Investigating the modifiable areal unit problem in Baltimore, Maryland

Proposal, GES Independent Study

Harrison DeFord

2025-04-09

### Introduction

Spatial data is an important component of decision-making, providing utility to disciplines including public health, urban planning, and disaster management. Areal units—polygons that provide a framework for pattern identification—are central to the use of spatial data across all domains. Well-known examples of spatial units used for analysis include Census geographies and neighborhoods: these spatial units allow for the upwards aggregation of raw geographic data. The aggregated data can be used to identify hotspots or clusters, and are often more relatable and useful in policy-writing.

Across all domains, the choice of spatial units in analysis can change the patterns that emerge, influencing both the conclusions drawn from data and the decisions based on them. This issue, known as the Modifiable Areal Unit Problem (MAUP), pervades spatial analysis and must be carefully considered when planning the construction of the geography in the analysis being done. An analysis which does not consider the MAUP can misrepresent spatial patterns, leading to conclusions which are, at best, misinformed.

As recently described by Wong (2009), the MAUP arises whenever real-world processes are analyzed within the framework of spatial units, whose arbitrary definitions can lead to vastly different, even contradictory, results. MAUP manifests in two key ways: the scale problem, where patterns are exaggerated or diminished depending on the spatial scale of areal units, and the zoning (or boundary) problem, where arbitrary borders exclude or separate data in ways that may not reflect underlying spatial processes. With modern GIS tools, users—regardless of experience—can define spatial scales and boundaries with just a few clicks: arguably, the most powerful capability these software packages provide.

Given this, it is important to understand how the MAUP works its way into spatial data when GIS is used as the source for new areal units. While much research has focused on shifting conclusions based on existing administrative units, such as Census geographies, less attention has been devoted to the development of these novel spatial units, from definition to digitization to aggregation.

This analysis will explore different methods for defining novel spatial units and their impact on observed patterns, using 3-1-1 calls for service across Baltimore, Maryland, as a case study. 3-1-1 data lends itself well to expression of the MAUP for several reasons: first, because 3-1-1 calls are related to specific locations across urban space. Since 3-1-1 calls are related to a specific address, classic administrative boundaries, which are often drawn along geographic features such as roads and rivers, can easily split patterns: consider, for example, a rat infestation in which rodents nest on both sides of a street, the centerline of which demarcates the boundary between two spatial units. Additionally, 3-1-1 calls are important indicators of non-emergency needs for city services. Thus, identification of geographic areas for resource allocation is an important consideration for local government—however, these decisons can easily be influenced by the MAUP.

### Data

Baltimore City makes historic 3-1-1 customer service requests available via their Open Baltimore data portal as an Esri Feature Service. This paper will focus on the “311 Customer Service Requests 2024” dataset (Baltimore City 2024b), as it is the most recent complete calendar year at the time of publishing.The 3-1-1 call data are points features which include detailed attributes such as the type of request, the location of the service request, and the time of the request. In addition, the large sample size () is well-suited for aggregation across the geography; it is unlikely that a spatial unit will have zero observations.

In addition, Baltimore City Neighborhood Statistical areas (Baltimore City 2024a) will be used as a starting point for the creation of novel units. Internal-point centroids of these areas will be calculated to represent place names for further aggregation. These centroids allow for a point-based definition of neighborhoods as a starting point for creating custom spatial units, without depending on existing boundary definitions which exacerbate the zoning issue.

### Methods

While there is much research available regarding the effect of the MAUP across various existing administrative units, less attention has been given to distinct geographic areas that do not rely on predefined boundaries. The existing Neighborhood Statistical Areas (NSAs) of Baltimore City will serve to define place names, but their boundaries will not be the focus of this work. Instead, the study will focus on creating new areal units using point-based methods and conceptualized areas. R Statistical Software (R Core Team 2024) and ArcGIS Pro (ESRI 2022) will be used to generate novel aggregation units, which will summarize 3-1-1 calls for service. These aggregation units, as described in [Table 1](#tbl-agg), will vary in size and configuration and will include buffer zones and Voronoi polygons around neighborhood centers, buffers around 3-1-1 activity zone centers, and tessellation grids across the Baltimore region.

A sensitivity analysis will be conducted by comparing descriptive statistics, as well as geographic and numeric distributions, across each type of novel spatial unit. By examining how each areal unit summarizes the call data, the analysis will identify how different definitions of space influence the interpretation of geographic patterns: this will demonstrate how the building of spatial units can introduce the Modifiable Areal Unit Problem (MAUP), as different units may lead to varying results and conclusions based on the same data.

### Deliverable

The goal for this independent study is to produce a research paper suitable for publication in an academic journal. This paper will focus on methods for creating novel spatial units for geographic analysis, specifically exploring techniques for defining new areal units that can provide more meaningful insights into spatial data. The study will investigate various spatial unit creation methods, including buffer-based approaches (both symmetrical and walkshed buffers), tessellation-based units (with different grid resolutions), and clustering methods (such as k-means clustering with varying numbers of clusters). Additionally, Voronoi polygons generated from existing neighborhood centers will be tested to assess their effectiveness in capturing spatial patterns. Detailed definitions of these areal units, including their specific characteristics and parameters, are outlined in [Table 1](#tbl-agg).

A sensitivity analysis will compare how these spatial unit definitions affect the aggregation of 3-1-1 call data, exploring the impact of buffer and isochrone sizes, tessellation resolution, and clustering configurations. This sensitivity analysis will reveal how the choice of spatial unit can introduce the Modifiable Areal Unit Problem (MAUP) and how different spatial definitions can lead to varying interpretations of geographic phenomena.

This work is scheduled to be carried out over the summer, with an estimated completion in late August–early September, 2025. Maintaining this timeline will allow the authors to collaborate and edit the written portion of the document before submission to a peer-reviewed journal.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: The novel spatial units to be compared.   |  |  |  | | --- | --- | --- | | Extent | Method | Description | | Baltimore city NSA centroids | Symmetrical buffer | The creation of a simple buffer around each of the NSA centroids (which act as place names) relates to the distance-decay function of 3-1-1 calls–the events which occur closest to a neighborhood have a stronger influence than those which occur further away. | | Baltimore city NSA centroids | Walkshed buffer | By creating isochrones to better represent the pedestrian-scale distance from a neighborhood, the above symmetrical buffer can be extended. This method posits a different distance-decay function: one which considers physical barriers to movement for neighborhood residents. | | Baltimore city NSA centroids | Voronoi polygons | This method assigns each event to its nearest neighborhood center by creating polygons around NSA centroids. This approach assumes that 3-1-1 calls are only associated with the closest named neighborhood, and that each event only matters to one neighborhood. | | Baltimore city boundaries | Regular tessellation | By creating a regular pattern of grid cells across the entire region, the influence of social or political factors in existing boundaries can be mitigated. Because each cell is the same size and shape, it eliminates variability introduced by differences in area, shape, or population across existing spatial units. | | Baltimore city 3-1-1 call locations | Centers of 3-1-1 activity | In this method, k-means clustering is used to group 311 call data into activity zones based on geographic proximity. From these activity centers, a buffer can be computed to identify an area of influence of the cluster–this method suggests that 311 calls reflect the needs of people near the cluster, rather than being tied to specific place definitions. | |

### References

Baltimore City. 2024a. “Neighborhood Statistical Area (NSA) Boundaries.” *Open Baltimore*. Baltimore City. <https://data.baltimorecity.gov/datasets/baltimore::real-property-information-2/explore>.

———. 2024b. “311 Customer Service Requests 2024.” *Open Baltimore*. Baltimore City. <https://data.baltimorecity.gov/datasets/68a1136acff444bba6c93e845dfc00e1_0/explore>.

ESRI. 2022. *ArcGIS Pro*. Esri Inc. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Wong, David. 2009. “The Modifiable Areal Unit Problem (MAUP).” In, 105–23. SAGE Publications, Ltd. <https://doi.org/10.4135/9780857020130.n7>.