

# Interactive spatial aggregation

Amelia McNamara and Aran Lunzer

August 10, 2015

## 1 Introduction

In this note, we present a fun little tool that allows a user to manipulate the size and orientation of spatial aggregation units, in order to explore the effect of those parameters on the visual pattern presented.

## 2 Background

Aggregated data is notoriously hard for novices to grasp Konold et al. (2014); Hancock et al. (1992). However, it is used in many situations, both symbolic (the mean) and visual (the histogram).

When we began showing our LivelyR prototypes to statisticians and teachers, we were met with excitement about the ability to simply manipulate the bin width and bin offset of a histogram. Also popular was the ‘histogram cloud’ feature that superimposed a set of slightly different histograms of the same data McNamara (2015). Both interactions give a user the ability to easily manipulate parameters that they often would leave as defaults, and compare the results of different parameter choices. Based on this popularity, I began thinking about a 2-dimensional analogue, particularly in the context of spatial data.

Much like the case of histograms, the visual pattern in a choropleth map is imbued with meaning by readers who forget the arbitrariness of the chosen aggregation units. In a spatial context, geographers call this phenomenon the Modifiable Areal Unit Problem (MAUP). Essentially, map-making and statistical analysis is highly sensitive to the area of aggregation (e.g. Census tracts, zip codes, neighborhoods) that is used. For data that comes as points, this problem can be skirted by using the raw point data to fit models, but much spatial data is distributed in already-aggregated forms. Again, the US Census is a prime example. Because of anonymity issues, the Census Bureau can only provide data at the level of Census blocks.

This work does not try to address the underlying statistical issue of generating more robust estimates from pre-aggregated data, but rather seeks to make visible the possible effects of slight variations in aggregation.

There are two varieties of MAUP that are commonly discussed—scale MAUP and zone MAUP. The scale problem is related to the scale of the areal unit you have chosen to use. Is it large (like a state) or small (like a Census block)? The zone problem has more to do with the shape of the units. Similarly sized, but differently shaped units (imagine squares versus hexagons) can provide very different patterns.

Notice that I am saying “imagine” and “like.” This is because there are surprisingly few demos that actually show this phenomenon in action. Sometimes, information about MAUP will show two or three discrete possibilities of aggregations of the same data, but usually only with toy data Penn State GEOG 486 (2014).

The most publicized examples of the MAUP tend to be related to political gerrymandering of electoral boundaries. Depending on how districts are drawn, election results can be swung wildly in favor of one candidate or another. However, even in the case of gerrymandering, theoretical language like “Suppose there’s a state that [...]” is often used, rather than concrete interactives Cohn (2015). Even in pieces that use visuals, toy examples are king Ingraham (2015).

Gerrymandering is a more complex version of the MAUP, because there is more at stake than simply the location of voters. Even people in the same political party can have very different ideas about how districts should be drawn—should they contain many like-minded people, or those with similar demographics? Or should they seek to be diverse in every sense of the word? Should they follow geographic boundaries or strive for geometric compactness?

Again, we are not trying to solve this problem. Rather, we aim to present a tool that allows users to see the effects of various spatial aggregation levels.

### 3 The tool

The tool, as it stands, is an HTML page using javascript (including d3.js and leaflet.js) to provide interaction. An OpenStreetMaps map is used as the base, and spatial point data is layered on top. Then, regular polygon aggregation units are provided, and colored according to the number of points they contain.

It supports either squares or hexagons as the polygon aggregation units, and provides 8 possible sizes, which can be moved between smoothly.

The polygons can be scaled, moved, and rotated. The base map can also be zoomed and moved, although the interface does not support rotation.

Bird data aggregated in 100-km squares Sullivan et al. (2008).

Disser Martin-Anderson (2014).

## References

Nate Cohn. Voting case has potential to put house further out of reach for democrats. *The New York Times*, June 2015.

- Chris Hancock, James J Kaput, and Lynn T Goldsmith. Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3):337–364, 1992.
- Christopher Ingraham. This is the best explanation of gerrymandering you will ever see. *The Washington Post*, March 2015.
- Clifford Konold, Traci Higgins, Susan Jo Russell, and Khalimahtul Khalil. Data seen through different lenses. *Educational Studies in Mathematics*, 2014.
- Brandon Martin-Anderson. Building precise maps with disser. <http://conveyal.com/blog/2014/04/08/aggregate-disser/>, April 2014.
- Amelia McNamara. *Bridging the Gap Between Tools for Learning and for Doing Statistics*. PhD thesis, University of California, Los Angeles, June 2015.
- Penn State GEOG 486. Choropleth maps. <https://www.e-education.psu.edu/geog486/node/1864>, 2014.
- Brian Sullivan, Steven T Kelling, Christopher Wood, Marshall Iliff, Daniel Fink, Mark Herzog, Doug Moody, and Grant Ballard. Data exploration through visualization tools. *Proceedings of the fourth international partners in flight conference: Tundra to topics*, pages 415–418, 2008.