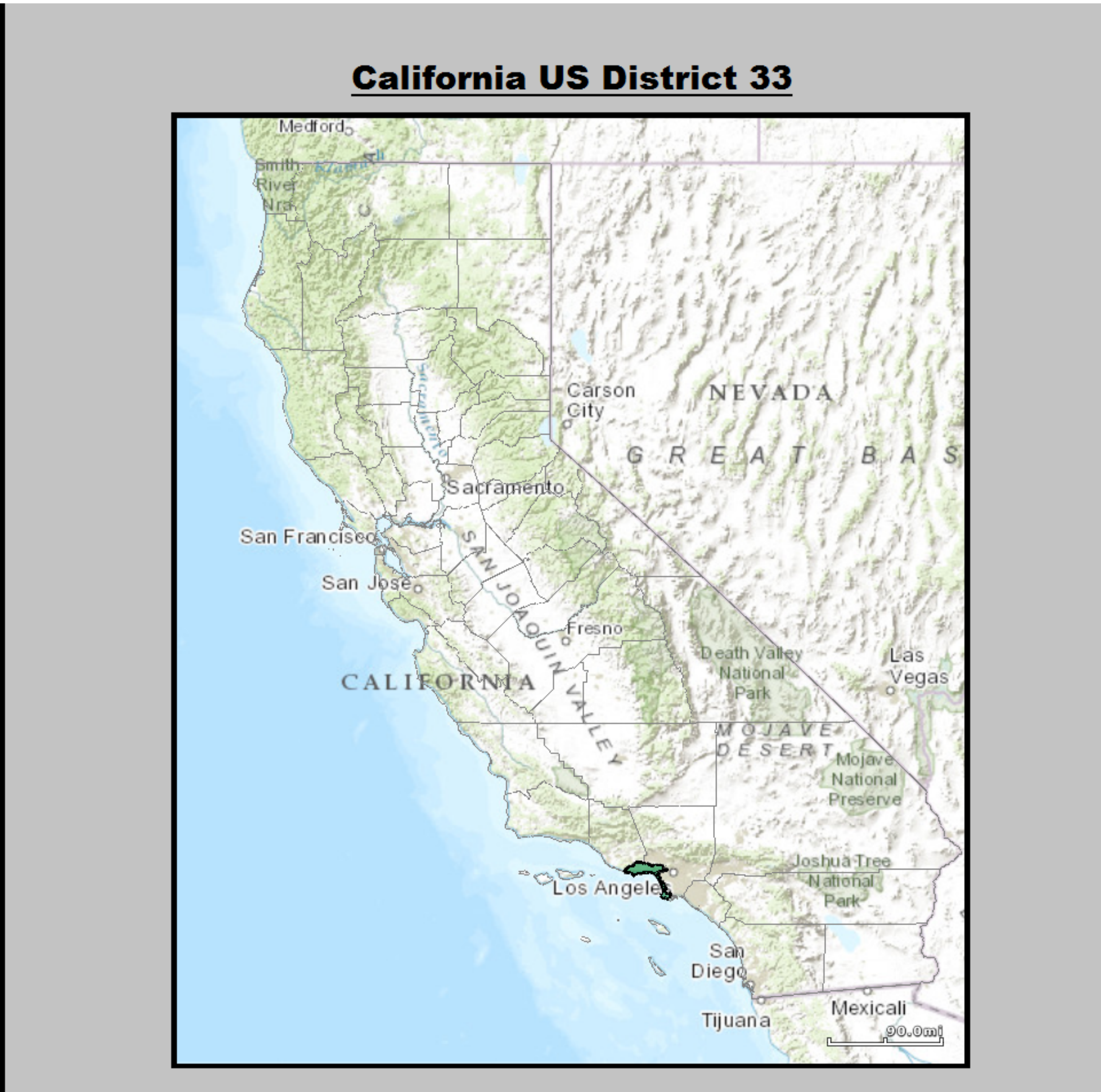
US Congressional districts since 2013
Source: http://nationalatlas.gov, 1 Million Scale project.

https://en.wikipedia.org/wiki/North_Carolina%27s_12th_congressional_district

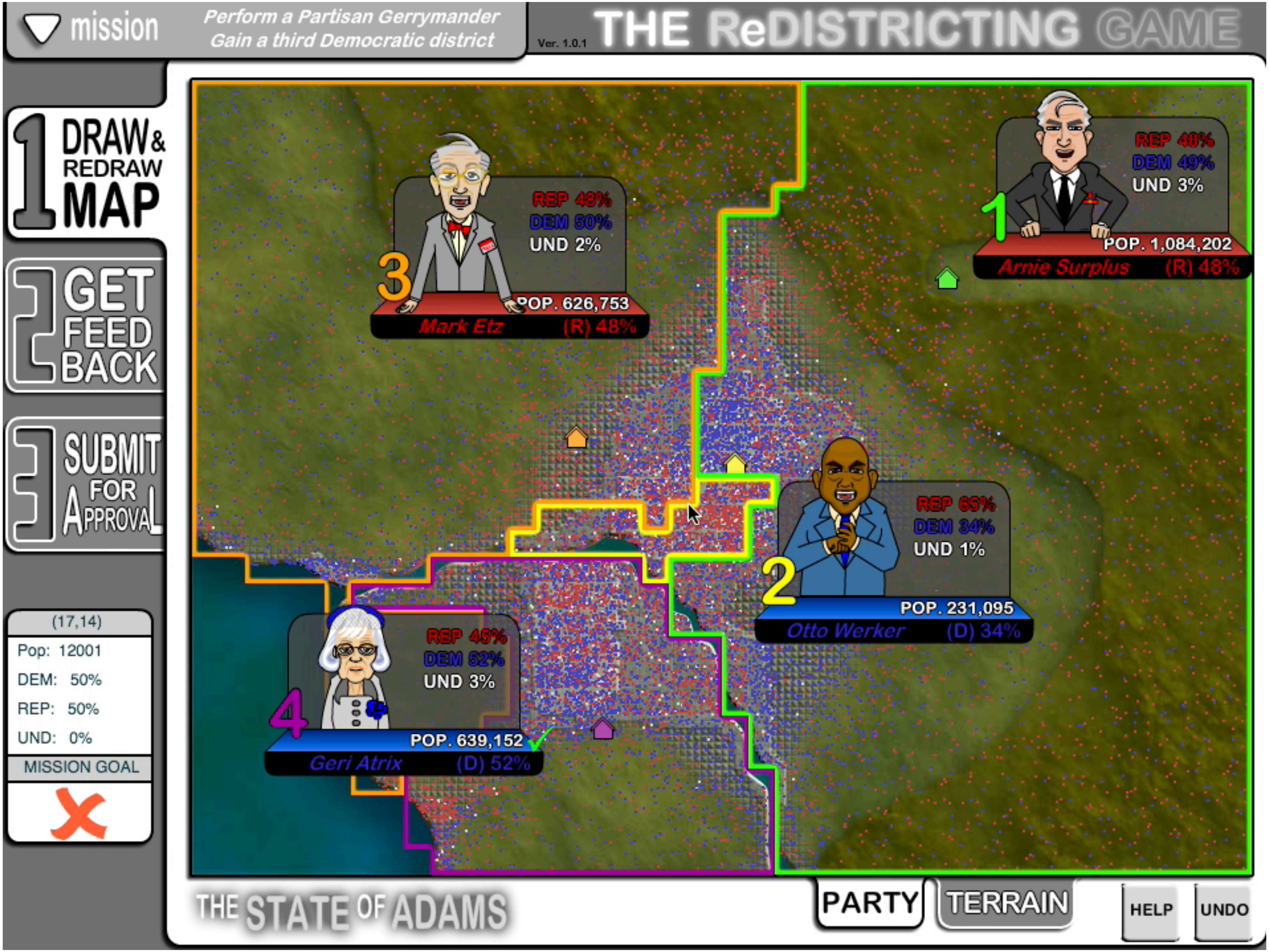
California’s 33rd district



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 — Arnie Surplus (R) 48% — REP 48% DEM 49% UND 3% — POP. 1,084,202

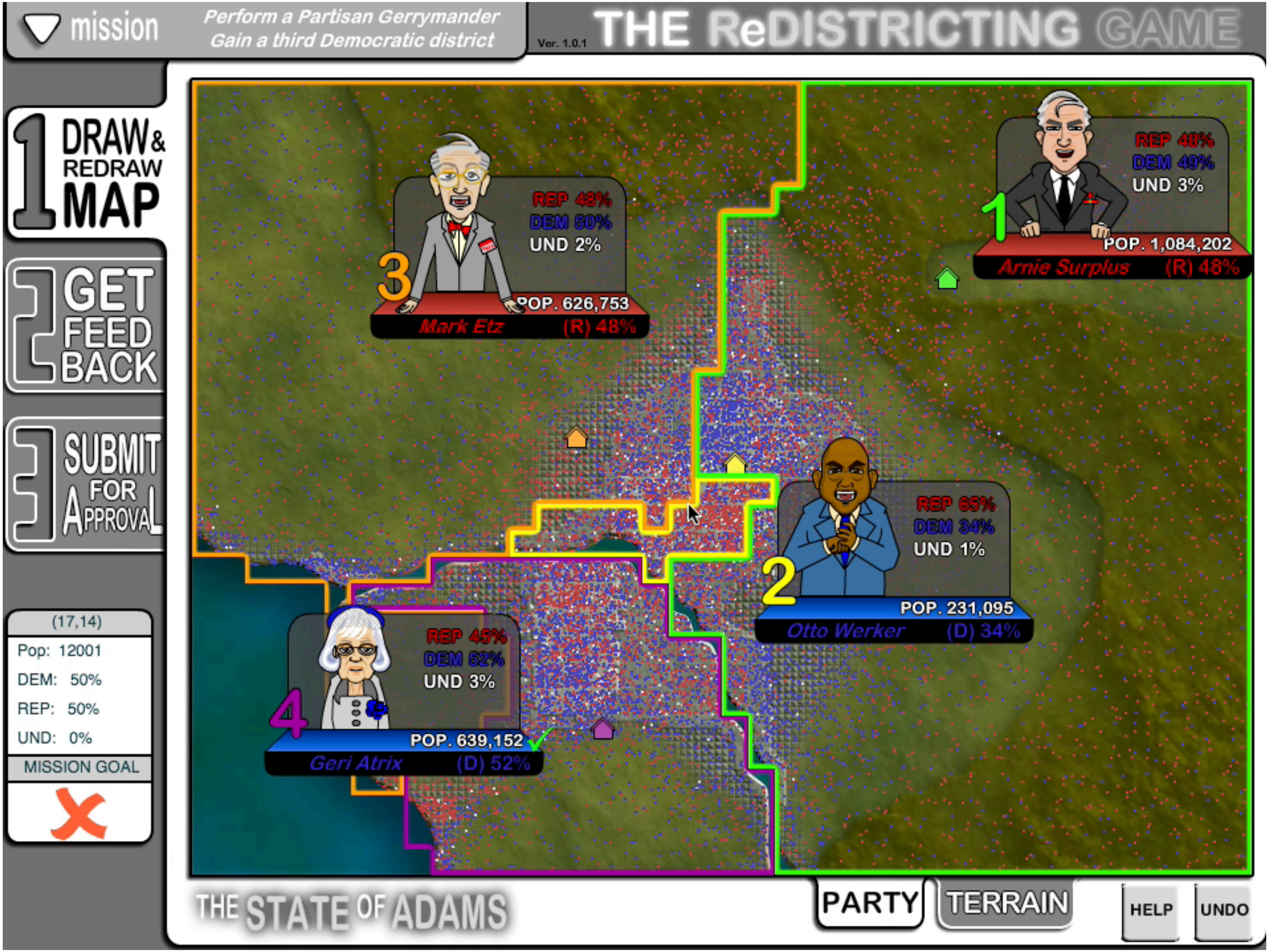
2 — Otto Werker (D) 34% — REP 65% DEM 34% UND 1% — POP. 231,095

3 — Mark Etz (R) 48% — REP 48% DEM 50% UND 2% — POP. 626,753

4 — Geri Atrix (D) 52% ✓ — REP 45% DEM 52% UND 3% — POP. 639,152

(17,14)
Pop: 12001
DEM: 50%
REP: 50%
UND: 0%
MISSION GOAL
✗

http://redistrictinggame.org/



Perform a Partisan Gerrymander
Gain a third Democratic district

(17,14)
Pop: 12001
DEM: 50%
REP: 50%
UND: 0%
MISSION GOAL
✗

1
REP 48%
DEM 49%
UND 3%
POP. 1,084,202
Arnie Surplus (R) 48%

2
REP 65%
DEM 34%
UND 1%
POP. 231,095
Otto Werker (D) 34%

3
REP 48%
DEM 50%
UND 2%
POP. 626,753
Mark Etz (R) 48%

4
REP 45%
DEM 52%
UND 3%
POP. 639,152 ✓
Geri Atrix (D) 52%

http://redistrictinggame.org/

Data CC-By-SA by OpenStreetMap

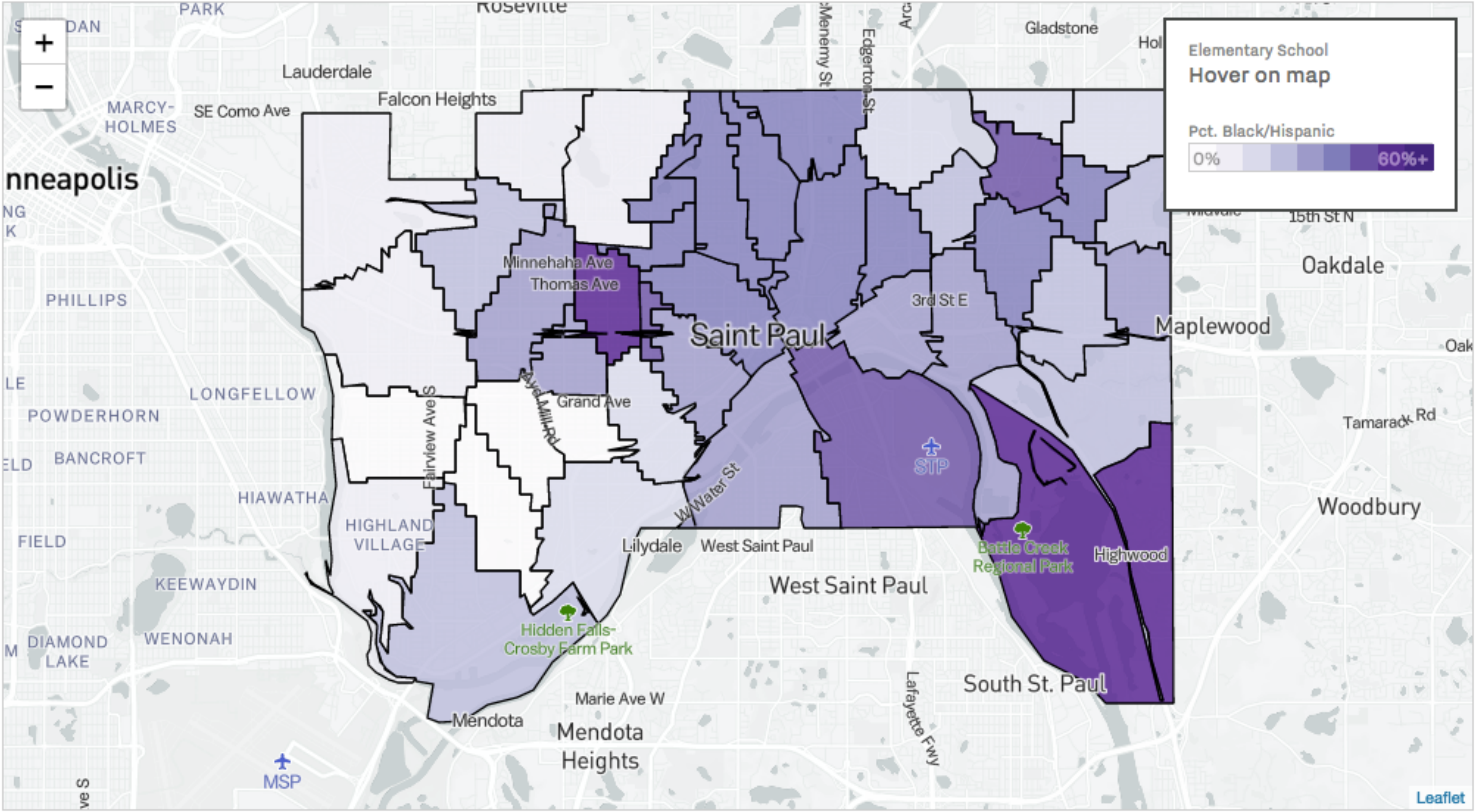
Toward a Talismanic Redistricting Tool: A Computational Method for Identifying Extreme Redistricting Plans.

Wendy Tam Cho and Yan Liu

http://bit.ly/TalismanicMaps

flickr: mulad

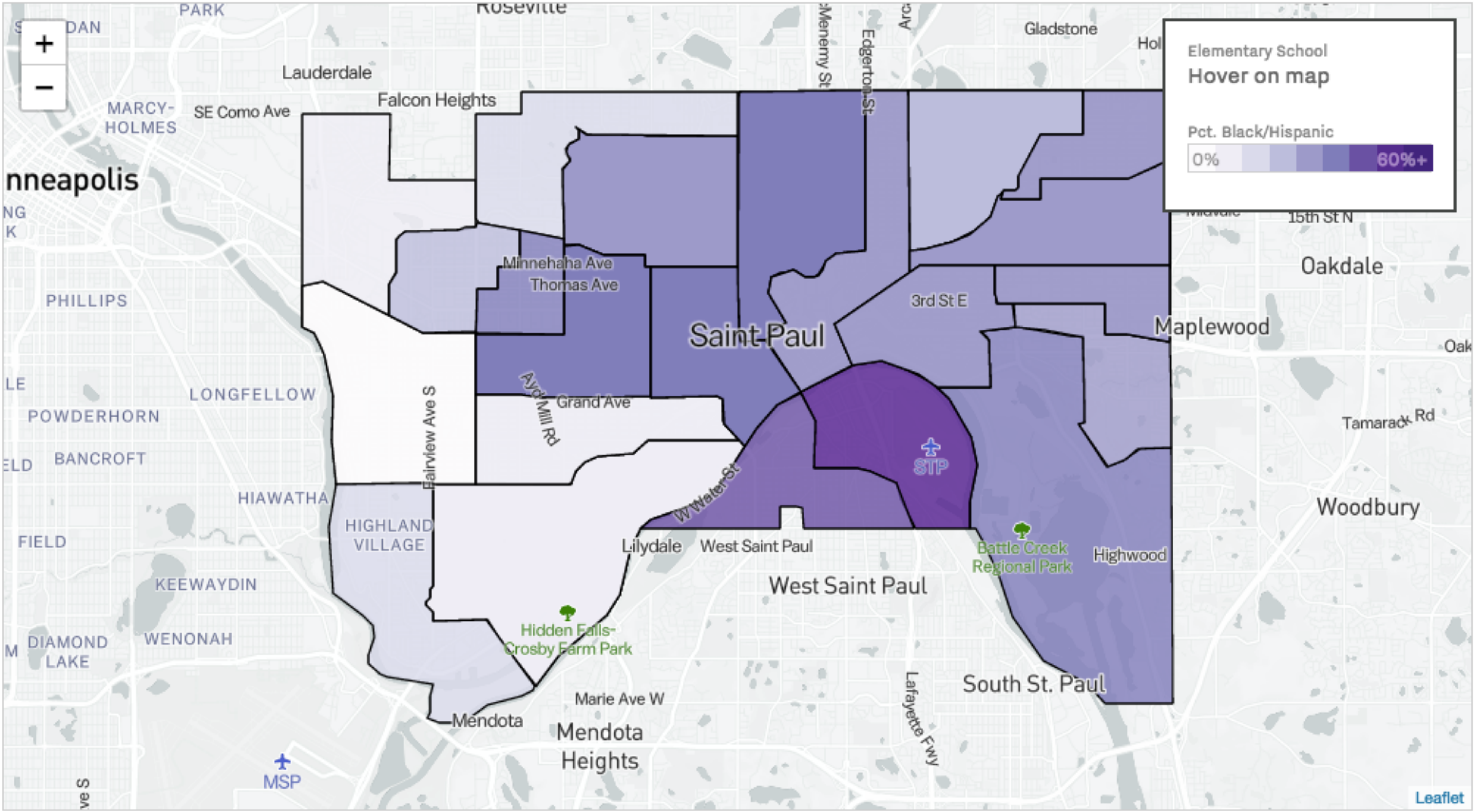
Gerrymandering school districts



Data from research by Tomas E. Monarrez, an economics PhD candidate at the University of California, Berkeley

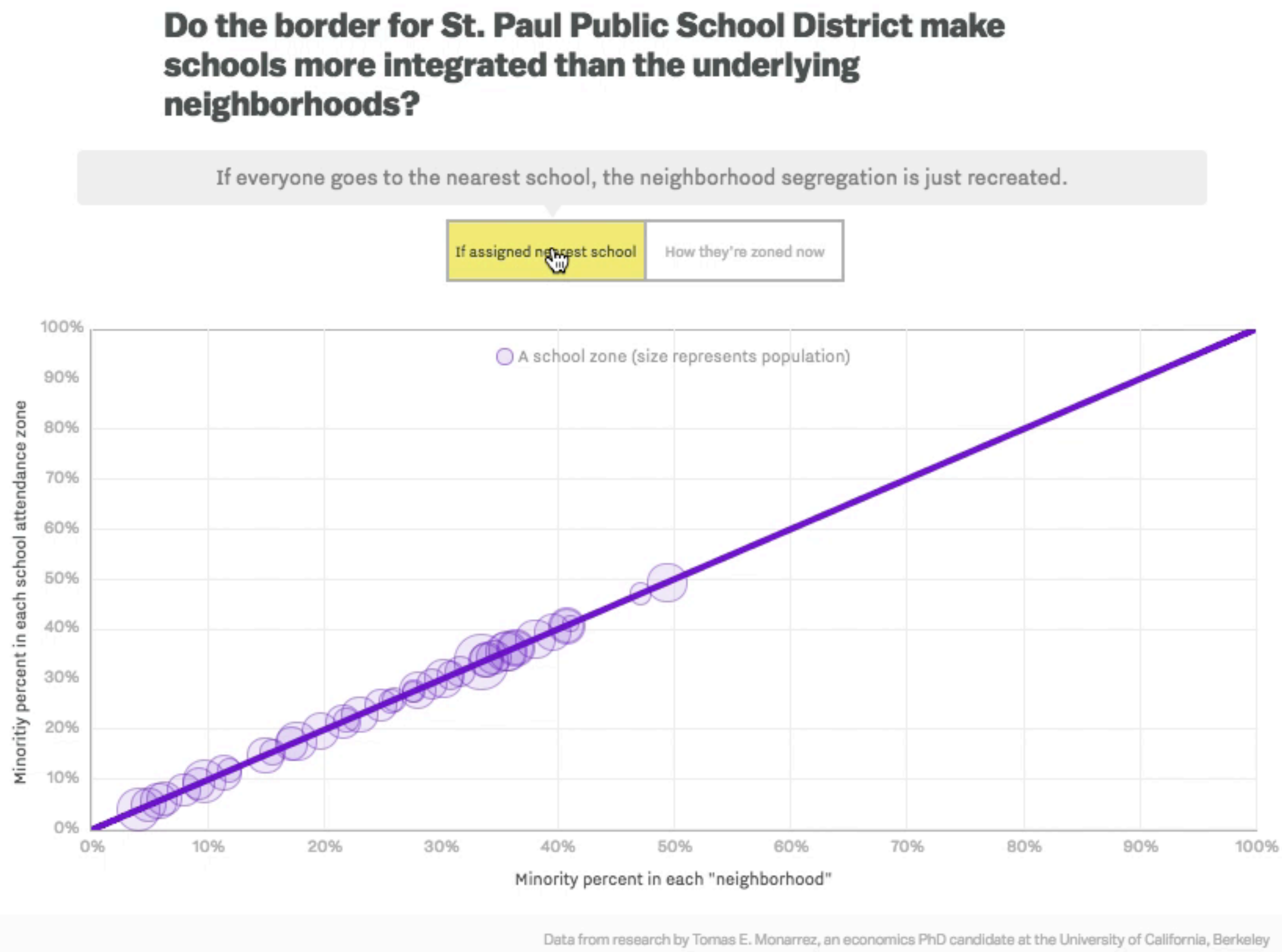
https://www.vox.com/2018/1/8/16822374/school-segregation-gerrymander-map

Gerrymandering school districts



Data from research by Tomas E. Monarrez, an economics PhD candidate at the University of California, Berkeley

https://www.vox.com/2018/1/8/16822374/school-segregation-gerrymander-map



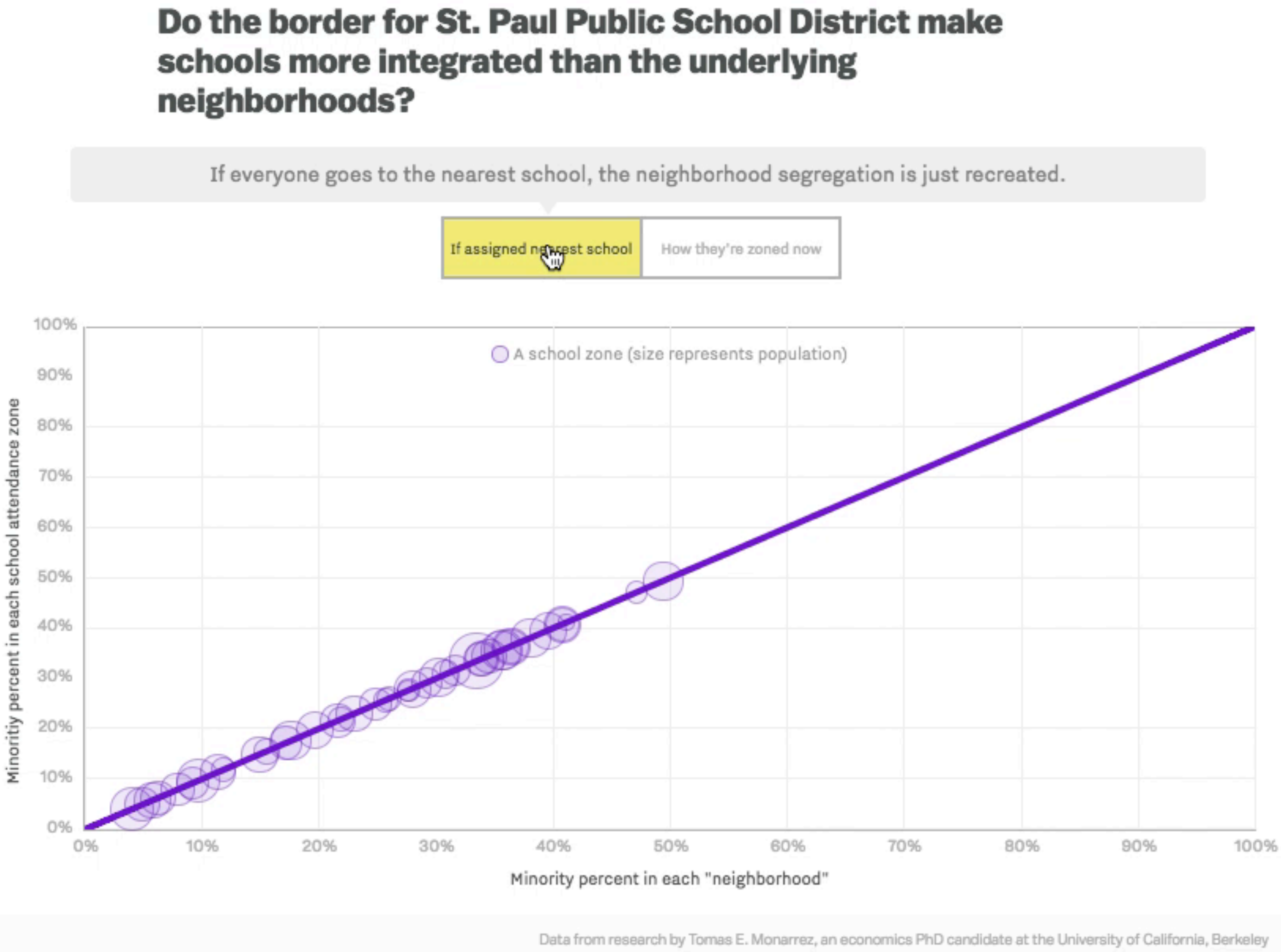
Do the border for St. Paul Public School District make schools more integrated than the underlying neighborhoods?

If everyone goes to the nearest school, the neighborhood segregation is just recreated.

If assigned nearest school
How they're zoned now
A school zone (size represents population)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Minoritiy percent in each school attendance zone
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Minority percent in each "neighborhood"

Data from research by Tomas E. Monarrez, an economics PhD candidate at the University of California, Berkeley

https://www.vox.com/2018/1/8/16822374/school-segregation-gerrymander-map



Do the border for St. Paul Public School District make schools more integrated than the underlying neighborhoods?

If everyone goes to the nearest school, the neighborhood segregation is just recreated.

If assigned nearest school
How they're zoned now
A school zone (size represents population)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Minoritiy percent in each school attendance zone
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Minority percent in each "neighborhood"

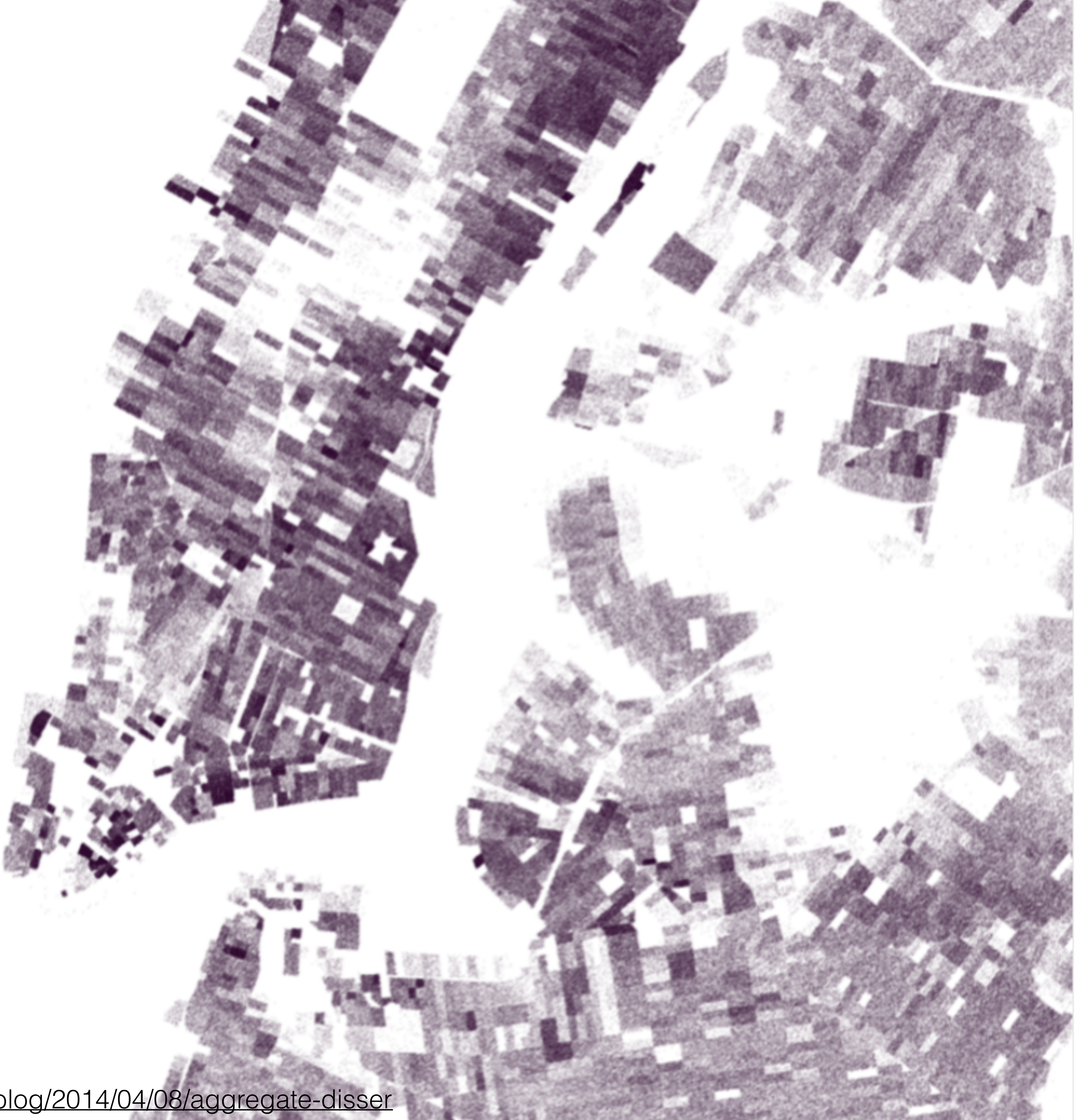
Data from research by Tomas E. Monarrez, an economics PhD candidate at the University of California, Berkeley

https://www.vox.com/2018/1/8/16822374/school-segregation-gerrymander-map

down-scaling

flickr: MattHagedorn

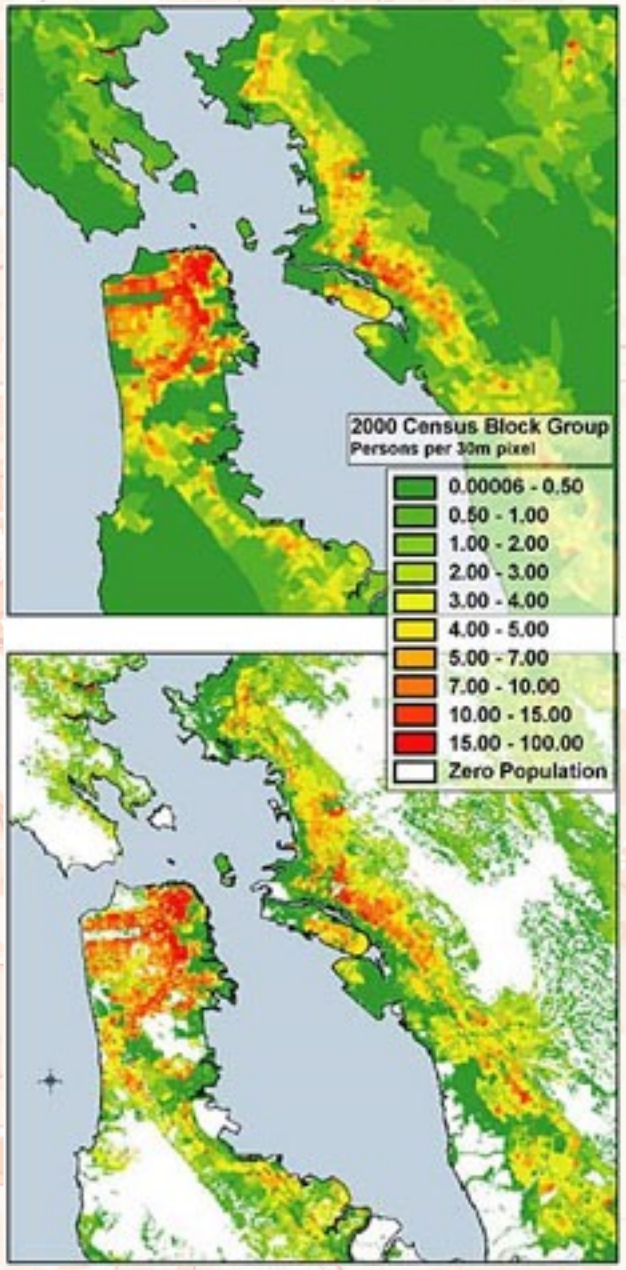
http://conveyal.com/blog/2014/04/08/aggregate-disser



http://conveyal.com/blog/2014/04/08/aggregate-disser

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



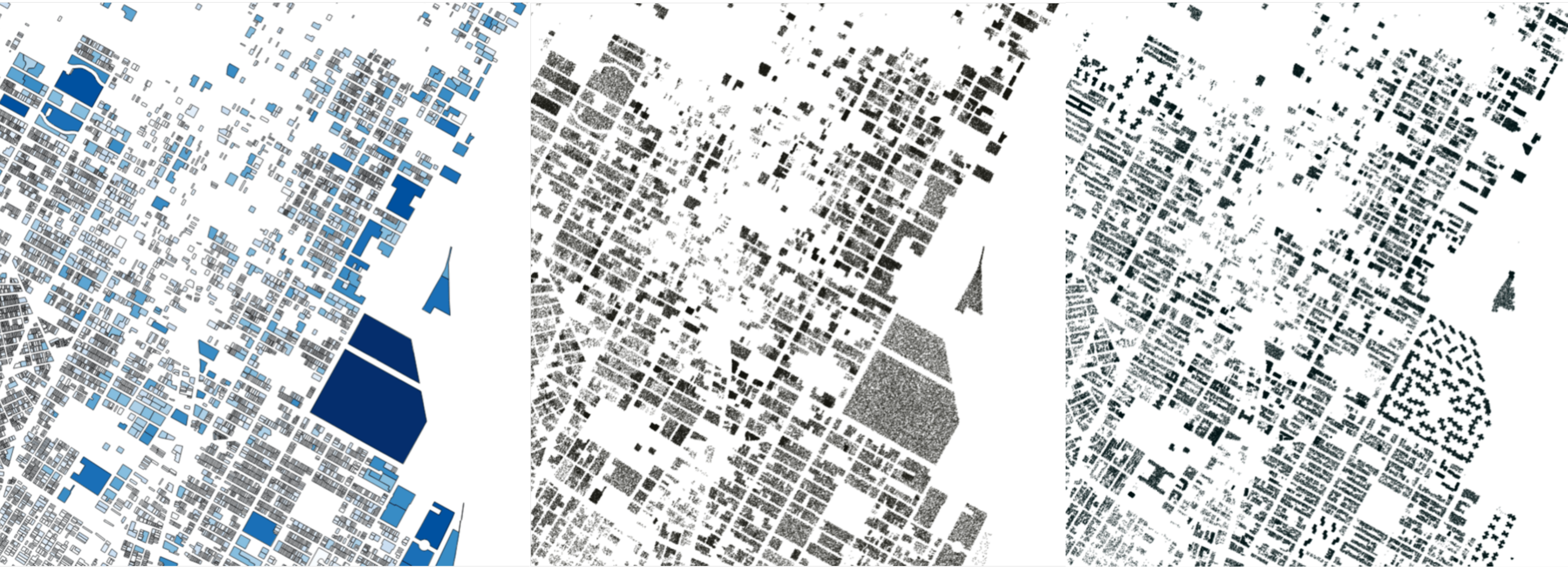
wikipedia: dasymetric map

flickr: mulad

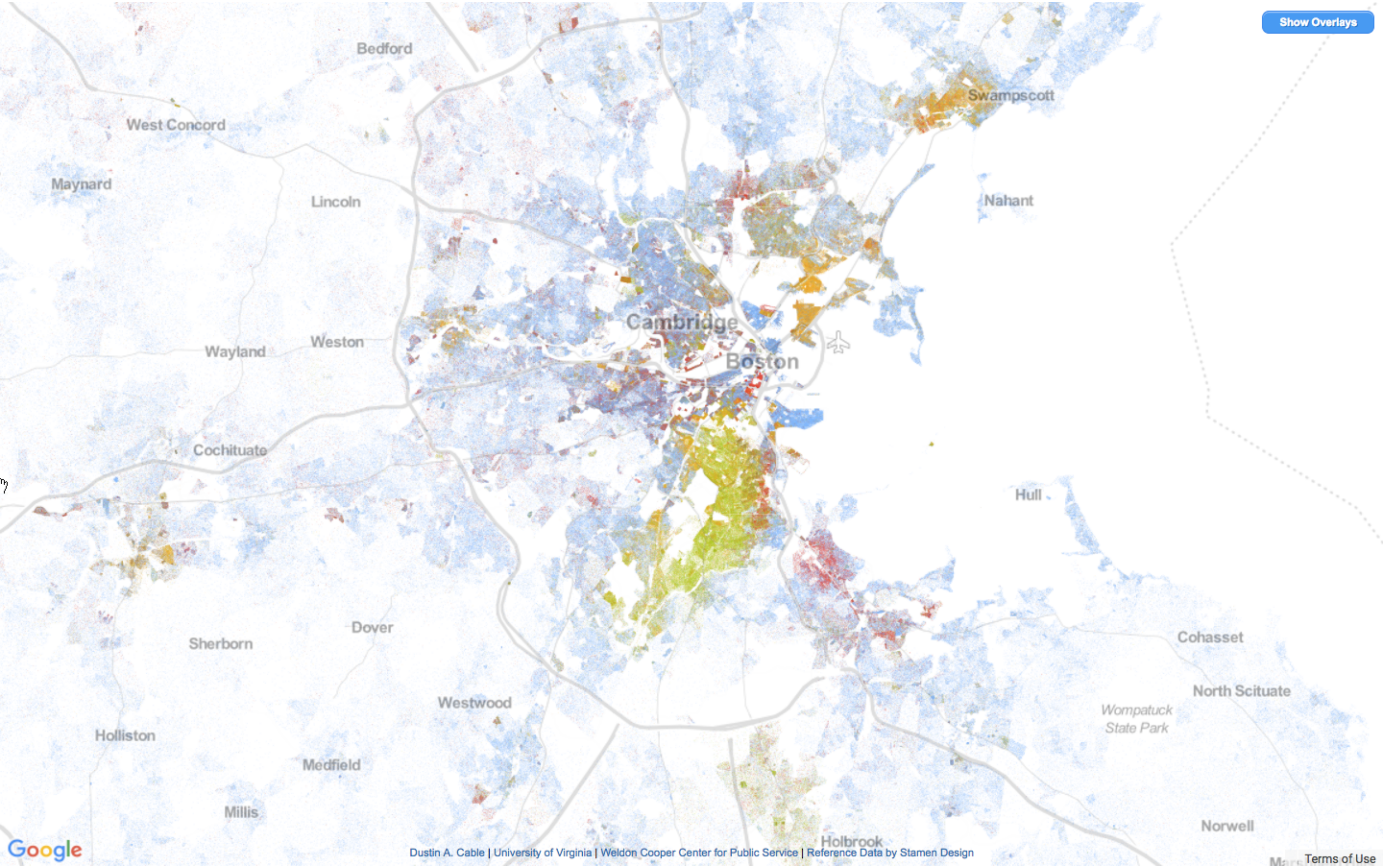
http://conveyal.com/blog/2014/04/08/aggregate-disser

http://conveyal.com/blog/2014/04/08/aggregate-disser

http://conveyal.com/blog/2014/04/08/aggregate-disser



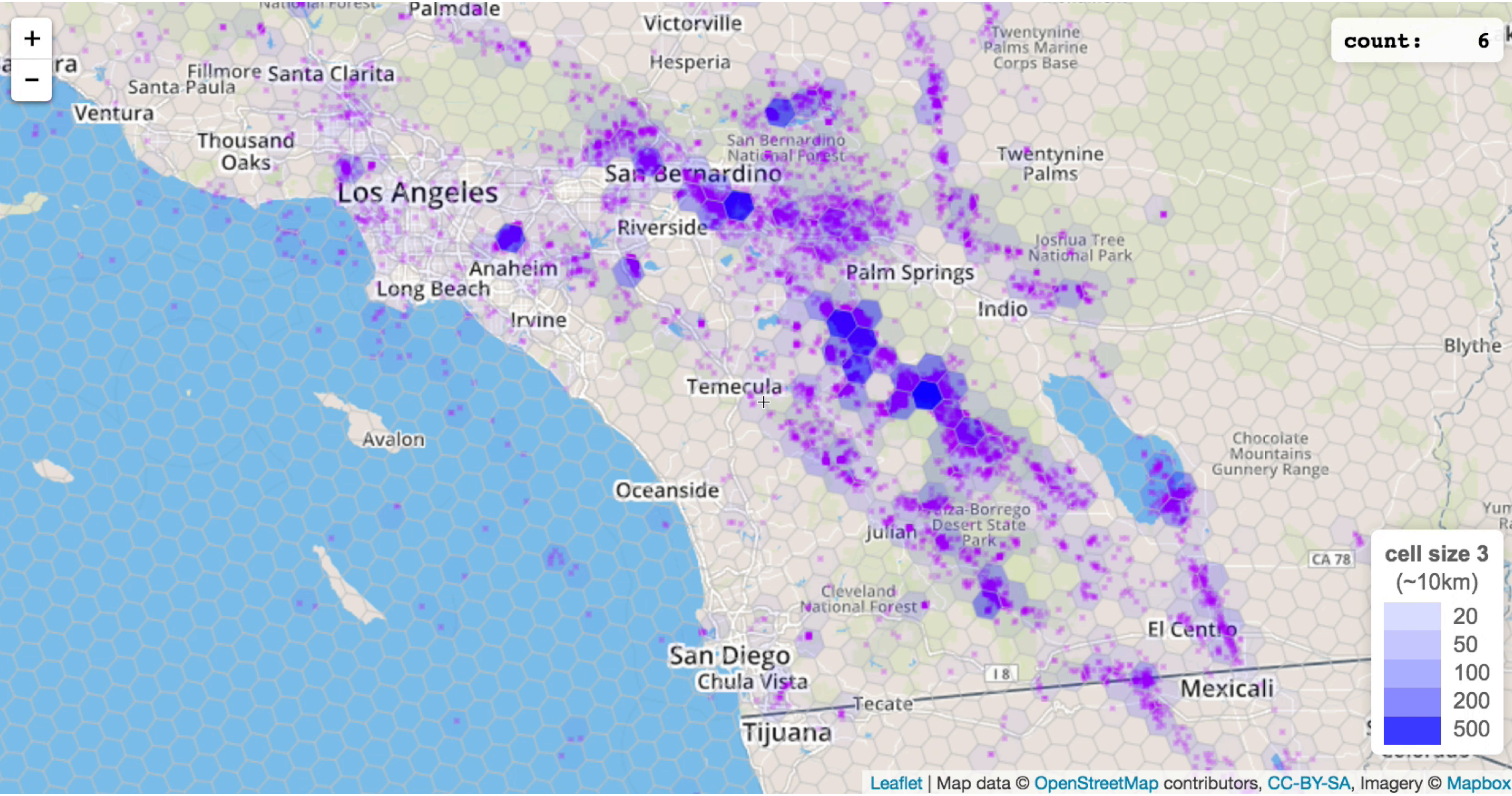
http://conveyal.com/blog/2014/04/08/aggregate-disser



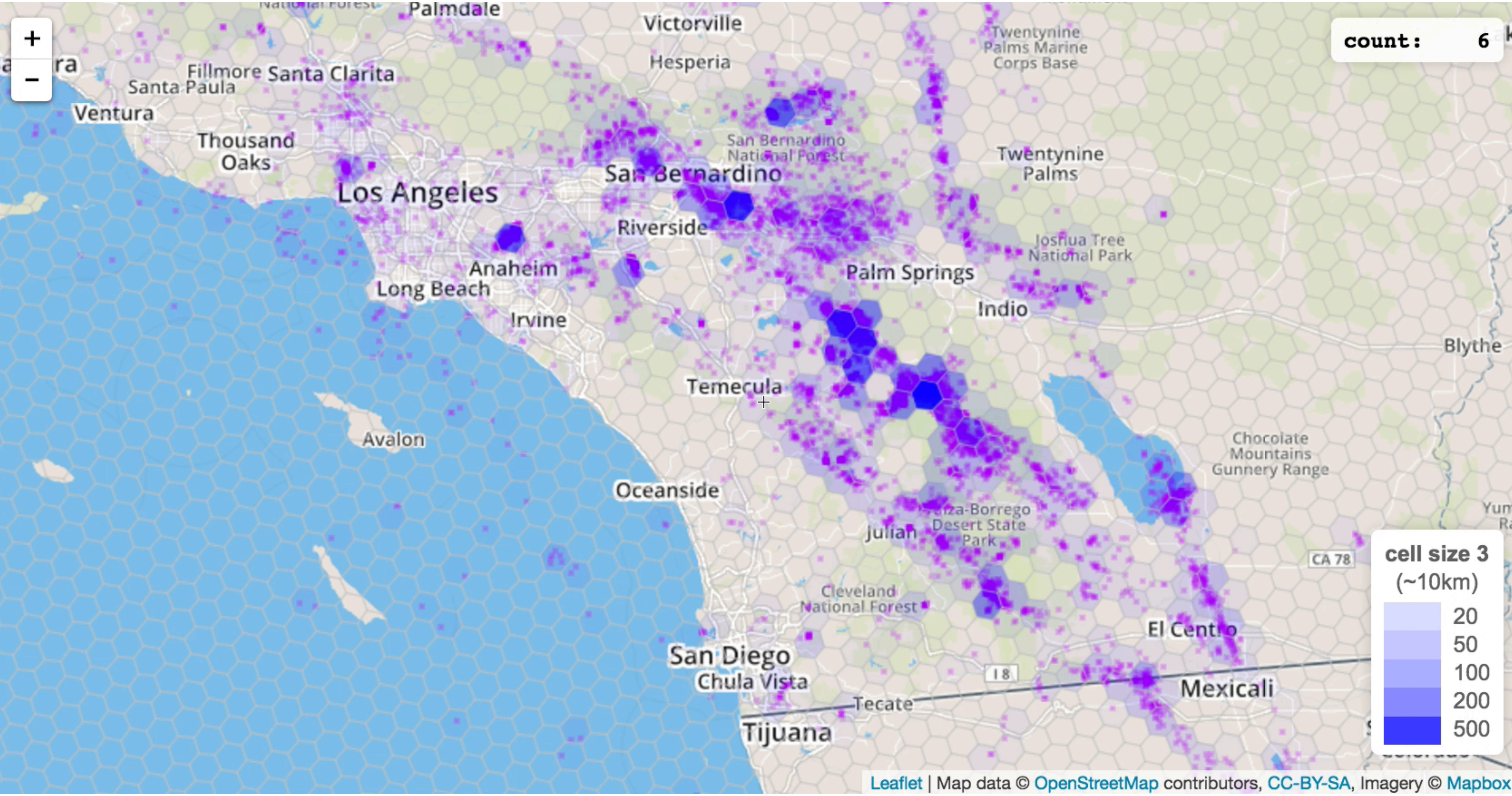
http://bit.ly/CensusDotMap

side-scaling (the hardest)

flickr: MattHagedorn



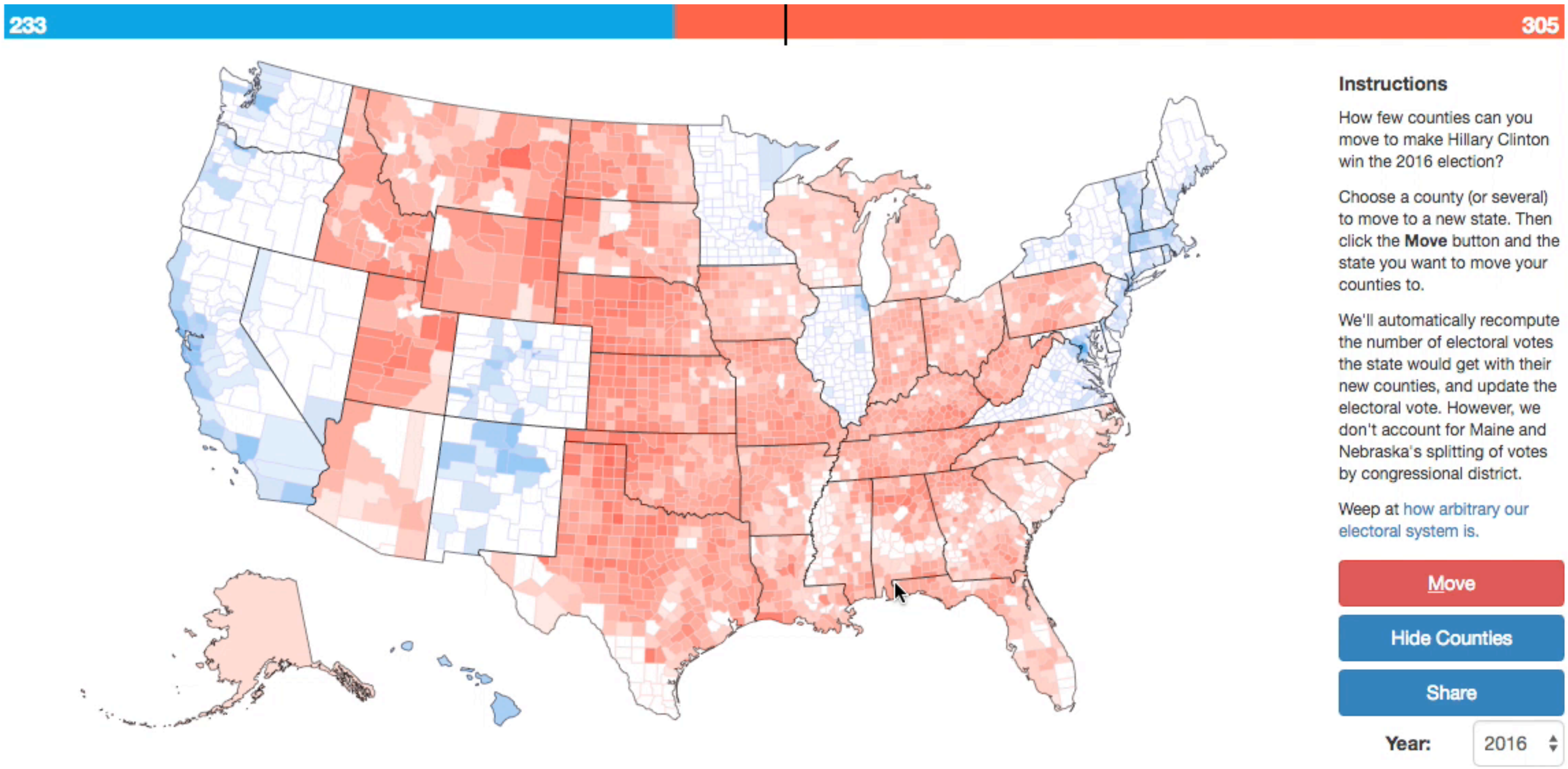
Lunzer and McNamara, www.bit.ly/spatial_agg



Lunzer and McNamara, www.bit.ly/spatial_agg

with nested polygons, not so bad

flickr: MattHagedorn



Instructions

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

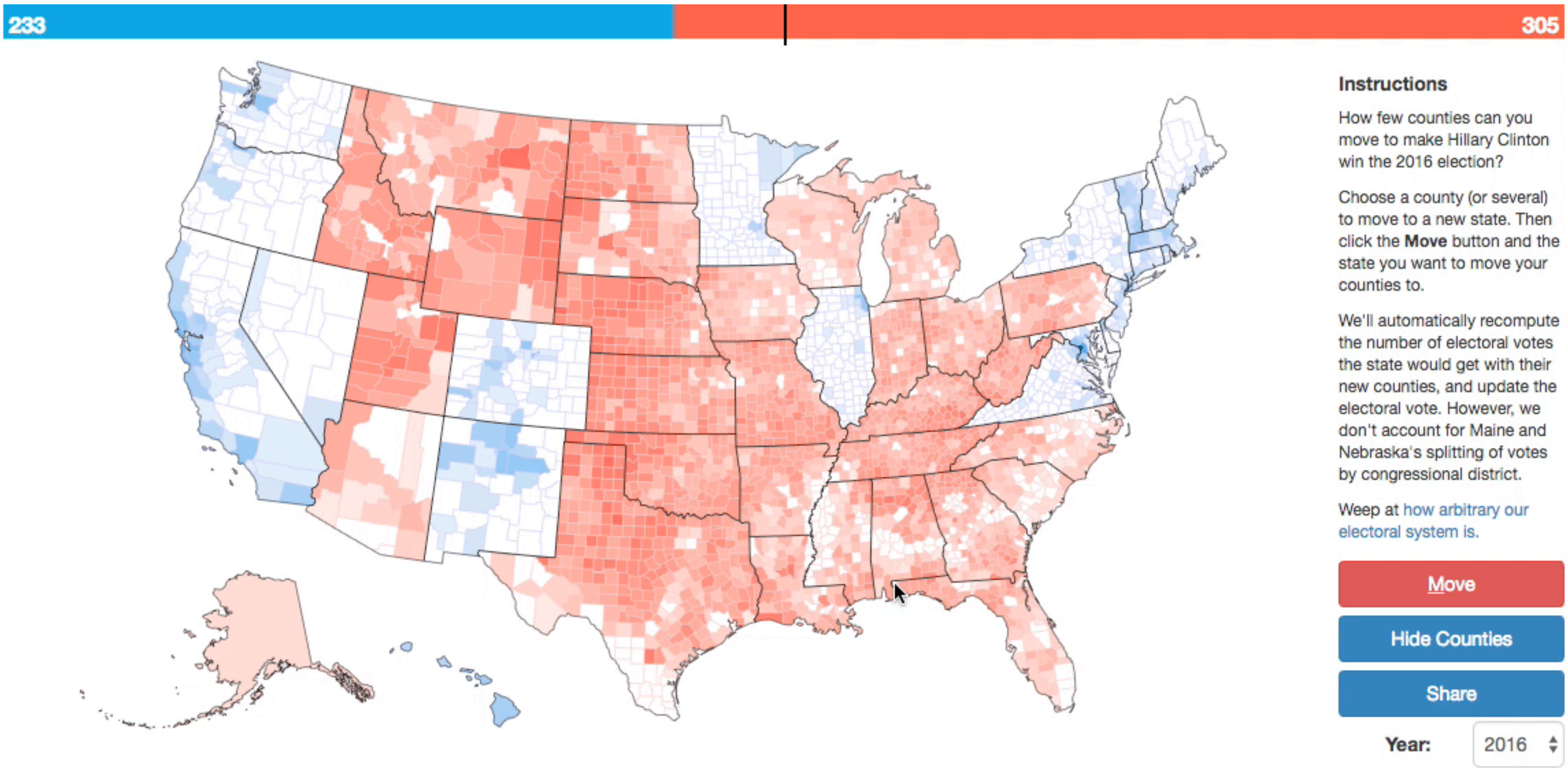
Choose a county (or several) to move to a new state. Then click the **Move** button and the state you want to move your counties to.

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.

Year: 2016

http://kevinhayeswilson.com/redraw/



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.

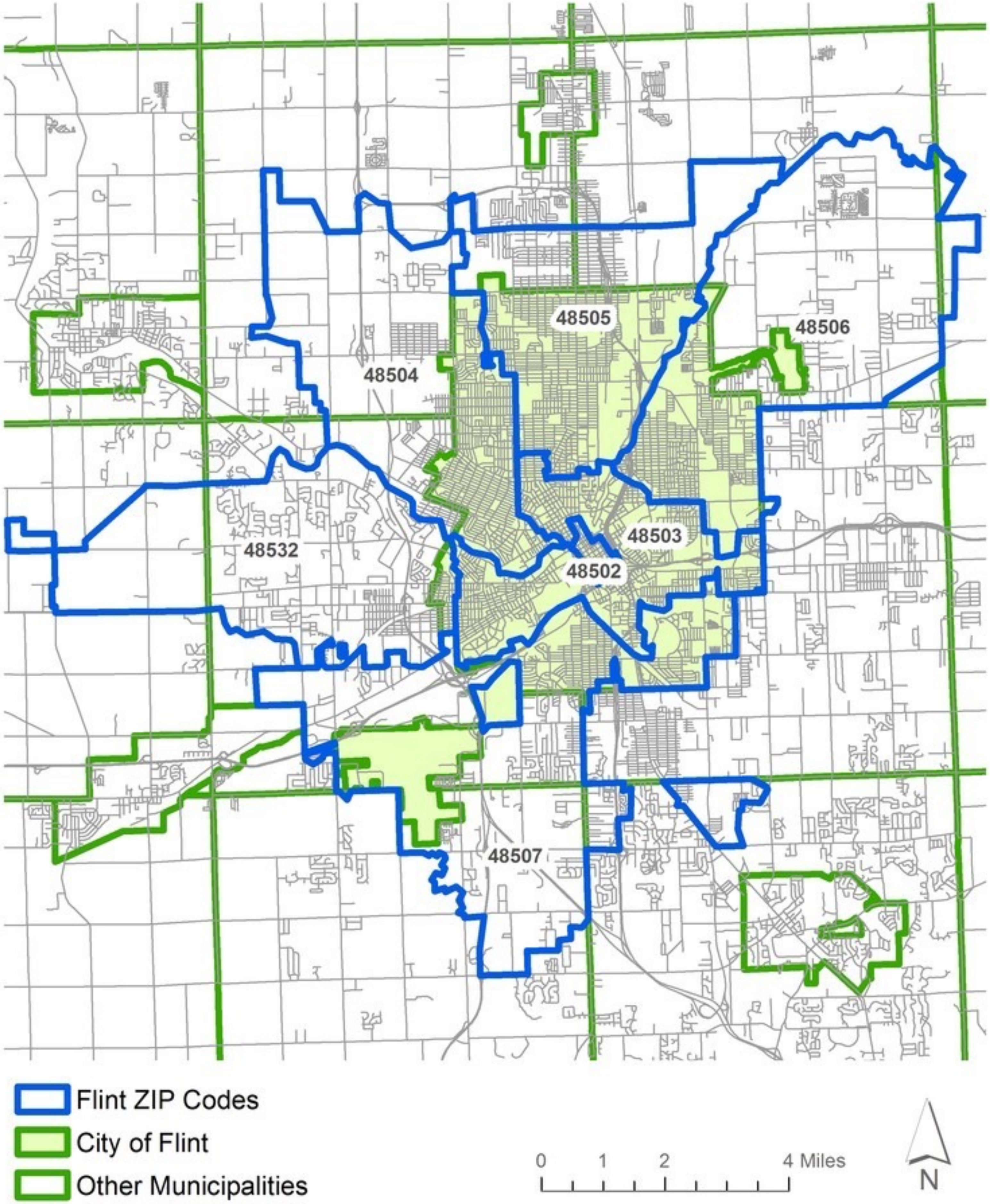
Year: 2016

http://kevinhayeswilson.com/redraw/

misaligned polygons are the big problem

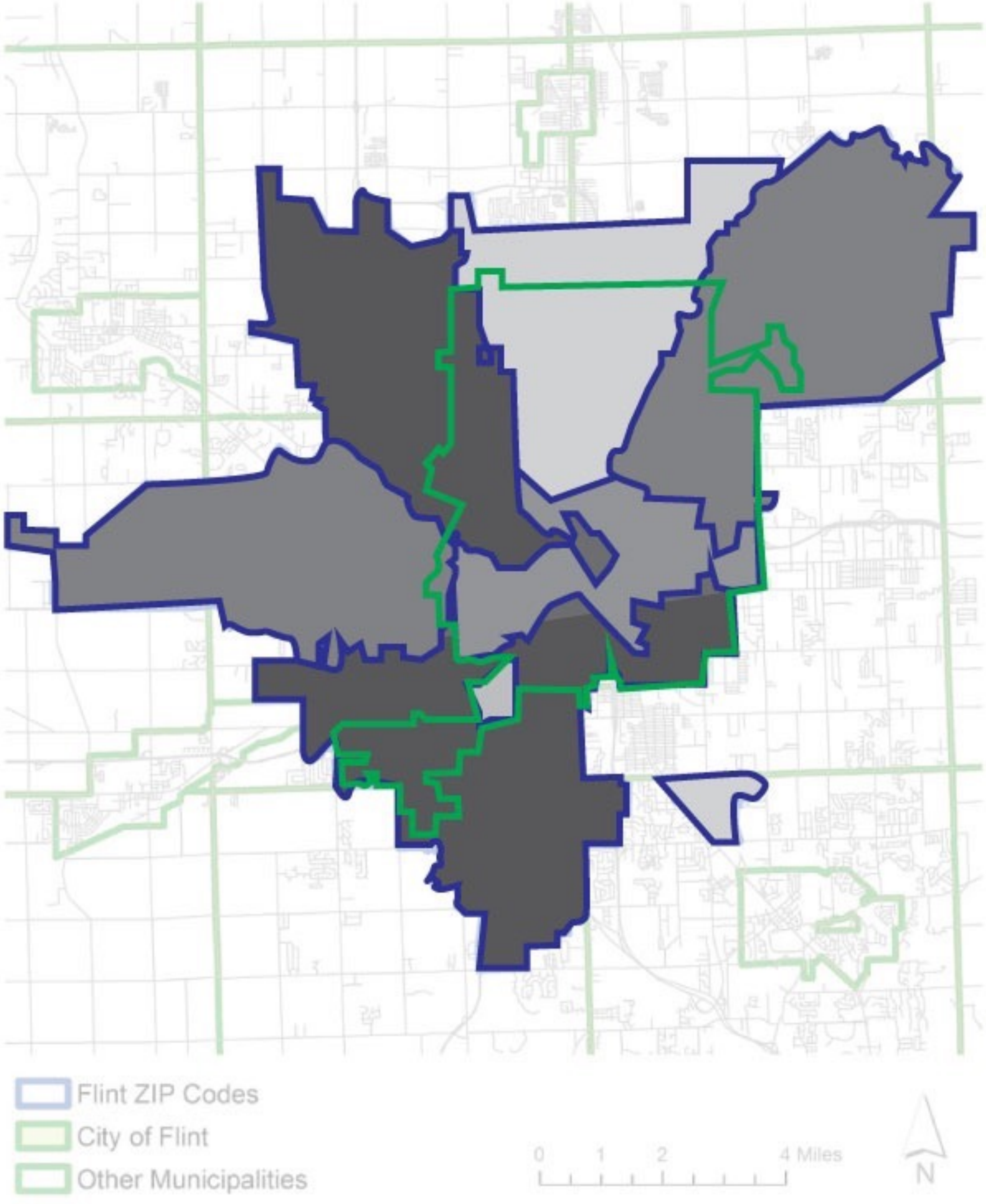
flickr: MattHagedorn

Misalignment between Flint ZIP Codes and City of Flint



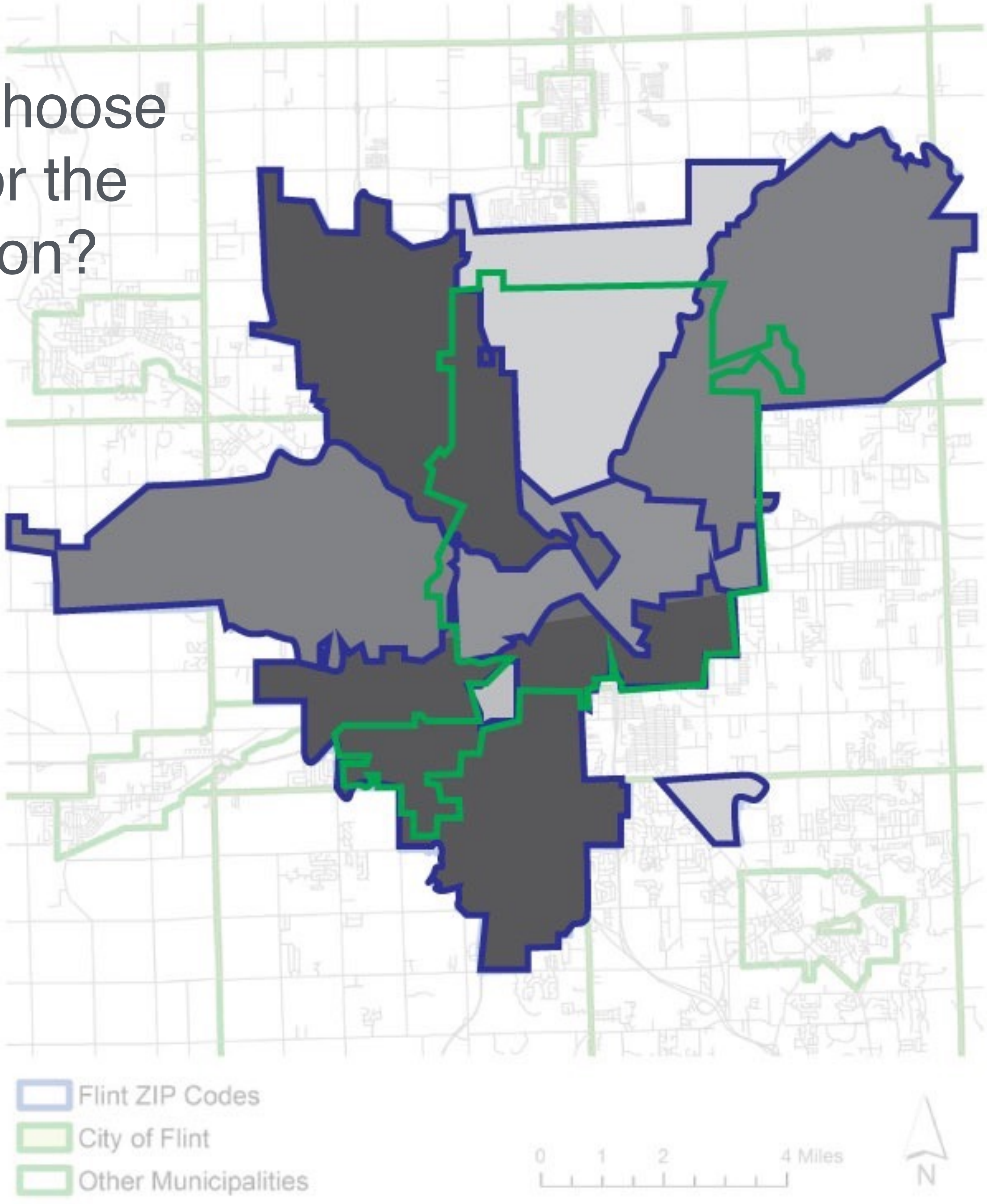
http://theconversation.com/how-zip-codes-nearly-masked-the-lead-problem-in-flint-65626

Misalignment between Flint ZIP Codes and City of Flint

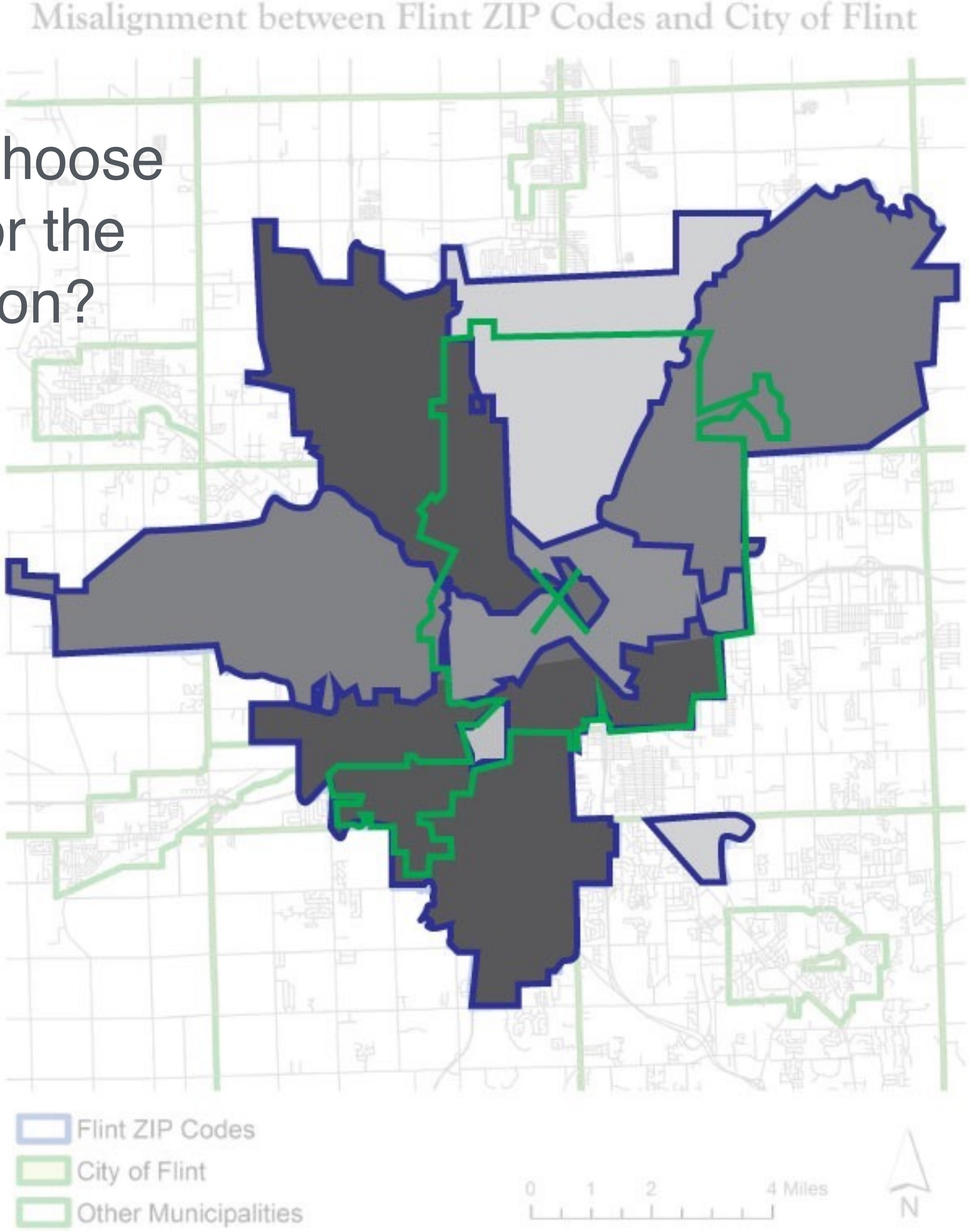


How do we choose the value for the green region?

Misalignment between Flint ZIP Codes and City of Flint

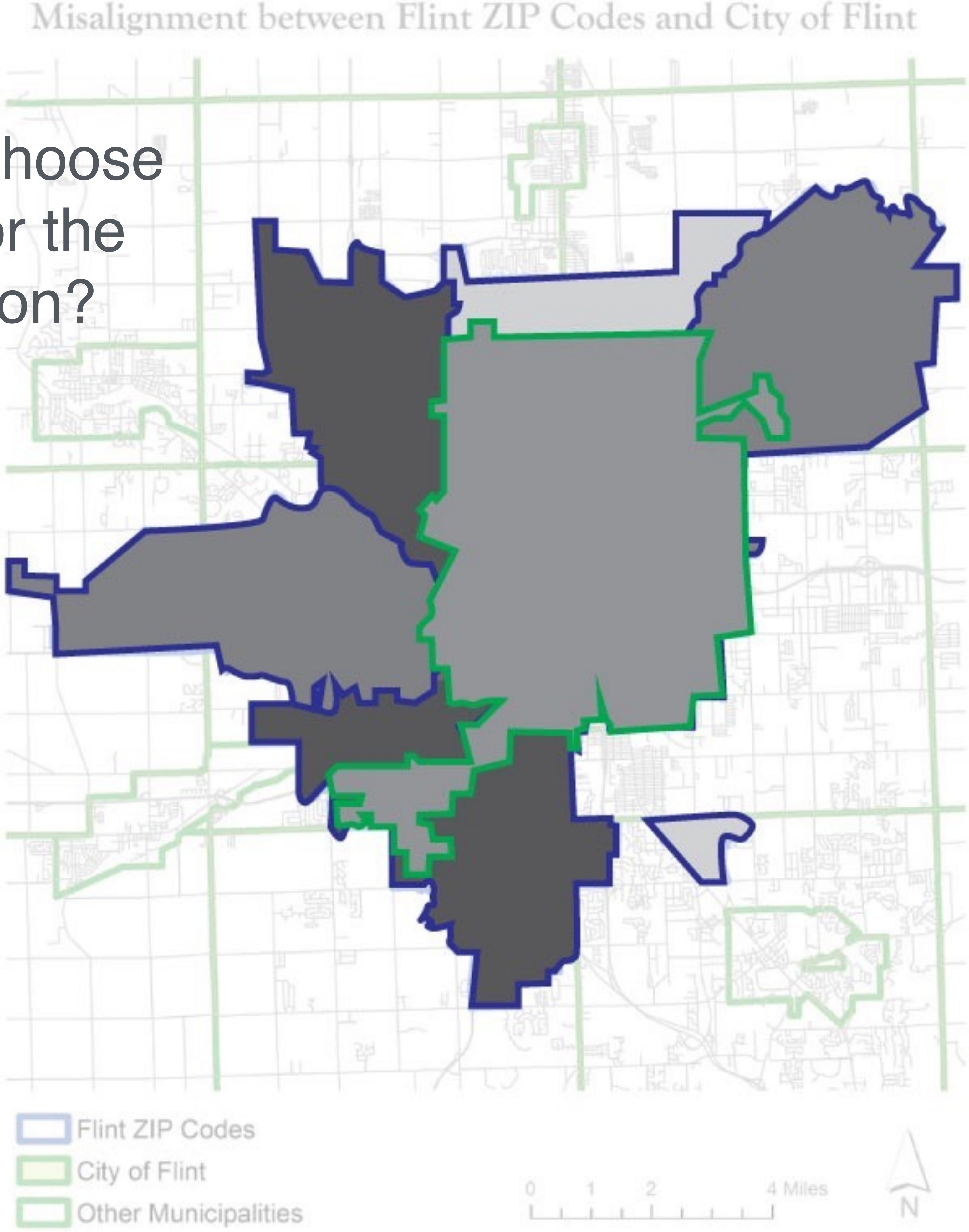


How do we choose the value for the green region?



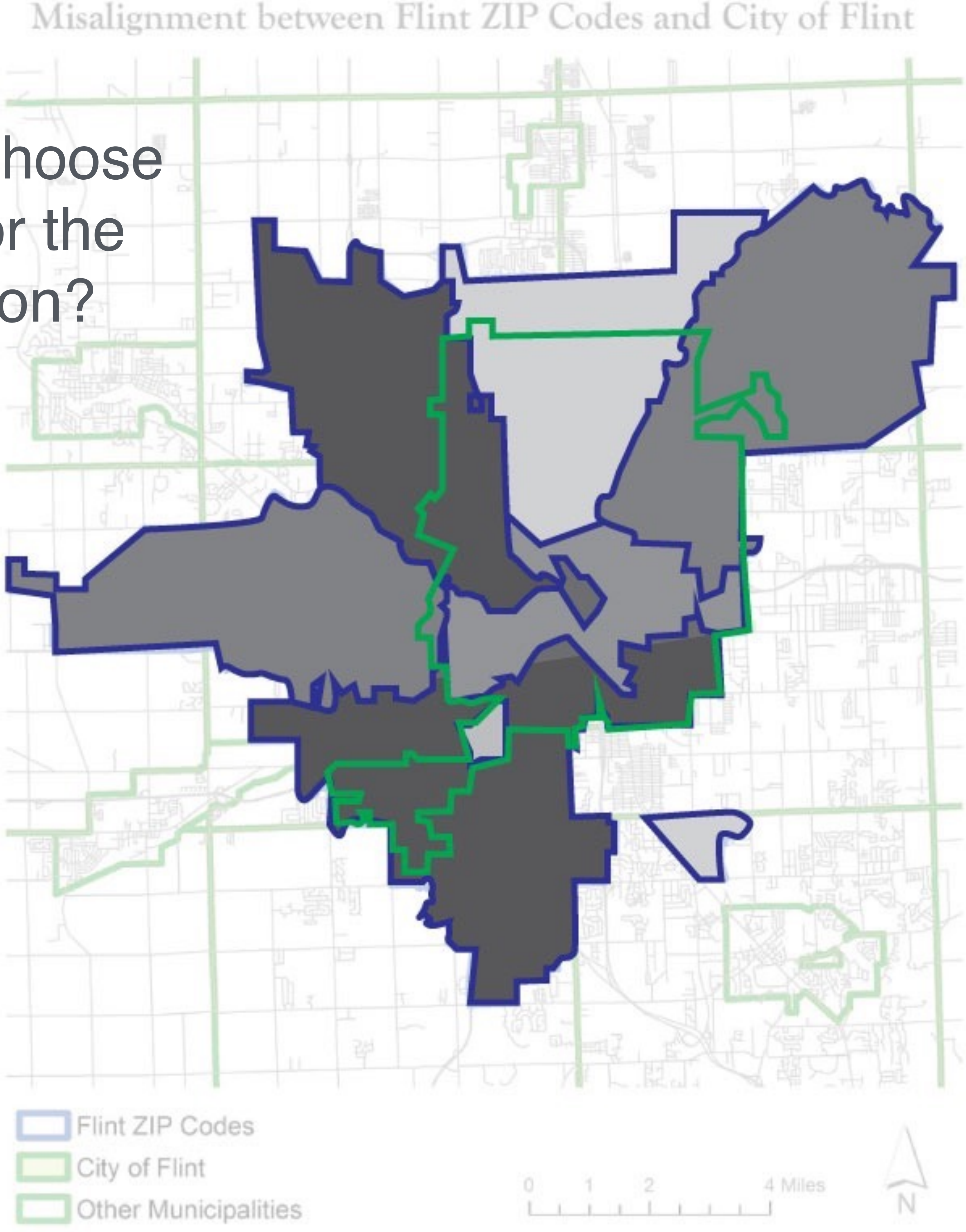
Misalignment between Flint ZIP Codes and City of Flint

How do we choose the value for the green region?



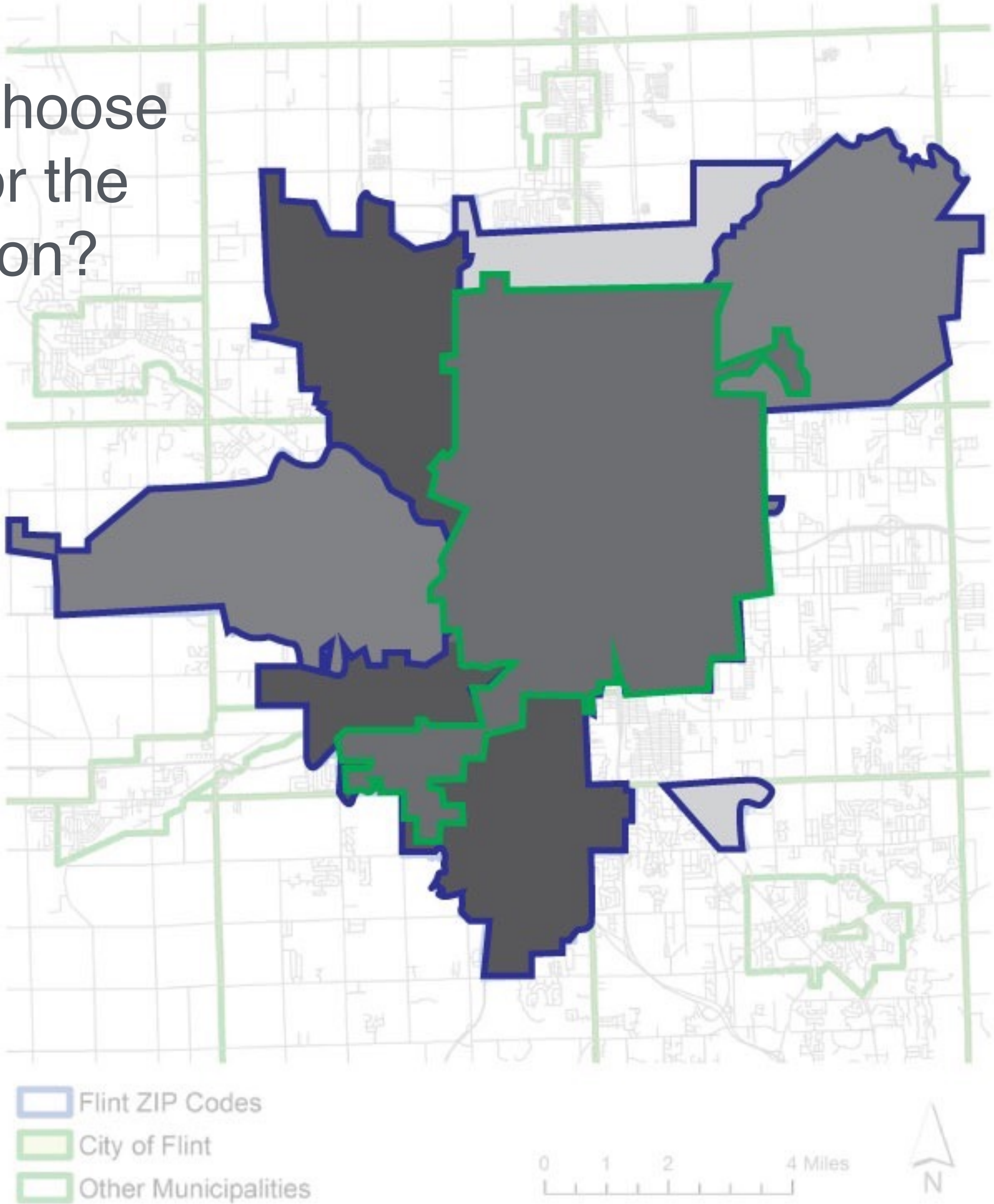
Misalignment between Flint ZIP Codes and City of Flint

How do we choose the value for the green region?



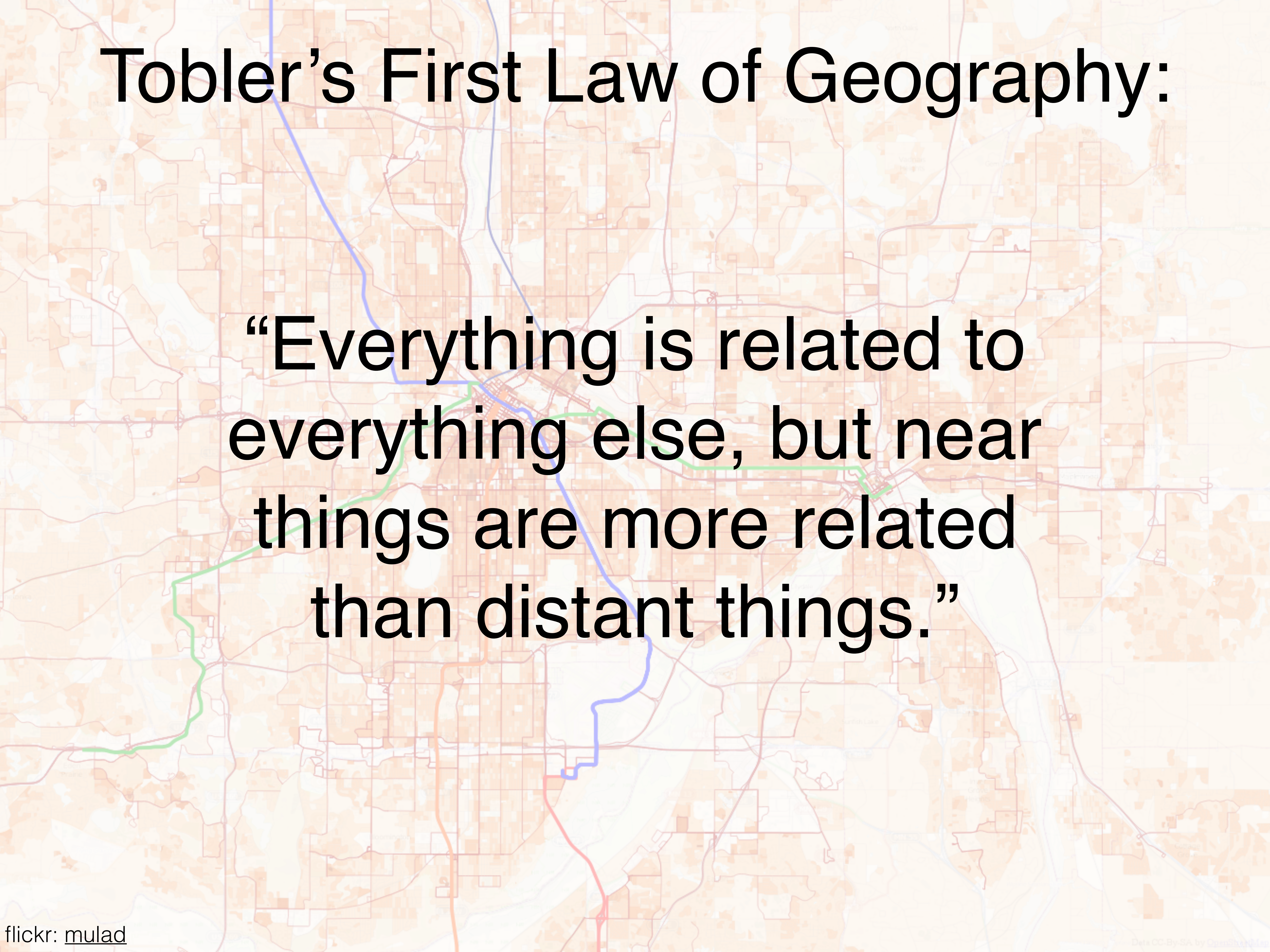
How do we choose the value for the green region?

Misalignment between Flint ZIP Codes and City of Flint



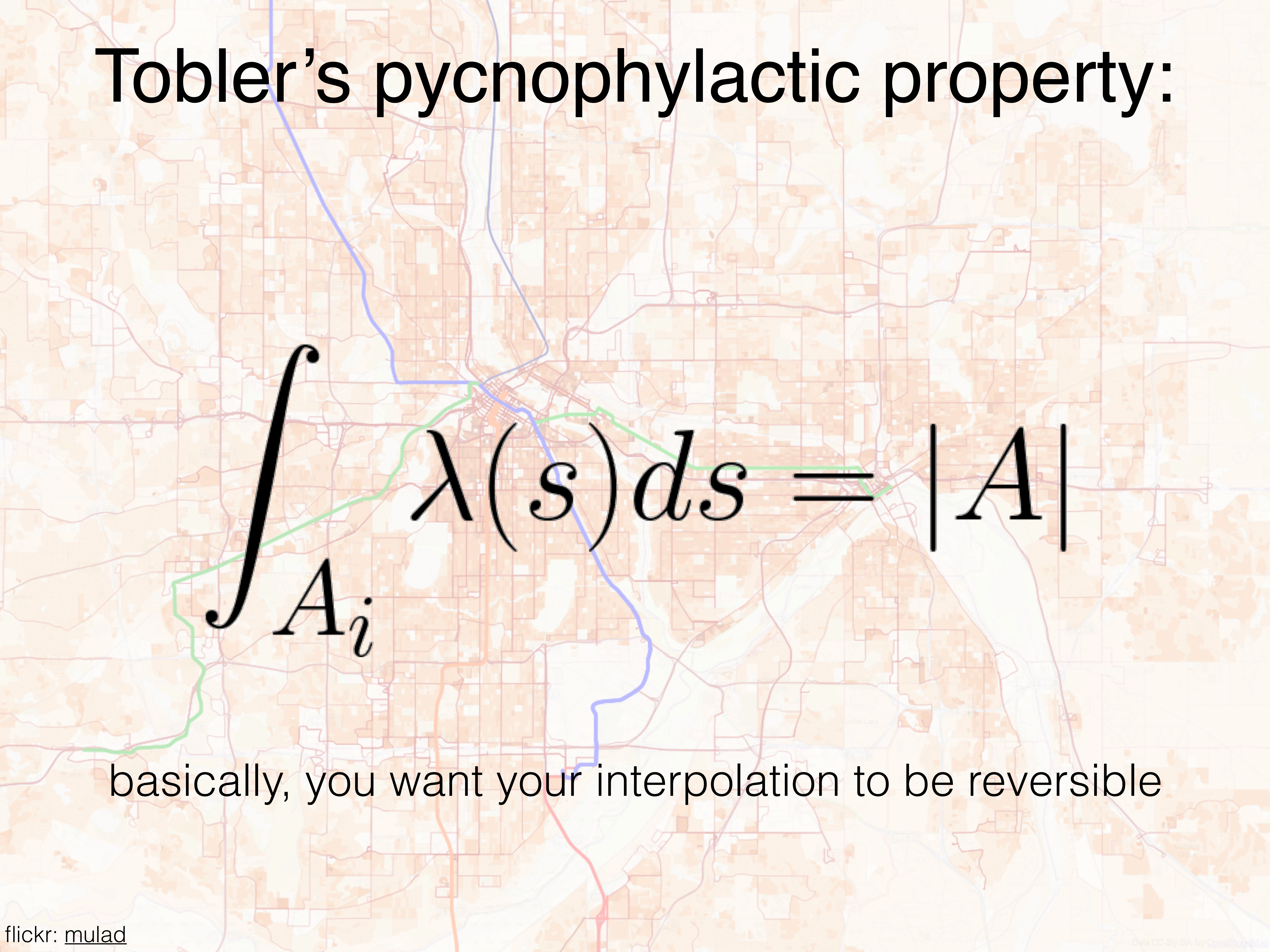
Tobler’s First Law of Geography:

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



flickr: mulad

Tobler’s pycnophylactic property:

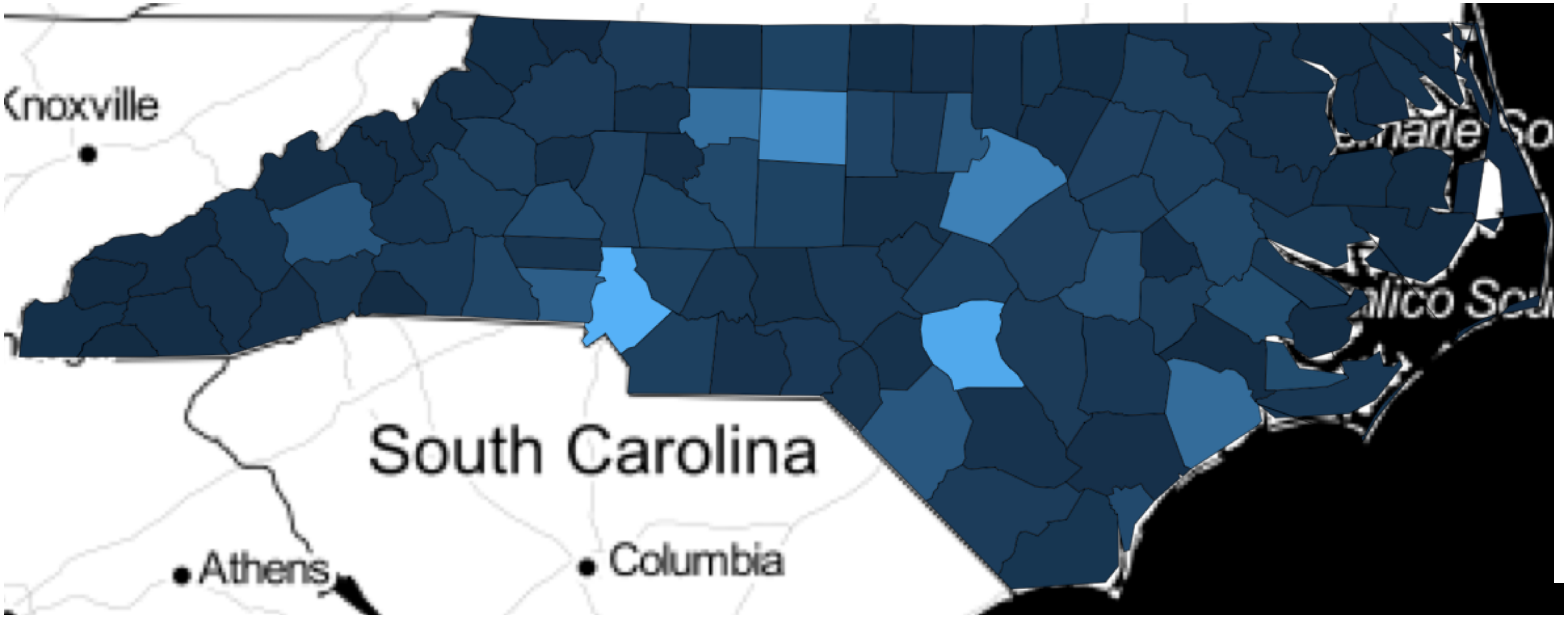


$$\int_{A_i} \lambda(s) ds = |A|$$

basically, you want your interpolation to be reversible

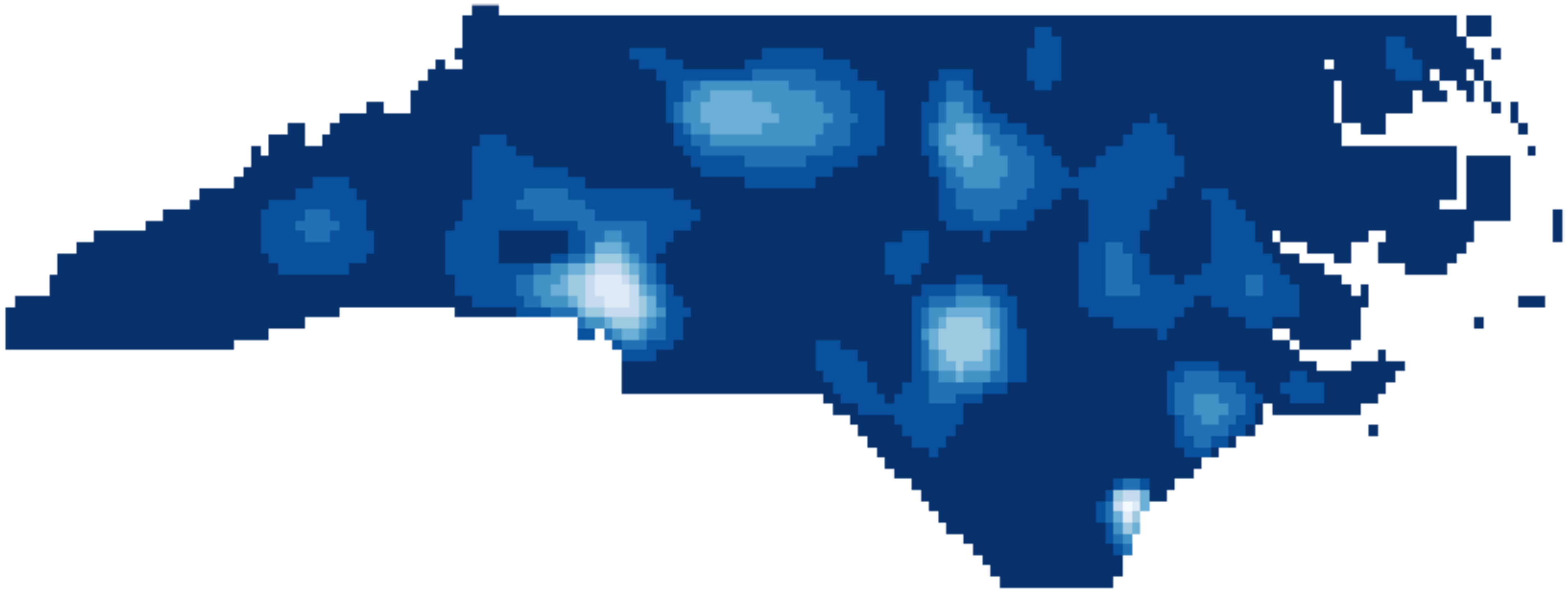
flickr: mulad

Working with the `pycno` package in R



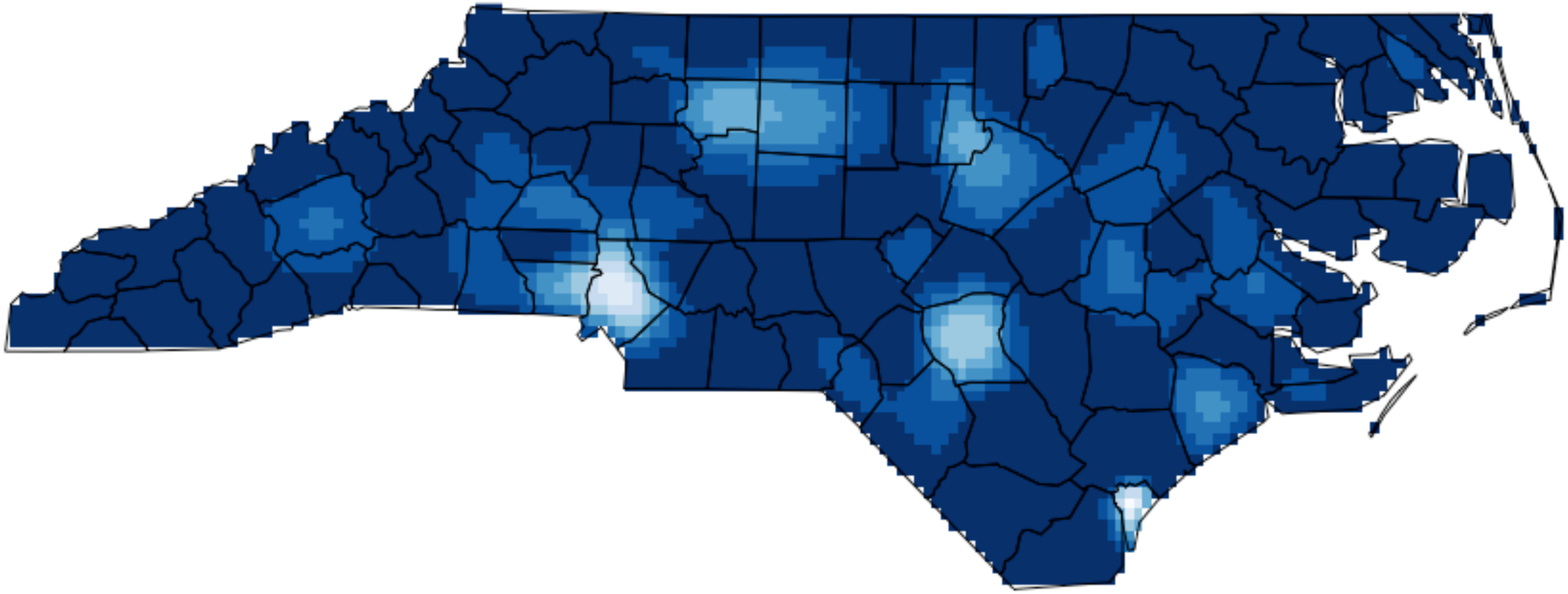
Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems

Working with the `pycno` package in R

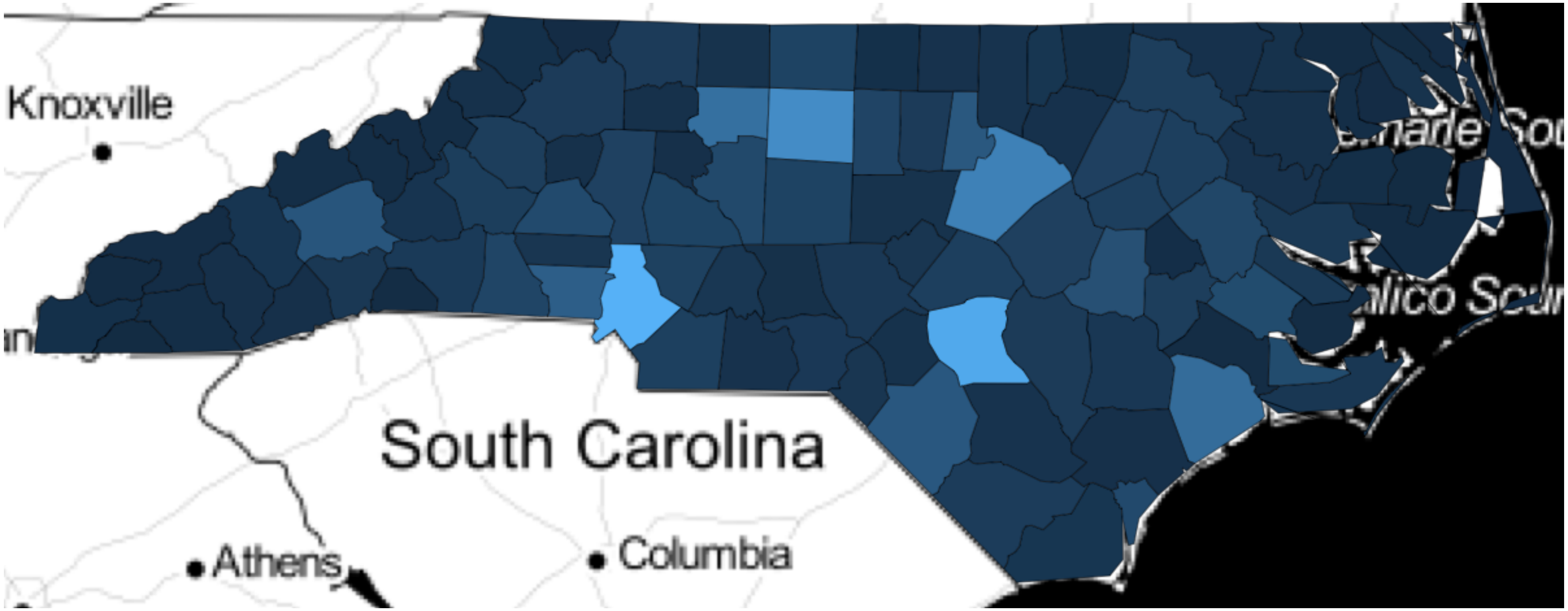


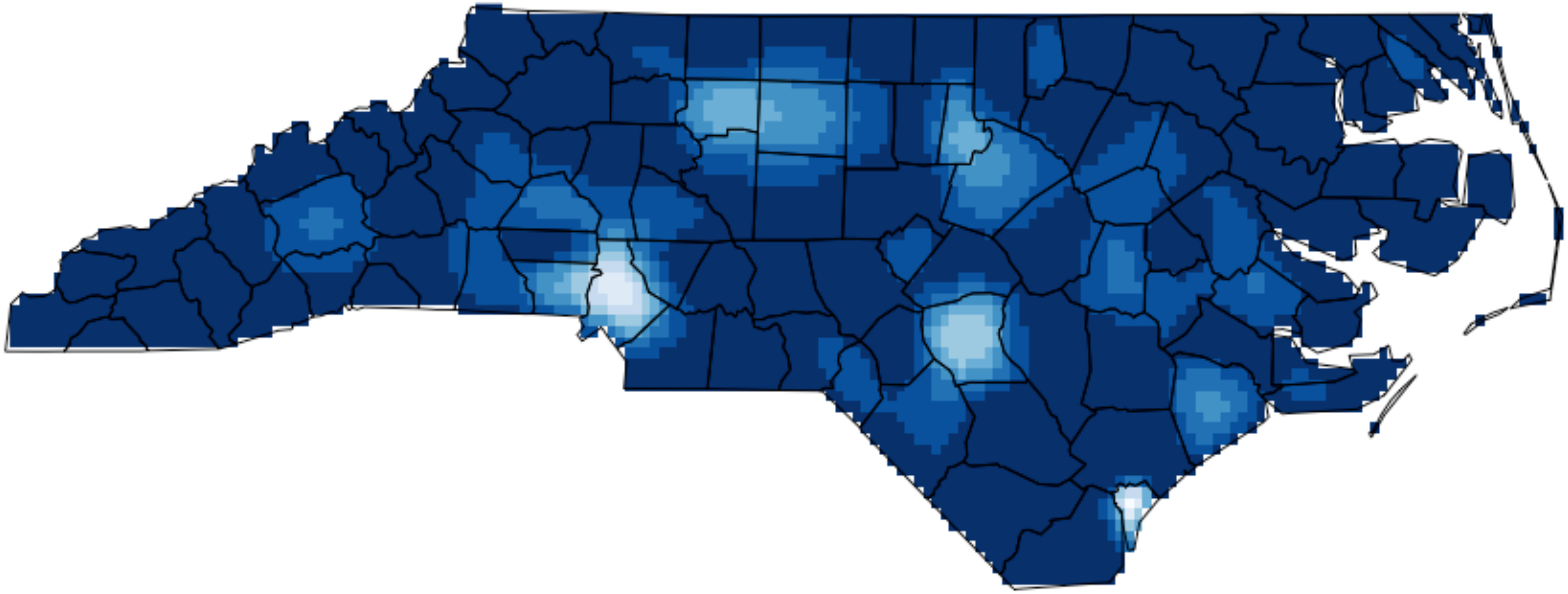
Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems

Working with the **pycno** package in R

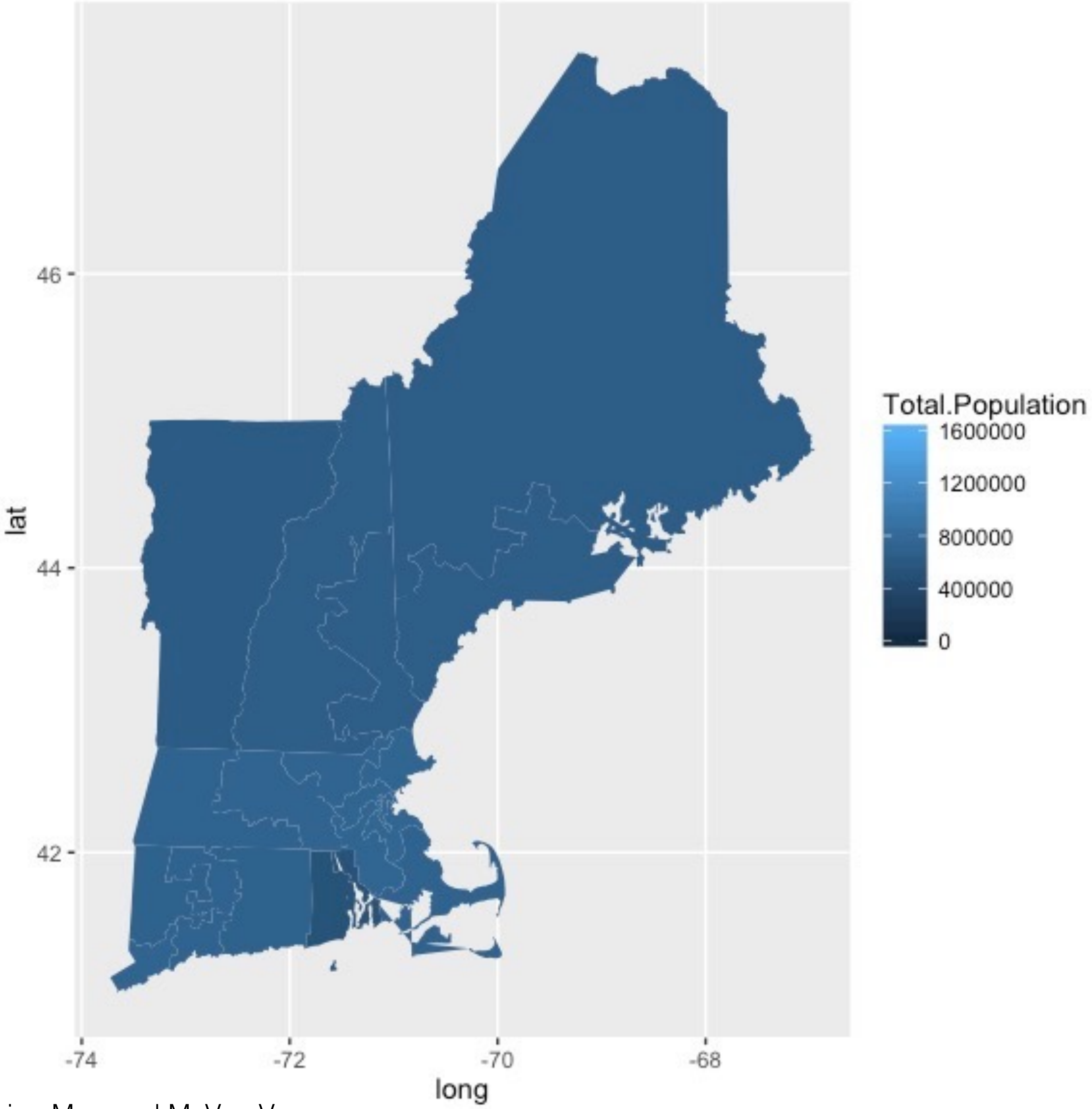


Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems

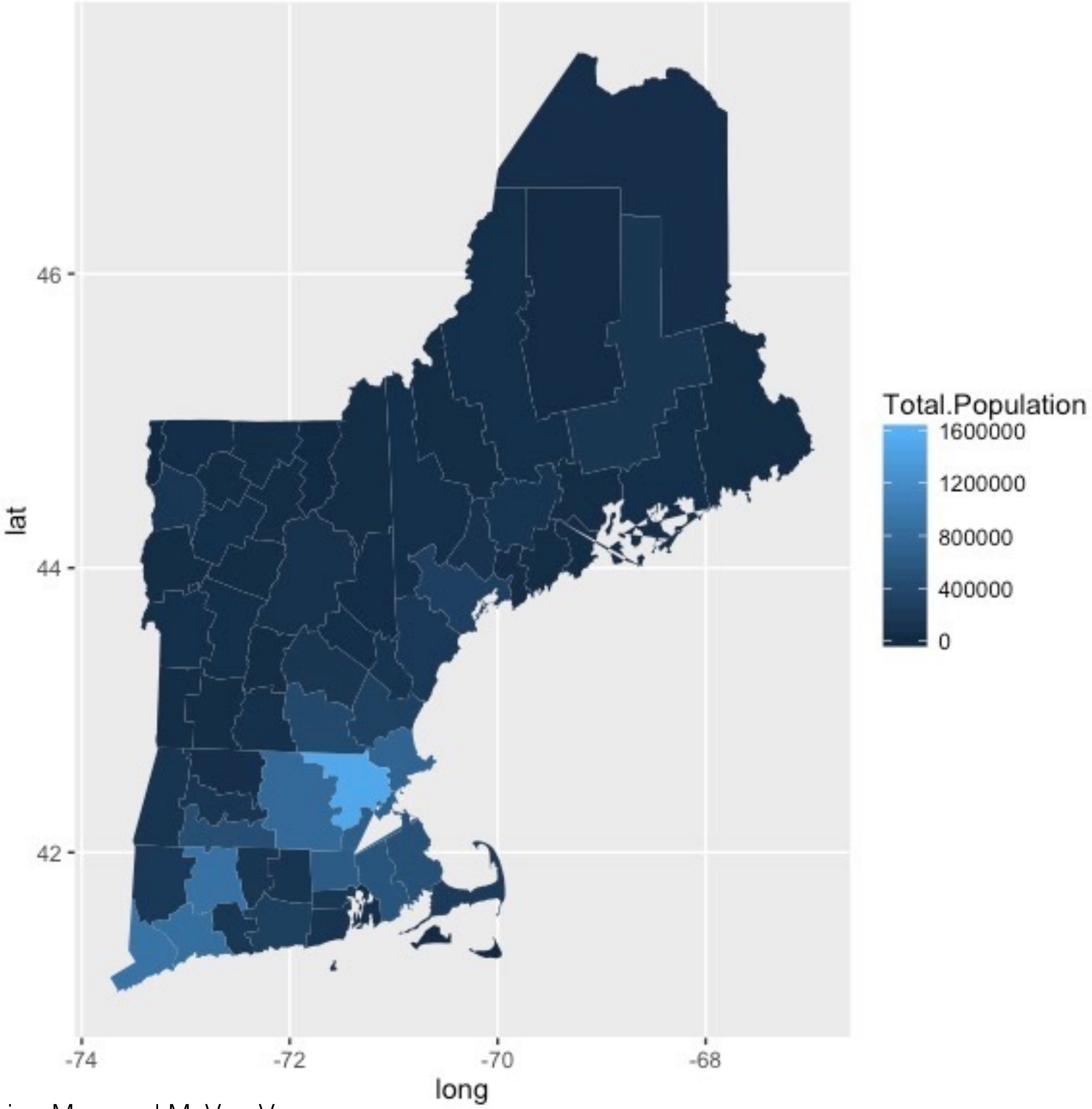




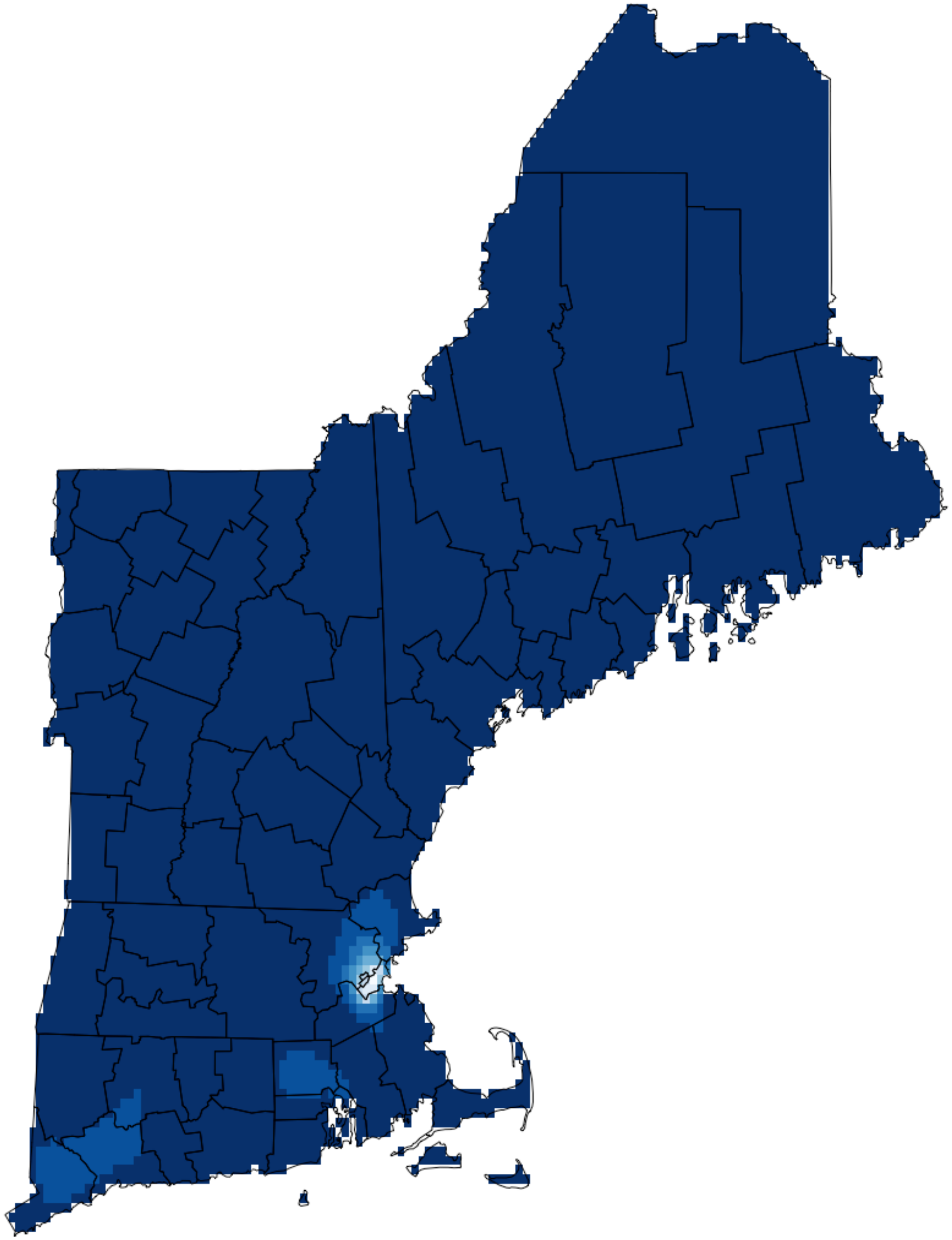
Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems



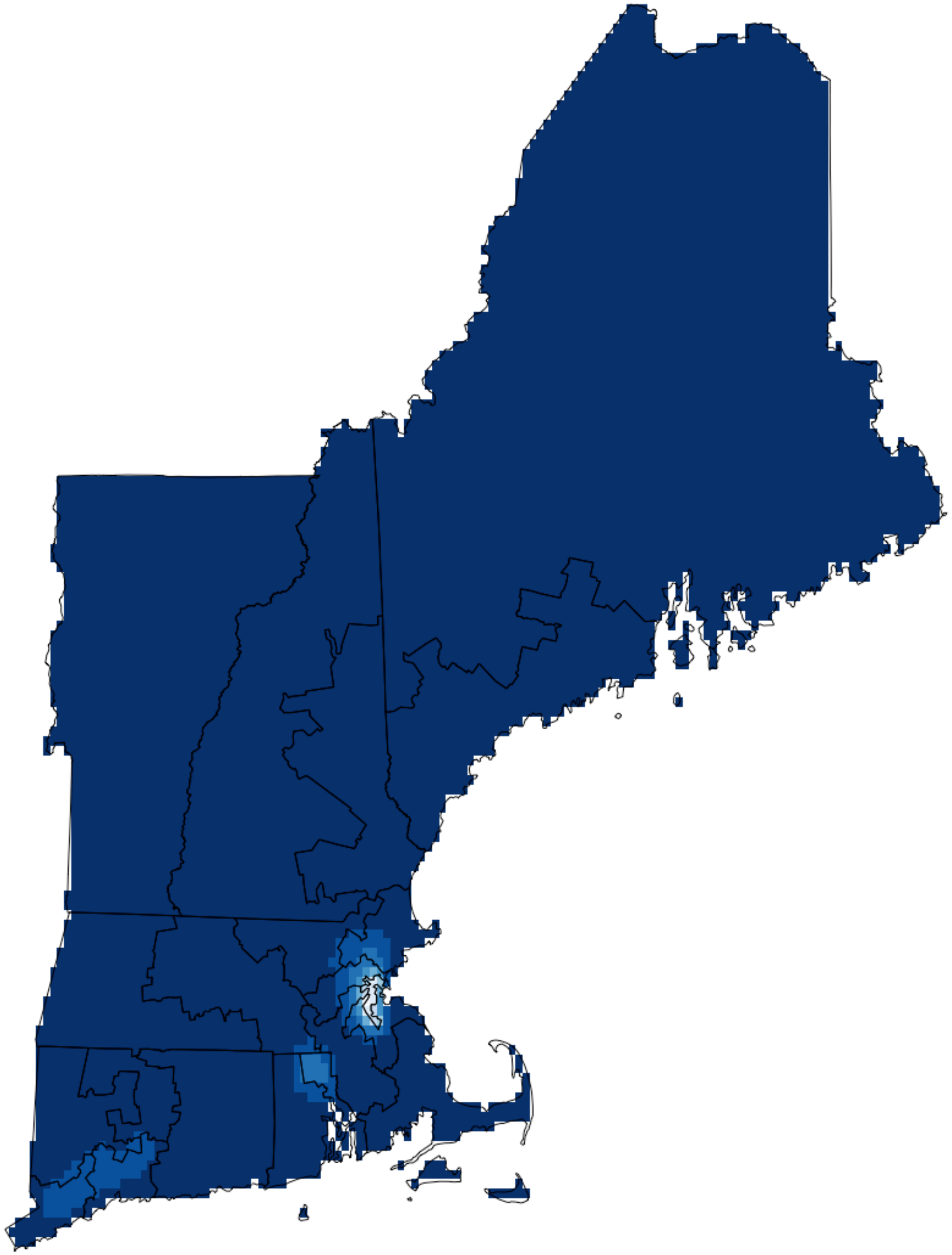
Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems



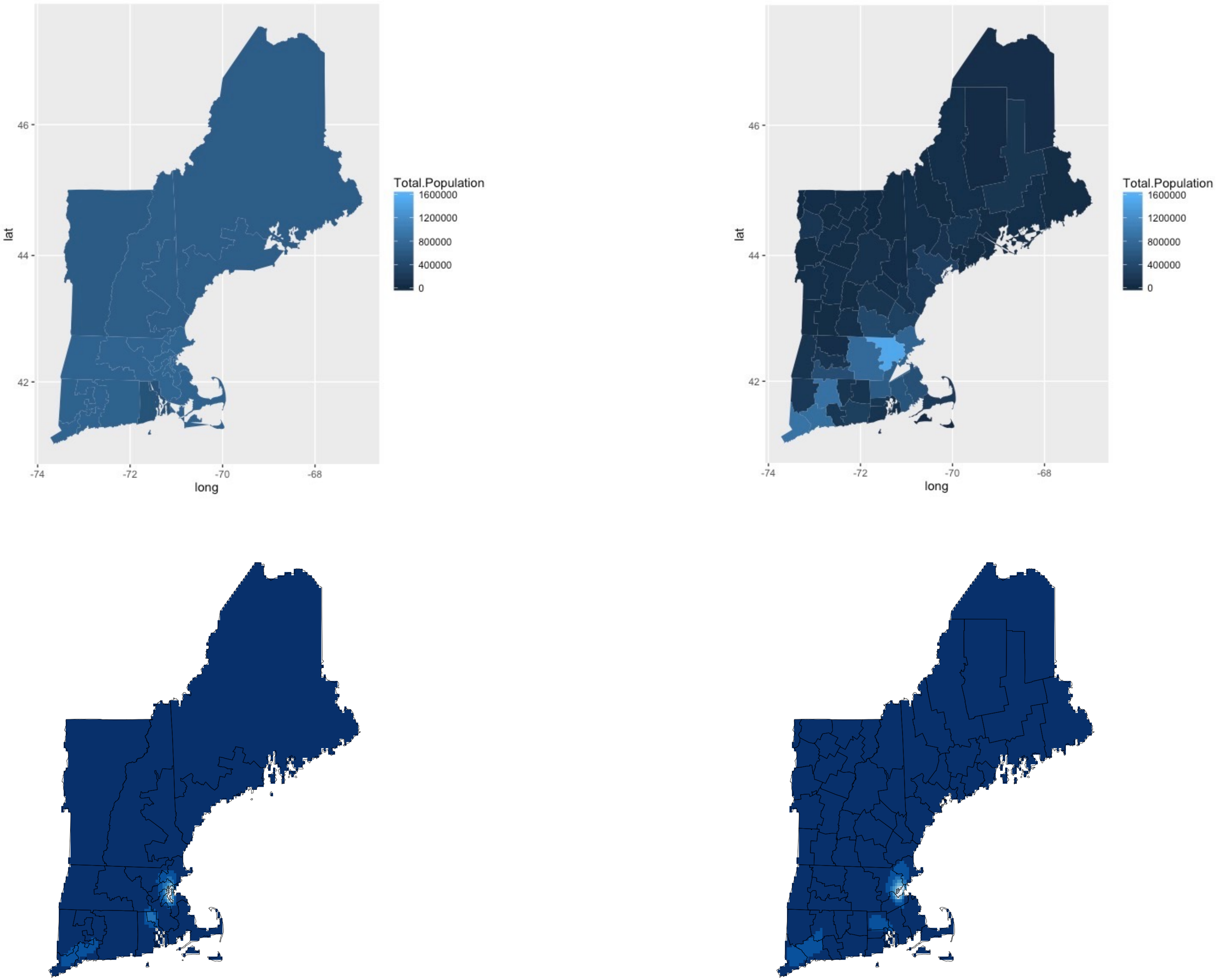
Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems



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

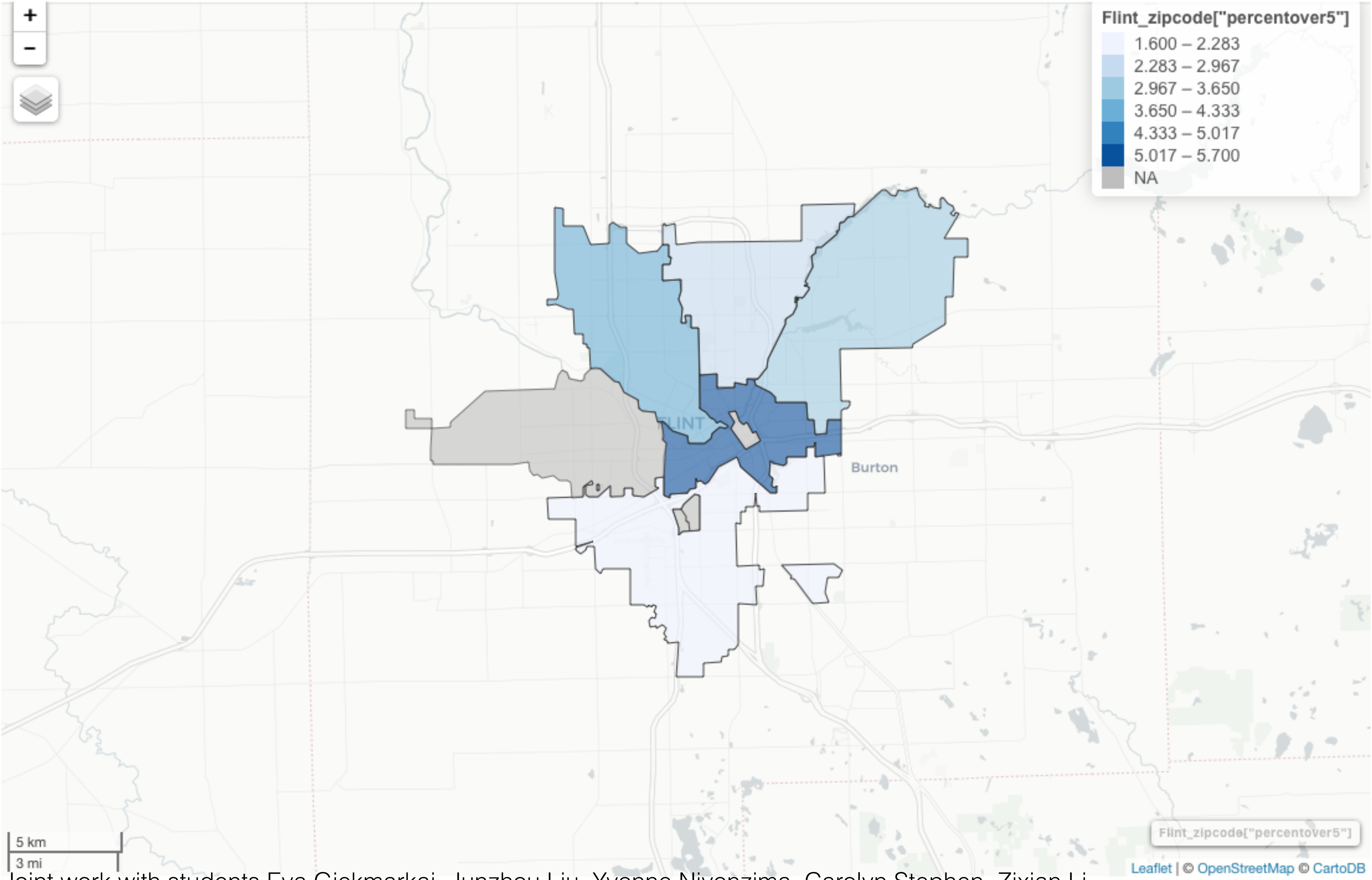


Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems

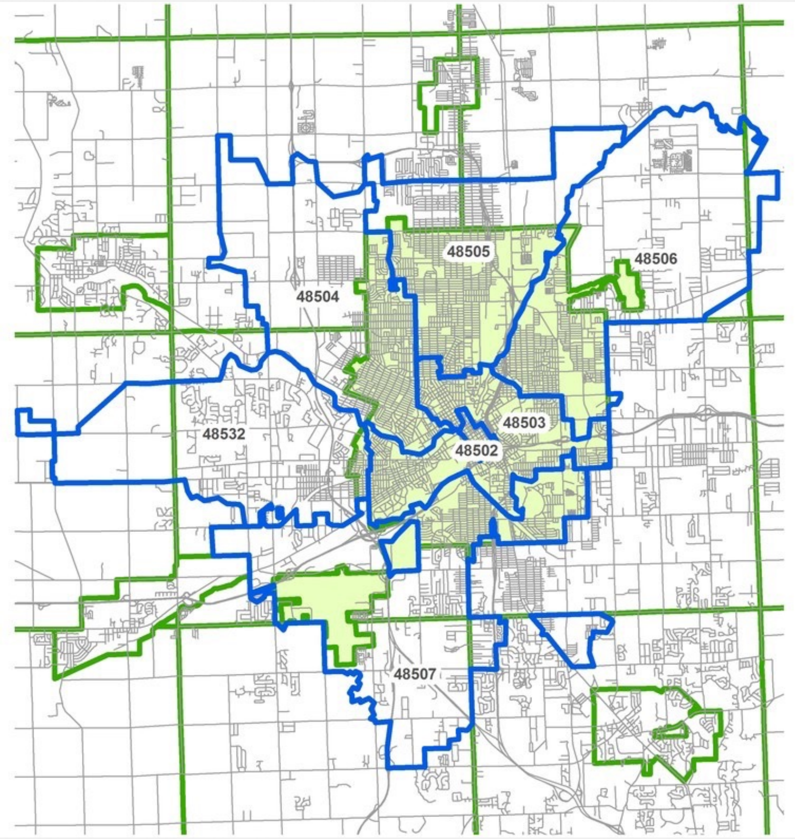


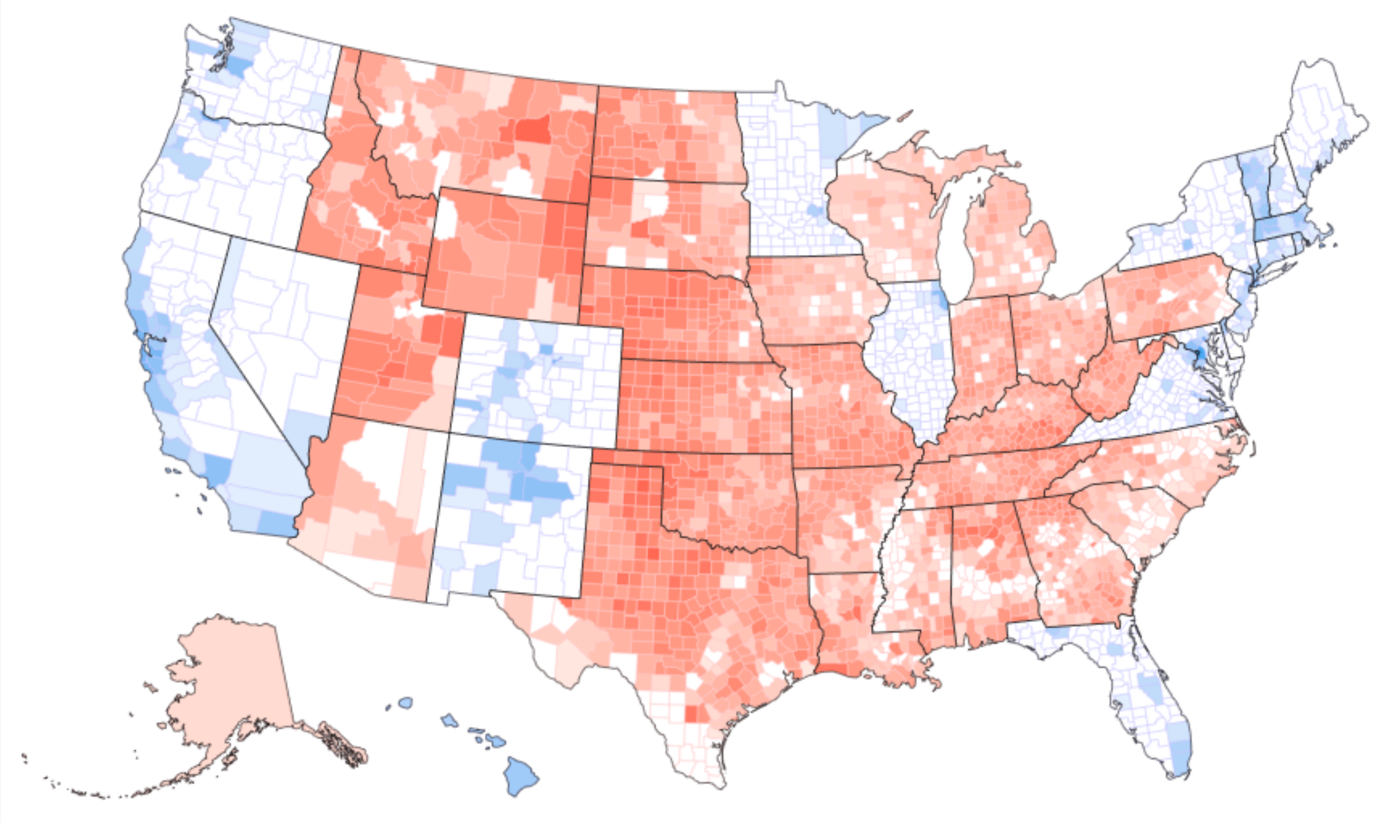
Joint work with students Jessica Mao and MyVan Vo.
Methods to Address Area-to-Area Change of Support and Modifiable Areal Unit Problems

Working with the **sf** and **mapview** packages



Joint work with students Eva Gjekmarkaj, Junzhou Liu, Yvonne Niyonzima, Carolyn Stephen, Zixian Li
Exploring MAUP using Flint water data

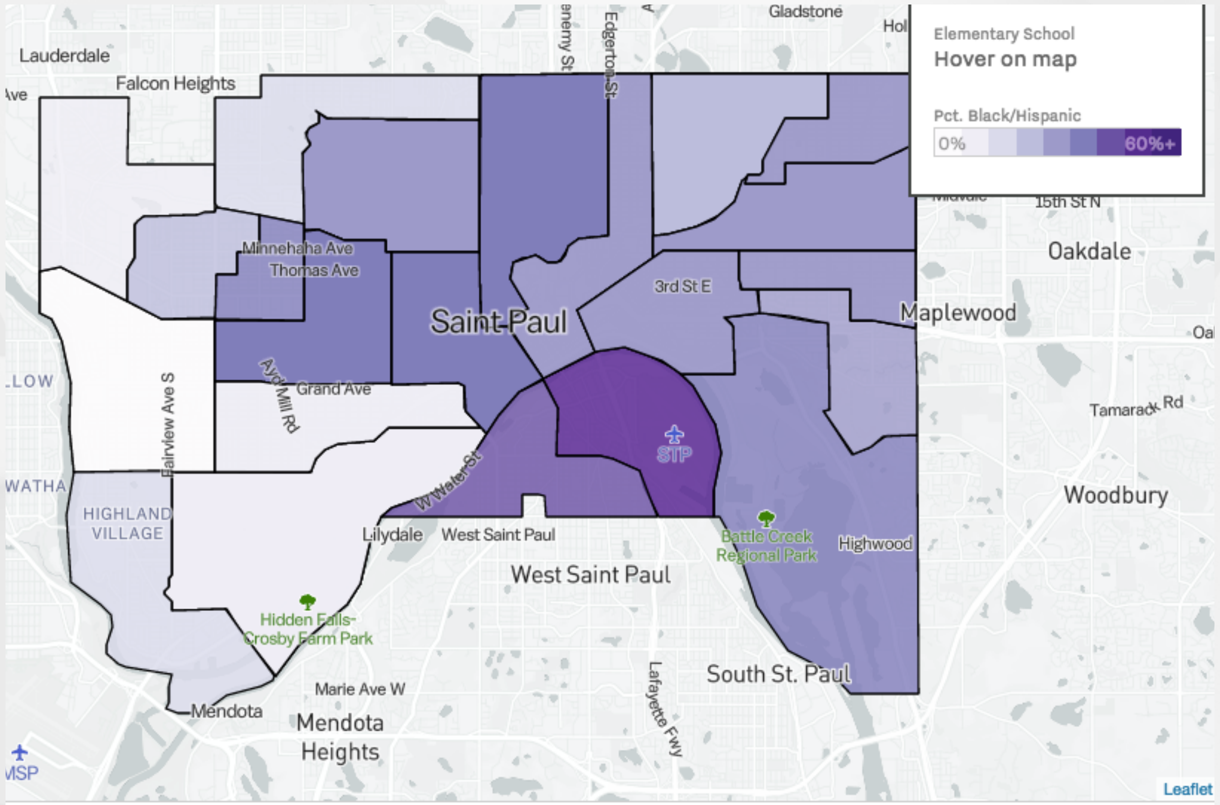




- don’t aggregate if you don’t have to

- pay attention to your spatial polygons

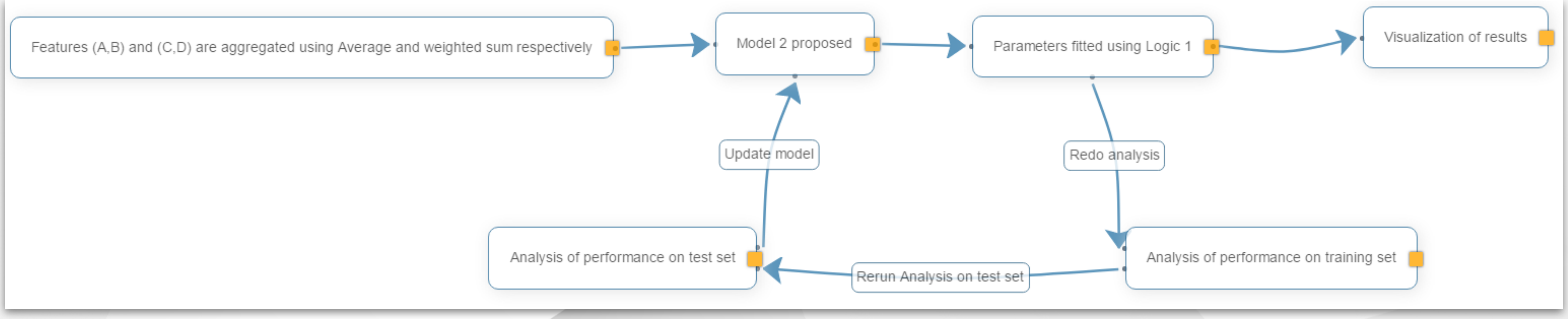
- use auxiliary information if you have it



Data from research by Tomas E. Monarrez, an economics PhD candidate at the University of California, Berkele

flickr: MattHagedorn

- remember researcher degrees of freedom



Features (A,B) and (C,D) are aggregated using Average and weighted sum respectively

Model 2 proposed

Parameters fitted using Logic 1

Visualization of results

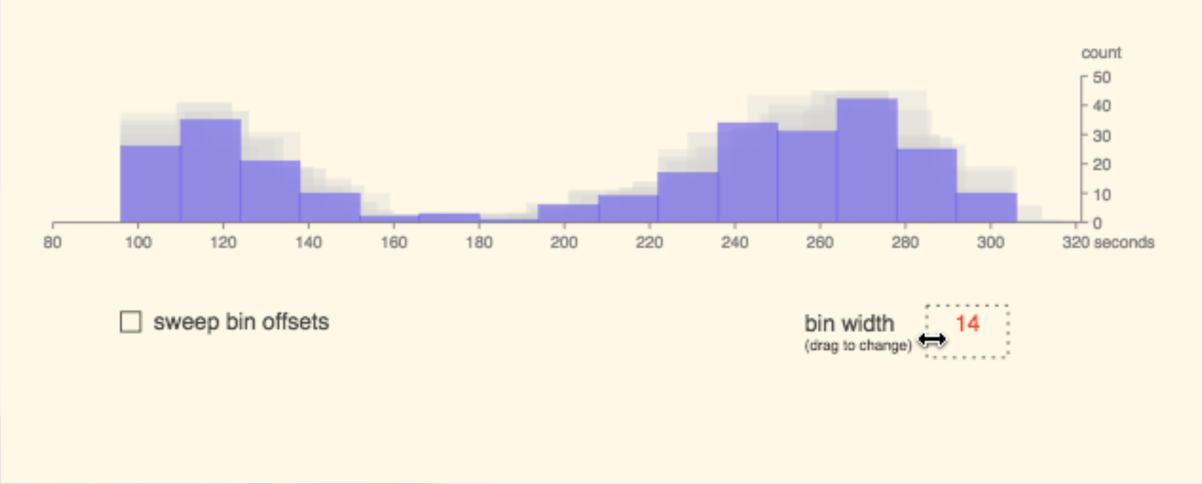
Redo analysis

Analysis of performance on training set

Rerun Analysis on test set

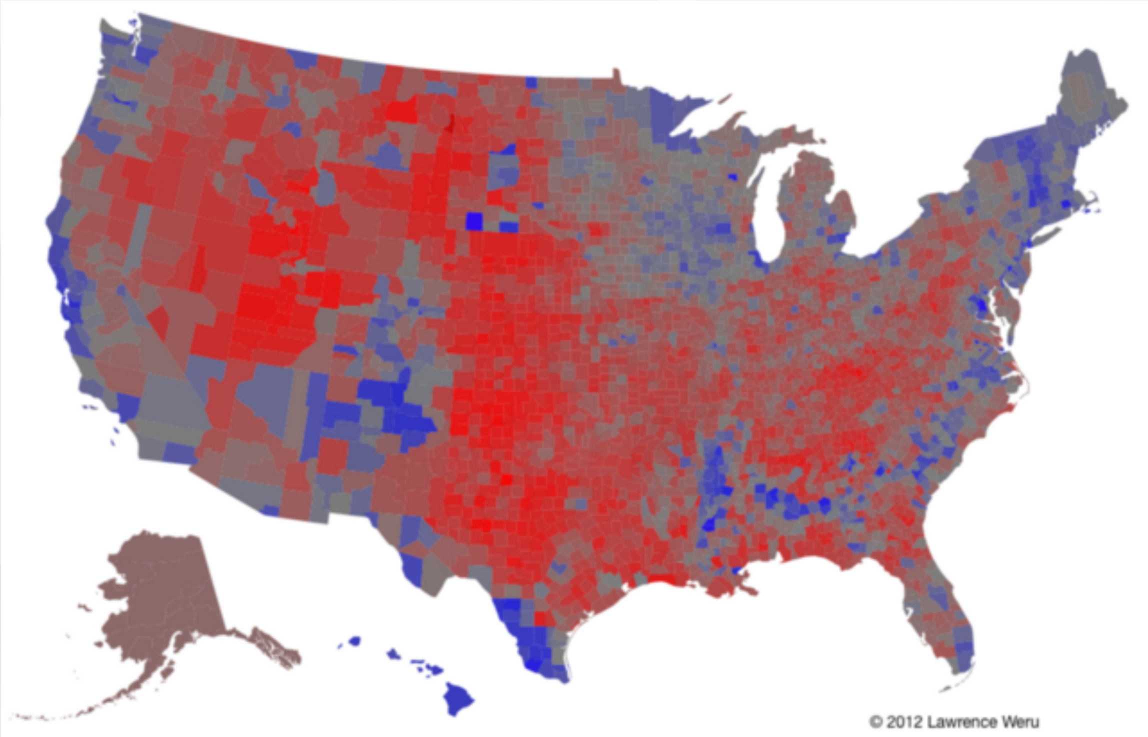
Analysis of performance on test set

Update model



☐ sweep bin offsets

bin width 14
(drag to change)



flickr: MattHagedorn

Thank you

flickr: bering_land_bridge